NASA CR 70851

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Space Programs Summary No. 37-36, Volume IV

for the period October 1, 1965 to November 30, 1965

Supporting Research and Advanced Development

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

December 31, 1965

Preface

The *Space Programs Summary* is a six-volume, bimonthly publication that documents the current project activities and supporting research and advanced development efforts conducted or managed by JPL for the NASA space exploration programs. The titles of all volumes of the *Space Programs Summary* are:

Vol. I. The Lunar Program (Confidential)

Vol. II. The Planetary–Interplanetary Program (Confidential)

Vol. III. The Deep Space Network (Unclassified)

Vol. IV. Supporting Research and Advanced Development (Unclassified)

Vol. V. Supporting Research and Advanced Development (Confidential)

Vol. VI. Space Exploration Programs and Space Sciences (Unclassified)

The *Space Programs Summary*, Vol. VI consists of an unclassified digest of appropriate material from Vols. I, II, and III; an original presentation of technical supporting activities, including engineering development of environmental-test facilities, and quality assurance and reliability; and a reprint of the space science instrumentation studies of Vols. I and II.

W. H. Pickering, *Director*
Jet Propulsion Laboratory

Space Programs Summary No. 37-36, Volume IV

Contents

SYSTEMS DIVISION

GUIDANCE AND CONTROL DIVISION

Contents (Cont'd)

Contents (Cont'd)

SPACE SCIENCES DIVISION

Contents (Cont'd)

Contents (Cont'd)

Contents (Cont'd)

SYSTEMS DIVISION

I. Systems Analysis

A. Drag Perturbations of a Mars Satellite

R. J. Richard

1. Introduction

An artificial satellite placed in orbit around Mars, like one placed in orbit around the Earth, encounters diverse forces which cause it to follow an orbit which cannot be simply described. These acting forces can be classified as: (1) the main force, i.e., the gravitational attraction of an assumed spherical central body (in our case, Mars); and (2) perturbing forces, the most important of which are the aerodynamic drag, the gravitational oblateness of the central body, the solar gravitation, and the solar radiation pressure.

In the absence of perturbative forces, a Mars satellite would traverse a planar conic orbit. Such an orbit could be completely specified by: (1) its size, designated by the semimajor axis a; (2) its shape, given by the eccentricity e; (3) the spatial orientation of the reference plane, given by: (a) the inclination i to some inertially fixed reference plane, and (b) the longitude Ω of the ascending mode of the orbit on this reference plane, measured in the reference plane from some inertial direction; and (4) the in-plane orientation of the orbit, specified by the argument of periapsis ω. The position of the spacecraft in the orbit is then given by the true anomaly f. These six orbital parameters relate simply to the position and velocity of the satellite (e.g., Ref. 1, p. 33 et seq.).

Briefly, the above perturbative forces have the following effects on the orbit. Aerodynamic drag produces a secular decrease in both the semimajor axis and the eccentricity. If the atmosphere is assumed to be rotating with the planet, then the orientation angles also change. Solar gravitation can cause the eccentricity and orientation angles to increase or decrease, depending on the orientation of the orbit. Oblateness has no direct secular or long-period effects on the size or shape of the orbit; however, it does change some of the orientation parameters and thus influences the manner in which solar gravitation changes the eccentricity. If there is no eclipsing of the spacecraft by the planet, radiation pressure produces no secular change in the semimajor axis, but it does change the eccentricity. When eclipsing occurs, the

situation becomes very difficult to analyze without making detailed computations; in general, however, there are secular changes in both parameters. When all of the effects are considered, the situation becomes extremely complicated, and there is no simple way to predict what the results will be, except to say that eventually drag will cause the spacecraft to impact if the initial orbit is not so large as to be outside of any sensible atmosphere. The following analysis deals with the effects produced by aerodynamic drag. The effects of the other forces will be discussed in a future SPS, Vol. IV.

2. Analysis

Because drag is a nonconservative force, it produces a secular decrease in the energy of the orbiter. Since the magnitude of the semimajor axis of the orbit is a direct function of the energy, drag produces a secular decrease in this parameter also, causing the orbit to decay. The drag caused by a nonrotating atmosphere will be treated in this analysis.

The drag force is a function of the shape and mass of the satellite, the velocity of the spacecraft with respect to the atmosphere, and the mass density of the atmosphere. The instantaneous speed depends on the size of the orbit and the altitude; i.e.,

$$V^2 = \frac{2\mu}{(h + R_0)} - \frac{\mu}{a} = \frac{2\mu}{r} - \frac{\mu}{a}, \qquad (1)$$

where V is the speed; μ, the gravitational constant of the planet; R_0, the radius of the planet; a, the semimajor axis of the orbit; and h, the altitude. The atmosphere density is also altitude-dependent, decreasing with altitude. In general, this altitude-dependence is complex and thus serves to complicate an already difficult problem. At low altitudes, the density decreases approximately exponentially with altitude; i.e.,

$$\rho = \rho_0 e^{-\beta h} = \rho_0 e^{-h/s}, \qquad (2)$$

where ρ_0 is the surface density and s is the scale height. With increasing altitude, the scale height tends to increase, and the rate of density decrease correspondingly becomes smaller.

The problems encountered in an attempt to determine orbit variations due to drag can perhaps be best demonstrated by starting with the following simple analysis: The drag force, directed along the velocity vector, is given by

$$D = \frac{1}{2} \rho C_D A V^2, \qquad (3)$$

and the acceleration is

$$A_D = -\frac{D}{m} = \frac{d\overset{\backslash}{V}}{dt} = -\frac{1}{2}\rho \frac{C_D A}{m} V^2, \qquad (4)$$

where C_D is the drag coefficient, A is the cross-sectional area, and m is the mass of the orbiter. We have used perturbative derivative notation, since the drag is a perturbative acceleration. From Eq. (1), we obtain

$$2V \frac{d\overset{\backslash}{V}}{dt} = -\frac{\mu}{a^2} \frac{d\overset{\backslash}{a}}{dt}. \qquad (5)$$

Combining Eqs. (1), (4), and (5),

$$2 \frac{\mu}{a^2} \frac{d\overset{\backslash}{a}}{dt} = -\rho \frac{C_D A}{m} \left(\frac{2\mu}{r} - \frac{\mu}{a} \right)^{3/2}. \qquad (6)$$

For a circular orbit, we have $r = a = r_c$; thus,

$$\frac{\mu}{r_c^2} \frac{dr_c}{dt} = -\rho B \left(\frac{\mu}{r_c} \right)^{3/2}, \qquad (7)$$

$$\frac{dr_c}{dt} = -\rho B (\mu r_c)^{1/2}. \qquad (8)$$

Now, since $r_c = R_0 + h_c$, we have

$$\frac{dh_c}{dt} = -\rho B [\mu(R_0 + h_c)]^{1/2}. \qquad (9)$$

Defining the lifetime of the satellite, T_L, as the time it takes for h_c to go to zero from an initial value h_i, we obtain

$$T_L = \frac{1}{B(\mu R_0)^{1/2}} \int_0^{h_i} \frac{dh_c}{\rho \left(1 + \frac{h_c}{R_0} \right)^{1/2}}. \qquad (10)$$

Finally, using Eq. (2), we can write

$$T_L = \frac{1}{\rho_0 B(\mu R_0)^{1/2}} \int_0^{h_i} \frac{e^{h_c/s} dh_c}{\left(1 + \frac{h_c}{R_0} \right)^{1/2}}, \qquad (11)$$

where s is some (often complicated) function of h_c. Thus, the lifetime problem has been reduced to quadratures for this special case. However, in general, the integration must still be performed numerically. There is one special

case which is directly and easily integrable: if s is a constant and $h_i \ll R_0$, then

$$\left(1 + \frac{h_c}{R_0}\right)^{-1/2} \approx 1 - \frac{h_c}{2R_0}. \qquad (12)$$

Thus, we can write

$$T_L \approx C_1 \int_0^{h_i} \left(1 - \frac{h_c}{2R_0}\right) e^{h_c/s} \, dh_c, \qquad (13)$$

which is easily integrated to give

$$T_L \approx C_1 s e^{h_i/s} \left(1 - \frac{h_i}{2R_0}\right), \qquad (14)$$

for which we have also assumed $s \ll h_i$ and $s \ll R_0$. In this case, the lifetime essentially increases exponentially with altitude.

Eq. (14) can be rewritten as

$$T_L \approx \left[\frac{S}{B(\mu R_0)^{1/2}} \left(1 - \frac{h_i}{2R_0}\right)\right] \frac{1}{\rho_i}. \qquad (14a)$$

Remembering that $h_i \ll R_0$, we see that, in the case of an exponential atmosphere, the lifetime of a low circular orbit is inversely proportional to the atmospheric density at the initial altitude. Fig. 2 is a plot of the lifetime as a function of altitude for the atmosphere shown in Fig. 1. This atmosphere is that which resulted from assuming

the atmosphere to be exponential with the ground level density and the scale height as determined by the *Mariner IV* occultation experiment (Ref. 2).

If, on the other hand, we have to deal with more complex atmosphere models such as given in Fig. 3, we then have to follow a different and more general approach to the problem. There are several methods of varying degrees of complexity which have been applied to this problem, but here we will use the one which is probably the most straightforward. A variation-of-parameters
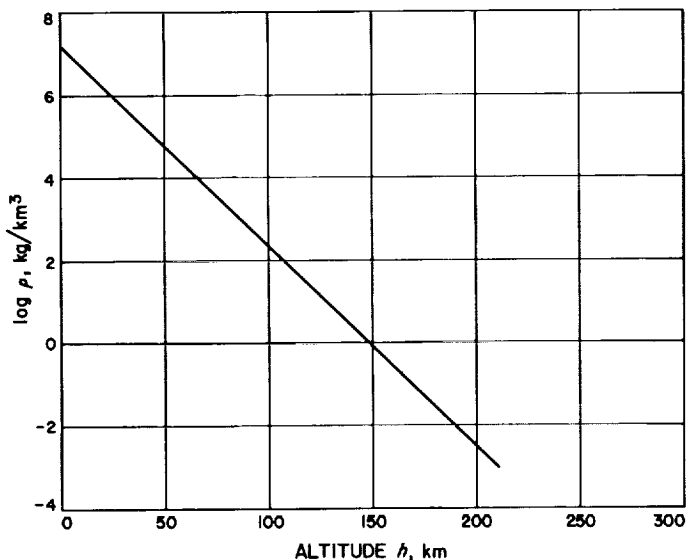
Fig. 1. Atmospheric density as a function of altitude for Mars Atmosphere A
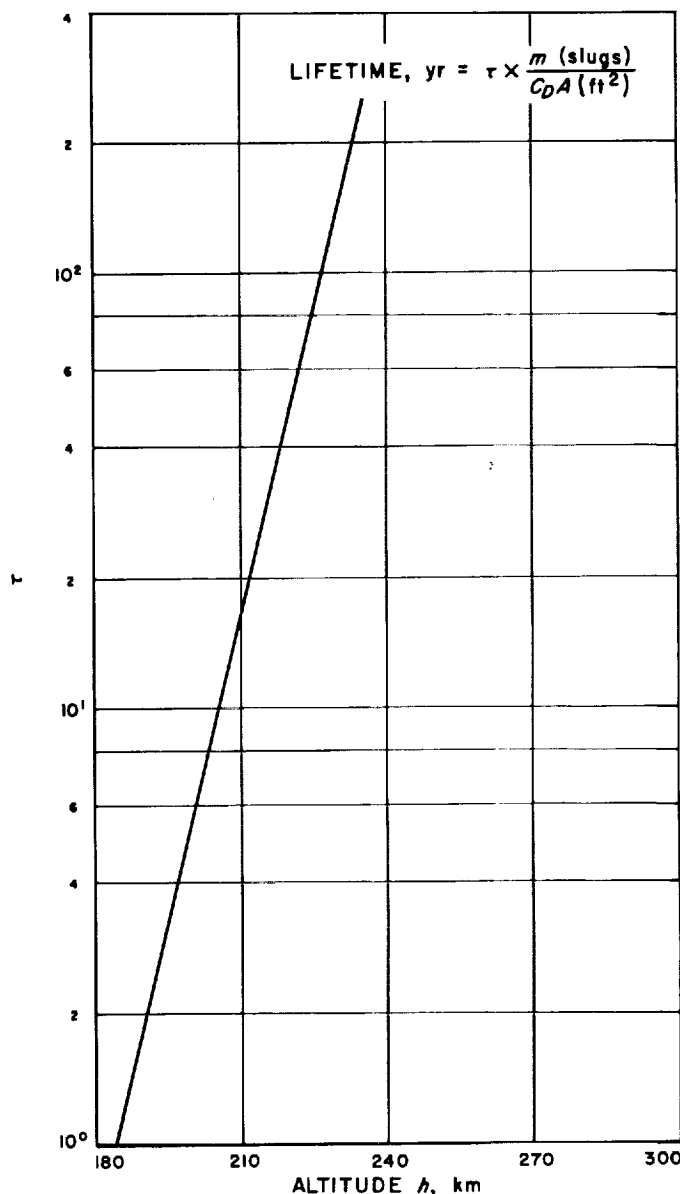
Fig. 2. Lifetime as a function of altitude for Mars Atmosphere A
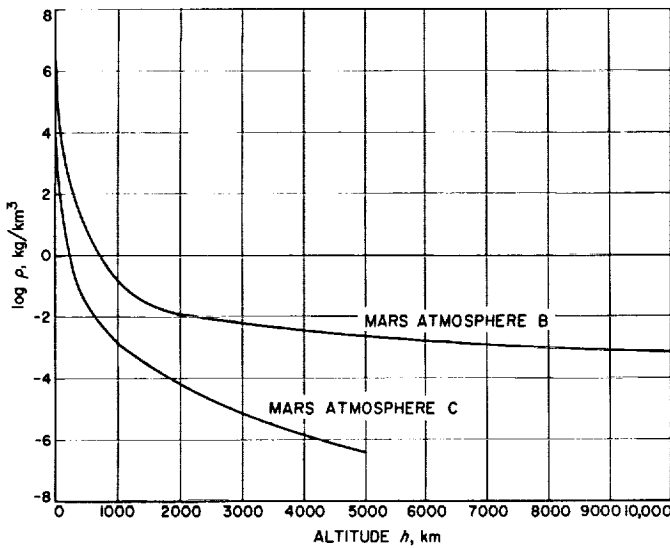
Fig. 3. Atmospheric density as a function of altitude for Mars Atmospheres B and C

technique making use essentially of the Krylov-Bogoliubiov (K-B) averaging method will be followed.

As a starting point, we will use the planetary equations in gaussian form:

$$\frac{da}{dt} = \frac{2}{n}\left(\frac{1 + e^2 + 2e\cos f}{1 - e^2}\right)^{1/2} T , \qquad (15a)$$

$$\frac{de}{dt} = \frac{2(e + \cos f)}{na}\left(\frac{1 - e^2}{1 + e^2 + 2e\cos f}\right)^{1/2} T , \qquad (15b)$$

$$\frac{di}{dt} = \frac{\cos(\omega + f)}{na(1 + e\cos f)}(1 - e^2)^{1/2} W , \qquad (15c)$$

$$\frac{dx}{dt} = \frac{(1 - e^2)}{nae(1 + e\cos f)}\left[2\sin f (e^4 - e^2 - 1 - e\cos f) T\right.$$

$$\left. + (e^2\cos f - 2e^4\cos f - \cos f - 2e^3)N\right] , \qquad (15d)$$

$$\frac{d\omega}{dt} = \frac{(1 - e^2)^{1/2}}{nae(1 + e\cos f)}\left[\frac{2\sin f (1 + e\cos f)}{(1 + e^2 + 2e\cos f)^{1/2}} T\right.$$

$$\left. + \frac{2e + e^2\cos f + \cos f}{(1 + e^2 + 2e\cos f)^{1/2}} N - e\cot i \sin(\omega + f)W\right] , \qquad (15e)$$

$$\frac{d\Omega}{dt} = \frac{(1 - e^2)^{1/2}\sin(\omega + f)}{na\sin i (1 + e\cos f)} W , \qquad (15f)$$

where $n = \mu^{1/2} a^{-3/2}$ is the mean motion, f is the true anomaly, $x = -nT_0$ is the mean anomaly at epoch, T is the component of the perturbing acceleration parallel to

the direction of motion, N is the component in the plane of the orbit normal to T (taken positive when directed to the interior of the orbit), and W is perpendicular to the plane of the orbit (taken positive in the direction of the angular momentum vector of the orbit). In the case of a nonrotating atmosphere, the drag force lies entirely along the T-direction; thus, Eqs. (15a) through (15f) become

$$\frac{da}{dt} = \frac{2}{n}\left(\frac{1 + e^2 + 2e\cos f}{1 - e^2}\right)^{1/2} T , \qquad (16a)$$

$$\frac{de}{dt} = \frac{2(e + \cos f)}{na}\left(\frac{1 - e^2}{1 + e^2 + 2e\cos f}\right)^{1/2} T , \qquad (16b)$$

$$\frac{di}{dt} = 0 , \qquad (16c)$$

$$\frac{dx}{dt} = \frac{2(1 - e^2)\sin f (e^4 - e^2 - 1 - e\cos f)}{nae(1 + e\cos f)} T , \qquad (16d)$$

$$\frac{d\omega}{dt} = \frac{2(1 - e^2)^{1/2}\sin f}{nae(1 + e^2 + 2e\cos f)^{1/2}} T , \qquad (16e)$$

$$\frac{d\Omega}{dt} = 0 . \qquad (16f)$$

Now, the disturbing acceleration is given by Eq. (4):

$$T = \frac{1}{2}\rho BV^2 . \qquad (4a)$$

The perturbation equations go into a more tractable form if we make the transformation from true anomaly to eccentric anomaly, E, where

$$\cos f = \frac{\cos E - e}{1 - e\cos E} , \qquad (17a)$$

$$\sin f = \frac{(1 - e^2)^{1/2}\sin E}{1 - e\cos E} . \qquad (17b)$$

Then, we obtain

$$\frac{da}{dt} = \frac{2}{n}\left(\frac{1 + e\cos E}{1 - e\cos E}\right)^{1/2} T , \qquad (18a)$$

$$\frac{de}{dt} = \frac{2(1 - e^2)}{na}\frac{\cos E}{(1 - e^2\cos^2 E)^{1/2}} T , \qquad (18b)$$

$$\frac{dx}{dt} = \frac{2(1 - e^2)^{3/2}}{nae}\frac{\sin E (e^3\cos E - e - 1)}{1 - e\cos E} T , \qquad (18c)$$

$$\frac{d\omega}{dt} = \frac{2(1 - e^2)^{1/2}}{nae}\frac{\sin E}{(1 - e^2\cos^2 E)^{1/2}} T . \qquad (18d)$$

Using the energy integral, Eq. (1), we have

$$V^2 = \frac{\mu}{a}\left(\frac{1 + e\cos E}{1 - e\cos E}\right).$$ (19)

Thus, the perturbation equations become

$$\frac{da}{dt} = \frac{\rho B}{n}\left(\frac{1 + e\cos E}{1 - e\cos E}\right)^{3/2},$$ (20a)

$$\frac{de}{dt} = \frac{\rho B(1-e^2)}{na}\frac{\cos E(1 + e\cos E)^{1/2}}{(1 - e\cos E)^{3/2}},$$ (20b)

$$\frac{dx}{dt} = \frac{\rho B(1-e^2)^{3/2}}{nae}\frac{\sin E(e^3\cos E - e - 1)(1+e\cos E)}{(1-e\cos E)^2},$$ (20c)

$$\frac{d\omega}{dt} = \frac{\rho B(1-e^2)^{1/2}}{nae}\frac{\sin E(1 + e\cos E)^{1/2}}{(1-e\cos E)^{3/2}}.$$ (20d)

Because of the generally complicated form of the dependence of ρ on the altitude and, hence, the eccentric anomaly, Eqs. (20a) through (20d) are not amenable to closed-form solutions. However, since they are slowly varying, we can treat the parameters a, e, and n on the right-hand side of the equations as constants and use the method of averages. The method consists of averaging over the mean anomaly M for one orbit (essentially the same as averaging over time) and then using this average rate of change for many orbits to compute the change in the elements. Thus, the averaged rate is given by

$$\left(\frac{\overline{dc}}{dt}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{dc}{dt}\,dM,$$ (21)

and

$$\Delta c = \left(\frac{\overline{dc}}{dt}\right)\Delta t,$$ (22)

where c represents any orbital element. Now,

$$M = E - e\sin E - x,$$ (23)

and

$$dM = (1 - e\cos E)\,dE,$$ (24)

since the orbital elements are treated as constants on the right-hand side of Eqs. (20) through (22). The averaged variation equations thus become

$$\left(\frac{\overline{da}}{dt}\right) = \frac{B}{2\pi n}\int_{-\pi}^{\pi}\rho(E)(1 + e\cos E)\left(\frac{1 + e\cos E}{1 - e\cos E}\right)^{1/2}dE,$$ (25a)

$$\left(\frac{\overline{de}}{dt}\right) = \frac{B(1-e^2)}{2\pi na}\int_{-\pi}^{\pi}\rho(E)\cos E\left(\frac{1 + e\cos E}{1 - e\cos E}\right)^{1/2}dE,$$ (25b)

$$\left(\frac{\overline{dx}}{dt}\right) = \frac{B(1-e^2)^{3/2}}{2\pi nae}\int_{-\pi}^{\pi}\rho(E)$$
$$\times\frac{\sin E(e^3\cos E - e - 1)(1 + e\cos E)}{1 - e\cos E}\,dE,$$ (25c)

$$\left(\frac{\overline{d\omega}}{dt}\right) = \frac{B(1 - e^2)^{1/2}}{2\pi nae}\int_{-\pi}^{\pi}\rho(E)\sin E\left(\frac{1 + e\cos E}{1 - e\cos E}\right)^{1/2}dE.$$ (25d)

The problem is thus reduced to quadratures, and the right-hand sides can be integrated numerically by the use of any of several quadrature formulas. A great simplification is introduced when we realize that the density is a function of the altitude only; hence the density depends on E only through the cosine, since

$$r = a(1 - e\cos E).$$ (26)

Thus, the arguments under the integral sign in Eqs. (25c) and (25d) are odd functions of E and integrate out to zero; i.e.,

$$\left(\frac{\overline{dx}}{dt}\right) = 0,$$ (27a)

$$\left(\frac{\overline{d\omega}}{dt}\right) = 0.$$ (27b)

Summarizing the results obtained, we see that, for a nonrotating atmosphere perturbing a satellite orbit, only the size and shape of the orbit change; the orientation angles of and in the plane are invariant after one complete revolution. This does not say that the orientation does not change; as a matter of fact, it does change continuously, but the changes are such that they sum up to zero over one complete revolution. We are now left with the following perturbation equations:

$$\left(\frac{\overline{da}}{dt}\right) = \frac{B}{2\pi n}\int_{-\pi}^{\pi}\rho(\cos E)(1 + e\cos E)\left(\frac{1 + e\cos E}{1 - e\cos E}\right)^{1/2}dE,$$ (25a)

$$\left(\frac{\overline{de}}{dt}\right) = \frac{B(1 - e^2)}{2\pi na}\int_{-\pi}^{\pi}\rho(\cos E)\cos\left(\frac{1 + e\cos E}{1 - e\cos E}\right)^{1/2}dE.$$ (25b)

In general, these equations must be solved by numerical integration on a computer. The altitude at any time is given by

$$h = a(1 - e\cos E) - R_0,$$

$$\text{LIFETIME, yr} = \tau \times \frac{m \text{ (slugs)}}{C_D A \text{ (ft}^2)}$$

ELLIPTICAL ORBIT:
$h = h_P$ = PERIAPSIS ALTITUDE
h_A = APOAPSIS ALTITUDE
CIRCULAR ORBIT:
$h = h_A = h_P$
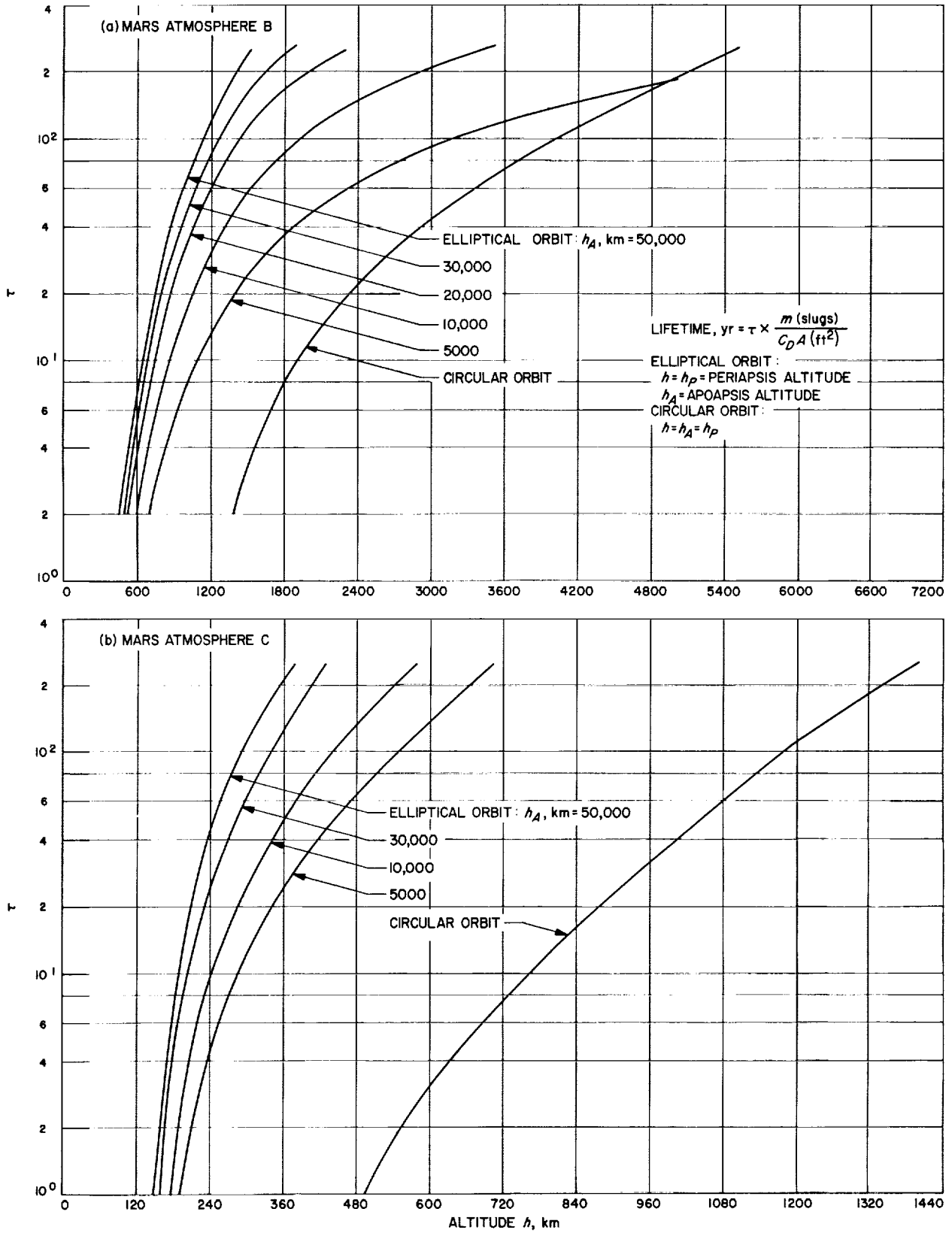
Fig. 4. Lifetime as a function of altitude for Mars Atmospheres B and C

where R_0 is the radius of the planet. In order to determine the lifetime, Eqs. (25a) and (25b) are integrated until $h = 0$. The time enters into the analysis through the relationship

$$M = E - e \sin E = nt + x.$$

3. Numerical Results

Extensive computer computations have been carried out based on the averaged planetary analysis equations presented above. The advantage of using averaged, rather than non-averaged, equations is that it is possible to use much larger integration step sizes. As a matter of fact, with unaveraged equations, the step size should be a small fraction of the orbital period; with the averaged equations, the step size is usually on the order of several orbital periods. (Actually, the use of a step size of less than one orbital period for the averaged equations is illogical, because the in-orbit variations have been averaged out.) Some results of the computations for the atmosphere models presented in Fig. 3 are given in Fig. 4. These results have been checked with programs using unaveraged planetary equations and direct methods of integration.

B. Effect of Separated Earth–Moon Masses on the Integrated Ephemeris of the Earth–Moon Barycenter

F. M. Sturms, Jr.

The current JPL ephemeris of the Earth–Moon barycenter is obtained by numerical integration of a fictitious body having mass equal to the sum of the masses of the Earth and Moon and considered to be at the barycenter (Ref. 3). This article gives the results from a study to evaluate the approximate difference between the motion of this fictitious body and that of the actual barycenter of the separated Earth and Moon.

1. Derivation of the Perturbing Accelerations

Let the real case be approximated by a two-dimensional, three-body model (Fig. 5). The equations of motion for
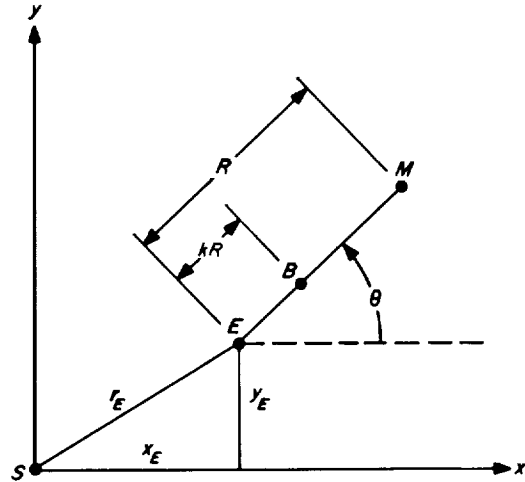
Fig. 5. Geometry of two-dimensional barycenter approximation

the Earth and Moon with respect to the Sun are:

$$\ddot{x}_E = - (\mu_S + \mu_E) \frac{x_E}{r_E^3} + \mu_M \left(\frac{x_M - x_E}{R^3} - \frac{x_M}{r_M^3} \right), \quad (1)$$

$$\ddot{x}_M = - (\mu_S + \mu_M) \frac{x_M}{r_M^3} + \mu_E \left(\frac{x_E - x_M}{R^3} - \frac{x_E}{r_E^3} \right), \quad (2)$$

plus the corresponding equations in y_E and y_M. The location of the barycenter is given by:

$$x_B = x_E + \left(\frac{\mu_M}{\mu_E + \mu_M} \right) (x_M - x_E). \quad (3)$$

To shorten notation, define:

$$\left. \begin{array}{l} k = \dfrac{\mu_M}{\mu_E + \mu_M}, \\[2mm] 1 - k = \dfrac{\mu_E}{\mu_E + \mu_M}, \\[2mm] \mu = \mu_S + \mu_E + \mu_M. \end{array} \right\} \quad (4)$$

Then, the equations of motion of the barycenter are:

$$\ddot{x}_B = (1 - k)\ddot{x}_E + k\ddot{x}_M, \quad (5)$$

which, from Eqs. (1), (2), and (4), becomes:

$$\ddot{x}_B = - \mu (1 - k) \frac{x_E}{r_E^3} - \mu k \frac{x_M}{r_M^3}. \quad (6)$$

In order to derive the radial and orthogonal accelerations, let the x-axis instantaneously pass through the
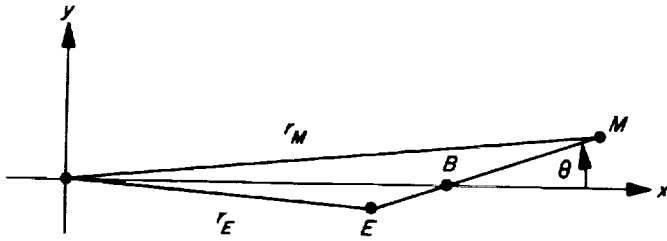
Fig. 6. Configuration for computing radial and orthogonal accelerations

barycenter, and measure θ from the Sun–barycenter line (Fig. 6). From the law of cosines:

$$r_E^2 = r_B^2 + k^2 R^2 - 2r_B kR \cos \theta$$

$$= r_B^2 \left[1 + k^2 \frac{R^2}{r_B^2} - 2k \frac{R}{r_B} \cos \theta \right], \tag{7}$$

$$r_M^2 = r_B^2 + (1 - k)^2 R^2 + 2r_B (1 - k) R \cos \theta$$

$$= r_B^2 \left[1 + (1 - k)^2 \frac{R^2}{r_B^2} + 2 (1 - k) \frac{R}{r_B} \cos \theta \right]. \tag{8}$$

Also,

$$\left. \begin{array}{l} x_E = r_B - kR \cos \theta, \\[2mm] x_M = r_B + (1 - k) R \cos \theta. \end{array} \right\} \tag{9}$$

Substituting Eqs. (7) through (9) into Eq. (6), expanding the denominators in terms of the small parameter R/r_B^2, and collecting terms yields:

$$\ddot{x}_B = - \frac{\mu}{r_B^2} - K [3 \cos 2\theta + 1], \tag{10}$$

where

$$K = \frac{3}{4} k (1 - k) \mu \frac{R^2}{r_B^4} \tag{11}$$

and terms of order $(R/r_B^2)^3$ and higher are dropped. A similar process with

$$\left. \begin{array}{l} y_E = - kR \sin \theta, \\[2mm] y_M = (1 - k) R \sin \theta, \end{array} \right\} \tag{12}$$

gives:

$$\ddot{y}_B = 2K \sin 2\theta. \tag{13}$$

Eqs. (10) and (13) are just those of a fictitious body with mass equal to the sum of the masses of the Earth and Moon, *plus* a radial perturbation

$$\ddot{r} = - K [3 \cos 2\theta + 1] \tag{14}$$

and an orthogonal perturbation

$$r\ddot{v} = 2K \sin 2\theta. \tag{15}$$

The effects of these perturbations are the desired differences between the actual barycenter motion and the current approximation.

2. Variation-of-Parameters Solution

For simplicity, assume that the approximate motion is circular. The motion of the actual barycenter, starting with initial osculating circular conditions, can easily be found by the variation-of-parameters method. To avoid the singularities associated with zero eccentricity, introduce the orbit elements

$$\left. \begin{array}{l} a_N = e \sin \omega, \\[2mm] a_M = e \cos \omega, \\[2mm] u = v + \omega. \end{array} \right\} \tag{16}$$

The radius is given by:

$$r = \frac{a (1 - a_N^2 - a_M^2)}{1 + a_N \sin u + a_M \cos u}. \tag{17}$$

The argument of latitude u (normally measured from the ascending node) can, in this two-dimensional case, be measured from any arbitrary reference line in the plane, which shall be selected here as the initial radius. For zero initial eccentricity, the perturbation equations may be shown to be:

$$\frac{da}{dt} = \frac{2}{n} (r\ddot{v}),$$

$$\frac{da_N}{dt} = \frac{1}{na} \left[- \cos u\, (\ddot{r}) + 2 \sin u\, (r\ddot{v}) \right], \tag{18}$$

$$\frac{da_M}{dt} = \frac{1}{na} \left[\sin u\, (\ddot{r}) + 2 \cos u\, (r\ddot{v}) \right],$$

where n is the mean motion. A first-order solution to Eqs. (18) can be obtained by assuming that K, a, and n are constant on the right-hand side and noting that:

$$\left. \begin{array}{l} u = nt, \\[2mm] \theta = \theta_0 + \omega t. \end{array} \right\} \tag{19}$$

The results are:

$$a = a_0 - \frac{2K}{n\omega} (\cos 2\theta - \cos 2\theta_0), \tag{20}$$

$$a_N = \frac{K}{na_0}\left[\frac{\sin u}{n} + \frac{\sin 2\theta_0 - \sin (u + 2\theta)}{2(n + 2\omega)}\right.$$

$$\left. + 7\,\frac{\sin 2\theta_0 + \sin (u - 2\theta)}{2(n - 2\omega)}\right], \qquad (21)$$

$$a_M = \frac{K}{na_0}\left[\frac{\cos u - 1}{n} + \frac{\cos 2\theta_0 - \cos (u + 2\theta)}{2(n + 2\omega)}\right.$$

$$\left. - 7\,\frac{\cos 2\theta_0 - \cos (u - 2\theta)}{2(n - 2\omega)}\right]. \qquad (22)$$

Expanding Eq. (17), we obtain

$$\Delta r = r - a_0 = (a - a_0) - a_0\, a_N \sin u - a_0\, a_M \cos u, \qquad (23)$$

which, after substitution of Eqs. (20) through (22), becomes

$$\Delta r = \frac{K}{n}\left\{\frac{\cos u}{n} + \left(\frac{2}{\omega}\cos 2\theta_0 - \frac{1}{n}\right)\right.$$

$$- \frac{2n^2 + 3\omega n}{\omega (n^2 - 4\omega^2)}\cos 2\theta$$

$$- \frac{1}{2(n + 2\omega)}\cos (u - 2\theta_0)$$

$$\left. + \frac{7}{2(n - 2\omega)}\cos (u + 2\theta_0)\right\} \qquad (24)$$

Fig. 7 shows plots of Δr versus time for three values of θ_0. From Eq. (24), the dominant effect is the annual term with amplitude K/n^2. This is modified by the two
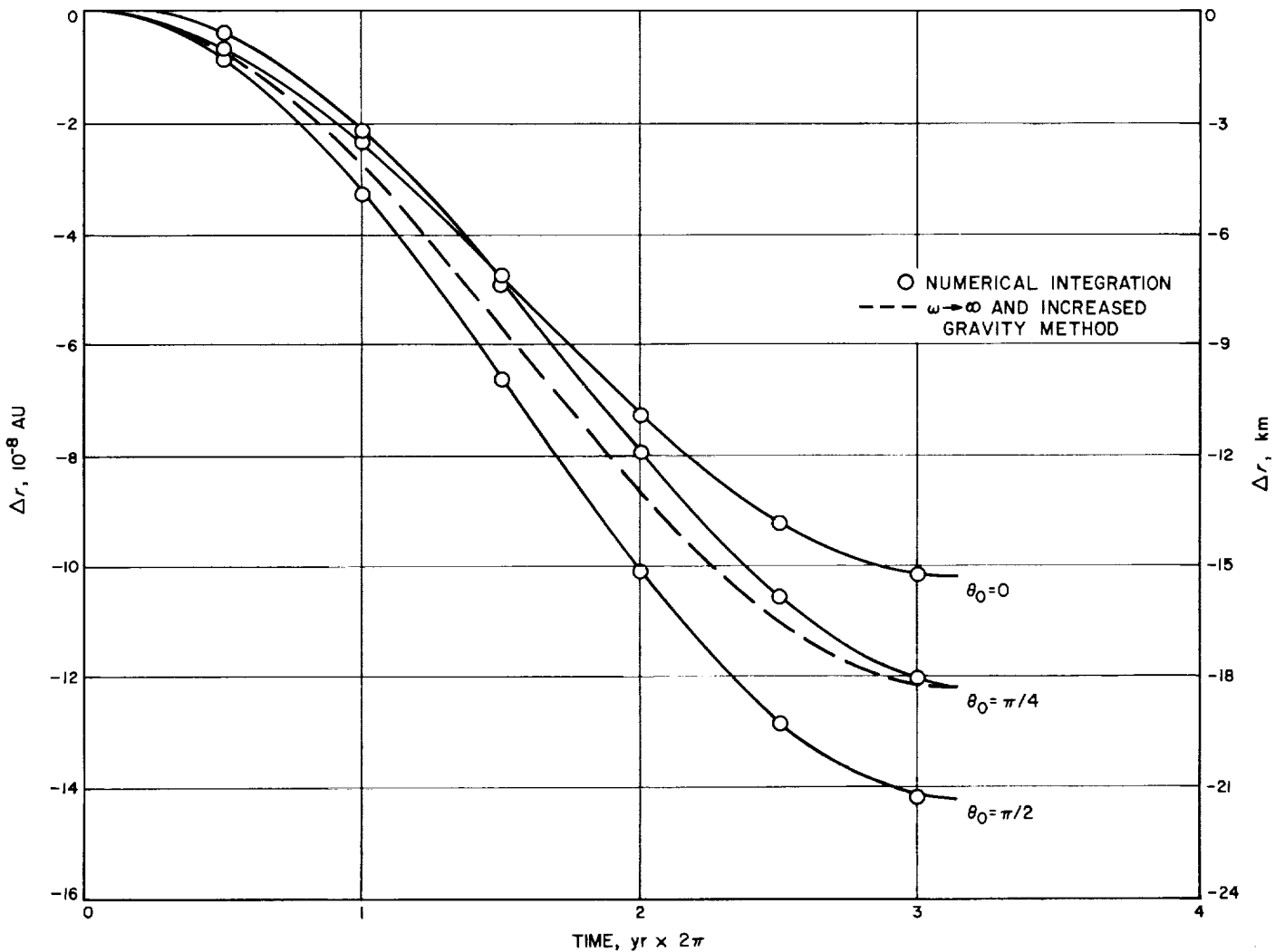
Fig. 7. Radial variation for zero initial osculating eccentricity

annual terms which are smaller in amplitude and out of phase by $\pm 2\theta_0$. This effect can be seen in Fig. 7. A small semimonthly term has amplitude

$$\frac{K(n + 3\omega)}{\omega(n^2 - 4\omega^2)} .$$

In characteristic units, we have

$$a_0 = r_B = 1 ,$$
$$n = 1 ,$$
$$R = 0.0026 ,$$
$$\omega = 12.4 ,$$
$$\mu = 1 ,$$
$$k = 1/82 = 0.0122 .$$

Then, from Eq. (11),

$$K = 0.0611 \times 10^{-6} .$$

The semimonthly amplitude is

$$\frac{K(n + 3\omega)}{\omega(n^2 - 4\omega^2)} = 306.5 \times 10^{-12} \text{ AU} ,$$

which, in metric units, is

$$(306.5 \times 10^{-12})(1.5 \times 10^8) = 0.046 \text{ km} ,$$

and the dominant annual amplitude is

$$\frac{K}{n^2} = 0.0611 \times 10^{-6} \text{ AU or } 9.16 \text{ km} .$$

The dominant characteristics of the motion can be most easily portrayed by recognizing that ω is large compared to n. As $\omega \to \infty$, we have

$$\Delta r \underset{\omega \to \infty}{\to} -\frac{K}{n^2}(1 - \cos u) . \tag{25}$$

Then, the maximum $|\Delta r|$ occurs at $u = \pi$, at which value $\Delta r = -18.3$ km.

The variation in the motion perpendicular to the radius can be evaluated by considering the angular momentum h. For zero initial eccentricity,

$$\frac{dh}{dt} = a_0(r\dot{v}) = \frac{na_0}{2}\frac{da}{dt} . \tag{26}$$

Then,

$$\Delta h = \frac{na_0}{2}\Delta a = -\frac{Ka_0}{\omega}(\cos 2\theta - \cos 2\theta_0) . \tag{27}$$

Also, since the horizontal velocity V_H is related to h by

$$h = V_H r , \tag{28}$$

we have

$$\Delta V_H = \frac{\Delta h}{a_0} - n\Delta r \tag{29}$$

and

$$\Delta V_H \underset{\omega \to \infty}{\to} \frac{K}{n}(1 - \cos u) . \tag{30}$$

The perturbed angular rate is

$$n = \frac{V_H}{r} . \tag{31}$$

Then,

$$\Delta n = \frac{\Delta V_H}{a_0} - \frac{n}{a_0}\Delta r , \tag{32}$$

$$\Delta n \underset{\omega \to \infty}{\to} \frac{2K}{na_0}(1 - \cos u) . \tag{33}$$

The separation along the orbit is

$$\Delta s = a_0 \int_0^t \Delta n \, dt , \tag{34}$$

$$\Delta s \underset{\omega \to \infty}{\to} \frac{2K}{n^2}(u - \sin u) . \tag{35}$$

After one complete revolution, the separation is

$$4\pi\frac{K}{n^2} = 76.8 \times 10^{-8} \text{ AU or } 115 \text{ km} .$$

This corresponds to a period shortening of 3.8 sec. It can be shown from the complete expression of Eq. (34) that the semimonthly variation in the separation along the orbit has an amplitude of 0.034 km.

The results of the variation-of-parameters solution are verified by numerical integration of Eq. (6) (and the corresponding equation in y_B), which is discussed in Section 4 below.

3. Increased-Gravity Solution

It is necessary to change the initial conditions of the actual barycenter motion in order to re-establish the known period. This could be done accurately in the variation-of-parameters solution by taking non-zero initial eccentricities; however, this results in very complex equations. An alternate method may be developed as follows:

The semimonthly periodic terms in the perturbing accelerations have only a small effect on the motion. If

they are ignored, there remains only a constant radial perturbing acceleration. Eq. (10) may be rewritten as

$$\ddot{x}_B = -\frac{(\mu + Kr_B^2)}{r_B^2}.$$ (36)

This is simply the acceleration resulting if the gravity constant were increased by

$$\Delta\mu = Kr_B^2 = \frac{3}{4}k(1-k)\mu\frac{R^2}{r_B^2}.$$ (37)

Therefore, the approximate motion of the barycenter is two-body, Keplerian motion about the Sun with constant increased gravity

$$\Delta\mu = Ka_0^2 = 0.0611 \times 10^{-6} \text{ solar masses}$$

$$= 0.81 \times 10^4 \text{ km}^3/\text{sec}^2.$$

From the vis-viva equation,

$$-\frac{1}{a} = \frac{V^2}{\mu} - \frac{2}{r}.$$ (38)

For a fixed initial position and velocity,

$$\frac{1}{a^2}\Delta a = -\frac{V^2}{\mu^2}\Delta\mu.$$ (39)

The initial conditions from the zero-eccentricity case are (in characteristic units): $\mu = 1$, $V = 1$, and $a = 1$. Therefore,

$$\Delta a = -0.0611 \times 10^{-6} \text{ AU}$$

$$= -9.16 \text{ km}.$$

The resulting two-body orbit starts at aphelion

$$r_a = a_0,$$

and has perihelion

$$r_p = a_0 + 2\Delta a.$$

Then, the eccentricity is

$$e = \frac{r_a - r_p}{r_a + r_p} = \frac{-2\Delta a}{2a_0 + 2\Delta a} \approx -\frac{\Delta a}{a_0}.$$ (40)

Now, since

$$r = \frac{a(1-e^2)}{1 + e\cos v} \approx a_0\left[1 + \frac{\Delta a}{a}(1 + \cos v)\right],$$ (41)

we have

$$\Delta r = \Delta a(1 + \cos v).$$ (42)

From Eq. (39), noting that

$$v = u - \pi,$$

$$V = na,$$

$$\mu = n^2 a^3,$$

we have finally

$$\Delta r = -\frac{K}{n^2}(1 - \cos u),$$ (43)

which is the same as that obtained in the variation-of-parameters solution in Eq. (25) for $\omega \rightarrow \infty$.

The period is given by

$$P = 2\pi\frac{a^{3/2}}{\mu^{1/2}}.$$ (44)

Then,

$$\Delta P = 2\pi\left(\frac{3}{2}\frac{a^{1/2}}{\mu^{1/2}}\Delta a - \frac{1}{2}\frac{a^{3/2}}{\mu^{3/2}}\Delta\mu\right)$$

$$= P\left(\frac{3}{2}\frac{\Delta a}{a} - \frac{1}{2}\frac{\Delta\mu}{\mu}\right).$$ (45)

In order to preserve the period ($\Delta P = 0$), we have

$$\Delta a = \frac{1}{3}\frac{a}{\mu}\Delta\mu = \frac{1}{3}\frac{K}{n^2}$$

$$= 2.04 \times 10^{-8} \text{ AU} = 3 \text{ km}.$$ (46)

The new orbit can be a circular orbit which is 3 km outside the original, in which case, for the circular velocity,

$$V^2 = \frac{\mu}{a},$$ (47)

$$2V\Delta V = \frac{1}{a}\Delta\mu - \frac{\mu}{a^2}\Delta a$$ (48)

$$= \frac{2}{3}\frac{\Delta\mu}{a},$$

$$\Delta V = \frac{1}{3}\frac{K}{n}$$ (49)

$$= 2.04 \times 10^{-8}.$$

In metric units,

$$\Delta V = 0.61 \times 10^{-6} \text{ km/sec}.$$

This value results from a ΔV of 0.91×10^{-6} km/sec required because of the increased mass [first term in Eq. (48)] and a ΔV of -0.30×10^{-6} km/sec required because of the increased radius [second term in Eq. (48)].

Table 1. Radius, velocity, and period increments obtained by numerical integration and the increased-gravity method

Increment	Circular orbit with $r = 1$				Circular orbit with $\Delta P = 0$			
	Numerical integration			Increased-gravity method	Numerical integration			Increased-gravity method
	$\theta_0 = 0$	$\theta_0 = \pi/4$	$\theta_0 = \pi/2$		$\theta_0 = 0$	$\theta_0 = \pi/4$	$\theta_0 = \pi/2$	
Δr, 10^{-8} AU	0	0	0	0	2.037	1.376	2.037	2.037
ΔV, 10^{-8} AU/t	2.569	3.066	3.556	3.055	1.550	2.378	2.538	2.037
ΔP, sec	−0.96	−0.65	−0.96	−0.96	0	0	0	0
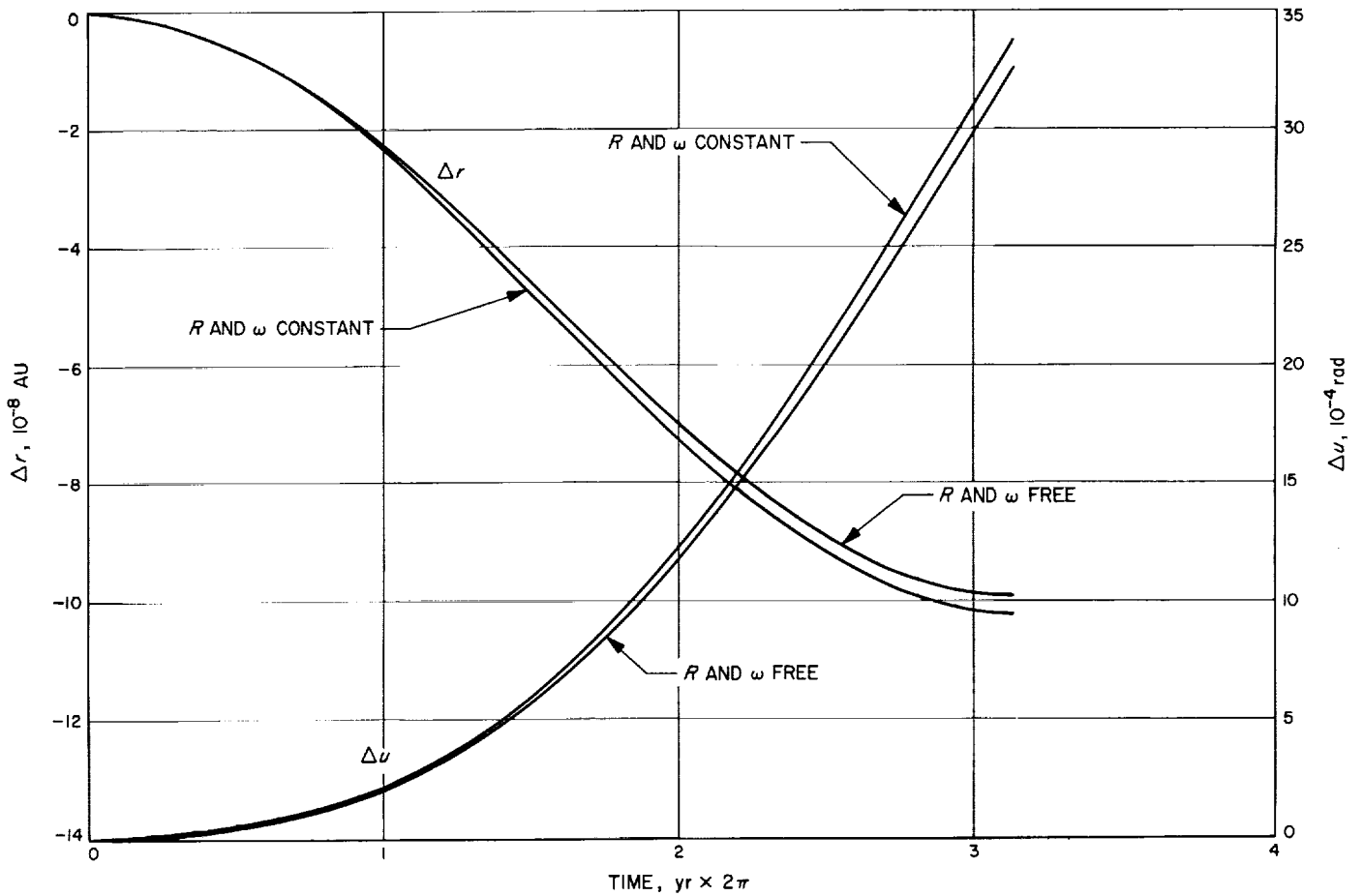
Fig. 8. Effect of holding R and ω constant for $\theta_0 = 0$

4. Numerical Integrations

A computer program has been written to integrate Eq. (6) and the corresponding equation in y_B. All computations are done in double-precision. The parameters x_E, x_M, y_E, and y_M are computed from Eqs. (9) and (12), and the assumption is made that the Earth–Moon distance R and angular rate ω are constant. This technique is analogous to that being used to generate a new barycenter ephemeris with PLOD (planetary orbit determination program; Ref. 4), i.e., assuming that the Moon's motion with respect to the Earth is known. Of course, in that case, the assumed motion is much more exact than in this study.

The results of the numerical integrations agree very well with the analytical solution in Eq. (24). The differences are negligible on the scale used for plotting Fig. 7.

The approximate analytical results obtained by using the increased-gravity method and by letting $\omega \to \infty$ agree most closely with the cases where $\theta_0 = \pi/4$, as can be seen from the comparison in Fig. 7.

Circular orbits were obtained by numerical integration with either unit radius or period of 2π. The necessary velocity and radius increments are compared with the analytical values from the increased-gravity method in Table 1.

Another computer program was written to check the effects of the assumption that R and ω are constant. In this program, Eqs. (1) through (3) were used, in which the Earth and Moon are free to follow the three-body motion resulting from point-mass gravity. The initial conditions were chosen such that the initial R and ω were the constant values used in the other program. Fig. 8
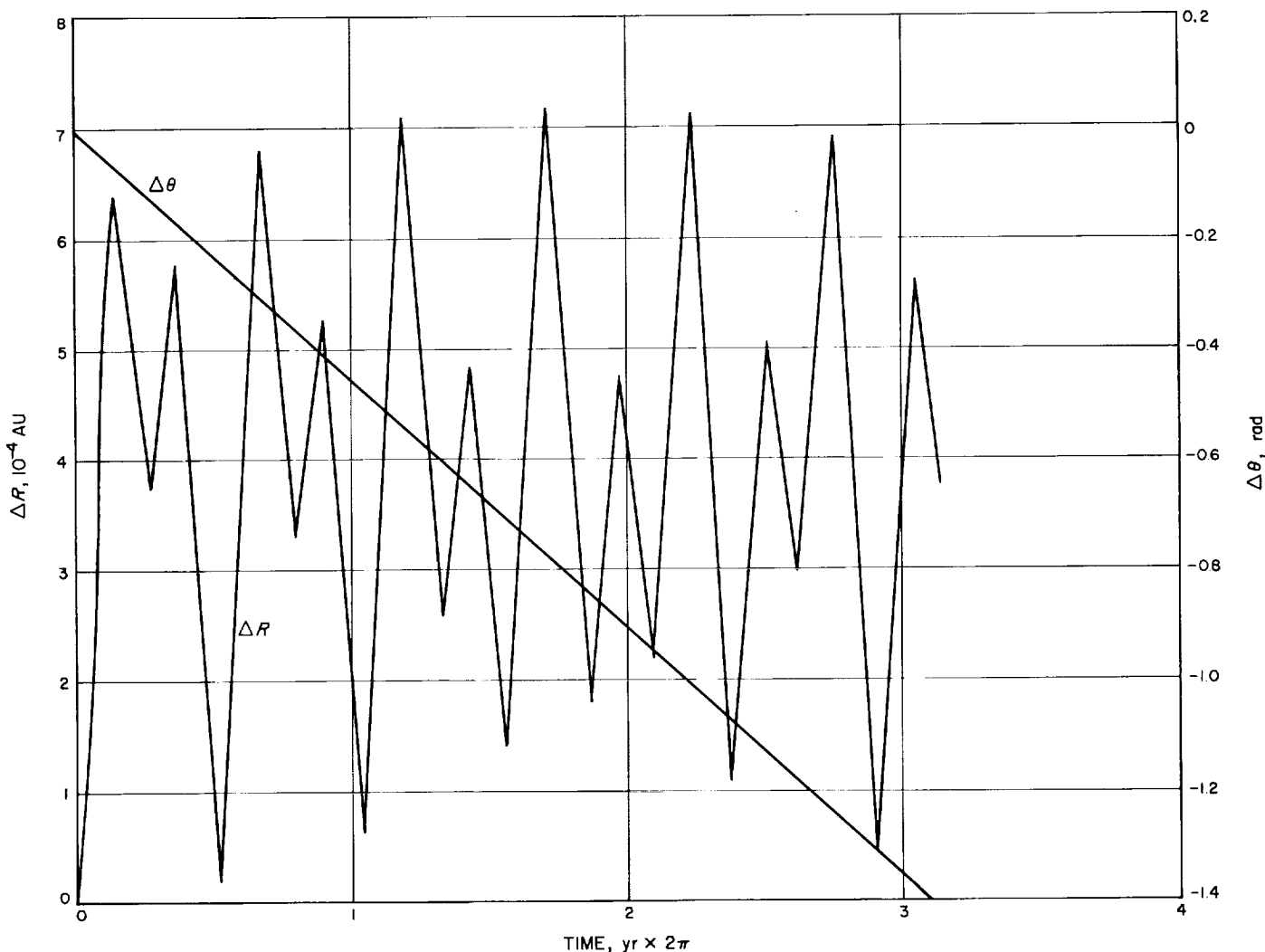
Fig. 9. Deviation of R and θ from circular values

shows plots comparing Δr and Δu for the two programs for the particular case $\theta_0 = 0$. Fig. 9 shows values of ΔR and $\Delta \theta$ which indicate the departure from circular motion for the Moon about the Earth. The results of this comparison indicate that the original assumption of constant R and ω did not seriously affect the nature of the results being sought.

5. Conclusions

The study assumes that the approximate barycenter motion which results from combining the Earth–Moon masses is circular at the radius necessary to give a 1-yr period. The actual barycenter motion has the following characteristics with respect to the approximate motion:

(1) There exist small oscillations of 46 m in radius and 34 m along the orbit, both with semimonthly periods.

(2) If the barycenter is initiated with the radius and velocity of the circular approximate motion, the annual variation of the actual motion is a decrease in radius to a maximum Δr of about -18 km after half an orbit, and a shortening of the period by about 4 sec.

(3) An almost circular orbit can be obtained for the actual barycenter at the original radius by increasing the initial velocity by about 1 mm/sec. This will result in a period about 1 sec shorter than that in the approximate case.

(4) An almost circular orbit, with a period equal to that in the approximate case, can be obtained for the actual barycenter by increasing the initial radius by about 3 km, and increasing the initial velocity by about 0.6 mm/sec.

According to p. 5 of Ref. 5, the 3-km bias and the 46-m semimonthly oscillation in radius were obtained by Hill in 1878.

If it can be assumed that the current Earth–Moon barycenter ephemeris has been obtained with a fit that removes secular departures from observed motion, then the period must be that of the actual barycenter. Therefore, from Item (4) above, the corrected ephemeris should give heliocentric radii about 3 km greater and heliocentric velocities about 0.6 mm/sec greater than those of the current ephemeris. Preliminary results verify these expected values.

C. Performance Computations With Pieced Solutions of Planetocentric and Heliocentric Trajectories for Low-Thrust Missions

W .G. Melbourne and C. G. Sauer, Jr.

1. Introduction

Recently, there has been increased interest in more accurate methods for obtaining performance estimates from matching two-body planetocentric and heliocentric low-thrust trajectories. The use of a more exact physical model, coupled with an optimal control program, to obtain exact values of performance requirements for an over-all interplanetary trajectory has not, to the author's knowledge, been reported upon in the open literature. Indeed, it is questionable whether such an extensive effort is warranted at this time, in view of the preliminary nature of current studies and the acceptable approximations currently in use.

The two primary objectives of this article are: (1) to modify the asymptotic velocity-intercept method described in Ref. 6 for constant thrust–acceleration trajectories to make it applicable to a constant thrust profile with a finite propellant loss, and (2) to establish a technique for obtaining performance estimates of both constant thrust and constant thrust–acceleration trajectories which depart from near a planet with parabolic or hyperbolic energy conditions and travel outward under low thrust. A set of analyses is presented which lends insight into the numerically obtained results. The symbols used in this discussion are defined in Table 2.

The asymptotic velocity-intercept method of matching two-body planetocentric and heliocentric trajectories is based on the observation that, for the case of a thrusting body in the gravitational field of a planet exclusively, the planetocentric velocity of the body escaping from the planet approaches an asymptotic form as the gravitational effects of the planet become negligible. The extrapolation of this asymptotic form to a zero value of planetocentric velocity defines a time T_i. By definition, the same thrusting body departing at the time T_i with zero planetocentric velocity from a massless planet would have this asymptotic form as its velocity profile.

Table 2. Definition of symbols

μ	planetary gravitational constant
R_0	initial radius
t	dimensionless time ($t = 0$ initially) $=$ time $\div$ ($R_0^{3/2}/\mu^{1/2}$)
A	thrust acceleration
a_0	normalized initial thrust acceleration $= A_0 R_0^2/\mu$
c	exhaust velocity of low-thrust propulsion system
ν	$(\mu/R_0)^{1/2}/c$
a	normalized thrust acceleration $= a_0/(1 - \nu a_0 t)$
r	normalized radius vector $= R/R_0$
θ	central angle
ξ	$4a_0\theta$
y	normalized semimajor axis of osculating conic
v	normalized velocity $=$ velocity $\div$ $(\mu/R_0)^{1/2}$
v_∞	normalized hyperbolic excess velocity
q	equivalent free-space velocity
E	normalized total energy per unit mass [see Eq. (5)]

The above suggests a method for performance calculations based upon a two-body analysis for interplanetary low-thrust trajectories involving a planetocentric spiral portion and a heliocentric transfer. On the departing planetocentric spiral, the propellant requirements are calculated from the initial time to T_I for a tangential thrust program. This is then added to the propellant requirements for the heliocentric transfer obtained from a trajectory departing at T_I with the heliocentric position and velocity of the planet, but in which the effect of the mass of the planet is ignored. The above procedure applies to an arrival or capture spiral also. This method provides over-all propulsion requirements and flight time estimates without the restrictions introduced by a more exact physical model. The above considerations apply equally well when the vehicle departs with an initial parabolic or hyperbolic energy during the planetocentric phase.

An obvious source of error in this method is the fact that the positions of the actual escaping body and the fictitious one starting from rest at T_I do not precisely match at large distances. However, experience with numerical cases has shown that this difference is small and that the sensitivity of the performance to distance variations is much smaller than the sensitivity to velocity variations. Actually, for spiral trajectories, the mismatch in position at asymptotic distances is of the order of $R_0 a^{-1/2}$. Other sources of error are the perturbation effects of the other planets, the gravity gradient effect of the Sun on the planet and the thrusting mass during the transition from the planetocentric to the heliocentric sphere of influence, and the fact that the tangential thrust program employed in the asymptotic region may not precisely coincide with the thrust program used in the two-body heliocentric model. All of these effects on performance estimates are considered small. In the case of the last effect, it is known that, for interplanetary transfer, the initial optimal thrusting direction for the two-body heliocentric model is fairly stationary and is closely aligned with the optimal asymptotic direction of a spiral trajectory escaping from the planet.

It should be observed that the performance estimates obtained by this method will agree closely with the results from an exact n-body formulation in which the "asymptote" of the planetocentric trajectory is optimally directed.

2. Spiral Trajectories

In order to provide a quantitative description of the method, we begin by defining the quantity q by the relationship

$$q = -\frac{1}{\nu} \ln (1 - \nu a_0 t),\qquad (1)$$

which, for infinite I_{sp}, is seen to become the quantity $a_0 t$. The planetocentric secular velocity of a body under constant tangential thrust in a low-thrust spiral trajectory is approximately given in normalized form by the expression (Ref. 7):

$$v = 1 - q.\qquad (2)$$

This expression is extremely accurate, except near escape. Furthermore, for a field-free region, v is given by the equation

$$v = q - q_I.\qquad (3)$$

The normalized velocity v has been plotted against q in Fig. 10 for initial thrust acceleration levels $a_0 = 10^{-2}$, 10^{-3}, 10^{-4}, and 10^{-5} in normalized units for $\nu = 0$. The asymptotic lines are shown for each value of a_0. Further, the nature of these relationships suggests that $v(q)$ and, hence, q_I should not be very sensitive to variations in ν. This is indeed the case. The values of q_I for a range of values of a_0 and ν are given in Table 3. Negative values of ν indicate a mass gain along the spiral and correspond to capture trajectories. The value of q_I is approximated to within 0.2% for the range of values of ν and a_0 given in Table 3 by the equation

$$q_I = 1 - 1.757\, a_0^{1/4} \exp\left[\nu(q_I + 1)/8\right].\qquad (4)$$

Table 3. Values of q_I for various a_0 and ν

a_0, g	Value of q_I for indicated ν								
	−0.5	−0.303	−0.2	−0.1	0	0.1	0.2	0.303	0.5
10^{-5}	—	0.9081	—	—	0.9012	—	—	0.8938	—
10^{-4}	0.8436	0.8362	0.8322	0.8283	0.8243	0.8202	0.8161	0.8117	0.803
10^{-3}	0.7195	0.7072	0.7006	0.6941	0.6876	0.6808	0.674	0.667	0.653
10^{-2}	0.4931	0.4740	0.4643	0.4548	0.4451	0.4351	0.425	0.413	—
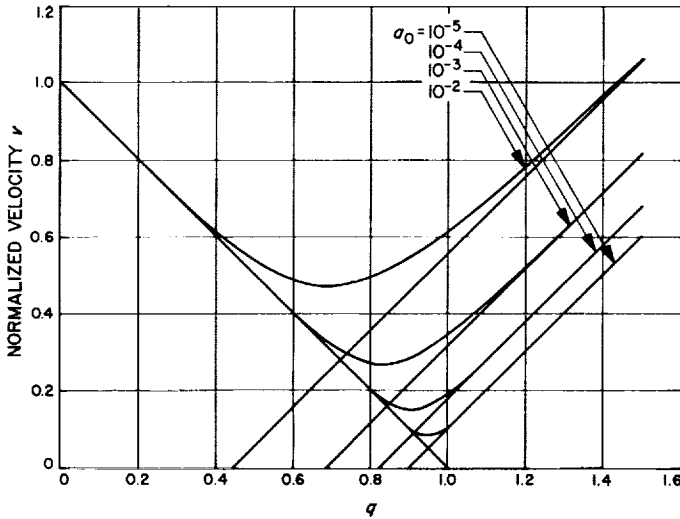
Fig. 10. Normalized velocity v as a function of q for several values of a_0 in normalized units for $\nu = 0$

Since $\nu/8$ is usually small, a first-order series expansion of the exponential term in Eq. (4) provides an approximate expression for q_I explicitly in terms of a_0 and ν which is adequate for most cases.

A justification for the form of Eq. (4) can be obtained by the following analysis: From first principles, the rate of change of the total energy E per unit mass of a tangentially thrusting body in an inverse-square gravity field is given by

$$\frac{dE}{dt} = av = \frac{d}{dt}\left(\frac{1}{2}v^2 - \frac{1}{r}\right), \qquad (5)$$

where the units of E are in terms of the square of the initial velocity. When dv/dt becomes zero, we have

$$r^{-2} = a\left[1 + r^2(d\theta/dr)^2\right]^{1/2}. \qquad (6)$$

In Ref. 7, it is shown that the semimajor axis of the osculating conic associated with the spiral trajectory is given in normalized form by the approximate equation

$$y = (1-q)^{-2}, \qquad (7)$$

and that the normalized central angle ξ is related to y by

$$\frac{dy}{d\xi} = \frac{1}{2}y^3 \exp(\nu q). \qquad (8)$$

If we replace r in Eq. (6) by y (which is an excellent approximation for initially circular low-thrust spirals up to about the last turn before escape), we can use Eqs. (7) and (8) in Eq. (6) to solve for a value, $\tilde{q}$, associated with the point on the spiral where v becomes a minimum. This is given by

$$\tilde{q} = 1 - \left(\frac{4}{3}\right)^{1/8} a_0^{1/4} \exp(\nu\tilde{q}/4)$$

$$= 1 - \left(\frac{4}{3}\right)^{1/8} a_{\tilde{q}}^{1/4} \qquad (9)$$

If we now assume: (1) that $\tilde{q}$ marks the intersection of the straight line asymptotes generated by Eqs. (2) and (3), and (2) that the $v(q)$ curve connecting the asymptotes is symmetric about this point, we can calculate q_I to be

$$q_I = 1 - 2(1 - \tilde{q})$$

$$= 1 - 2\left(\frac{4}{3}\right)^{1/8} a_0^{1/4} \exp\left[\frac{\nu}{8}(q_I + 1)\right]. \qquad (10)$$

We would not expect the coefficient $2(4/3)^{1/8}$ in Eq. (10) to correspond to the computed numerical value (1.757) in Eq. (4) because of the approximations made in the analysis. But, we would expect the form of Eq. (4) to hold for a wide range of values of a_0 and ν. Table 3 shows that this is the case.

3. Parabolic and Hyperbolic Energy Conditions

At this point, a relatively good approximation is available for low-thrust spiral escape from an initial circular orbit. Now, however, consider a tangential, constant thrust–acceleration mission having an initial energy close to zero, where the low-thrust planetocentric phase commences from the normalized coordinates $r = 1$, $\dot{r} = 0$, and $v = (2 + 2E)^{1/2}$. Because of the escape or near-escape conditions, the vehicle rapidly leaves the vicinity

of the planet, the position and velocity initially being little affected by the low-thrust conditions. As the radial distance increases, the magnitude of the retarding acceleration due to the inverse-square gravitational attraction of the central body decreases until, at approximately

$$r = a_0^{-1/2}, \qquad (11)$$

the velocity of the vehicle reaches a minimum of about

$$v_{\min} = (4a_0^{1/2} + 2E)^{1/2}. \qquad (12)$$

At a large radial distance relative to $a_0^{-1/2}$, the velocity approaches an asymptotic form having a rate equal to the thrust acceleration of the vehicle. As discussed previously, the point where the velocity asymptote crosses zero defines an intercept time T_I such that a vehicle departing from a massless planet at this time with zero initial planetocentric velocity would have this asymptotic form as its velocity profile. An example of the

velocity profile for a thrust acceleration of 10^{-4} g and initial parabolic energy is shown in Fig. 11; the corresponding velocity along a parabolic path resulting from a ballistic trajectory is shown by the dashed curve.

An analytic expression which gives the intercept time t_I is not easily determined. However, since the motion of the vehicle becomes nearly radial after a short period of time, the case of pure radial motion, which is amenable to analysis, can be investigated; thus, an approximation of intercept time t_I can be made which, hopefully, is valid for the more realistic model.

The time t along the path for purely radial motion under an outward-directed constant-thrust acceleration is given by

$$t = \int_0^r \frac{dr}{v} = \int_0^r \frac{r^{1/2} \, dr}{(2a_0 \, r^2 + 2E_0 \, r + 2)^{1/2}}, \qquad (13)$$
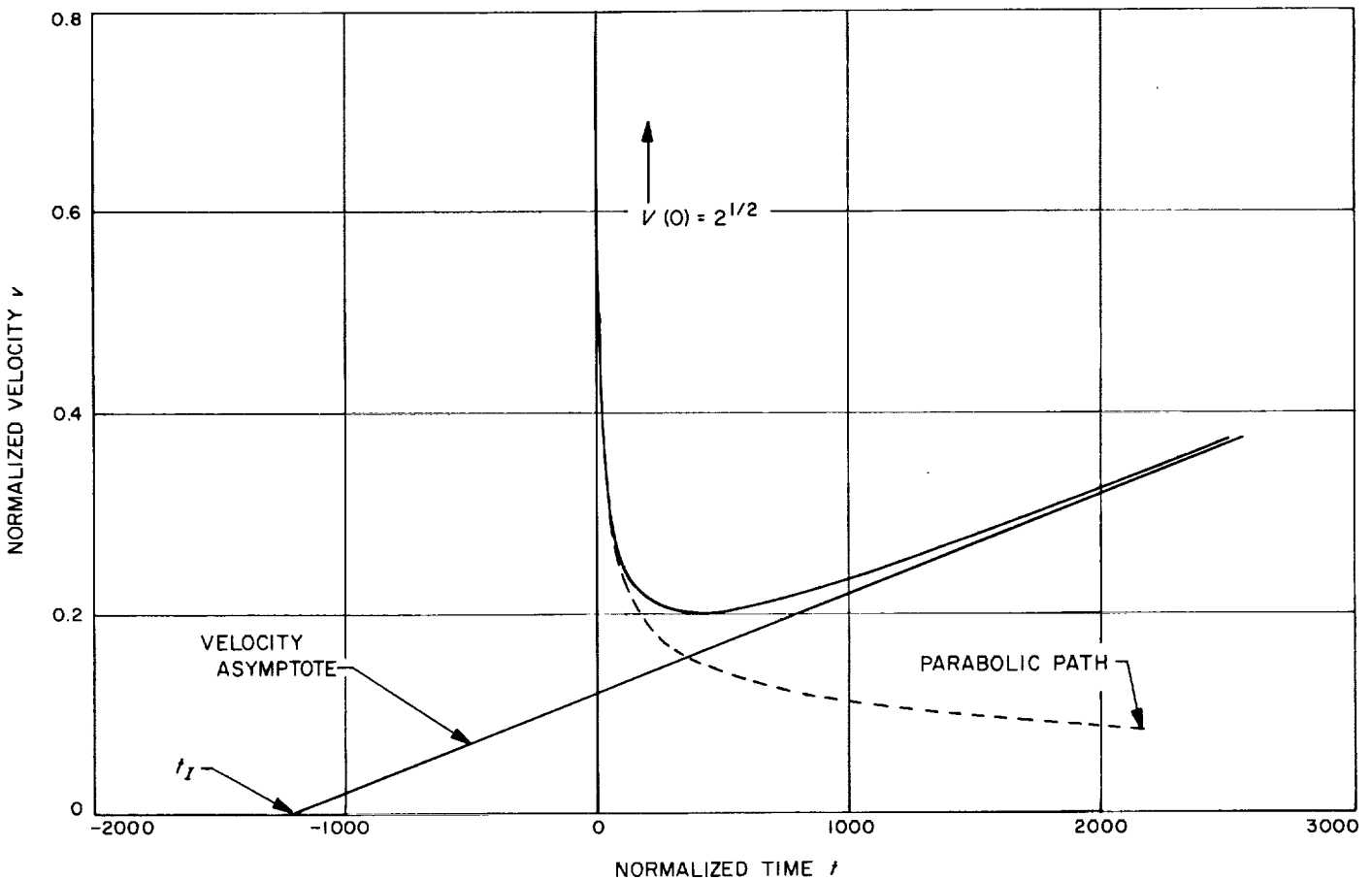
Fig. 11. Normalized velocity v vs normalized time t for $a_0 = 10^{-4}$ g and initial-parabolic-energy conditions

where we have considered the motion to start from $r = 0$ at $t = 0$ with an initial energy E_0. Along the asymptotic velocity path, the position and velocity are given by

$$r = \frac{1}{2} a_0 (t - t_I)^2 - \frac{E_0}{a_0} , \qquad (14a)$$

$$v = a_0 (t - t_I) , \qquad (14b)$$

where t_I is the intercept time with respect to $t = 0$ for purely radial motion. The time $(t - t_I)$ as a function of r is found from Eq. (14a) to be

$$(t - t_I) = \left(\frac{2r}{a_0} + \frac{E_0}{a_0^2} \right)^{1/2} . \qquad (15)$$

The integral in Eq. (13), when integrated, gives a result in terms of elliptic integrals, and the quantity in Eq. (15) can be subtracted to yield the intercept time t_I as a function of radial distance r. Since we are primarily interested in an asymptotic value of intercept time, the radial distance is allowed to go to infinity and a solution for the intercept time t_I is derived which is a function of the initial thrust acceleration a_0 and the initial energy E_0; i.e.,

$$t_I = G(x) a_0^{-3/4} , \qquad (16)$$

where

$$x = \frac{E_0}{2a_0^{1/2}} . \qquad (17)$$

The functional form of $G(x)$ depends on the roots of the quadratic expression appearing in the denominator of Eq. (13). When $-1 < x \leq 1$, the roots are imaginary, and the functional form of $G(x)$ is

$$G(x) = 2^{1/2} (F - 2E) , \qquad -1 < x \leq 1 , \qquad (18)$$

where F and E are complete elliptic integrals of the first and second kinds with modulus k equal to

$$k^2 = \frac{1}{2} (1 - x) . \qquad (19)$$

When $x > 1$, the roots are real and negative, and the form of $G(x)$ is

$$G(x) = - 2^{1/2} \left[x + (x^2 - 1)^{1/2} \right]^{1/2} E , \qquad x \geq 1 , \qquad (20)$$

with inverse modulus k' of the complete elliptic integral E being given by

$$k'^2 = 2x \left[x - (x^2 - 1)^{1/2} \right] - 1 . \qquad (21)$$

A graph of $G(x)$ as a function of x is given in Fig. 12. $G(x)$ is a monotonically decreasing function of x, with a value at $x = 0$ of 1.198140 for initial-parabolic-energy
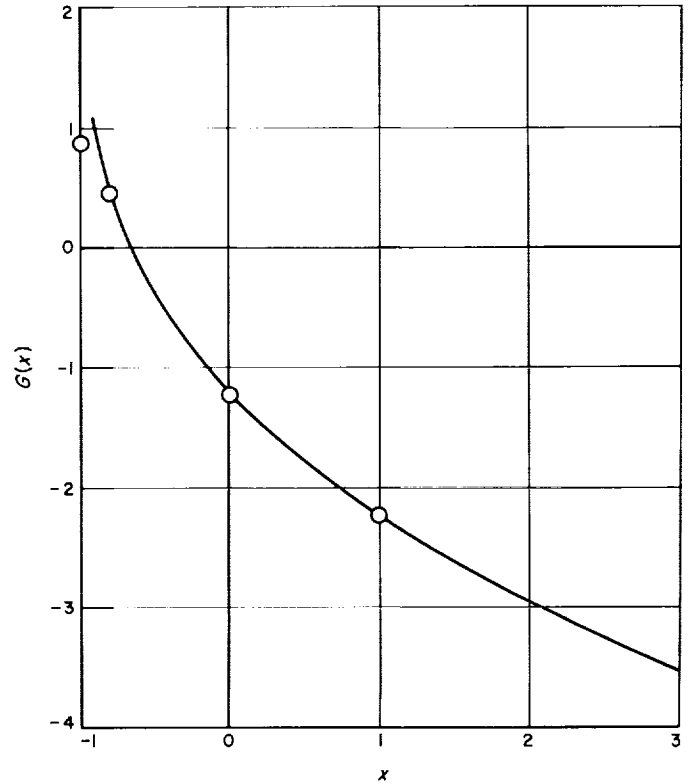
Fig. 12. G(x) vs x

conditions. The form of the function $G(x)$ eventually approaches $2x^{1/2}$ for large values of x. The quantity $a_0 t_I$ (which, for this case, is q_I) calculated from Eq. (16) is seen to be nearly independent of the thrust acceleration for large values of x and is approximately equal to the hyperbolic excess velocity v_∞. Note that, when the value of x is greater than approximately $-2/3$, the value of $G(x)$ is negative such that the intercept time T_I must be subtracted from the heliocentric portion of the trajectory. As an alternative to subtracting T_I, the asymptotic velocity associated with $T_I = 0$ may be used as the initial planetocentric velocity in the heliocentric portion which, in this case, commences at $T_I = 0$. The latter procedure has the advantage of reducing position errors in the matching method, particularly for large v_∞ where T_I is large. On the other hand, this method has the disadvantage of requiring an optimization of the direction of the initial planetocentric velocity in the heliocentric program; also, the method is cumbersome to use when one is optimizing the thrust level employed in the heliocentric phase.

For $x < -1$, the combination of initial energy and thrust acceleration is insufficient to allow direct escape

from the attracting body. In this case, the two roots of the quadratic in Eq. (13) are real and positive, and the velocity reaches zero at a radial distance equal to the smaller root. At this point, the thrust acceleration is less than the gravity acceleration; the motion and thrust acceleration are now reversed, and the radial distance subsequently decreases to zero. Then, if the gain in energy is sufficient to allow $x > -1$, direct escape will occur from the attracting body. If $x < -1$, the process will be repeated until a value of $x > -1$ occurs which will allow direct escape. At a value of $x = -1$, the roots of the quadratic are both real and equal. At a radial distance equal to the roots, the velocity will reach a minimum of zero, with the magnitudes of thrust acceleration and gravity acceleration being exactly equal. The value of $G(x)$ at $x = -1$ is infinity and does not represent a good approximation to motion using a more realistic trajectory model.

An indication of the adequacy of the purely radial case in predicting the value of t_I for a more physically realizable model is obtained by comparing the results from numerically integrated two-dimensional trajectories with those calculated from Eq. (16). The results of several cases with different thrust accelerations and with initial parabolic energy are shown in Table 4. The value of q_I, rather than t_I, is presented, since two of the examples have a finite exhaust velocity and mass loss. The comparison in Table 4 shows a maximum error of 1¾% for the 10^{-2}-g trajectory, the error decreasing at lower thrust acceleration levels. Points for some 10^{-4}-g thrust acceleration trajectories are also shown in Fig. 12 for different values of initial energy. The rectilinear motion assumed in the analysis thus provides a relatively accurate approximation of the more realistic motion for values of x greater than about -0.8.

Table 4. Comparison of exact and calculated q_I values for initial-parabolic-energy conditions

σ_0, g	v	Value of q_I	
		Exact	Calculated from Eq. (16)
10^{-2}	0	-0.3857	-0.3789
10^{-3}	0	-0.2162	-0.2131
10^{-4}	0	-0.1207	-0.1198
10^{-5}	0	-0.0675	-0.0674
10^{-4}	0.2	-0.1209	-0.1198
10^{-4}	-0.2	-0.1205	-0.1198

Finally, when t_I in Eq. (16) for initial parabolic energy is transformed to more conventional units, the intercept time T_I becomes

$$T_I = -1.198140\,\mu^{1/4}A_0^{-3/4} \qquad (22)$$

and is independent of the initial radial distance of the vehicle from the planet during the planetocentric phase.

D. The Covariance Equation in Continuous-Estimation Problems

T. Nishimura

1. Introduction

The covariance equation to be derived in this article corresponds to the one of discrete form used in the current orbit determination program when the continuous estimation method is employed. Continuous estimation implies that a space probe is continuously tracked from the Earth, having the integrated (counted) doppler data $\rho(t)$ which is contaminated by noise, and estimation of orbit parameters is carried out continuously rather than intermittently, as in the current orbit determination program. The equations for continuous estimation are derived in Ref. 8 based on the Gauss-Markov theorem.[1] If the noise is integrated white noise, the modified Kalman's continuous-estimation technique can be applied to the same problem. The covariance equation derived by this method agrees with the one whose derivation is based on the Gauss-Markov theorem.

2. Derivation of Covariance Equation

The observation $\delta\rho(t)$, which is the deviation of the integrated (counted) doppler data from the corresponding ephemeris data, consists of a term due to the initial deviation $\delta\bar{x}$ of the parameters being estimated and a noise term $n(t)$ which takes a form of integrated white noise (Wiener process noise). Thus,

$$\delta\rho(t) = A_1(t)\,\delta\bar{x} + n(t), \qquad (1)$$

[1]Also, Pfeiffer, C. G., *Application of the Continuous Sequential Estimation Technique to Some Orbit Determination Problems*, Interoffice Technical Memorandum No. 312-546, Jet Propulsion Laboratory, Pasadena, California, April 6, 1965.

where $\delta \bar{x}$ is an $(n \times 1)$ column vector when $\bar{x}$ is a parameter vector being estimated, and A_1 is a $(1 \times n)$ row matrix composed of partials of ρ with respect to $\bar{x}$.

$$A_1 = \left[\frac{\partial \rho}{\partial \bar{x}}\right], \qquad (2)$$

and $n(t)$ is described as

$$n(t) = n(0) + \int_0^t w(s)ds, \qquad (3)$$

where $w(t)$ is white noise having a power spectral density Φ_w (two-sided). Then[2]

$$E\left[n(t)\,n(s)\right] = \sigma_0^2 + \Phi_w s, \qquad t \geq s,$$

where σ_0^2 is the mean-squared value of the initial condition $n(0)$.

Since the process being estimated is sequentially correlated, the continuous-estimation theory of Ref. 8 can be applied. According to Ref. 8 and Pfeiffer (Footnote 1), the covariance $\Lambda^*(t)$ associated with the estimate $\delta \bar{x}^*$ of the parameter $\delta \bar{x}$ and $n^*(t)$ of the noise $n(t)$ is defined as

$$\Lambda^*(t) = E\left[\begin{pmatrix} \delta \bar{x}^*(t) - \delta \bar{x} \\ n^*(t) - n(t) \end{pmatrix}(\delta \bar{x}^*(t)' - \delta \bar{x}', n^*(t) - n(t))\right]$$

$$= \left[\begin{array}{c|c} \Lambda_{11}^*(t) & \Lambda_{12}^*(t) \\ \hline \Lambda_{21}^*(t) & \Lambda_{22}^*(t) \end{array}\right]. \qquad (4)$$

Then, it may be obtained as a solution of the following Riccati-type differential equation:

$$\frac{d\Lambda^*}{dt} = F\Lambda^* + \Lambda^*F' - \Lambda^*G\Lambda^* + H, \qquad (5)$$

with the initial condition

$$\Lambda^*(0) = \Lambda(0) - \Lambda(0)\,A'(0)(A(0)\,\Lambda(0)\,A'(0))^{-1}A(0)\,\Lambda(0). \qquad (6)$$

The quantities A and Λ and the coefficient matrices F, G, and H are defined as follows: First, the observation matrix A is obtained by augmenting A_1 to an $(n + 1) \times 1$ row matrix to add a corresponding element for the noise.

$$A = [A_1, 1] \qquad (7)$$

The *a priori* covariance matrix $\Lambda(t)$ is

$$\Lambda(t) = E\left[\begin{pmatrix} \delta \bar{x} \\ n(t) \end{pmatrix}(\delta \bar{x}', n(t))\right] = \left[\begin{array}{c|c} \Lambda_x & 0 \\ \hline 0 & \sigma_0^2 + \Phi_w t \end{array}\right]. \qquad (8)$$

Then, the normalized correlation matrix becomes

$$R(t, s) = E\left[\begin{pmatrix} \delta \bar{x} \\ n(t) \end{pmatrix}(\delta \bar{x}', n(s))\right]\Lambda^{-1}(s) \qquad (9)$$

$$= I,$$

where I is an identity matrix. Hence,

$$\frac{dR}{dt} = \left[\frac{\partial R(t, s)}{\partial t}\right]_{t=s} = 0, \qquad (10)$$

$$S = \frac{d\Lambda}{dt} - \Lambda\frac{dR'}{dt} - \frac{dR}{dt}\Lambda = \frac{d\Lambda}{dt}$$

$$= \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \Phi_w \end{array}\right]. \qquad (11)$$

Also,

$$ASA' = \Phi_w. \qquad (12)$$

The F matrix in Eq. (5) is defined by Pfeiffer (Footnote 1) in the following equation:

$$F = \frac{dR}{dt} - SA'(ASA')^{-1}\left(\frac{dA}{dt} + A\frac{dR}{dt}\right). \qquad (13)$$

This can be computed using the results of Eqs. (8) through (12). Then, it may be partitioned into four blocks as follows:

$$F = \left[\begin{array}{c|c} 0 & 0 \\ \hline -A_1 & 0 \end{array}\right] = \left[\begin{array}{c|c} F_{11} & F_{12} \\ \hline F_{21} & F_{22} \end{array}\right], \qquad (14)$$

where F_{11} is an $(n \times n)$ square matrix, F_{12} is an $(n \times 1)$ column matrix, F_{21} is a $(1 \times n)$ row matrix, and F_{22} is a scalar. Then G in Eq. (5) is computed as

$$G = \left[\frac{dA'}{dt} + \frac{dR'}{dt}A'\right][ASA']^{-1}\left[\frac{dA}{dt} + A\frac{dR}{dt}\right]$$

$$= \frac{1}{\Phi_w}\left(\frac{dA}{dt}\right)'\left(\frac{dA}{dt}\right), \qquad (15)$$

which is partitioned in a way similar to that for F as

$$G = \left[\begin{array}{c|c} \frac{1}{\Phi_w}\dot{A}_1'\dot{A}_1 & 0 \\ \hline 0 & 0 \end{array}\right] = \left[\begin{array}{c|c} G_{11} & G_{12} \\ \hline G_{21} & G_{22} \end{array}\right]. \qquad (16)$$

Finally, H is computed as follows:

$$H = S - SA'(ASA')^{-1}AS$$

$$= 0. \qquad (17)$$

Let $Y(t)$ and $Z(t)$ be $(n \times 1) \times (n \times 1)$ matrices where $Z(t)$ is invertible, and let

$$\Lambda^*(t) = Z^{-1}(t)\,Y(t). \qquad (18)$$

---

[2]Pfeiffer, C. G., A *Simple Comparison of Alternate Ways of Processing Continuously-Counted Doppler Data*, Interoffice Technical Memorandum No. 312-565, Jet Propulsion Laboratory, Pasadena, California, June 29, 1965.

Then, the nonlinear differential Eq. (5) is decomposed to two linear differential equations of Y and Z.

$$\frac{dY}{dt} = YF' + ZH, \qquad (19)$$

$$\frac{dZ}{dt} = -ZF + YG, \qquad (20)$$

with initial conditions

$$Y(0) = \Lambda^*(0), \qquad (21)$$

$$Z(0) = I. \qquad (22)$$

Since H is zero in this particular problem, the above two differential equations are decoupled from each other, and their solutions become obtainable.

First, Y and Z matrices are partitioned as follows:

$$Y = \left[\begin{array}{c|c} Y_{11} & Y_{12} \\ \hline Y_{21} & Y_{22} \end{array}\right], \qquad (23)$$

$$Z = \left[\begin{array}{c|c} Z_{11} & Z_{12} \\ \hline Z_{21} & Z_{22} \end{array}\right], \qquad (24)$$

where Y_{11} and Z_{11} are $(n \times n)$; Y_{12} and Z_{12}, $(n \times 1)$; Y_{21} and Z_{21}, $(1 \times n)$, and Y_{22} and Z_{22}, (1×1) matrices, respectively.

The two linear differential Eqs. (19) and (20) are separated into the following eight linear differential equations of submatrices. Then, they are simplified considerably by taking advantage of many zeros in the F, G, and H matrices.

$$\frac{dY_{11}}{dt} = 0, \qquad (25)$$

$$\frac{dY_{21}}{dt} = 0, \qquad (26)$$

$$\frac{dY_{12}}{dt} = Y_{11}F'_{21}, \qquad (27)$$

$$\frac{dY_{22}}{dt} = Y_{21}F'_{21}, \qquad (28)$$

$$\frac{dZ_{11}}{dt} = -Z_{12}F_{21} + Y_{11}G_{11}, \qquad (29)$$

$$\frac{dZ_{21}}{dt} = -Z_{22}F_{21} + Y_{21}G_{11}, \qquad (30)$$

$$\frac{dZ_{12}}{dt} = 0, \qquad (31)$$

$$\frac{dZ_{22}}{dt} = 0. \qquad (32)$$

These differential equations are easily solved using the initial conditions of Eqs. (21) and (22).

$$Y_{11}(t) = \Lambda^*_{11}(0), \qquad (33)$$

$$Y_{21}(t) = \Lambda^*_{21}(0), \qquad (34)$$

$$Y_{12}(t) = \Lambda^*_{12}(0) + \Lambda^*_{11}(0) \int_0^t F'_{21}(s)ds, \qquad (35)$$

$$Y_{22}(t) = \Lambda^*_{22}(0) + \Lambda^*_{21}(0) \int_0^t F'_{21}(s)ds, \qquad (36)$$

$$Z_{11}(t) = I + \Lambda^*_{11}(0) \int_0^t G_{11}(s)ds, \qquad (37)$$

$$Z_{21}(t) = -\int_0^t F_{21}(s)ds + \Lambda^{21}_*(0) \int_0^t G_{11}(s)ds, \qquad (38)$$

$$Z_{12}(t) = 0, \qquad (39)$$

$$Z_{22}(t) = 1. \qquad (40)$$

As a result, the Z matrix is now described as

$$Z = \left[\begin{array}{c|c} Z_{11} & 0 \\ \hline Z_{21} & 1 \end{array}\right]. \qquad (41)$$

The inverse of Z becomes

$$Z^{-1} = \left[\begin{array}{c|c} Z_{11}^{-1} & 0 \\ \hline \bar{Z}_{21} & 1 \end{array}\right], \qquad (42)$$

where $\bar{Z}_{21}$ represents the first n elements in the last row of Z^{-1}.

Then, the submatrices of Λ^* are derived as

$$\Lambda^*_{11}(t) = Z_{11}^{-1}(t) \Lambda^*_{11}(0), \qquad (43)$$

$$\Lambda^*_{12}(t) = Z_{11}^{-1}(t) Y_{12}(t), \qquad (44)$$

$$\Lambda^*_{21}(t) = \bar{Z}_{21}(t) \Lambda^*_{11}(0) + \Lambda^*_{21}(0), \qquad (45)$$

$$\Lambda^*_{22}(t) = \bar{Z}_{21}(t) Y_{12}(t) + Y_{22}(t). \qquad (46)$$

In particular, $\Lambda^*_{11}(t)$ is rewritten as in the following equation by substituting Eq. (37) into Eq. (43):

$$\Lambda^*_{11}(t) = \left[I + \Lambda^*_{11}(0) \int_0^t G_{11}(s)ds \right]^{-1} \Lambda^*_{11}(0)$$

$$= \left[\Lambda^*_{11}(0)^{-1} + \frac{1}{\Phi_w} \int_0^t \dot{A}'_1(s) \dot{A}_1(s)ds \right]^{-1}. \qquad (47)$$

This is the continuous form of the error covariance matrix corresponding to the one used in the current orbit determination program.

Simple manipulation of Eq. (6) yields the initial condition $\Lambda_{11}^*(0)$ as

$$\Lambda_{11}^*(0) = \Lambda_{11}(0) - \Lambda_{11}(0) A_1'(0)$$
$$\times \left[A_1(0) \Lambda_{11}(0) A_1'(0) + \sigma_0^2\right]^{-1} A_1(0) \Lambda_{11}(0) .$$
(48)

3. Derivation from Kalman's Formula

In the case of Kalman's continuous-estimation formula (Ref. 9), which is derived from the orthogonal projection theorem, it is essential that the observation noise is white. Since the observation noise is integrated white noise in this example, it is not possible to apply Kalman's formula directly to the problem. However, the extended method by Bryson and Johansen (Ref. 10) makes it possible to treat the problem because the derivative of data $\delta\dot{\rho}$ contains a white noise w. According to their method, the actual differentiation of data is unnecessary, but an intermediate state is introduced which consists of derivatives of the observation containing white noise. Then, it follows Kalman's method and finally eliminates the requirement of differentiation by absorbing it into the integrators of the Kalman filter. The final results are presented here.

The covariance matrix of $\delta\bar{x}^*$ is denoted as P in Refs. 9 and 10.

$$P(t) = E\left[(\delta\bar{x}^* - \delta\bar{x})(\delta\bar{x}^* - \delta\bar{x})'\right] .$$
(49)

Then, $P(t)$ can be obtained as a solution of the following Riccati-type differential equation:

$$\frac{dP}{dt} = F_{11}(t)P(t) + P(t)F_{11}'(t)$$
$$- P(t)\dot{A}_1'(t) R_w^{-1}(t)\dot{A}_1(t)P(t) ,$$
(50)

where $F_{11}(t)$ is an $(n \times n)$ matrix describing the transition of the system. It is equal to zero in this case because $\delta\bar{x}$ is constant.

$$F_{11}(t) = 0 .$$
(51)

Furthermore, $R_w(t)$ is the weight of the covariance matrix of the observation noise. Hence,

$$R_w(t) = \Phi_w .$$
(52)

Resorting to a similar decomposition method as that in the previous case, the nonlinear differential Eq. (50) can easily be solved and the following result can be derived:

$$P(t) = \left[P(0^+)^{-1} + \frac{1}{\Phi_w}\int_0^t \dot{A}_1'(s) \dot{A}_1(s) ds\right]^{-1} , \quad (53)$$

where $P(0^+)$, which is different from $P(0)$ in general, is given in Example 1 of Ref. 10 as

$$P(0^+) = P(0) + P(0) A_1'(0)$$
$$\times \left[A_1(0)P(0)A_1'(0) + \sigma_0^2\right]^{-1} A_1(0) P(0) .$$
(54)

Eq. (53) derived by the modified Kalman's method exactly agrees with the result of Eq. (47) because, by definition,

$$\Lambda_{11}^*(t) = P(t), \quad t > 0 , \tag{55}$$

$$\Lambda_{11}(0) = P(0) , \tag{56}$$

and the initial condition $\Lambda_{11}^*(0)$ in Eq. (48) corresponds to $P(0^+)$ of the above equation. The result of Eq. (53) is analogous to the one derived by the author[3] for the discrete-estimation problem for Kalman's discrete-estimation formula.

The discontinuity in the covariance equation between $t = 0$ $(\Lambda_{11}(0), P(0))$ and $t = 0^+$ $(\Lambda_{11}^*(0), P(0^+))$ should be noted. This results because exact measurements are carried out on the augmented states $(\delta\bar{x}', n(t))$ at the instant the observation $\delta\rho$ becomes available, and Λ_{11} (or P) is updated from $t = 0$ to $t = 0^+$.

---

[3]Nishimura, T., *On the Solution of Covariance Difference Equation by Means of Linear Decomposition and z-Transform*, Interoffice Technical Memorandum No. 312-595, Jet Propulsion Laboratory, Pasadena, California, September 1965.

References

1. Brouwer, D., and Clemence, G. M., *Methods of Celestial Mechanics,* Academic Press, New York, 1961.

2. Kliore, A. J., Cain, D. L., Levy, G. S., Eshleman, V. R., Fjeldbo, G., and Drake, F. D., *Science,* Vol. 149, p. 1243, 1965.

3. Peabody, P. R., Scott, J. F., and Orozco, E. G., *JPL Ephemeris Tapes E9510, E9511, and E9512,* Technical Memorandum No. 33-167, Jet Propulsion Laboratory, Pasadena, California, March 2, 1964.

4. Devine, C. J., *PLOD II: Planetary Orbit Determination Program for the IBM 7094 Computer,* Technical Memorandum No. 33-188, Jet Propulsion Laboratory, Pasadena, California, April 15, 1965.

5. *Physics and Astronomy of the Moon,* Kopal, Z. (Ed.), Academic Press, New York, 1962.

6. Fimple, W. R., and Edelbaum, T. N., *Applications of SNAP-50 Class Powerplants to Selected Unmanned Electric Propulsion Missions,"* American Institute of Aeronautics and Astronautics Paper No. 64-494, 1964.

7. Melbourne, W. G., *Interplanetary Trajectories and Payload Capabilities of Advanced Propulsion Vehicles,* Technical Report No. 32-68, Jet Propulsion Laboratory, Pasadena, California, 1961.

8. Pfeiffer, C. G., *Sequential Estimation of Correlated Stochastic Variables,* Technical Report No. 32-445, Jet Propulsion Laboratory, Pasadena, California, July 1, 1963.

9. Kalman, R. E., and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," *ASME Transactions, Journal of Basic Engineering,* Vol. 83D, pp. 95–108, March 1961.

10. Bryson, A. E., and Johansen, D. E., "Linear Filtering for Time-Varying Systems Using Measurements Containing Colored Noise," *IEEE Transactions on Automatic Control,* Vol. AC-10, No. 1, pp. 4–10, January 1965.

Erratum

The following should be noted for *SPS 37-35,* Vol. IV: In computing the trajectory results given on pp. 12–22, erroneous ephemeris data were used for the outer planets Saturn and Uranus. The errors in celestial longitude — approximately 30 deg for Saturn and 10 deg for Uranus — naturally affect the plots given in the figures involving these planets; however, it is believed that these errors do not affect the conclusions concerning the feasibility of these missions, but merely change the optimum launch years and the bending angles at the intermediate planets. Corrected data will be published in a future issue of the *SPS,* Vol. IV.

II. Scientific Programming

A. Numerical Integration of Planet Positions When Considering the Effects of the Earth and Moon Separately

D. Dunham

The derivation of the differential equation for the motion of the Earth–Moon barycenter and the method for computing the acceleration in the JPL planetary orbit determination program PLOD II (Ref. 1) are presented herein. The subscripts used and the quantities they denote are as follows:

0	center of inertial reference frame
1	center of Sun
2	center of Earth
3	center of Moon
B	center of mass of Earth–Moon system, i.e., Earth–Moon barycenter
j	center of Mercury, Venus, Mars, or one of the five outer planets

Consider the n-body problem in an inertial reference frame. The equation of motion of the ith body, according to Newton's law of gravitation, is as follows:

$$\ddot{\mathbf{r}}_{0i} = G\sum_{\substack{k=1 \\ k \neq i}}^{(n)} m_k \frac{\mathbf{r}_{ik}}{r_{ik}^3} \qquad (1)$$

where $\mathbf{r}_{ik}$ is the vector from the ith body or point to the kth body or point, r_{ik} is the magnitude of $\mathbf{r}_{ik}$, G is the universal constant of gravitation, and m_k is the mass of the kth body. Using this equation, the case of the solar system can be considered, and the inertial-frame equations of motion of the Sun, Moon, and Earth can be written; however, care must be taken with the algebraic signs.

For the Sun (mass S), Earth (mass E), and Moon (mass M), respectively,

$$\ddot{\mathbf{r}}_{01} = GE\frac{\mathbf{r}_{12}}{r_{12}^3} + GM\frac{\mathbf{r}_{13}}{r_{13}^3} + G\sum_j P_j \frac{\mathbf{r}_{1j}}{r_{1j}^3} \qquad (2)$$

$$\ddot{\mathbf{r}}_{02} = -GS\frac{\mathbf{r}_{12}}{r_{12}^3} + GM\frac{\mathbf{r}_{23}}{r_{23}^3} + G\sum_j P_j \frac{\mathbf{P}_{2j}}{P_{2j}^3} \qquad (3)$$

$$\ddot{\mathbf{r}}_{03} = -GS\frac{\mathbf{r}_{13}}{r_{13}^3} - GE\frac{\mathbf{r}_{23}}{r_{23}^3} + G\sum_j P_j \frac{\mathbf{r}_{3j}}{r_{3j}^3}, \qquad (4)$$

where P_j represents the mass of the planet (either Mercury, Venus, Mars, or one of the five outer planets) at j. The vector $\mathbf{r}_{23}$ will be used throughout this discussion since it is available directly from the lunar theory; $\mathbf{r}_{32}$ will not be used.

For PLOD, we need the equations of motion of the Earth and Moon in a heliocentric frame. The equation for translating to the heliocentric frame for the kth body is as follows:

$$\mathbf{r}_{1k} = \mathbf{r}_{0k} - \mathbf{r}_{01} \,. \tag{5}$$

Differentiating Eq. (5) twice yields the following:

$$\ddot{\mathbf{r}}_{1k} = \ddot{\mathbf{r}}_{0k} - \ddot{\mathbf{r}}_{01} \,. \tag{6}$$

Letting $k = 2$ for the Earth and 3 for the Moon, and using Eqs. (6), (2), (3), and (4), we obtain the desired equations. The heliocentric differential equations for the Earth and Moon are, respectively,

$$\ddot{\mathbf{r}}_{12} = \ddot{\mathbf{r}}_{02} - \ddot{\mathbf{r}}_{01} = - G(S+E)\frac{\mathbf{r}_{12}}{r_{12}^3}$$
$$+ GM\left(\frac{\mathbf{r}_{23}}{r_{23}^3} - \frac{\mathbf{r}_{13}}{r_{13}^3}\right) + G\sum_j P_j\left(\frac{\mathbf{r}_{2j}}{r_{2j}^3} - \frac{\mathbf{r}_{1j}}{r_{1j}^3}\right)$$
$$\tag{7}$$

$$\ddot{\mathbf{r}}_{13} = \ddot{\mathbf{r}}_{03} - \ddot{\mathbf{r}}_{01} = - G(S+M)\frac{\mathbf{r}_{13}}{r_{13}^3}$$
$$- GE\left(\frac{\mathbf{r}_{23}}{r_{23}^3} + \frac{\mathbf{r}_{12}}{r_{12}^3}\right) + G\sum_j P_j\left(\frac{\mathbf{r}_{3j}}{r_{3j}^3} - \frac{\mathbf{r}_{1j}}{r_{1j}^3}\right)$$
$$\tag{8}$$

The forms of Eqs. (7) and (8) can be recognized as simply those of the heliocentric equation for the Earth, perturbed by the j planets and by an additional "planet," the Moon, and the equation for the Moon, perturbed by the j planets and the Earth, respectively. It is undesirable to integrate Eqs. (7) and (8) directly since both contain the term $\mathbf{r}_{23}/r_{23}^3$. This lunar theory term has too short a period and too large an amplitude (about half as large as the solar term in the case of the Moon) to be considered a perturbation. Therefore, Eqs. (7) and (8) should be combined to eliminate these terms. This can be done by investigating the motion of the Earth–Moon barycenter, the heliocentric coordinate of which is given below:

$$\mathbf{r}_{1B} = \frac{E}{E+M}\mathbf{r}_{12} + \frac{M}{E+M}\mathbf{r}_{13} \,. \tag{9}$$

Differentiating Eq. (9) twice, we obtain the acceleration of the barycenter:

$$\ddot{\mathbf{r}}_{1B} = \frac{E}{E+M}\ddot{\mathbf{r}}_{12} + \frac{M}{E+M}\ddot{\mathbf{r}}_{13} \tag{10}$$

If the expressions for $\ddot{\mathbf{r}}_{12}$ and $\ddot{\mathbf{r}}_{13}$ in Eqs. (7) and (8) are substituted into Eq. (10), the troublesome lunar theory terms cancel.

$$\ddot{\mathbf{r}}_{1B} = - \frac{GE(S+E)}{E+M}\frac{\mathbf{r}_{12}}{r_{12}^3} + \frac{GEM}{E+M}\frac{\mathbf{r}_{23}}{r_{23}^3} - \frac{GEM}{E+M}\frac{\mathbf{r}_{13}}{r_{13}^3}$$
$$+ \frac{GE}{E+M}\sum_j P_j\left(\frac{\mathbf{r}_{2j}}{r_{2j}^3} - \frac{\mathbf{r}_{1j}}{r_{1j}^3}\right) - \frac{GM(S+M)}{E+M}\frac{\mathbf{r}_{13}}{r_{13}^3} \tag{11}$$
$$- \frac{GEM}{E+M}\frac{\mathbf{r}_{23}}{r_{23}^3} - \frac{GEM}{E+M}\frac{\mathbf{r}_{12}}{r_{12}^3} + \frac{GM}{E+M}\sum_j P_j\left(\frac{\mathbf{r}_{3j}}{r_{3j}^3} - \frac{\mathbf{r}_{1j}}{r_{1j}^3}\right)$$

The $[GEM/(E+M)](\mathbf{r}_{23}/r_{23}^3)$ terms are the ones which cancel. After some manipulation, Eq. (11) can be reduced to the following form:

$$\ddot{\mathbf{r}}_{1B} = - G(S+E+M)\left(\frac{E}{E+M}\frac{\mathbf{r}_{12}}{r_{12}^3} + \frac{M}{E+M}\frac{\mathbf{r}_{13}}{r_{13}^3}\right)$$
$$+ G\sum_j P_j\left(\frac{E}{E+M}\frac{\mathbf{r}_{2j}}{r_{2j}^3} + \frac{M}{E+M}\frac{\mathbf{r}_{3j}}{r_{3j}^3} - \frac{\mathbf{r}_{1j}}{r_{1j}^3}\right). \tag{12}$$

Eq. (12) is now used by PLOD. It is important to realize that, although Eq. (12) is an equation for the motion of the barycenter, the masses of the Moon and Earth are *not* considered to be at that point (as in the old version of PLOD). To the contrary, Eq. (12) rigorously takes into account the effects of the Earth and Moon separately in their true positions.

In PLOD, the motion of $\mathbf{r}_{1B}$ is integrated. But, the expressions on the right-hand side of Eq. (12) require the heliocentric coordinates of the Earth $(\mathbf{r}_{12})$ and the Moon $(\mathbf{r}_{13})$. These are obtained as follows:

$$\mathbf{r}_{12} = \mathbf{r}_{1B} - \frac{M}{E+M}\mathbf{r}_{23} \,, \tag{13}$$

$$\mathbf{r}_{13} = \mathbf{r}_{1B} + \frac{E}{E+M}\mathbf{r}_{23} \,, \tag{14}$$

where $\mathbf{r}_{1B}$ is taken from the integration, and $\mathbf{r}_{23}$ is obtained from the lunar theory (coordinates of the Moon with respect to the Earth).

The effects of the Earth and Moon are separated when integrating the positions of all the other planets (Mercury, Venus, Mars, and the outer planets) by treating the Earth and Moon as two separate perturbing planets.

Therefore, Eq. (7) of Ref. 1, p. 3, is used where the summation is over nine "planets": the Earth, the Moon, and seven others (excluding the perturbed planets whose position is being integrated from the group of Mercury, Venus, Mars, and the five outer planets). The heliocentric coordinates of the Earth and Moon are obtained directly from a subroutine which computes them using Eqs. (13) and (14), with both r_{1B} and r_{23} being obtained from JPL ephemeris tape E9510.

B. Ephemerides of the Earth–Moon Barycenter, Venus, and Mars When Considering the Earth and Moon as Separate Bodies

C. J. Devine

1. Introduction

Currently, the JPL ephemeris tapes E9510, E9511, and E9512 contain JPL planetary orbit determination program PLOD II-generated position–velocity ephemerides of the planets (with the exception of Mercury and Neptune) which were constructed by considering the Earth–Moon system as a point mass located at the barycenter of the Earth–Moon system (Refs. 1–3); i.e., the following second-order nonlinear differential equation was solved by the numerical integration method of PLOD II:

$$\ddot{X}^i = -k^2(1+M)\frac{X^i}{r^3} + \sum_j k^2 m_j \left(\frac{X_j^i - X^i}{P_j^3} - \frac{X_j^i}{(r_j)^3} \right), \quad (1)$$

where

$$i = 1, 2, 3,$$

$$j = 1, 2, \cdots, 8,$$

$$(r)^2 = \sum_{i=1}^{3} (X^i)^2,$$

$$(r_j)^2 = \sum_{i=1}^{3} (X_j^i)^2,$$

$$(P_j)^2 = \sum_{i=1}^{3} (X_j^i - X^i)^2,$$

M is the mass of the planet integrated, m_j are the masses of the other perturbing planets, and X_j^i are the positions of the perturbing planets. In the integration of the Earth–Moon barycenter, M would equal the mass of the Earth

plus that of the Moon; for any other planet, M would equal the mass of that planet, and lower case m_3 would equal the mass of the Earth plus that of the Moon; i.e., the planet would be perturbed by the barycenter approximation.

A more precise model of the motion of the Earth–Moon barycenter is obtained by considering the Earth and Moon as separate bodies. The derivation of differential equations to represent this model is reported in the preceding article (Section A). This model was considered, but not adopted, in the production of the JPL ephemeris. More recent interest in the possibility of planetary radar experiments as tests of general relativity and the need for more accurate ephemerides for spacecraft data analysis have renewed interest in the refinement of this model. This article will give the results of a study of this separation effect on the ephemerides of the Earth–Moon barycenter, Venus, and Mars for the period from December 30, 1949, to August 11, 1968 (Julian dates 2 433 280.5 to 2 440 048.5).

2. Method

Given a source ephemeris of the body in question, a unique solution of a differential equation describing the orbit of the body is obtained by fitting to the source ephemeris, in a least-squares sense, a numerical solution of the differential equation. If the differential equations of two least-squares fits to the same source ephemeris differ, the effect of this difference on the generated ephemerides may be seen by plotting the difference between the two fitted ephemerides. An alternative method would be to fit one model of the differential equation to an ephemeris generated by a different model. Both of these methods were used in the present study and were found to give the same results to the accuracy appropriate to the study. Here, a PLOD A fit is defined as a least-squares fit to a source ephemeris using Eq. (1) as the differential equation model, and a PLOD B fit is defined as a least-squares fit to a source ephemeris using the differential equation model of the preceding article (Section A).

3. Results

a. Earth–Moon barycenter. Using PLOD II, PLOD A and PLOD B fits were made to the current source ephemeris of the Earth–Moon barycenter. The maximum and minimum values of the residuals of these two runs may be seen in Table 1, and plots of the residuals are presented in Figs. 1 and 2. Quantities plotted were δx, δy, δz, the residuals in position in the source-minus-

Table 1. Maximum, minimum residual values for PLOD fits

PLOD Fit	Multiplication factor for all values	δx, AU	δy, AU	δz, AU	$\delta \dot{x}$, AU	$\delta \dot{y}$, AU	$\delta \dot{z}$, AU	$\delta R/R$	$(\cos\beta)\delta\lambda$, rad	$\delta\beta$, rad
PLOD A to Earth–Moon barycenter source ephemeris (Fig. 1)	10^{-7}	4.5, −5.2	5.0, −5.4	3.3, −3.0	0.12, −0.13	0.16, −0.14	0.070, −0.083	1.9, −4.3	4.9, −5.9	2.5, −2.0
PLOD B to Earth–Moon barycenter source ephemeris (Fig. 2)	10^{-7}	4.1, −4.9	5.1, −5.4	3.4, −2.8	0.13, −0.12	0.16, −0.14	0.086, −0.091	1.4, −4.5	4.5, −6.0	2.4, −2.2
PLOD A to Earth–Moon barycenter source ephemeris minus PLOD B to Earth–Moon barycenter source ephemeris (Fig. 3)	10^{-3}	9.8, −10.0	7.8, −8.3	9.4, −9.5	0.12, −0.12	0.18, −0.17	0.17, −0.18	—	—	—
PLOD B to PLOD A for Earth–Moon barycenter (Fig. 4)	10^{-8}	9.8, −10.0	7.8, −8.1	9.5, −9.4	0.11, −0.12	0.17, −0.17	0.17, −0.18	2.9, −6.7	9.4, −9.8	9.3, −9.3
PLOD A to Venus source ephemeris	10^{-7}	4.5, −6.0	3.7, −3.8	2.9, −2.4	0.17, −0.15	0.15, −0.14	0.077, −0.091	2.8, −3.3	6.5, −8.6	3.8, −3.2
PLOD B to PLOD A for Venus	10^{-11}	3.9, −3.8	3.1, −4.0	1.7, −2.3	0.110, −0.072	0.085, −0.085	0.05, −0.06	2.0, −1.8	6.0, −6.1	2.8, −2.9
PLOD B to PLOD A for Mars	10^{-11}	5.1, −3.6	4.8, −4.2	2.2, −2.1	0.034, −0.055	0.040, −0.042	0.017, −0.017	0.92, −1.30	4.2, −3.4	0.40, −0.41
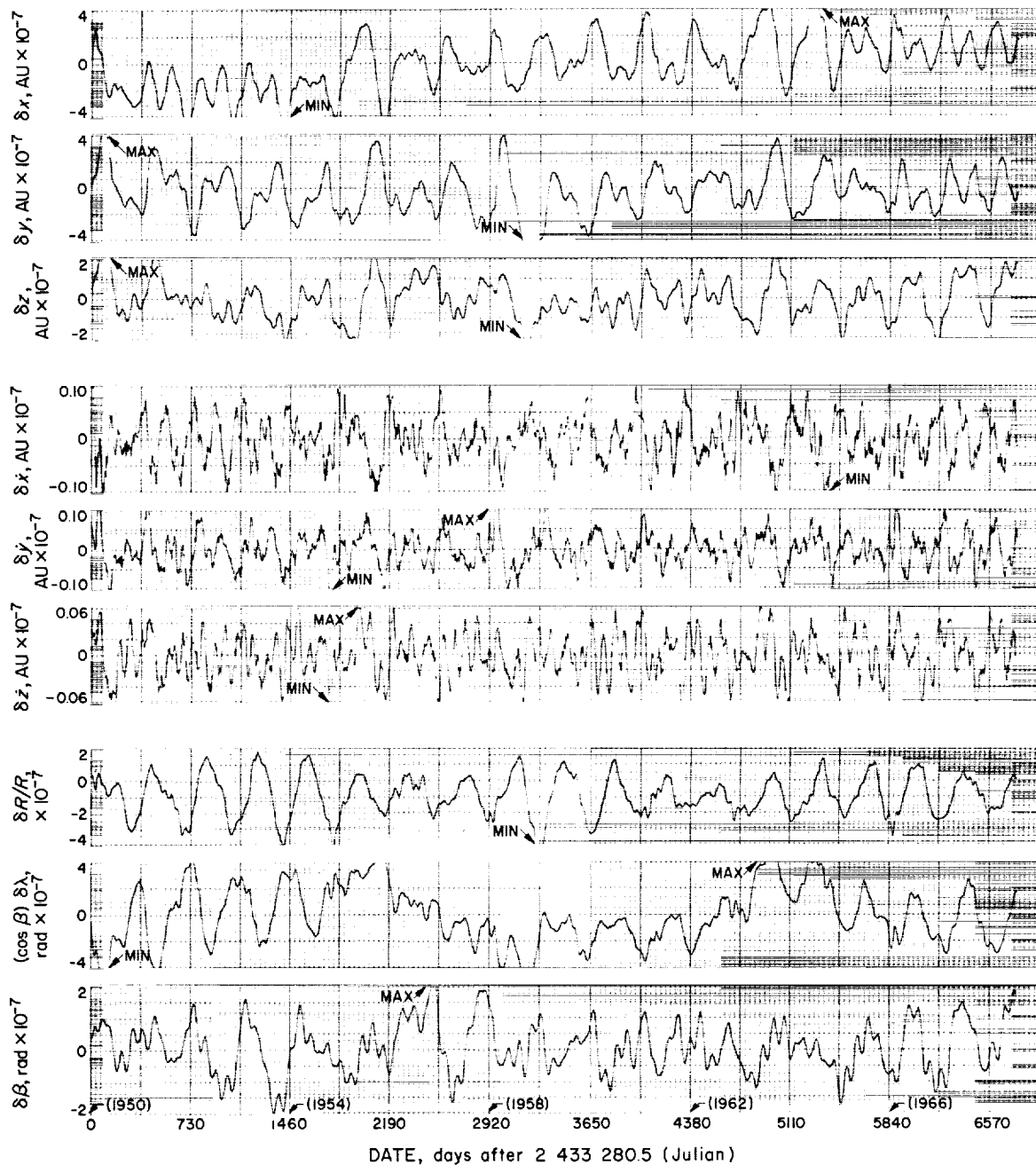
Fig. 1. Residual plots for PLOD A fit to Earth–Moon barycenter source ephemeris
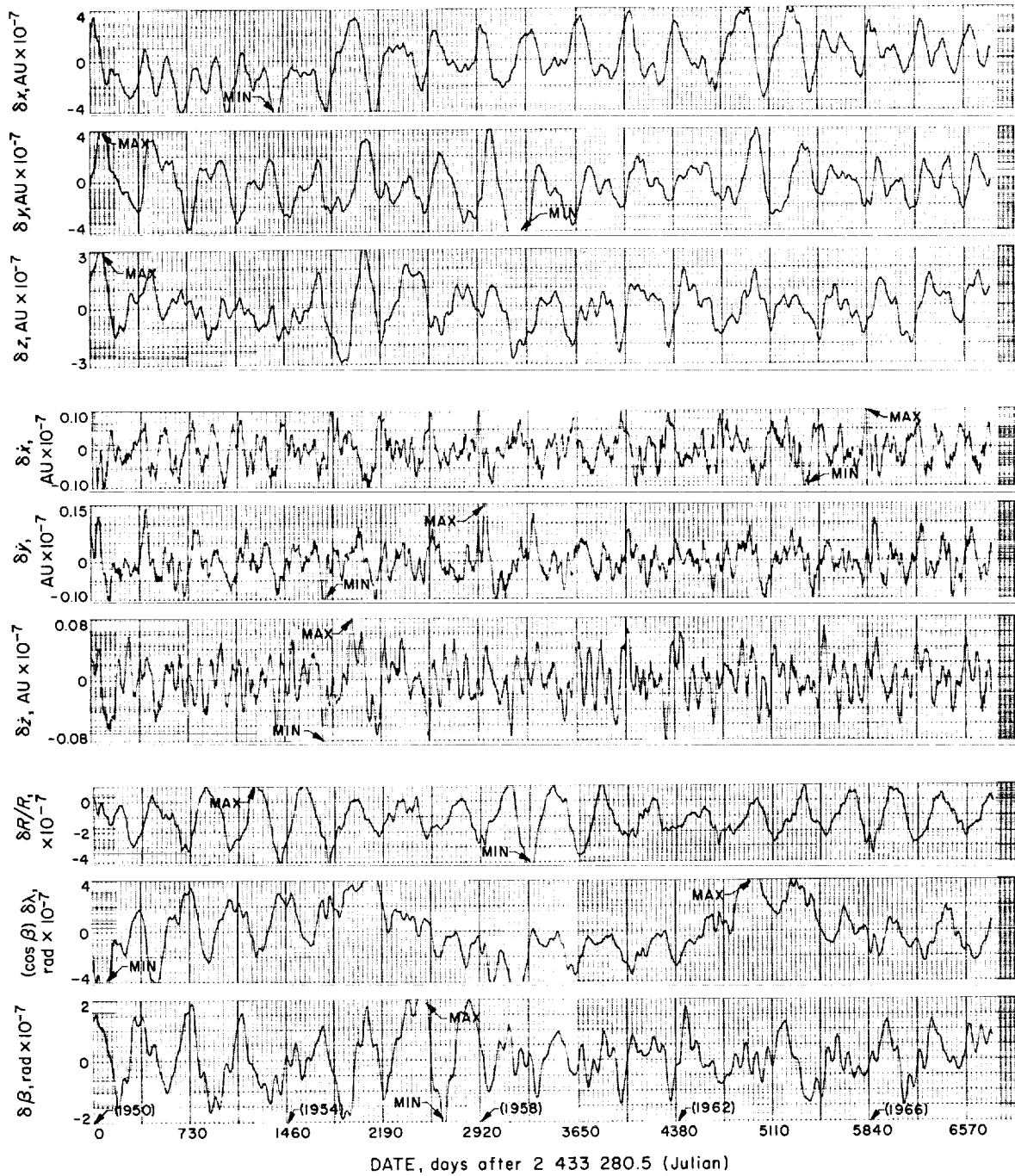
Fig. 2. Residual plots for PLOD B fit to Earth—Moon barycenter source ephemeris

computed sense; $\delta\dot{x}$, $\delta\dot{y}$, $\delta\dot{z}$, the residuals in velocity; $\delta R/R$, the relative residual in the radius vector; $(\cos\beta)\,\delta\lambda$, cosine of the latitude times the residual in longitude; and $\delta\beta$, the latitude residual.

The PLOD A fitted ephemeris minus the PLOD B fitted ephemeris was then plotted to indicate the effect of the separation of the Earth and Moon (Table 1 and Fig. 3). In order to conserve machine time in computation and to obtain more information, it was decided to use the alternative method of obtaining the separation effect as described above; i.e., after a PLOD A fit has been made to a source ephemeris, a PLOD B fit is made to the PLOD A fit and the residuals are plotted. In the

case of the Earth–Moon barycenter, this gave essentially the same results, as can be seen in Table 1 and by comparing Fig. 3 with Fig. 4. The significant fact to be noted in Fig. 4 is that a mean of the residuals $\delta R/R$ gives a value of approximately -2×10^{-8}, as was predicted by Hill in 1878 (Ref. 4) in indicating the principal deviation of the Sun's orbit from a Keplerian ellipse. This value should also be compared with the results reported in Section I, A. *Effect of Separated Earth–Moon Masses on the Integrated Ephemeris of the Earth–Moon Barycenter*, pp. 7–14, of this SPS.

Also of interest is the fact that the residuals between a PLOD A or a PLOD B run and the source ephemeris
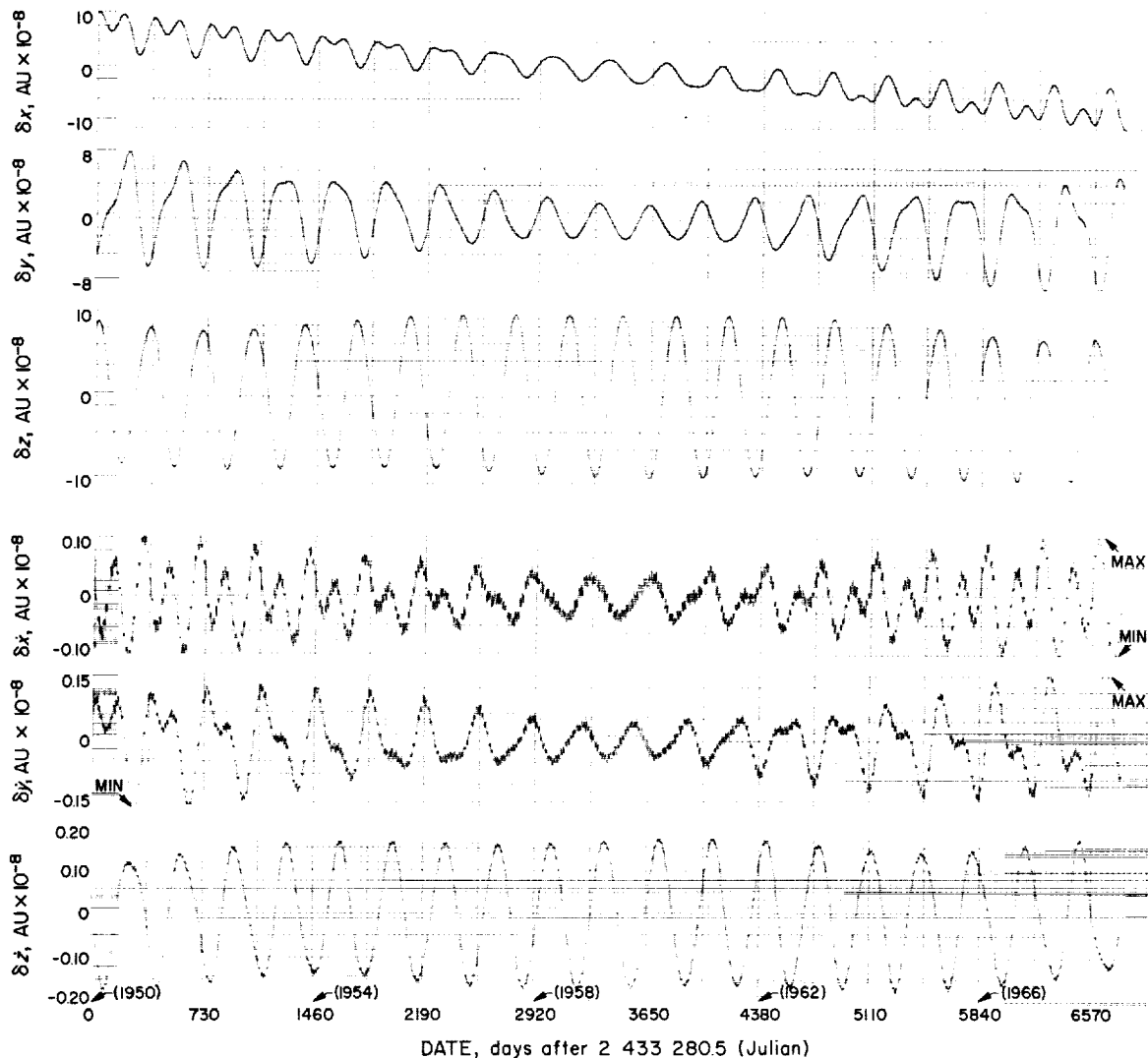
Fig. 3. Residual plots for PLOD A fit to Earth–Moon barycenter source ephemeris minus PLOD B fit to Earth–Moon barycenter source ephemeris
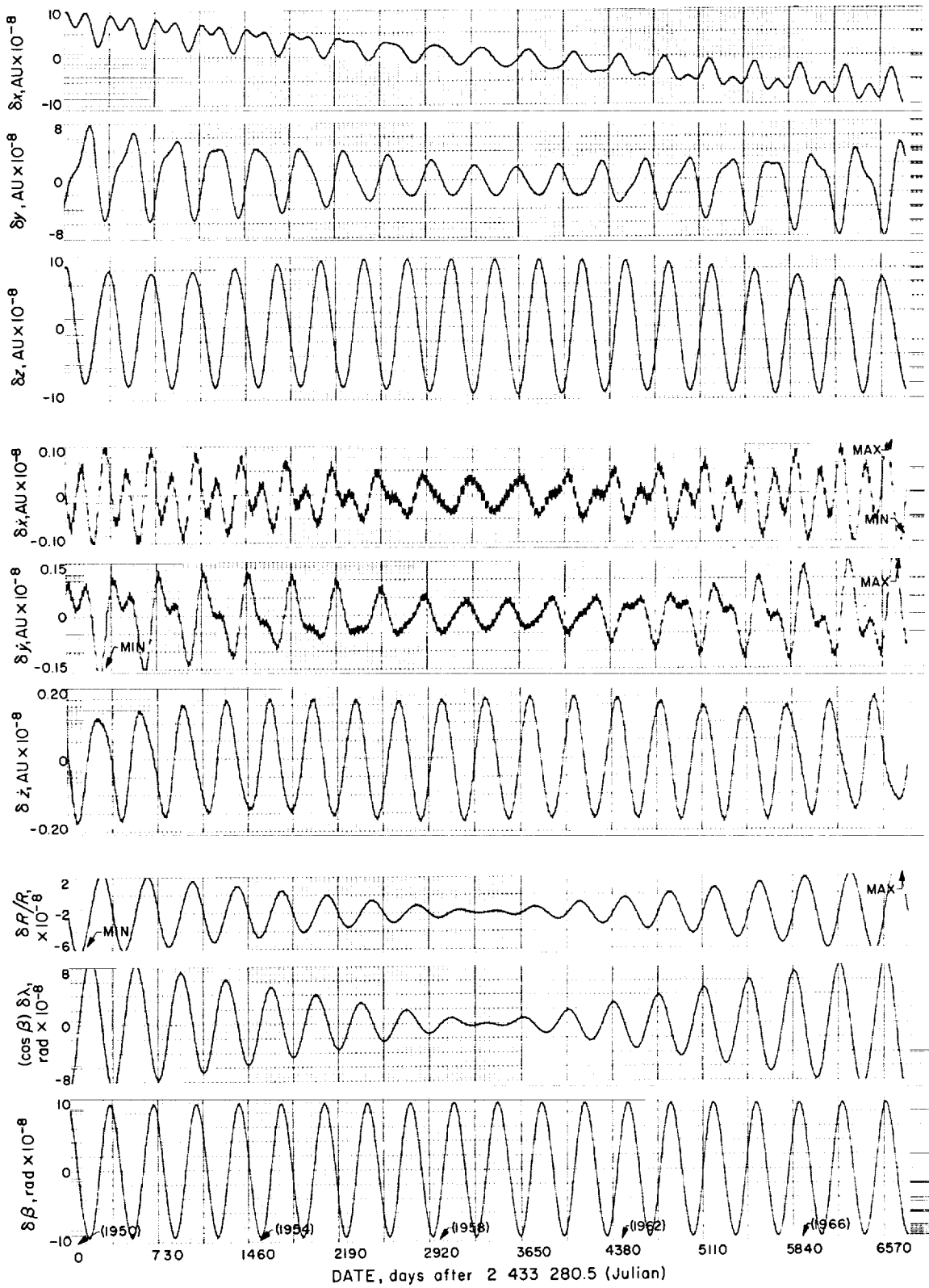
Fig. 4. Residual plots for PLOD B fit to PLOD A fit for Earth–Moon barycenter

are larger than the residuals attributed to the different models of the differential equation used (i.e., PLOD A minus PLOD B). Hence, it appears that the errors in the source ephemeris are larger than the difference in the models being studied here.

b. Venus. Again using PLOD II, a PLOD A fit was made to a current Venus source ephemeris, obtaining the maximum and minimum residuals given in Table 1.[1] Then, a PLOD A fit was made to the PLOD B fit and the residuals were again plotted (Table 1), indicating a goodness of fit well above the accuracy currently required by ephemeris users. Therefore, the effect of the separation of the Earth and Moon on Venus is negligible when considering the present requirements of the ephemeris system.

c. Mars. A PLOD B fit was made to a previous PLOD A fit to a Mars source ephemeris, and the maximum and minimum values of the residuals obtained were approximately the same order of magnitude or even smaller than those of Venus (Table 1).[1] Hence, the separation effect of the Earth and Moon on Mars is very small and thus also negligible when considering the present requirements of the ephemeris system.

C. FORTRAN IV and MAP Subroutines to Digitally Filter and Compute the One-Side Power Spectral Density of a Time-Sampled Function

W. Silsby

Subroutines, written in FORTRAN IV and MAP, to digitally filter and compute the one-side power spectral density of a time-sampled function are described herein. For this usage, special input and output subroutines must be supplied in order to handle data communication. The main emphasis here is on the use and logic of these subroutines. Equations are presented without proof or amplification.

---

[1]Plots of the residuals of the PLOD fits for Venus and Mars will be presented by Devine, C. J., and Dunham, D., in *The Ephemerides of the Earth–Moon Barycenter, Venus, and Mars, Considering the Earth and Moon as Separate Bodies,* Technical Memorandum No. 33-232, Jet Propulsion Laboratory, Pasadena, California (To be published).

1. Power Spectral Density Formulas

The power spectral density (PSD) of a function $X(t)$ is defined as the Fourier transform of the autocorrelation function of $X(t)$ (Ref. 5). The autocorrelation is

$$C(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} X(t) X(t + \tau) \, dt .$$

The PSD is then

$$P(f) = \int_{-\infty}^{\infty} C(\tau) \exp\left(-i2\pi f\tau\right) d\tau .$$

Since $C(\tau)$ is even, this reduces to

$$P(f) = 2 \int_{0}^{\infty} C(\tau) \cos 2\pi f\tau d\tau .$$

In practice, $X(t)$ is time-sampled at $t_0, t_1, \cdots, t_m$, where $t_i - t_{i-1} = t_{i+1} - t_i$ for all i and $2T = t_m - t_0$. Thus, $X(t)$ can be denoted by the sequence $X_0, X_1, \cdots, X_m$. The autocorrelation is approximated numerically by

$$\phi_i = \frac{1}{n-i} \sum_{j=0}^{n-i} X_j X_{j+i}, \, i = 0, M .$$

Where M is the lag number. The PSD function is calculated at $M + 1$ frequencies.

$$f_k = \frac{kf_c}{M}, \, k = 0, 1, \cdots, M .$$

The numerical formula for this calculation is

$$P_k = P\left(\frac{kf_c}{M}\right) = \Delta t \left[\phi_0 + 2 \sum_{i=1}^{M-1} \phi_i \cos \frac{\pi i k}{M} + (-1)^k \phi_M \right].$$

These values can be "smoothed" as follows:

$$\hat{P}_0 = \tfrac{1}{2}(P_0 + P_1) ,$$

$$\hat{P}_k = \tfrac{1}{4}(P_{k-1} + 2P_k + P_{k+1}), \, k = 1, M-1 ,$$

$$\hat{P}_M = \tfrac{1}{2}(P_{M-1} + P_M) .$$

The sample rate is $s = 1/\Delta t$, where $\Delta t = t_i - t_{i-1}$. The function $X(t)$ must have a maximum frequency $f_c = s/2$. It is impossible to determine any frequency if there are not at least two samples for each cycle of the frequency. A frequency that is sampled at less than two samples per cycle looks like a lower frequency. It may be seen below that a function with a frequency of 4 cps looks like it has a frequency of 4/3 cps when sampled at a rate of 8/3 samples/sec. This effect is called aliasing.

|← 1/4 sec →|
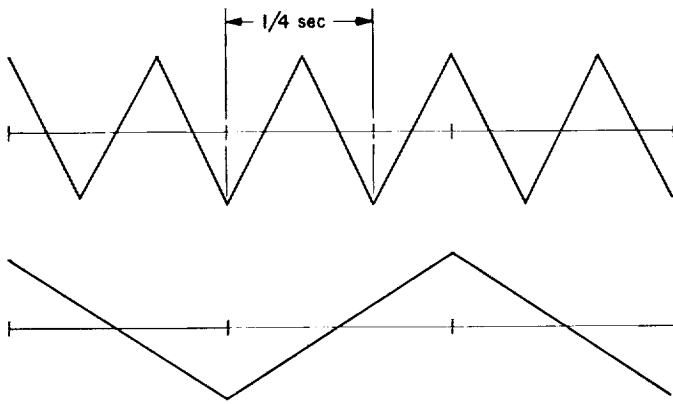
2. Digital Filter Formulas

The basic formula to digitally filter a time-sampled function $X_0, X_1, \cdots, X_m$ is (Ref. 6):

$$Y_j = \sum_{i=0}^{k} A_i \left(X_{j+i} + X_{j-i} \right), \qquad (1)$$

where Y_j is the filtered function defined for $j = k, m-k$ and $A_0, A_1, \cdots, A_k$ are called filter coefficients. The shape and accuracy of the filter determine the values and the number of the coefficients.

The filter shape is defined by a set of points (f_i, g_i) shown in Fig. 5. Note that $g_m = 0$. Going from $i = m-1$ to $i = 0$, a sequence of numbers $a_1, a_2, \cdots, a_m$ is determined from

$$g_i = \sum_{j=i+1}^{m} a_j \left(1 - \frac{f_i}{f_j} \right).$$

The coefficients are then directly computed from

$$A_0 = \frac{s}{2} \sum_{j=1}^{m} a_j f_j,$$

$$A_u = \frac{2}{(2\pi u)^2 s} \sum_{j=1}^{n} \frac{a_j}{f_j} \left(1 - \cos 2\pi f_j s u \right),$$
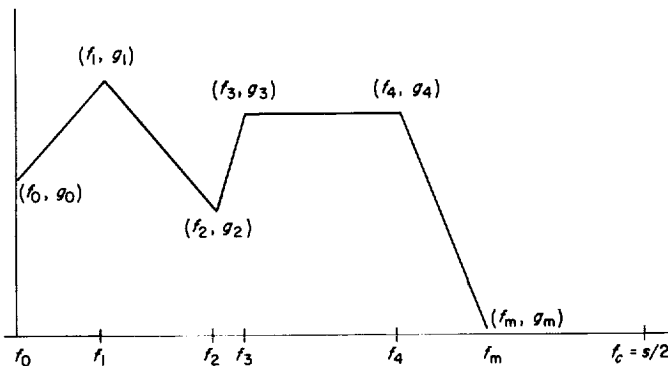
Fig. 5. Set of points (f_i, g_i) defining filter shape

where s denotes the rate at which $X_0, X_1, \cdots, X_m$ was sampled.

A convenient error criteria was developed at JPL. Let

$$Q = \frac{s}{3} \sum_{j=1}^{m} (f_j - f_{j-1}) \left(g_{j-1}^2 + g_{j-1} g_j + g_i^2 \right).$$

Then,

$$E = \left[\frac{Q - \left(2A_0^2 + \sum_{j=1}^{k} A_j^2 \right)}{2A_0^2 + \sum_{j=1}^{n} A_j^2} \right]^{1/2}. \qquad (2)$$

If the filter is placed on the frequency intervals $[f_1, f_2]$, $[f_3, f_4], \cdots, [f_{l-1}, f_l]$, and if $P(f)$ denotes the unfiltered PSD while $P'(f)$ denotes the filtered PSD, then error E is interpreted as

$$E = \frac{\displaystyle\int_{f_1}^{f_2} |P(f) - P'(f)| \, df + \cdots + \int_{f_{l-1}}^{f_l} |P(f) - P'(f)| \, df}{\displaystyle\int_{f_1}^{f_2} P(f) df + \cdots + \int_{f_{l-1}}^{f_l} P(f) df}$$

3. Use of SPEC Subroutine

The subroutine SPEC performs the calculation for the PSD described in Section 1. The calling sequence for SPEC is:

CALL SPEC (WORK, IORK, MEAN, IBT, ISK, S, AUTO, M, POW)	
WORK	area used by the program for computation purposes
IORK	size of WORK
MEAN	mean or sample mean in fixed point of the sampled function X_i $i = 0, n$
IBT	absolute value of the largest X_i
ISK	every ISK point to be processed
S	sample rate
AUTO	array where autocorrelation function is stored
M	number of lagged products, [restriction on $M : IORK > 4 (M + 1)$]
POW	array where the $M + 1$ spectral estimates are stored (also for use by the subroutine to read in the sampled function X_i $i = 0, n$)

The user must supply a routine for use by the subroutine SPEC to obtain the sampled function X_i $i = 0, n$. The calling sequence is:

CALL TRANS (POW, NO, IFLAG)	
POW, NO	On return from TRANS, the subroutine SPEC expects $|NO|$ fixed-point values of X_i at B35 in the array POW. This array is the same as the one that will contain the spectral estimates on return from PSD. If NO is negative, all X values have been supplied.
IFLAG	The first time SPECT calls TRANS, IFLAG will be zero. TRANS will be called until NO < 0. Each time, SPECT will expect the next $|NO|$ X values.

4. Use of Filter Subroutines

Two subroutines are described here: (1) subroutine COEF, which calculates filter coefficients A_0, A_1, $\cdots$, A_k, using the formulas described in Section 2, and (2) subroutine FILTER, which performs the basic filter calculation, Eq. (1). The calling sequence to COEF is:

CALL COEF (G, F, A, ABUF, L1, K1, SAMP, EPS)	
G	array where L1 g_i values are stored; $g_i = G(I + 1)$
F	array where L1 frequencies f_i are stored; $f_i = F(I + 1)$
A	array where, on exit, K1 filter coefficients are stored; $A_i = A(I + 1)$
ABUF	array of length L1 used by the subroutine
L1	number of frequencies given. Note L1 = $m + 1$
K1	on entrance, maximum number of filter coefficients desired; on exit, number of filter coefficients computed. Computation ends when the error E [Eq.. (2)] is less than EPS or when K1 coefficients have been calculated. Note K1 = $k + 1$ [Eq. (1)].
SAMP	sample rate
EPS	on entrance, error tolerance desired; on exit, error tolerance achieved

The calling sequence to FILTER is:

CALL FILTER (Y, NUMA, NUMB, A, X, MM, L, K1, IDEL)	
X, MM	array X: a buffer of size MM. The subroutine I0, provided by the user and described below, reads the sampled function into this array.
A, K1	K1 filter coefficients supplied by A array. Note K1 = $k + 1$ [Eq. (1)] and 2K1 ≤ MM − 1.
IDEL	every IDEL point to be filtered
Y	array used by the subroutine to form the filtered points from each block of X values provided through the subroutine IO. Its size should be at least as large as any MM/IDEL + 2.
NUMA	array used by the subroutine of dimension as large as that of the Y array
NUMB, L	two-dimensional array used by the subroutine of size (L, 2), where L is at least as large as any K1/IDEL + 2

Two user-provided input routines are needed. To read *floating point* X values, the subroutine uses:

CALL IO (IOF, X)

When FILTER first calls IO, IOF (1) = 0; on subsequent calls, IOF (1) = 1. The number of X values provided by the subroutine is in IOF (2). If IOF (2) is less than MM, the subroutine assumes that it has received the last block of values. If IDEL is very large and MM is relatively small, FILTER must skip blocks of input data. It expects to skip J blocks of size MM by:

CALL IO1 (J)

Two output routines must be provided for this subroutine also. When filtering of a point is completed, the subroutine will:

CALL COV (Y(I))

where Y(I) is the filtered point in floating point form. When all points provided have been filtered, it will:

CALL COVL
RETURN

As stated above, when less than MM sampled points are supplied through the subroutine IO, FILTER assumes they are the last points of the sample.

5. Use of Filters

High resolution of low frequencies when higher frequencies are present can be obtained by a low-pass filter. Suppose it is desired to examine frequencies up to 500 cps when there exists frequencies up to 5000 cps with a sample rate of 10,000 samples/sec. The user would then place the filter at 500 cps and process every tenth point. Resolution of 1 cps is obtained using 500 lags, but, without the low-pass filter, 1-cps resolution would require 5000 lags.

High resolution can be obtained for high frequencies by the use of a band-pass filter. Suppose high resolution is desired between frequencies f_a and f_b, where $f_b - f_a > f_a$. All frequencies outside the interval $[f_a, f_b]$ are filtered. Also, points are skipped to achieve a lower sampling rate; thus, the interval $[f_a, f_b]$ is aliased down to a lower frequency interval $[\hat{f_a}, \hat{f_b}]$. Since $\hat{f_b}$ is considered the highest frequency present, high resolution can be obtained with fewer lags.

Consider a function with frequency f and period P sampled every Δt sec:

$$F(n\,\Delta t) = \sin(2\pi fn\,\Delta t).$$

Since aliasing occurs when Δt is greater than $P/2$, Δt can be expressed as

$$\Delta t = (M + \epsilon)P,$$

where M is an integer and $\epsilon < 1$.
Substituting for Δt,

$$\sin[2\pi fn(MP + \epsilon P)] = \sin[2\pi fn\,MP + 2\pi fn\,\epsilon P]$$

$$= \sin[2\pi nM + 2\pi n\epsilon]$$

$$= \sin[2\pi n\epsilon].$$

Dividing and multiplying the argument by Δt gives

$$\sin\left[\frac{2\pi n\epsilon\,\Delta t}{\Delta t}\right] = \sin[2\pi \hat{f} n\,\Delta t],$$

where $\hat{f} = \epsilon/\Delta t$ is the new frequency. But,

$$\Delta t = (M + \epsilon)P$$

or

$$\epsilon = \frac{\Delta t}{P} - M.$$

Substituting for ϵ and setting $f = 1/P$ gives

$$\hat{f} = f - \frac{M}{\Delta t}.$$

The frequencies f_a, f_b that bound the high-frequency interval must satisfy the following relationships:

$$f_a\Delta t - M \le 0, \qquad f_b\Delta t - M \le 0.5.$$

Thus, the ϵ associated with f_a must be ≥ 0, and the ϵ associated with f_b must be ≤ 0.5.

As an example, suppose a function is sampled at 1000 cps. A 1-cps resolution is desired for the interval [500,550]. Normally, this would require 1000 lag products. If 10 points are skipped, Δt becomes 0.01 $M = (0.01)\,500 = 5$. Thus, the new interval is [0,50], and 1-cps resolution can be obtained with 50 lags.

PSD plots. To demonstrate the effects of the use of filters, plots of PSD for vibrational test data and for recorded telemetry data from a spacecraft are shown in Figs. 6 and 7, respectively. Data for both figures are given in Table 2.

It seems obvious that filtering is efficient only when more than a few points are skipped. This is demonstrated in Fig. 6. If the number of lags was to be increased by a factor of 4, a 0.5-cps resolution would be achieved for the interval [0,400]. The computer time would be increased to about 100 sec. This compares favorably with the 72-sec total computer time using the filter. The example of the spacecraft telemetry data (Fig. 7) is quite clear. Using a small amount of computer time, high resolution around the principal frequency can be obtained by using the filter.
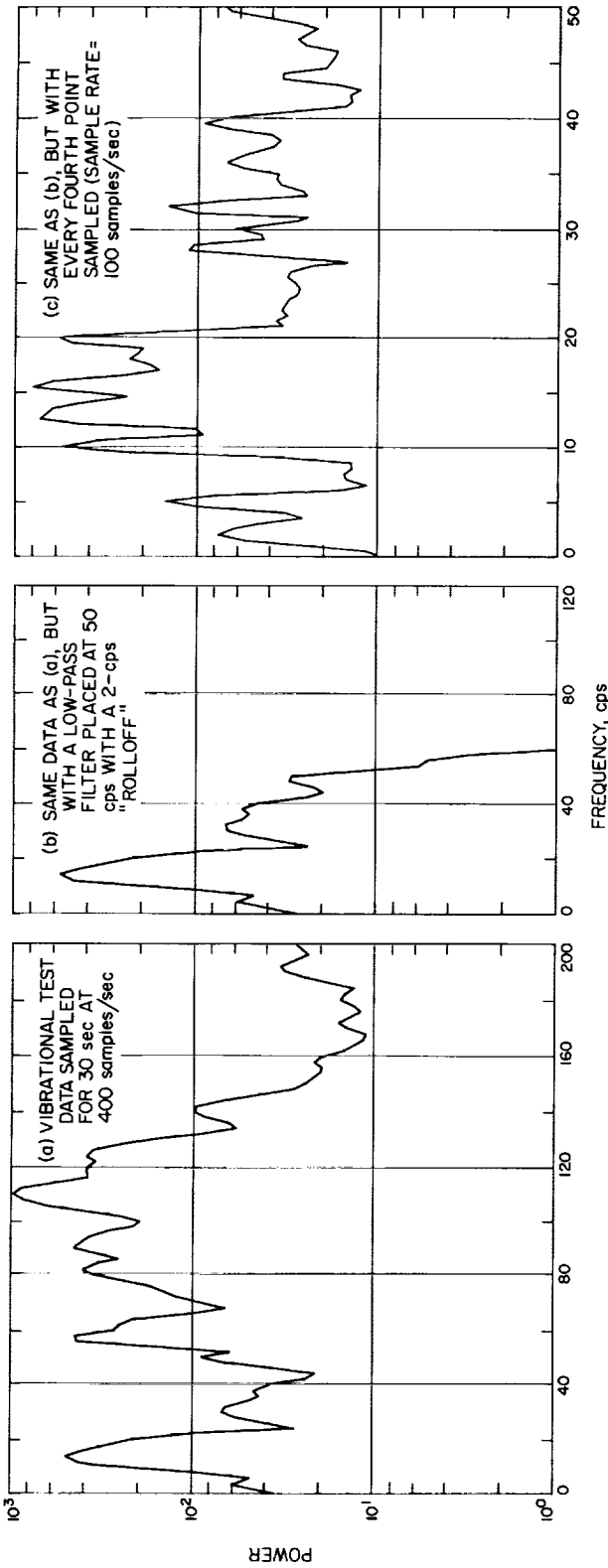
Fig. 6. PSD plots for vibrational test data
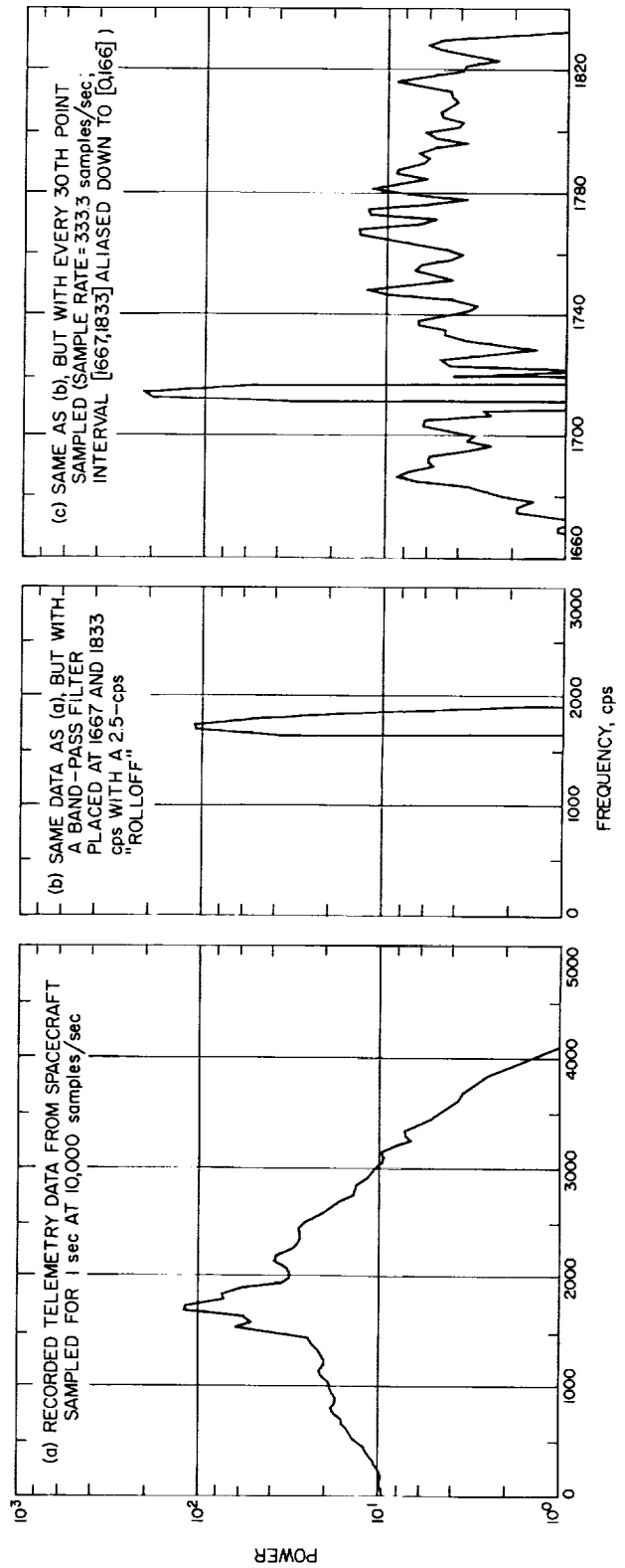
Fig. 7. PSD plots for recorded telemetry data from a spacecraft

Table 2. Effects on PSD data of using filters and skipping points

Item	Fig. 6			Fig. 7		
	(a)	(b)	(c)	(a)	(b)	(c)
Relative power error	—	0.02	0.02	—	0.02	0.02
Number of filter coefficents	—	19	19	—	263	263
Number of points skipped	0	0	3	0	0	29
Effective sample rate, samples/sec	400	400	100	10,000	10,000	333.3
Number of lags	100	100	100	100	100	100
Resolution, cps	2	2	0.5	5	5	1.66
Computer time for filter, sec	—	72	72	—	300	11
Computer time for PSD, sec	26	26	9	41	41	5

6. Coding Logic for SPECT Subroutine

The main portion of this subroutine was coded in the FAP language. It consisted of a collection of subroutines that computed the lagged products. To convert to MAP, these subroutines were meshed into one large closed subroutine and were linked together by TSX commands, i.e., open subroutines. The SPECT subroutine calls the closed subroutine TRANS, supplied by the user and described in Section 3, and a FORTRAN subroutine POWR, that computes the PSD from the lagged products. The calling sequences for the open subroutines are:

```
TSX    SETFUL, 4
TSX    L(WORK)
TSX    C(IORK)
TSX    L(MEAN)
TSX    C(M)
TSX    L(AUTO)
TSX    C(ISK)
TSX    C(IBT)

TSX    CODFUL, 4
TSX    L(POW)
TSX    C(NO)

TSX    ENDFUL, 4
```

The names of the above variables are consistent with those defined in Section 3. On return from ENDFUL, the following floating point information is available: the number of X values processed, in the AC; the true mean, in the MQ; and the lagged products, in the array AUTO.

SETFUL does the preparation for the lagged product computation. It is entered once. Each time CODFUL is

entered, it expects the next NO fixed-point X values in the array POW. ENDFUL is entered after the last X values have been given to CODFUL. After ENDFUL, SPEC enters the closed subroutine POWR, which computes the PSD (Fig. 8).

As CODFUL is supplied the X values, it stores them in the array WORK. Computation begins when IQ = IORK − 2(M + 1) values have been stored. As much is done to form the M + 1 lagged products as is possible with these values; then, the next block of values is read into WORK and computation continues. To "bridge" the computation between blocks, some of the previous X values must be left in core. This is done through the 2(M + 1) locations left in WORK. Thus, in one pass through the data, all lagged products are computed.

CODFUL uses IBT, the absolute value of the largest X value, to compute MBIT, the maximum number of bits for any X. The basic loop for the computation of the Jth lagged product is:

```
         STZ    SUM
         AXT    C(L), 1
FL01     LDQ    X + C(L), 1
         VLM    X + C(J) + C(L), 1, C(MBIT)
         LLS    C(MBIT)
         ADD    SUM
         TOV    OUT
         STO    SUM
         TIX    FL01, 1, 1
```

The fixed-point number in SUM is floated and added to the partially formed J^{th} lagged product computation. Transfer to OUT does the same, but returns control to the loop.

The calling sequence for POWR is:

CALL POWR (AUTO, POW, TRMN, S, M, WORK)

AUTO	on entrance, the M + 1 lagged products; on exit, the autocorrelation
POW, M	array containing the M + 1 spectral estimates
TRMN	total number of observations
S	sample rate
WORK	array used by POWR as a table of cosine values generated in the subroutine

POWR is written in FORTRAN IV. However, it makes two entries to a closed MAP subroutine that generates a table of cosines used to compute the spectral estimates. The entries are:

CALL COSTBL (PIM, WORK)	
PIM	floating point number equaling $\pi/\text{FLOAT(M)}$
WORK	array used to store cosine table

and

CALL COSX (FRG, COXX)	
FRG, COXX	COXX = COS (FRG * PIM)

COSTBL is called once to set up the table.

7. Coding Logic for Filter Subroutine

A flow chart of the filter subroutine is given in Fig. 9. The subroutine expects a buffer of size MM into which the samples of the unfiltered function X_i are to be read. In general, the number of samples, n, will be greater than size MM, and the subroutine must compute the filtered samples Y_i from consecutive blocks of X values of length MM. If this be the case, when computing Y_i, if X_i is within $k + 1$ values of the first X of the block or within $k + 1$ values of the last X of the block, then Y_i must use X values from two blocks. Fig. 10 pictorially describes the calculation of Y_i when two blocks are needed.

The basic filter Eq. (1) can be rewritten

$$Y_i = \sum_{j=\alpha+1}^{k} A_j X_{i+j} + \sum_{j=0}^{\alpha} A_j (X_{i+j} + X_{i-j}) + \sum_{j=\alpha+1}^{k} A_j X_{i-j}$$
(3)

or

$$Y_i = S_1 + S_2 + S_3 .$$

Suppose X_i is in Block J and X_{i+j} $\alpha < j \leq k$ is in Block J + 1 (Fig. 10a). Since MM, the size of the blocks, is greater than k, all X_{i-j} values are in Block J. The calculation for S_3 is called "initial calculation for current Y values" (ICCY). The calculation for S_2, called the "one block calculation" (OBC), is made when Block J is in core. When Block J + 1 is read in, the calculation for S_1,

Fig. 8. Flow diagram of SET values from calling sequence to CALL POWR
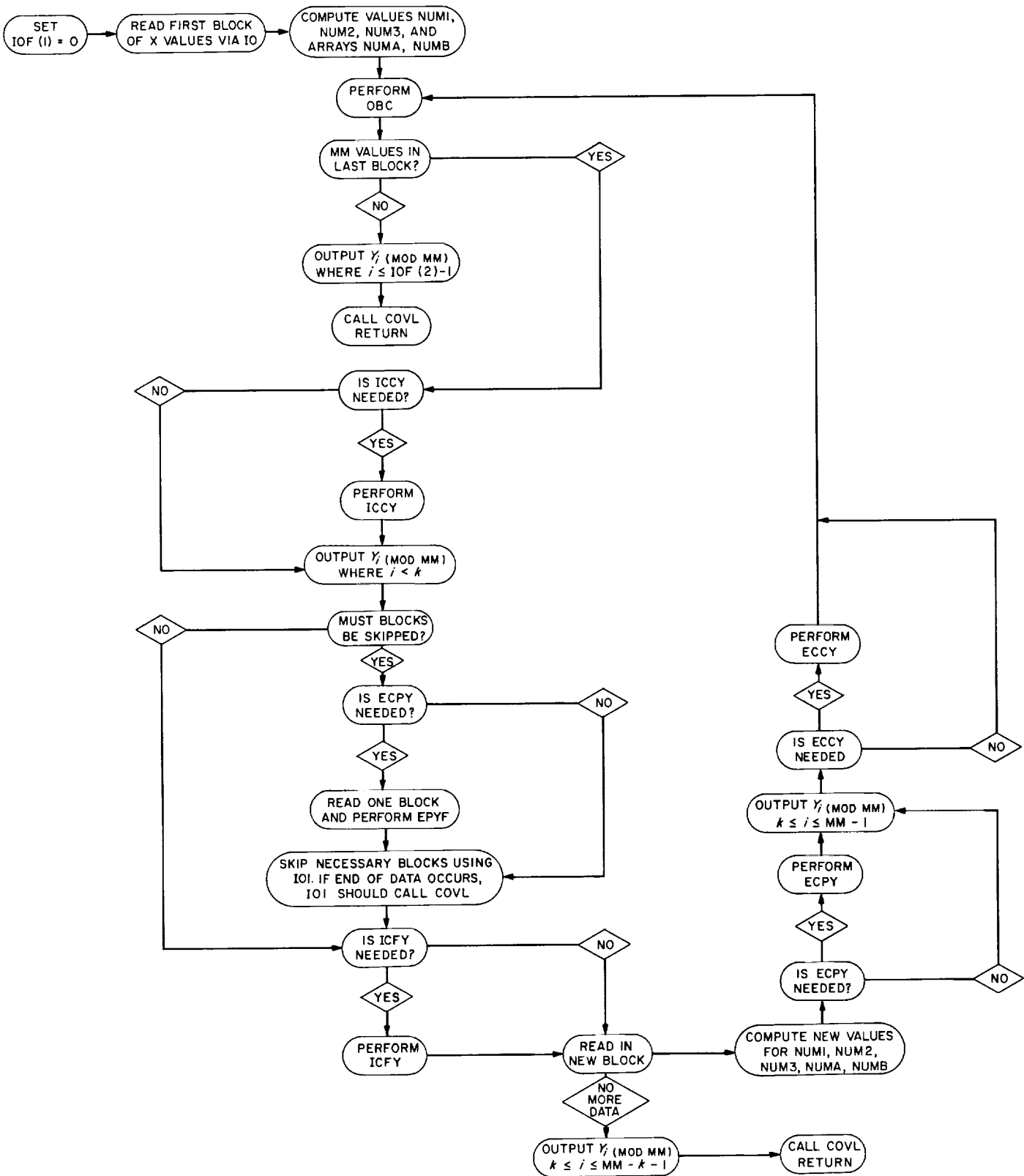
Fig. 9. Flow diagram of filter subroutine
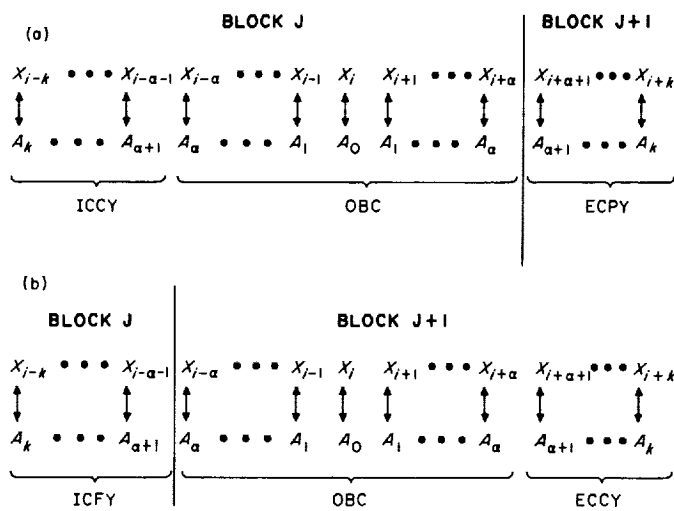
Fig. 10. Calculation of Y_i when two blocks are required

called the "end calculation for prior Y values" (ECPY), is made.

Suppose X_i is in Block J + 1 and X_{i-j} $\alpha < j \leq k$ is in Block J (Fig. 10b). The calculation S_3 called "initial calculation for future Y values" (ICFY), is made when Block J is in core. When Block J + 1 is read, the calculation for S_2, OBC, and the calculation for S_1, called "end calculation for current Y values" (ECCY), are made.

For every block of X values in core, the program constructs three fixed-point arrays: NUMA(I,1), NUMB(I), and NUMA(I,2). The array NUMA(I,1) contains the subscripts of the Y values for which an ECPY calculation needs to be made. NUMB(I) contains the subscripts for which ICCY, ECCY, and OBC calculations need to be made. NUMA(I,2) contains the subscripts for which an ICFY calculation needs to be made. The subroutine DEF computes subscripts mod MM.

References

1. Devine, C. J., PLOD II: Planetary Orbit Determination Program for the IBM 7094 Computer, Technical Memorandum No. 33-188, Jet Propulsion Laboratory, Pasadena, California, April 15, 1965.

2. Peabody, P. R., Scott, J. F., and Orozco, E. G., JPL Ephemeris Tapes E9510, E9511, and E9512, Technical Memorandum No. 33-167, Jet Propulsion Laboratory, Pasadena, California, March 2, 1964.

3. Peabody, P. R., Scott, J. F., and Orozco, E. G., Users Description of JPL Ephemeris Tapes, Technical Report No. 32-580, Jet Propulsion Laboratory, Pasadena, California, March 2, 1964.

4. Kopal, Z., Physics and Astronomy of the Moon, p. 5, Academic Press, New York, 1962.

5. Blackman, R. B., and Tukey, J. W., The Measurement of Power Spectra, pp. 34–36, Dover Publications, Inc., New York, 1958.

6. Andres, E. B., et al., Digital Filters, NASA Contract Report CR-136, p. 36, 1964.

III. Deep Space Network Systems

A. Evaluation of Operator Requirements for Real-Time Direct Command Capability From the SFOF

W. Wong and K. Heftman

In the past, the command and control requirements of spacecraft (*Ranger* and *Mariner*) have been met primarily by extensive preprogramming of the anticipated command inside the spacecraft. Most commands were initiated either automatically, with ground initiation serving as backup, or by premission preparation of commands for anticipated command sequences. Future missions (e.g., *Surveyor* and *Voyager*) present more difficult problems due to: (1) the increased requirement for commands in real or near-real time, (2) an increase by an order of magnitude in the quantity of command instruction, (3) the now-meaningful time delay (transmission delay and operator response time) due to real-time control, and (4) the desire for direct command capability from the SFOF.

A study was begun October 25, 1965, under JPL contract with Serendipity Associates, for:

(1) Development of a basic study method to provide a framework for identifying and evaluating functional requirements for the operator in the command/ control loop during JPL Space Flight Operations Facility (SFOF) operations.

(2) Derivation of a conceptual design of equipments and operations which satisfy these operator requirements, i.e., a control station for the SFOF to illustrate the application of the method. (The derivation of this operator station encompasses identification of the station functional requirements, determination of an operational concept, and the conceptual design of equipment and personnel subsystems for display and control.)

A study plan[1] was prepared to describe the approach to the study and to establish the performance schedule for the study tasks. The approach, in general, consists of the following steps: definition of input/output and means-limiting states (prerequisites, requirements, and constraints), identification of criterion measures to evaluate the system or its parts, identification of alternate means, selection of means for optimal solution, and verification of selections. Initially, existing and potential missions and mission experiments are being reviewed and analyzed to define an "object system," i.e., a generalized spacecraft system which will form the basis for the functions required of the operator control station.

_____
[1]*Program Plan for a Study to Develop Conceptual Design for Ground Control Station*, TR 34-65-23, Contract No. 951313, Serendipity Associates, Chatsworth, California, November 1965.

GUIDANCE AND CONTROL DIVISION

IV. Spacecraft Power

A. High-Power Low-Saturation Voltage Silicon Switching Transistor

T. J. Williams

1. Introduction

One of the major problems in the utilization of thermionic generators involves the unavailability of a power conditioning device that will efficiently and reliably convert the low-voltage high-current generator output to a level suitable for use in a spacecraft. The voltage outputs of these generators may typically vary from 0.6 v to 4 v, depending upon the number and electrical connections of the thermionic diodes used. Therefore, the current levels that must be switched are very high. This imposes severe requirements on the devices used as power switches in the converter. To be efficient, they must switch the high current rapidly and with little saturation loss. In addition, they must be capable of reliable operation in a moderately high temperature environment, even with effective isolation from the high temperature zone of the thermionic generator.

Low input voltage DC-to-DC converters that have been investigated and used to date generally employ germanium transistors as power switches. These devices are attractive because of their excellent saturation characteristics. The largest unit available is capable of switching 150-amp collector current with a collector-to-emitter saturation voltage of less than 0.1 v. A major disadvantage of these units is their slow switching speed (specification of 58 μsec maximum), which limits the upper frequency of operation of the converter if efficiency is not to suffer. Also, because of the low operating frequency, magnetic components in the converter are large and heavy. A second disadvantage, one that more or less precludes their use in practical thermionic space power systems, is that of limited temperature capability. The junction temperature is limited to a maximum of 110°C, and reliable operation is seriously degraded above 50°C. Since it is unlikely that operating temperatures less than this could be obtained in a practical system, a more suitable device must be found.

The problem of high temperature environment can be relieved somewhat through the use of silicon transistors as power switches. They are usable to junction temperatures of 175°C. Unfortunately, none have been developed to date with the combined properties of high current capability and extremely low saturation drop that is required in low voltage conversion. Consultation with semiconductor specialists both at JPL and in industry reveal that these properties are obtainable in silicon,

however. The purpose of this effort is to develop a silicon power switching transistor suitable for use in power conditioners operating from low voltage sources.

2. Specification

When a development is undertaken, a set of parameters must be established to which a design can be made. In this development it is desired to obtain a high-current silicon transistor with saturation characteristics approaching those made of germanium but with a faster switching speed. Table 1 is a preliminary specification that establishes the basic electrical and thermal requirements that the device must meet. All requirements, with the exception of saturation voltages, are moderate and well within the state-of-the-art. The extremely low collector-to-emitter saturation voltage is the most important requirement and will be the most difficult to meet.

Table 1. Specifications for high-power, low-saturation voltage silicon switching transistor

Absolute maximum ratings	
Collector-to-emitter voltage, BV_{CEO}, minimum	20 volts
Emitter-to-base voltage, BV_{EBO}, minimum	4 volts
Collector current, I_C, minimum	100 amp
Base current, I_B, minimum	15 amp
Collector dissipation, $T_C = 100°C$, minimum	150 watts
Thermal resistance, junction to case, θ_{JC}, maximum	0.5°C/W
Junction temperature range, T_J	−65 to +175°C

Electrical characteristic[a]	Test conditions	Min.	Max.	Units
Breakdown voltage, BV_{CEO}	$I_C = MS^b$	20	—	volts
Breakdown voltage, BV_{EBO}	$I_{EB} = MS^b$	4	—	volts
Collector cutoff current, I_{CEX}	$V_{CE} = 20v$ $V_{BE} = MS^b$	MS[b]	MS[b]	ma
Emitter cutoff current, I_{EBO}	$V_{EB} = 4v$	MS[b]	MS[b]	ma
DC current gain, h_{FE}	$I_C = 75a$ $V_{CE} = 1v$	20	—	—
Saturation voltage, $V_{CE(sat)}$	$I_C = 75a$ $I_B = 5a$	—	0.2[c]	volt
Saturation voltage, $V_{BE(sat)}$	$I_C = 75a$ $I_B = 5a$	—	1.4	volts
Total switching time $(t_d + t_r + t_s + t_f)$	$I_C = 75a$ $I_B = 5a$ $V_{BE} = -1.5v$ on turnoff	—	15	μ sec

[a] Electrical characteristics at 100°C case temperature.
[b] Manufacturer's specifications are acceptable.
[c] A saturation voltage of 0.1 volt under the above conditions shall be a design goal.
[d] Device is to be used in DC-to-DC converter of parallel configuration. Duty cycle is 50%, and operating frequency may be up to approximately 5 kc. Operating conditions are either fully saturated or cut off.

3. General Requirements

Detailed discussion of the technology necessary to make a transistor of the type desired is beyond the scope of this article. It is informative, however, to consider a few of the more general requirements in simplified terms. First, the total junction area will be physically large to provide for the 100-amp collector current capability. It is estimated that an entire silicon wafer, 1-in. diameter, will be required to obtain the necessary area. Only a portion of the wafer is used to form an active junction, however. One approach is to make a single large-area junction device on the wafer. This can result in poor yield if the material is not flawless and if extremely precise manufacturing techniques are not used. The alternate approach is to make smaller devices and parallel the chips within the same transistor case. Yield can be greater with this approach, but so is the manufacturing complexity. Both approaches are being investigated.

A low collector-to-emitter saturation is one of the more important requirements for this transistor. This quantity is made up of an intrinsic voltage across the junctions in the transistor plus bulk resistance and contact resistance drops. Bulk resistance drops are reduced by using low resistivity (i.e., highly doped) material in the collector, base and emitter regions. Low contact drop requires physically large contact area. The intrinsic voltage drop across a junction is determined by the carrier density on either side of it. In a transistor with collector and emitter junctions adjacent to each other, this voltage is related to the normal and inverted alphas, i.e., current gains. To reduce this drop to a minimum requires almost equal doping in the collector and emitter regions. The normal and inverse alphas are maximized, and the base width is made extremely narrow.

4. Design Approach

As stated above, a symmetrical transistor geometry, i.e., one where the collector and emitter doping profiles are almost equal, is necessary in order to achieve the desired performance. There are several methods by which this type structure can be made. One of the most convenient is an alloyed process. This method is difficult when a very narrow base width over a large area is required, however. It also results in a large variation in current gain versus emitter current and in slow switching speed. Therefore, this approach is not being considered in this development.

A method that has previously demonstrated good results and which might be used to approach the voltage

design requirement of $V_{CE(sat)} = 0.2$ v is a two-sided diffusion process. Here, n-type dopant is deeply diffused into a p-type parent material simultaneously from both sides to form collector and emitter junctions. Its weakness is that the deep diffusion required for narrow base width results in low net doping in the base, and lateral base resistance is higher than desired. Since ohmic drop in the base is a significant part of the saturation voltage, the $V_{CE(sat)}$ cannot be minimized with this approach.

A more modern approach promises better results than the above. A p-type base is grown epitaxially on a heavily doped $N+$ substrate. The surface is then masked and etched. An n-type dopant can then be diffused to form an emitter junction. There are several variations of this method, all of which can be used to approach the design goal of $V_{CE(sat)} = 0.1$ v.

Both of the above techniques are under investigation. Several variations of the latter are being tried and results are promising.

5. Parallel Design Effort

The design and development of this transistor is being carried out by two separate companies under a parallel design effort. These companies are Westinghouse Semiconductor Division in Youngwood, Pennsylvania, and ITT Semiconductors/Shockley Laboratories in Palo Alto, California. This procedure offers a better chance of success and may result in two separate sources for the device. Work began in August, 1965, and is scheduled for completion by May, 1966.

B. Thermal Energy Storage

R. A. Boring

A program has been initiated to evaluate the feasibility of using thermal energy storage techniques to enhance the performance of a solar thermionic power system during periods in which the concentrator is not Sun-oriented. The program has as its goal the successful fabrication and test of a solar thermionic converter-thermal energy storage feasibility model. This model is obtained by integrating a container of a suitable Thermal Energy Storage (TES) material with the emitter of a thermionic

converter. It has been postulated that the latent heat of fusion of the molten TES material will be capable of maintaining the supply of thermal energy to the thermionic converter when shadow or disorientation periods occur. It is not expected that TES will completely eliminate the requirement for electrochemical (battery) storage requirements for every mission but it could substantially (as much as 5-to-1) reduce the weight of the total power sub-system by requiring a battery only for initial start-up and/or peak load demands.

During the second quarter of 1965 JPL, under NASA sponsorship, initiated an effort with Thermo-Electron Engineering Corporation, Waltham, Mass. (TEECO) to evaluate the feasibility of the TES concept. Prior to contract release, JPL and TEECO independently performed preliminary analyses to determine how the solid/liquid TES material interface could be expected to change during operation. The results of these analyses indicated that there should be a significant extension of a solar thermionic TES converter operation after the actual heat input to the assembly is terminated.

Using the results of this preliminary analysis, the formal model design effort was contracted with TEECO in March 1965. The initial effort of this program was to evaluate analytically the effect of enhancing the assumed low oxide thermal conductivity. This enhancement took the form of adding various percentages, by weight, of rhenium metal to the bulk oxide. The rhenium would be interspersed throughout the oxide when the slug was fabricated. The results of this analysis indicated that the specific power is highest with zero enhancement regardless of container size (Fig. 1).
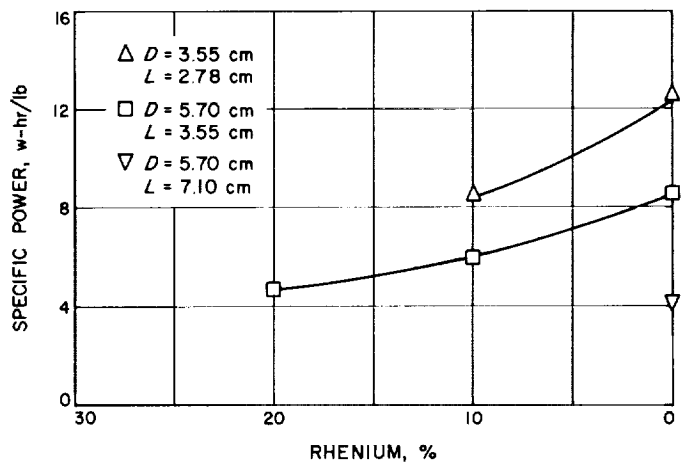
Fig. 1. Effect of rhenium metal content on specific power
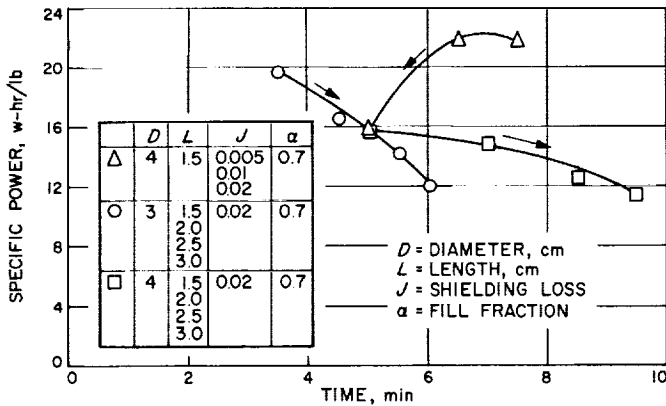
Fig. 2. Performance of various thermal energy storage configurations

Fig. 2 gives the specific power with respect to operating time for various conditions. (D and L are container dimensions, J is the percentage of black body losses through the shielding, and 2 is the room temperature volume percentage of oxide to which the container is filled.) The oxide for this model is 2MgO-3BeO and was selected due to its apparent compatibility with rhenium and the extensive thermophysical property data reported for it. From the calculations which led to Fig. 2, a decision was made that a container with $D = 4$ cm and $L = 3$ cm would yield a model that had both an interesting specific power and length of operating time. If shield losses can be kept below 2% of black body, it would yield a model having an average specific power of 11.5 w-hr/lb over a 9.5 min operating time.

The thermionic converter used in this study was a modified version of the JPL/TEECO series VIII diode. This converter was modified to incorporate a TES device similar to the design shown in Fig. 3. Fig. 4 is a plot of the test data for this converter. Converter performance was measured prior to attachment of the TES container to evaluate changes due to fabrication and to calibrate for model tests. It is significant to note the variation in optimum cesium temperatures for the various data points. If the model is operated at a constant voltage output the optimum cesium temperature varies widely. If, on the other hand, constant current operation is used the cesium temperature variation is much less. Thus for this model the constant current mode of operation was selected. Fig. 3 is a cutaway schematic of the model including electron bombardment heater and thermal shielding. The

Fig. 3. Modified thermionic converter incorporating a thermal energy storage device
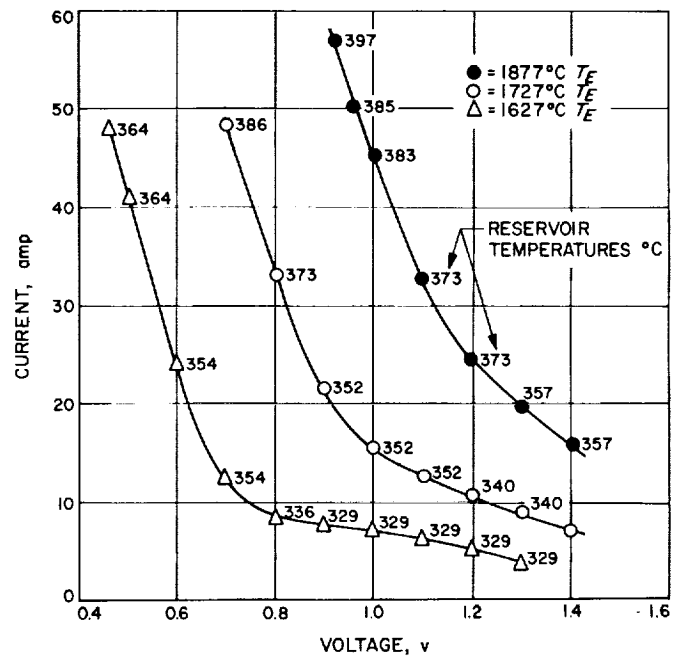
Fig. 4. Performance of thermal energy storage device shown in Fig. 3

converter is instrumented with chromel-alumel thermo-couples and an 8:1 hohlraum for pyrometric measurement of emitter temperatures. A hole in the shielding allows emitter temperature monitoring during operation.

Fig. 5 is a photograph of the shielding assembly for the model while Fig. 6 shows the electron bombardment gun designed and developed for use in heating the TES container. Only refractory metals are used for the shields and electron gun. Two legs carry the filament current which then is distributed to tungsten wires by two heavy tantalum rings.

Fig. 5. Thermal shield for thermal energy storage model
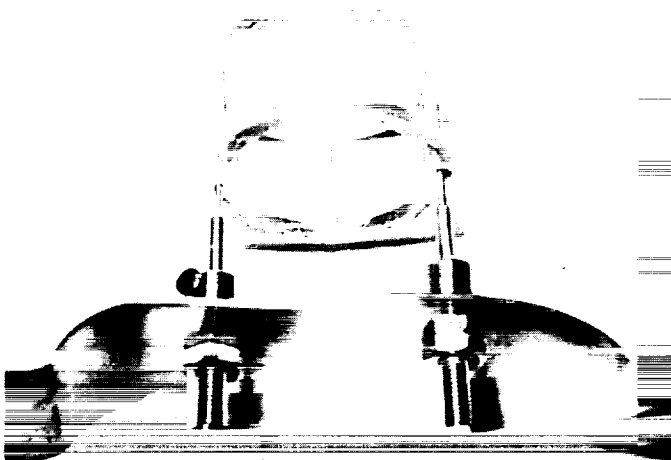
Fig. 6. Electron bombardment gun for heating TES container

Fig. 7 is a schematic of the TES model setup for testing within a 12-in. belljar, oil diffusion, vacuum station. After some initial adjustments, the model was slowly heated to the oxide melting temperature. As the oxide melting temperature was reached a large amount of vapor was observed being collected on the belljar. Power to the model was immediately removed, and cooldown began. Upon opening of the vacuum system and removal of the shields, it was obvious that the container had developed an oxide leak. Careful microscopic examination of the upper weld region of the TES container showed small pinholes. Important data on the converter's electrical performance were obtained, however, during the TES material cooldown (see Fig. 8). Although the data are not conclusive, they show that the oxide was at least partially melted prior to heat removal and a thermal energy storage effect is evident for several minutes. The monitored performance of the model approximated the design performance. Table 2 presents the raw data plotted in Fig. 8. The constant current during the first 8 min of operation was 14.4 amp. This resulted in an indicated thermal

Table 2. Constant current cooling curve[a]
$(T_{cB} = 380°C = constant)$

Position	Voltage, v	Time, min.
1	1.211	0
2	1.135	0.6
3	1.063	1.0
4	1.00	1.4
5	0.962	1.5
6	0.942	2.2
7	0.926	2.7
8	0.908	3.1
9	0.896	3.5
10	0.884	4.1
11	0.866	4.2
12	0.850	4.6
13	0.784	5.4
14	0.695	5.8
15	0.576	6.4
16	0.437	6.7
17	0.309	7.2
18	0.203	7.6
19	0.068	8.4
20	0.047	8.8
21	0	9.5

[a] From curve of voltage vs time, the average voltage (for 8 min) is:
$\frac{126}{16} \times \frac{1}{10} = 0.788$ v.

Specific power $= 0.788$ v $\times 14.4$ amp $\times \frac{8 \text{ min}}{60 \text{ min}} \times \frac{454 \text{ gm/lb}}{65.6 \text{ gm}} = 10.45 \frac{\text{w-hr}}{\text{lb}}$

The output current remained constant from shutdown until 8 min had elapsed. Zero output occurred 9.5 min after shutdown. Output power at shutdown was observed to remain absolutely constant for 10 sec.
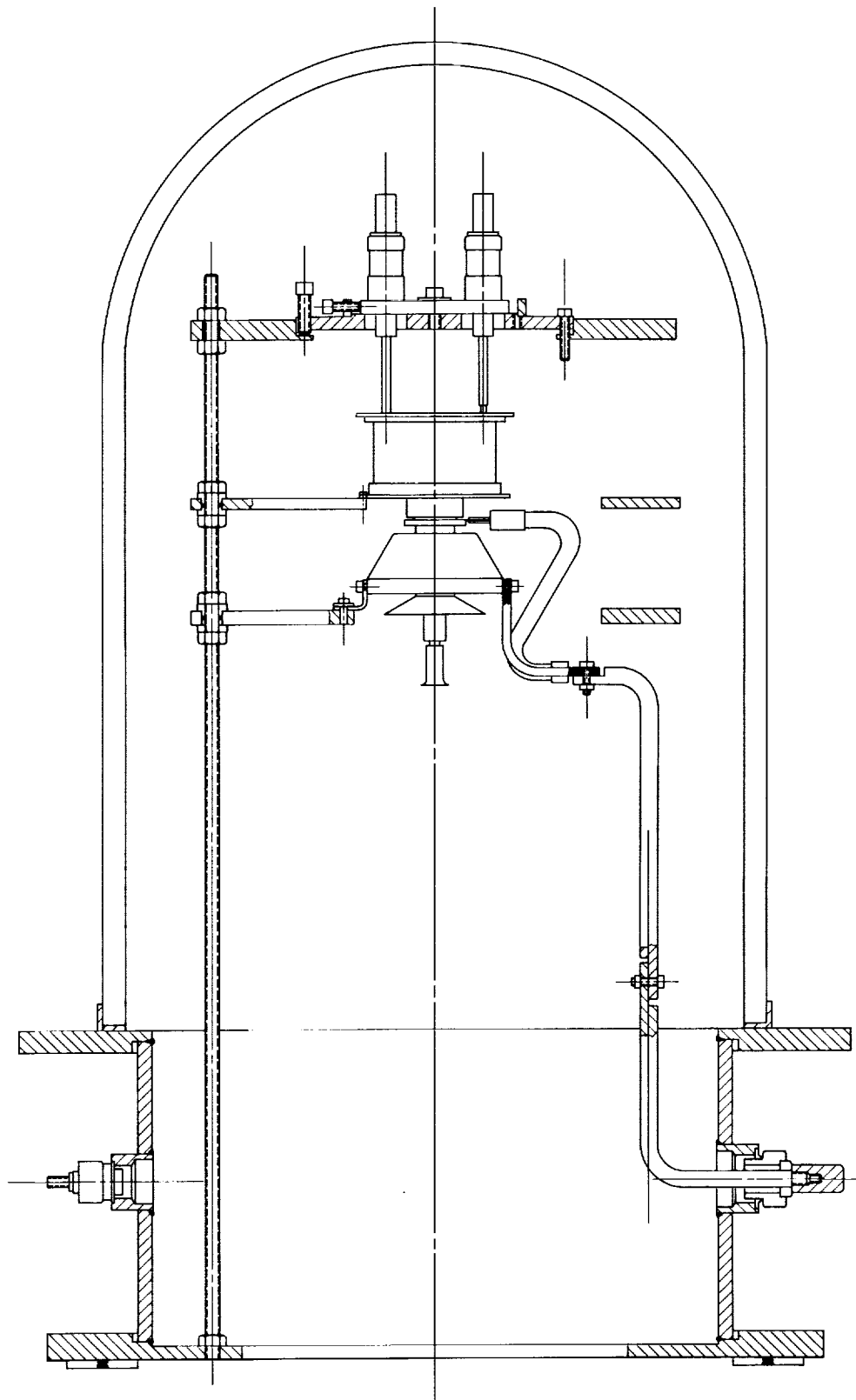
Fig. 7. Test model TES setup in belljar

energy storage specific power of 10.45 w-hr/lb with 9.5 min of converter operation.

The leakage problem experienced during this effort is attributed to weld poisoning. A modified version of the TES model is being fabricated to continue these studies.

This modified device will include some refinements of the heater, shielding, and oxide slug designs to improve device efficiency and to minimize future failures of a similar nature. If successful the improved model will undergo performance and life testing at JPL. This second model is expected to be ready for testing early in 1966.
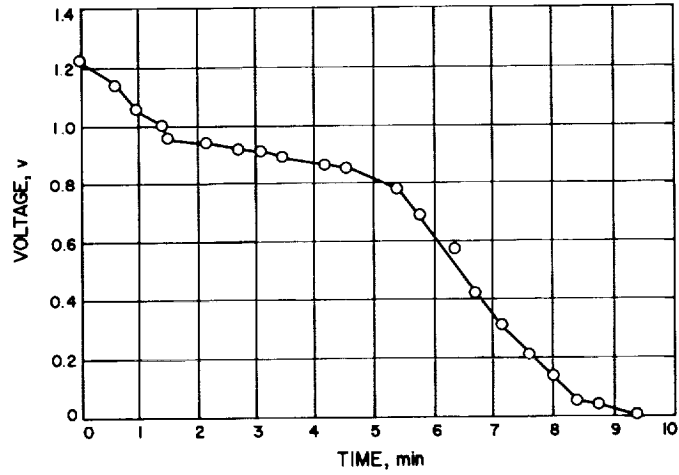
Fig. 8. Performance of TES model during material cooldown

V. Spacecraft Control

A. Hydrazine-Plenum Attitude Control System

J. C. Randall

As part of the *Voyager* support activity, a study was conducted of the various possible types of attitude control systems. It was concluded in the study that both a water electrolysis system and a hydrazine-plenum system competed favorably with the conventional nitrogen cold-gas system as used on the *Ranger* and *Mariner* spacecrafts.

Hydrazine has been used as a propellant in the past in the midcourse propulsion units for both *Ranger* and *Mariner*. Although it is basically a monopropellant, the use of nitrogen tetroxide was required to start the combustion. Shell Oil Company has recently, however, developed a catalyst which eliminates the need for the nitrogen tetroxide. Hydrazine immediately decomposes into nitrogen, hydrogen, and ammonia on contact with the catalyst, thus greatly simplifying the hardware required. It was first proposed that hydrazine be employed as an attitude control propellant with a catalyst bed at each exit nozzle. Upon further investigation, however, it was concluded that the response times would be too long and the thrust levels too high for *Voyager* use. The concept of using this catalyst technology for a gas generator was then proposed.

A functional schematic diagram of the hydrazine-plenum system is shown in Fig. 1. The hydrazine is stored as a liquid under relatively low pressure. A pressure-controlled valve senses the downstream pressure. At the entrance to a plenum volume, the hydrazine passes through a catalyst bed and decomposes into the plenum. The decomposed gases remain in the plenum as a cold gas until required by the attitude control jets. When the plenum gas is depleted to a certain pressure, hydrazine flows again raising the plenum pressure. The chief advantages of this type of system are that both the response times and the thrust levels are at the desired levels. In addition, the hydrazine gases have a specific impulse of 115 lb-sec/lb in comparison to 75 lb-sec/lb for nitrogen. Additional gains in system weight are also obtained because of the low-pressure hydrazine storage.

Construction of a breadboard hydrazine-plenum attitude control system was started in July, 1965, to learn more of the operating characteristics associated with this type of system. A conventional *Ranger* half-distribution system was connected to a gas generator for testing. Fig. 2 shows the entire breadboard system in a test pit.

The tests performed to date have demonstrated the ability of the gas generator to more than meet the demands of the attitude control system when operated under simulated *Voyager* conditions. The restart capability has also been partially demonstrated with several
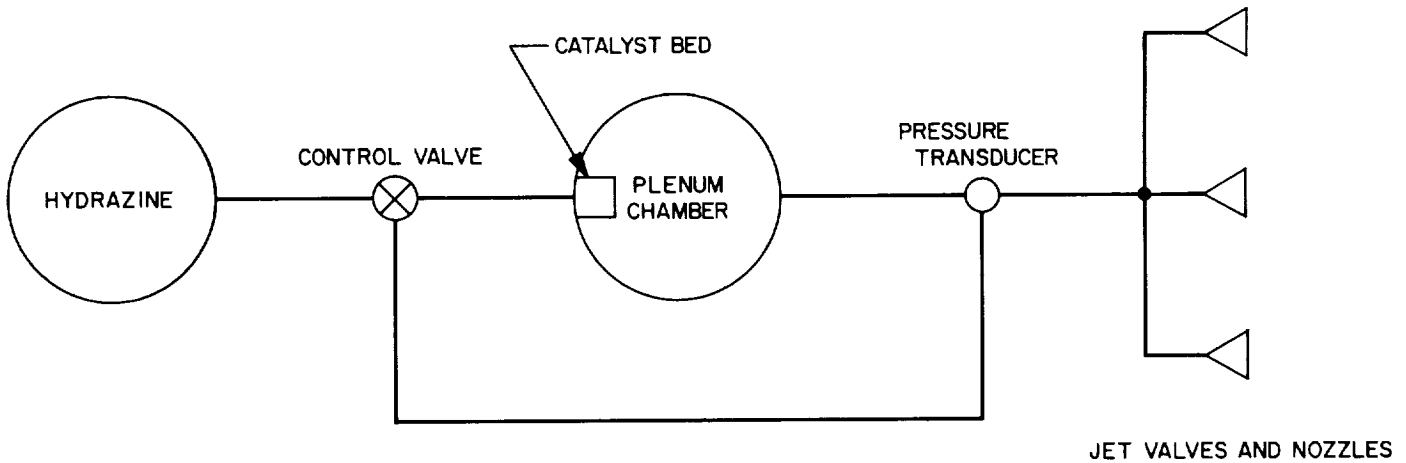
Fig. 1. Hydrazine-plenum attitude control system schematic
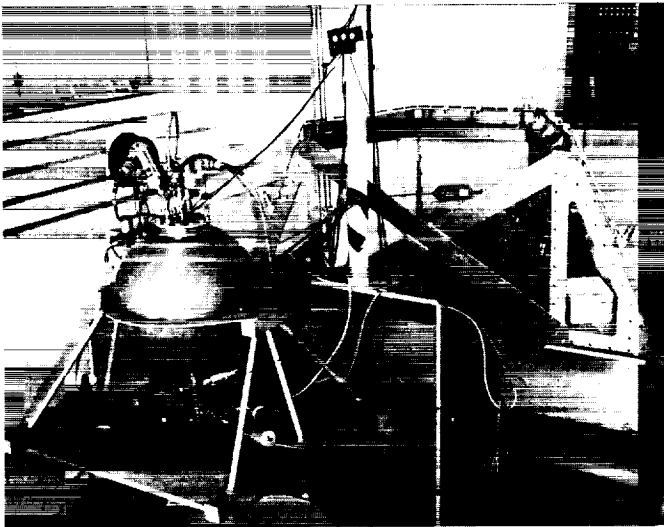
Fig. 2. Breadboard of hydrazine-plenum system

hundred restarts on the particular catalyst bed in use. It has been found, however, that the catalyst bed firings generate particles 50 to 100 μ in size. Since the attitude control jets require cleanliness down to 5 μ in size, it appears that filters will be required for any flight use of this system. Tests are now being conducted on labyrinth filters, with a large filter area, to check their susceptibility to clogging during long-term missions such as *Voyager*.

It was first observed that the catalyst bed had a tendency to pack as the unit was fired. Extreme packing would cause incomplete dissociation of the hydrazine and thereby lower the efficiency. After many firings, however, it appears that the catalyst bed packing begins to level off at a point which is acceptable.

To prove the long-term capabilities of the hydrazine-plenum system, it is planned to run an isolated 40-day duration test. During this test, the gas generator will be actuated approximately the same number of times as for a 450-day *Voyager* mission. The total mass flow of hydrazine for a *Voyager* mission will also be duplicated.

B. Electrically Suspended Gyro Development

T. C. Lear

In 1960, the Laboratory entered into a program to demonstrate the feasibility of constructing a gyroscope utilizing an electrically suspended free spherical rotor with optical readout of spin axis position. As the result of this program, and similar work conducted for the Department of Defense, a contract was let to Honeywell, Inc., in June of 1963 to develop an electrically suspended gyro, tailored specifically for strapdown application in spacecraft systems. The results of the gyro development program (Ref. 1), completed late in 1964, will be reported here.

1. Launch Survival Suspension System

Since the physical parameters of an electrically suspended gyro (ESG) are determined largely by its environmental requirements, the first study task undertaken was to determine the method by which the device could be

mechanized to withstand the launch phase of spacecraft operation while maintaining reasonable limits of size and power requirements. Three design variations were studied which would meet these criteria, as follows:

a. ESG hard-mounted to the spacecraft, with rotor spinup, damping, and alignment accomplished on the launch pad. It was determined that this technique would require use of relatively high voltage suspension electronics, operating near the limits of semiconductor technology, and presenting the attendant difficulties due to high voltages during the changing pressure environment of the launch phase. In addition, this design would require the use of a smaller rotor size, increasing the readout errors due to line definition limits. The higher electrode-to-rotor voltages would also mean increased electrical drift rates during the launch and injection phases of spacecraft operation, and decreased rotor rundown time.

b. ESG mounted to spacecraft with mechanical vibration isolation devices. This design approach allowed a reduction of approximately 33% in rotor-to-electrode voltages, reducing considerably the objectionable effects cited for Design *a* above. This design, however, raised the overall weight by 200% and the size by 400%. The spinup, damping and alignment of the rotor could still be done on the launch pad, but considerable evaluation would be required to determine the degradation of the alignment caused by failure of the isolation system to re-acquire perfect register after launch and separation of the spacecraft. The reduction in power attained in this configuration would not be of major significance.

c. ESG hard-mounted to spacecraft with rotor not suspended during launch and separation phases. This design technique minimized the high-voltage problems cited in Design *a* above and also removed the extra weight and size of the vibration isolation system required in *b*. An additional complexity was imposed by the requirement for a system of rotor spinup, damping, and alignment after spacecraft separation from the injection boost vehicle. A system for accomplishing these sequential operations has been outlined; it would require one onboard command from the central computer and sequencer, with a possible attendant ground command backup. Environmental tests have been made on an ESG rotor suitable to this design approach, which demonstrated no detectable rotor deformation or increase in pickoff signal-to-noise ratio due to the rotor rattling in the suspension electrode cavity.

The predicted performance and design parameters of the three design approaches are presented in Table 1. The third design configuration was selected as the most promising. It was apparent that the high voltages and small fragile rotor used in Design *a* would not meet the present launch and separation environments. The reductions in high-voltage effects and readout error presented by the vibration isolation scheme of Design *b* did not seem to be significant enough in view of the large increase in volume and weight.

Table 1. Design approach for ESG suspension system

Parameter	1.5-in. rotor 100-g suspension	2.0-in. rotor 30-g suspension	2.5-in. rotor 5-g suspension
Rotor diameter, in.	1.5	2.0	2.5
Rotor mass, gm	7	20	47
Rotor inertia, gm-cm^2	19	90	300
Voltage gradient, v/mil	1300	900	450
Performance (predicted)			
RSS drift at 1 g, deg/hr	0.10	0.035	0.025
RSS drift at 0.05 g, deg/hr	0.003	0.0025	0.002
Rundown time, days	465	712	950
Read-out resolution, arc sec	30	20	15
Systematic error, arc sec	100	75	60
Power required, w	28	24	6

2. Readout System Development

In order to determine the position of the ESG spin axis in relation to its electrode structure, and hence to the spacecraft frame, an optical readout system is presently necessary. This system utilizes three pickoffs (Figs. 3, 4) whose optical axes describe a case-fixed reference frame. These pickoffs determine the time of appearance in their field of view of a line pattern (Fig. 5) applied to the surface of the spinning rotor. The output of two of these pickoffs, after appropriate processing, is sufficient to define the rotor spin axis position. The third pickoff is necessary for use when one pickoff sees the region near the pole of the spinning rotor, where resolution of line position is too poor to use in the strapdown gyro.

The requirements of the readout system were determined to be 8.5 arc sec in resolution and 17 arc sec in accuracy. The characteristics of two gyro components control these parameters: the resolution of the optical system and the accuracy with which the line pattern is
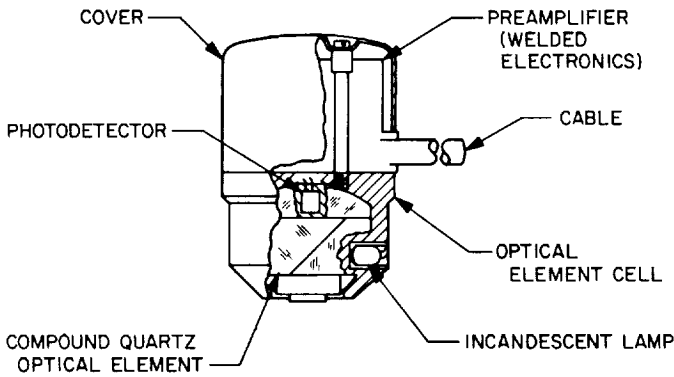
Fig. 3. Electrically suspended gyro pickoff assembly cross section

applied to the rotor. A new design was completed for the optical pickoff to meet these requirements (see Fig. 3). The characteristics of this device, including its pre-amplifier, are listed in Table 2.

Prior to this development, the readout pattern applied to ESG rotors for operation in the strapdown mode consisted of a great circle inclined to the rotor's spin axis at an angle of about 50 deg; see Fig. 5(a). This pattern required the computer associated with the gyro to perform complex trigonometric computations to derive the
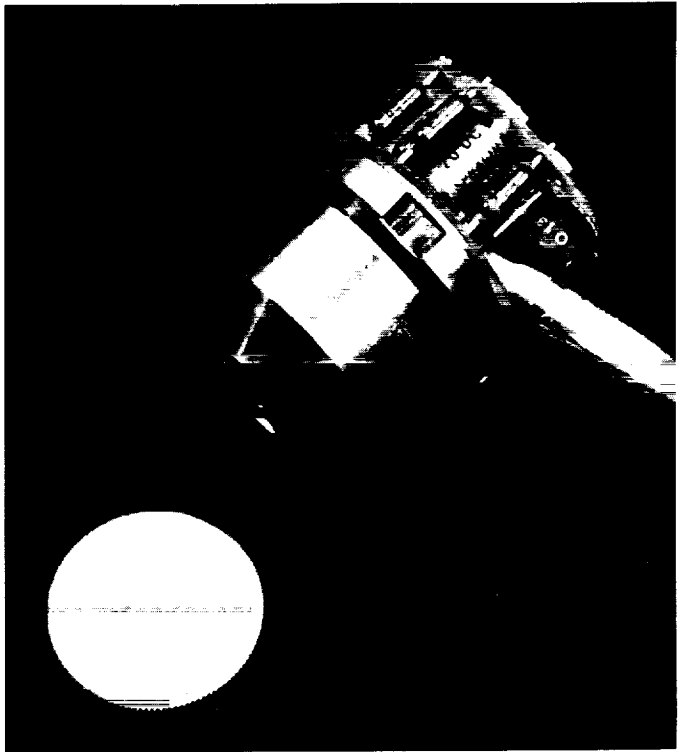
Fig. 4. Electrically suspended gyro pickoff assembly photograph

$$\cos\ \theta = \sin\left\{\tan^{-1}\left[\tan\ i\ \cos\left(\frac{4TC_1}{C_1+C_2}\right)\right]\right\}$$

i = INCLINATION OF GREAT CIRCLE TO EQUATOR

$$\cos\ \theta = \frac{4C_1}{C_1+C_2}-2$$

(a) GREAT CIRCLE LINE PATTERN   (b) COSINE LINE PATTERN
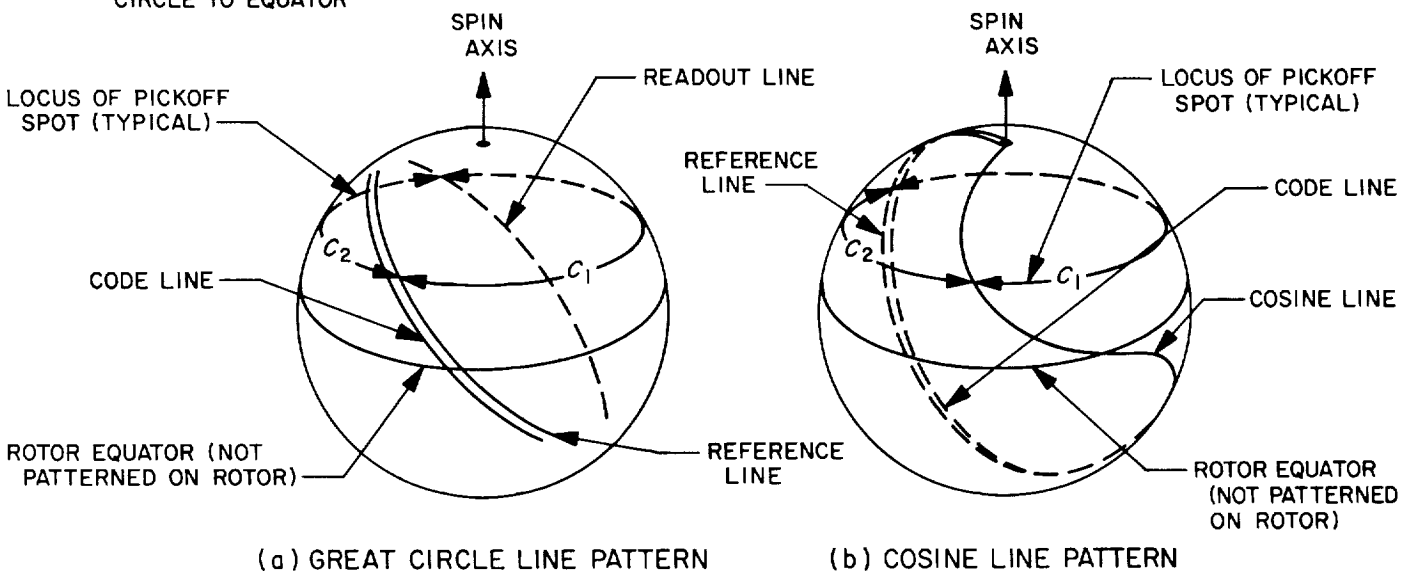
Fig. 5. Great circle line pattern and cosine line pattern

Table 2. Optical pickoff characteristics

Signal-to-noise ratio	77/1
Rise time, μsec	0.5
Output voltage, v	0.62
Working distance, in.	0.117
Optical aperture	f 0.5
Alignment sensitivity, arc sec	< 10
Output impedance, kΩ	5
Input power, total, w	0.4
Volume, in.^{3}	1

spin axis position. To reduce this computer complexity, a new readout rotor pattern was developed. This pattern, called the cosine pattern, consists of a reference line which is a great circle segment perpendicular to the rotor's equator plus a curved line on the opposite side of the ball from the great circle segment. This curved line is shaped such that the times between the line crossing signals are proportional to the latitude cosine of the pickoff with respect to the rotor equator; see Fig. 5(b). The cosine pattern is a much harder one to apply to the rotor since it requires a nonlinear motion of the tool in longitude with respect to latitude angle. The variation of brightness in the pickoff spot is also more critical with the cosine pattern since the angle of approach is different for the code line than for the reference line. It was decided that these difficulties had a good chance of improvement with advances in pattern application techniques and better pickoff lamp design. The pattern accuracies necessary to meet the readout requirements are as follows:

Pattern tilt	100 μin.
Pattern shape uncertainty	33 μin.
Pattern edge uncertainty	97 μin.

3. Vacuum System Study

Since power to spin the rotor is applied only during the first few hours of gyro operation, a vacuum level of 10^{-8} Torr is needed inside the ESG electrode structure. This vacuum level will give a rotor rundown time constant of about 1000 days.

A vacuum test unit, similar in construction, assembly, and material to an ESG was built and tested to determine the composition of the gases in a gyro cavity, the rate of accumulation of these gases, and the percentage of gas due to leakage of the enclosure. The gas in the enclosure was found to be 75% H_2O and 12% H_2, with the remainder made up of a number of other gases in trace amounts. The rate of outgassing from the gyro parts was 5×10^{-7} Torr liter/sec. Leakage through the enclosure was found

to be less than 0.2% of total gas evolution. The vacuum pump used on ESGs prior to this study consumed power at the rate of about 5 w. Since these pumps had a capacity far in excess of what was needed, a new pump was designed for this application which reduced the power consumption to approximately 0.5 w. Pump designs presently exist using cold filaments. These "self-sustaining" pumps require much higher grid voltages, on the order of 300 v, about the same amount of power to sustain emission as those using heated filaments. In addition, the base metal for these cold filaments is nickel, placing an undesirable magnetic element in the gyro.

4. Sterilization and Reliability Study

Typical components of an ESG were tested for temperature sterilization effects at 360°F for 36 hr. Components included in these tests were the electrode structure, rotors, photo diodes, optical pickoff, and the output transformers. The electrode surfaces made of electroless nickel, plated directly on the ceramic structure, and unplated beryllium rotors showed no effect from the sterilization cycling. Plated rotors and electrode structures with electrodes plated over copper showed considerable blistering and separation of the plating. No changes in electrode structure shape, except those caused by blistering, and no leakage increase was noted due to the sterilization tests. The optical assembly of the pickoff also withstood the temperature sterilization with no adverse effects. The photo diode and the pickoff preamplifier both showed significant changes in performance. The characteristics of the suspension output transformers indicated marginal operation also. Further development or a change in sterilization requirements is needed to qualify these electronic components for a sterilizable gyro.

Based on previous experience with the electronic components and arbitrary values for the mechanical components, a reliability prediction was made for a spacecraft ESG being unsuspended during launch with spinup in space (see Fig. 6). The suspension electronics in this case would be operating at only a fraction of their capability during the cruise portion of the flight, with readout assumed to be operating during 2% of the mission duration. The least reliable systems are the spinup and damping electronics and the control logic for these functions. Even though these operations are used only once during a mission, failure of any of them would cause catastrophic failure of the gyro system. Because of this, these one-shot systems should incorporate as much redundancy and derating as is possible within the limitations of size and weight.
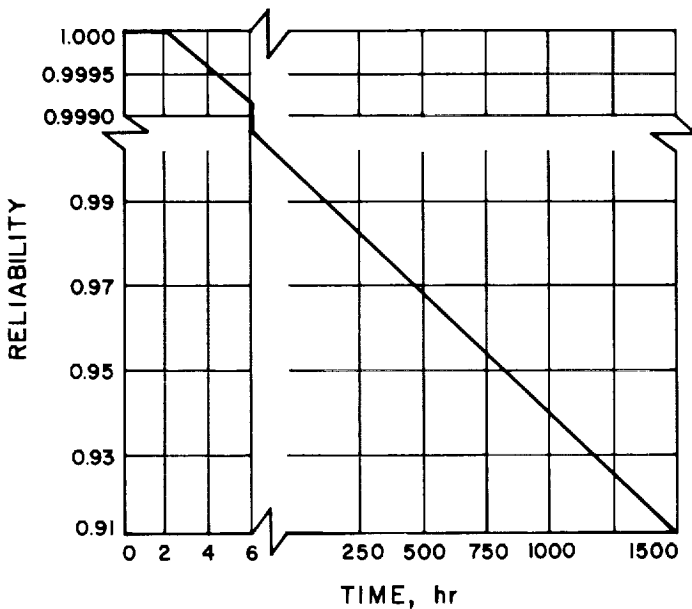
Fig. 6. Electrically suspended gyro reliability estimate

5. Breadboard Gyro and Electronics

As the result of the design and study efforts previously discussed, a breadboard gyro assembly and suspension system was built (see Fig. 7). This gyro used newly designed electrode assemblies to accommodate a 2-in. rotor, the new pickoff assembly discussed in Part 2 of this report, and new spin and damp coil assemblies. The suspension system was built on circuit boards with no effort made toward miniaturization.
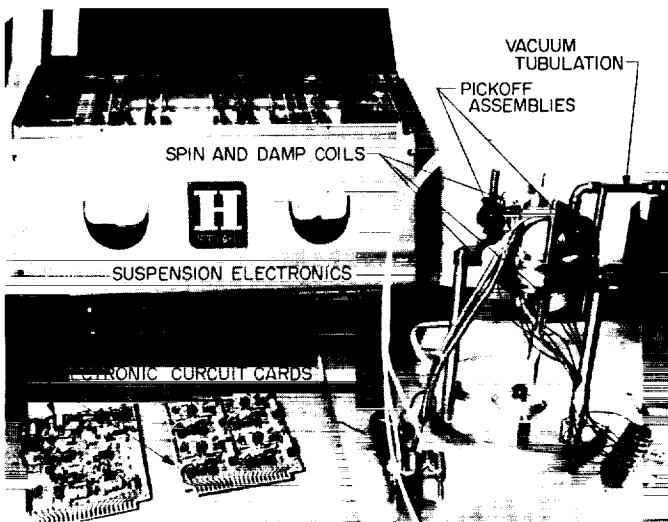
Fig. 7. Breadboard ESG showing suspension system

Testing of the gyro and suspension was very limited under this program. The operation of the suspension system was verified with both 15-g and 4-g capabilities. The suspension was shown to be capable of switching from the 15-g mode to the 4-g mode without losing rotor suspension. Power requirement was measured as 22 w for the 15-g mode, and 6 w for the 4-g mode. The rotor was spun, and the optical pickoffs energized. The pickoff system output was shown to be adequate, and the suspension system held the rotor position well enough to allow readout of the rotor spin axis position to the required accuracy.

6. Current ESG Development

The current development effort in ESG technology is based on the work done in the program described. Areas of effort are as follows:

(1) Build and test an automatic, remotely controlled ESG starting system.

(2) Perform complete evaluation test on the breadboard gyro.

(3) Build and incorporate into the gyro design a new low-power ion pump.

(4) Modify the spin coil structure to increase spinup rate.

(5) Build an additional gyro and additional rotors to insure an undelayed test program.

(6) Build the required auxiliary electronics to record and code the gyro output for the test program.

The progress of current and future development work will be reported in subsequent issues of the SPS.

C. Mechanization of a Strapdown Inertial Navigational System

B. M. Dobrotin and J. C. Nicklas

The Strapdown Electrostatic Gyro Aerospace Navigational System (SEAN) uses three body-mounted velocimeters, two electrostatic gyros and a digital computer. This article presents some of the numerical techniques that can be used in mechanizing the navigation equations.

The basic equation to be implemented in the SEAN system is:

$$\ddot{\mathbf{R}} = -T_I^B \, \mathbf{f}_B + T_I^L \, \mathbf{G}_L \qquad (1)$$

where

$\ddot{\mathbf{R}}$ = position vector of vehicle referred to an inertial coordinate inertial system with the origin at the center of Earth.

T_I^B = transformation matrix from a body-fixed coordinate system (B) to the inertial coordinate system (I).

$\mathbf{f}_B$ = specific force referred to a body-fixed coordinate system = $-a_B + G_B$.

a_B = inertial acceleration.

T_I^L = transformation between Earth-fixed (geographic) coordinate system (L) and the inertial reference system (I).

$\mathbf{G}$ = gravitational force per unit mass (assumed a function of latitude and altitude).

In a simple form, the computer must accept the data from the velocimeters and gyros, compute the transformation matrices, the acceleration and the gravity vector, solve Eq. (1), perform two integrations, and compute the latitude and longitude. A block diagram showing only the flow of basic information is presented in Fig. 8. In comparison with navigation systems using an inertial platform, it is seen that the mechanical complexity of the inertial platform is exchanged for the increased computer requirements. Additionally, the use of a digital computer, instead of an analog system, requires the use of discrete time computation as opposed to continuous computation. However, the use of the digital computer allows the use of digital instrumentation with an attendant increase in accuracy.

Since a digital computer works only in scalar equations and numbers, Eq. (1) must be adapted for solution by a digital computer. Generation of the scalar equation is inherent in the method of obtaining the acceleration in that the velocimeters are mounted in an orthogonal triad. The transformations in Eq. (1) preserve the form of the scalars.

The second portion of the problem of digital computations is that of solving the scalar equations numerically. This portion of the problem is concerned with obtaining the required data from the instrumentation, selecting the proper methods of integration and differentiation, and estimating the computer requirements (e.g., timing, computational errors, and scaling).

The first computational requirement is that of taking the output of the velocimeters and the ESG's and transforming the raw data into the required form. The velocimeters, as the name implies, produce an integrated output, which is the integrated specific force. It is convenient to define the integrated specific force as ΔV:

$$\Delta \mathbf{V}_B \triangleq \int_{t_1}^{t_2} \mathbf{f}_B = \int_{t_1}^{t_2} (-\mathbf{a}_B + \mathbf{G}_B) \qquad (2)$$

Eq. (2), having been integrated in the body-fixed reference frame, has very little physical meaning and cannot
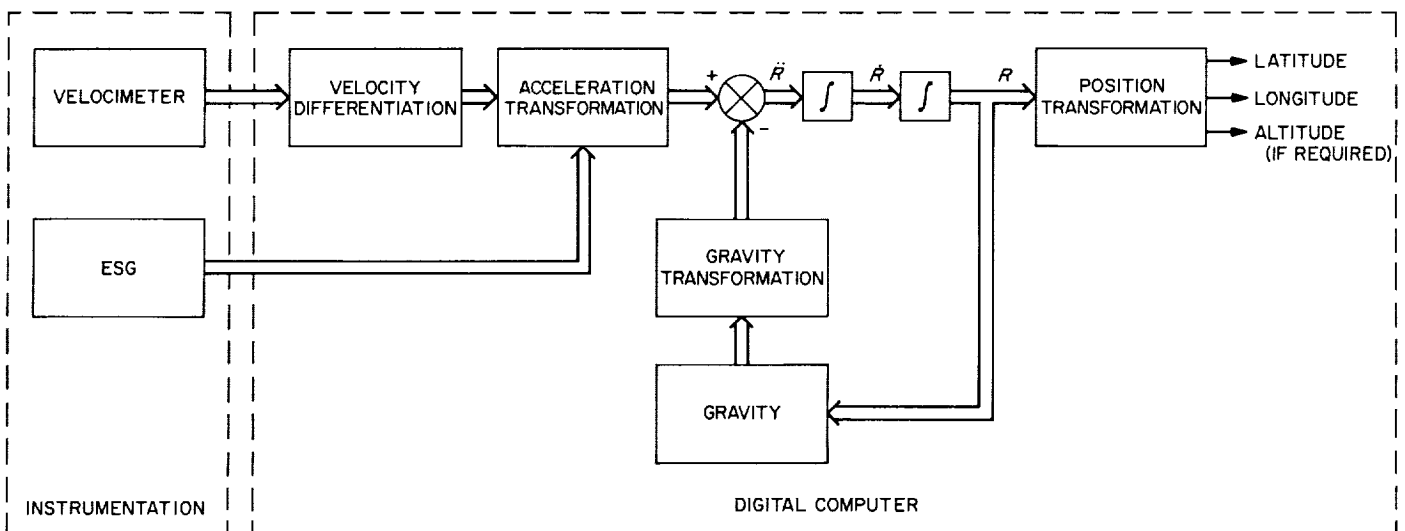
Fig. 8. Strapdown inertial navigation system block diagram

be transformed directly to the inertial coordinate system:

$$\int \ddot{R} = - \int T_I^B f_B + \int T_I^L G_L \qquad (3a)$$

$$\neq - T_I^B \int f_B + T_I^L \int G_L \qquad (3b)$$

Eq. (3b) is true only if the transformation matrices are constant. Since this is not generally true, the available specific force data, as shown in Eq. (2) must be changed to the form required by Eq. (3a). This requires that Eq. (2) be differentiated by some numerical method.

The technique selected for both this differentiation and the subsequent integrations is that of polynomial fitting. Briefly, this samples a function of time at finite intervals to generate a table of discrete data points. Next, a polynomial of any desired degree is fitted to the data points and the required operation (integration, differentiation, interpolation) is performed on this polynomial. In the use and tabulation of the data, two approaches are available. The first, known as the Lagrangian method, uses the sampled value (ordinate) of the function directly. The second method uses differences rather than the ordinate. The construction of a difference table is shown in Table 3.

where Δt is the sampling interval. For Eq. (4) to be exact, all data must be completely accurate and the entire infinite series must be used. If the data contains some inaccuracy, a roundoff error is introduced. Additionally, if the infinite series is truncated at the nth order difference, the terms remaining will form a truncation error. Both of these errors must be estimated in order to arrive at the value of Δt and the highest order difference to be retained. Unfortunately, as higher order differences are retained, any roundoff error present in the initial data is magnified. This is shown in Table 4, where all data is presumed accurate with the exception of $f(t_0)$ which contains an error $+\epsilon$ and $f(t_{-1})$ which contains an error $-\epsilon$. This represents a practical situation, as a digital system is usually accurate to within $\pm$ one count. However, if the truncation error is to be reduced, higher order differences are required. If both the truncation error and the roundoff error can be estimated, it is possible to find both the optimum sample interval and the maximum total error. At the present time it is difficult to assess the errors involved, and Eq. (4) has been truncated at the second difference (∇^2):

$$f_s(t_0) = \frac{1}{\Delta t}\left(\nabla + \frac{1}{2}\nabla^2\right)V + \frac{\Delta t^2}{3}V'''(\xi) \qquad (5)$$

Table 3. Backward difference table

Time	f(t)	Difference			
		1st	2nd	3rd	4th
t_{-4}	$f(t_{-4})$				
		$\nabla f(t_{-3})$			
t_{-3}	$f(t_{-3})$		$\nabla^2 f(t_{-2})$		
		$\nabla f(t_{-2})$		$\nabla^3 f(t_{-1})$	
t_{-2}	$f(t_{-2})$		$\nabla^2 f(t_{-1})$		$\nabla^4 f(t_0)$
		$\nabla f(t_{-1})$		$\nabla^3 f(t_0)$	
t_{-1}	$f(t_{-1})$		$\nabla^2 f(t_0)$		
		$\nabla f(t_0)$			
t_{-0}	$f(t_0)$				

Table 4. Error progagation in difference table

Time	f(t)	Difference		
		1st	2nd	3rd
t_{-3}	0			
		0		
t_{-2}	0		$-\epsilon$	
		$-\epsilon$		4ϵ
t_{-1}	$-\epsilon$		3ϵ	
		2ϵ		
t_0	$+\epsilon$			

The velocimeter data is particularly suited for the use of a difference table since the first difference can be obtained directly from the velocimeter output. The remaining portion of the difference table is easily formed by the computer. The differential, f_s, may then be found from the infinite series:

$$f_s(t_0) = \frac{1}{\Delta t}\left(\nabla + \frac{1}{2}\nabla^2 + \frac{1}{3}\nabla^3 + \frac{1}{12}\nabla^3 \cdots\right)V \qquad (4)$$

where the second term on the right-hand side of Eq. (5) represents the truncation error. $V'''(\xi)$ represents the third derivative of the velocimeter output evaluated at some time, ξ, which is contained in the sample interval. It is seen that the truncation error is $0(\Delta t)^2$.

Once the specific force is available, it may be transformed into the inertial coordinate system by the transformation matrix defined by the ESG's. After combining the gravity components with the specific force, the computer must then integrate Eq. (1) to obtain $\dot{R}(t)$ and $R(t)$. Again, a variety of numerical procedures are available.

Here as in differentiation, only those integration methods using differences are considered. This is done for several reasons. First, the difference tables are easy to generate by a computer using the index registers. Additionally, procedures are available to allow either successive integration or direct double integration of the second derivative (using the same difference table). Either form may be used depending on the information desired. If $\dot{R}$ as well as R is required, the following formulas may be used:

$$\dot{R}(t_0) = \dot{R}(t_{-1}) + \Delta t$$

$$\times \left(1 - \frac{1}{2} \nabla - \frac{1}{12} \nabla^2 - \frac{1}{24} \nabla^3 \cdots \right) \ddot{R}(t_0)$$

(6a)

$$R(t_0) = R(t_{-1}) + \Delta t \, \dot{R}(t_{-1})$$

$$- \Delta t^2 \left(\frac{1}{2} - \frac{1}{6} \nabla - \frac{1}{24} \nabla^2 - \frac{1}{45} \nabla^3 \cdots \right) \ddot{R}(t_0)$$

(6b)

This produces both the inertial position and the inertial velocity. The inertial velocity might be desired, for instance, for damping Eq. (1). If $\dot{R}$ is not desired, then R may be obtained directly from the difference table of $\ddot{R}$:

$$\nabla^2 R(t_0) = \Delta t^2 \left(1 - \nabla + \frac{1}{12} \nabla^2 - \frac{1}{240} \nabla^4 \cdots \right) R(t_0)$$

(7a)

and

$$R(t_0) = \nabla^2 R(t_0) + 2R(t_{-1}) - R(t_{-2})$$

(7b)

Regardless of which equation is selected to perform the integration, some estimate must be made of the truncation error. One guideline is that the truncation error introduced by differentiating the ΔV data is $0(\Delta t^2)$ so that an integration procedure corresponding to Simpson's rule, which gives a truncation error $0(\Delta t^5)$ should be satisfactory. This requires the use of the second difference. Assuming that the error introduced by repeated integration is less than the error which would be introduced by two separate integrations, the error term for Eqs. (6b) and (7a) is

$$E_{R(t)} = \frac{\Delta t^5}{40} \left[R^{(6)}(\xi) + R^{(5)}(\eta)\right]$$

(8a)

and for Eq. (6a)

$$E_{\dot{R}(t)} = \frac{\Delta t^5}{40} R^{(6)}(\phi)$$

(8b)

where $R^{(n)}(t)$ is the nth derivative of $R(t)$.

If it is assumed that the higher order derivatives of R decrease in magnitude quite rapidly, i.e., $R^{(6)}(t) \ll R^{(5)}(t)$, which also implies that $R^5(t)$ is nearly constant, the truncation error in $R(t)$ reduces to

$$E_{R(t)} \text{ (integration)} = \frac{\Delta t^5}{40} R^{(5)}(\eta)$$

(9)

This may be compared to the truncation error introduced in f_B by numerical differentiation, which when expressed in terms of $R(t)$ is:

$$E_{R(t)} \text{ (differentiation)} = \frac{t^2}{3} R^{(5)}(\mu)$$

This indicates that numerical differentiation should be avoided whenever possible.

Examining the scalar equations obtained from Eq. (1), it is seen that they are:

$$\ddot{R}_x + \frac{G_x}{R_x} R_x = f_x$$

$$\ddot{R}_y + \frac{G_y}{R_y} R_y = f_y$$

(10)

$$\ddot{R}_z + \frac{G_z}{R_z} R_z = f_z$$

where the subscripts x, y, and z are referenced to the basis vectors of the inertial coordinate system. The equations contained in (10) are seen to be harmonic oscillators with a forcing function. The solution to these equations will have two parts: the particular solution (due to the forcing function) and the complementary function (as a result of initial conditions). If the initial conditions are not exactly equal to zero, a sustained oscillation will result with a frequency of $(G/R)^{1/2}$ or a period of ~ 84 min. This effect is known as the Schuler effect. Since Eqs. (10) are linear, the origin may be shifted to any time with resultant initial conditions. Thus any error introduced into the system at any time will create a sustained oscillation. This includes computation errors (both truncation as well as roundoff) which become indistinguishable from the instrumentation errors. Either these errors must be tolerated, with an attendant degradation in accuracy, or damping by means of external information must be provided.

Reference

1. *Strapdown Electrostatic Gyro for Spacecraft Application,* Final Report, Contract 950607, Documents 1726-FR 1 and FR 2, Honeywell, Inc., 2600 Ridgeway Road Minneapolis, Minnesota, December 30, 1964.

VI. Guidance and Control Research

A. Flux Reversal in Shorted Superconducting Coils

J. T. Harding

The magnetic field which supports the rotor of the cryogenic gyroscope being investigated at JPL is produced by currents flowing in a pair of shorted, superconducting coils. A difficulty in working with such coils is the long time required to initiate, terminate or reverse their fields. These operations typically involve heating a portion of the shorted circuit while supplying the desired current. The time constant for the alteration of currents is equal to the inductance of the coil divided by its resistance. However, the resistance of a short segment of wire heated to the normal state is very tiny, so that time constants may be of the order of seconds.

In testing the cryogenic gyro it is useful to be able to reverse the field which levitates the rotor rapidly enough so that the rotor will not fall in the process. By causing the entire coil to go normal at once it has been possible to obtain flux reversals in a matter of milliseconds. The following procedure has proved successful.

Two geometrically similar sets of superconducting coils are used. (See Fig. 1.) The two coils in each set are connected in series. The outer set (1) is supplied with current

by leads, A; the inner set (2) is shorted upon itself at B and is fitted with a heater, C. When a current is fed into (1), a current is induced in (2) which tends to prevent any change in the field in the region interior to (2). If

a_1 = 2.64 cm
a_2 = 3.74 cm
N_1 = 700 TURNS EACH
θ = 49° (MAXWELL GRADIENT PAIR CONFIGURATION)
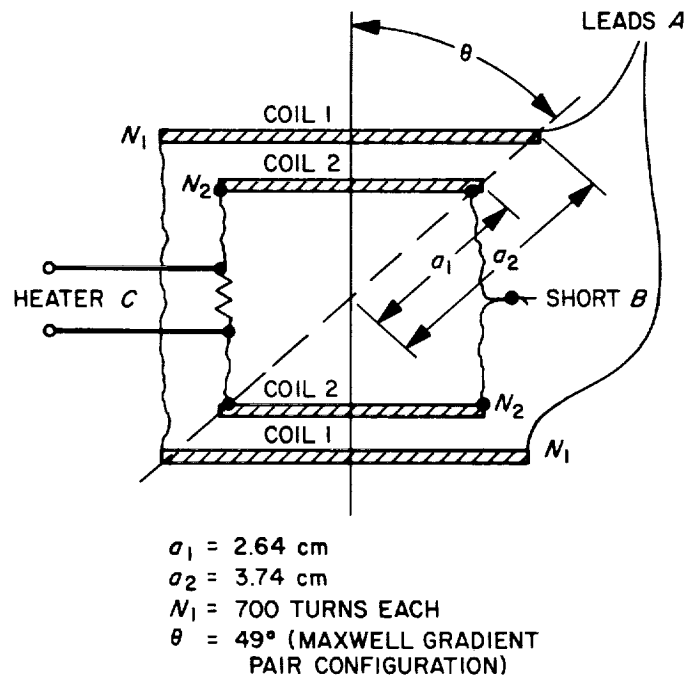
Fig. 1. Coil geometry

the heater is activated, flux due to (1) will leak into (2). (If a superconducting ball is located within (2) it will levitate at this point.) Now, if the heater is turned off and the current in (1) reduced to zero, coil (2) will tend to maintain the field constant (and the ball remains levitated). Next, the current is reversed in (1) and increased until the induced current in (2) exceeds the critical value and coil (2) suddenly goes normal and admits the field produced by (1). (If the reverse current is adequate and the time constant short enough, the ball will remain levitated.) After waiting a brief time for cooling to restore coil (2) to superconductivity, the reverse current in (1) may be reduced to zero while (2) maintains the reversed field. This completes the flux-reversal procedure.

In practice it has required a certain amount of cut and try, to achieve uninterrupted levitation. If there are too few turns of wire on (2) this coil will break down at too small a value of current to support the ball; if too many turns, the coil will not break down at all with the field available from (1). The use of a high-critical-field superconducting alloy accelerates the reversal: alloys have higher normal resistance than pure metals, while the higher critical field permits the use of fewer turns which means less inductance. Coils (1) and (2) were wound of 10-mil NbZr wire.

Data were obtained for several inner coils differing only in their number of turns. The procedure consisted

of trapping a field in (2) due to a current I_T supplied to (1). The current in (1) was reduced to zero, then reversed until (2) broke down. The value of reverse current in (1) needed to cause breakdown is denoted I_B. For a given coil, data indicate the sum $I_T + I_B$ is approximately constant. A 120-turn copper search coil wound adjacent and similar to the inner coil was used to measure the rate of change of flux in the inner coil. Oscilloscope traces of the transition (with the ball absent) are shown in Fig. 2 for three different inner coils for which $N_2 = 65$, 100 and 220 turns. Since the integral of the voltage is proportional to the total flux change, it is evident that the greater the number of turns the greater the flux that could be sustained. However, with more turns there is an increase in the transition time. Only the 220-turn coil could support enough flux to sustain a 1-in. niobium ball through a flux reversal. When using this coil, the field has been successfully reversed without dropping a suspended rotor.

B. Velocity of Sound in Liquid Helium Near 1°K

W. M. Whitney

The properties of liquid helium below 2.2°K can be related to a specific microscopic model in which it is assumed that the thermal energy resides in two types of localized excitations: phonons, which are quantized density fluctuations or sound waves; and rotons, which are correlated motions of a few helium atoms over regions several angstroms wide. Similar models, although with different types of energy carriers involved, are used in the description of the properties of many other types of materials, e.g., semiconductors (electrons and holes), metals (plasmons and quasi-particles), molecular crystals (excitons), and ferromagnetic materials (magnons); and the mathematical techniques and physical ideas invoked in the elaboration of these models have much in common. The phonon-roton model successfully predicts the equilibrium thermal properties of liquid helium from 1.6°K down to very low temperatures, and accounts for the transport properties above 0.8°K, but several types of measurements indicate that dissipative processes below this temperature are not well understood. We are studying the characteristics of sound propagation below 1°K to obtain experimental information and to develop ideas that will contribute to the solution of this problem.

Fig. 2. Oscilloscope traces of the flux reversal

As one part of the program, we are analyzing measurements made in collaboration with C. E. Chase (National Magnet Laboratory, M.I.T.) of the velocity of sound in liquid helium at 1.0, 3.91, and 11.9 Mc/sec, preparatory to publishing the results. This work shows that the velocity depends upon frequency over a temperature interval that extends from approximately 0.4°K to just above 1.1°K. To emphasize the variation with frequency, we subtract the velocity values at 1.00 Mc/sec from those at 4 and 12 Mc/sec and plot the (smoothed) differences against temperature in Fig. 3. These curves show that the dispersion is maximum near 0.9°K. Since it is here that the attenuation also reaches its highest values, it is a reasonable conjecture that both phenomena have the same microscopic origin. Very general physical considerations, such as those embodied in the Kramers-Kronig dispersion relations (Ref. 1), require, in fact, that the

change in absorption be reflected in the velocity. The question to be discussed is thus whether the amount of dispersion measured is compatible with what is already known about the absorption (Ref. 2).

Khalatnikov has developed a theory of sound propagation in liquid helium II (Refs. 3 and 4) that accounts quite successfully for the variation of the attenuation coefficient with temperature, frequency, and pressure above 0.9°K (Refs. 5 and 6). The wave velocity has not been explicitly determined, but an expression for it is easily derived from his equations. We will show here that calculated values of the velocity differences at temperatures near and above the peak compare well with the experimental values presented in Fig. 3. On the low-temperature side of the velocity maximum, however, the measured dispersion, like the measured absorption, is greater than predicted.
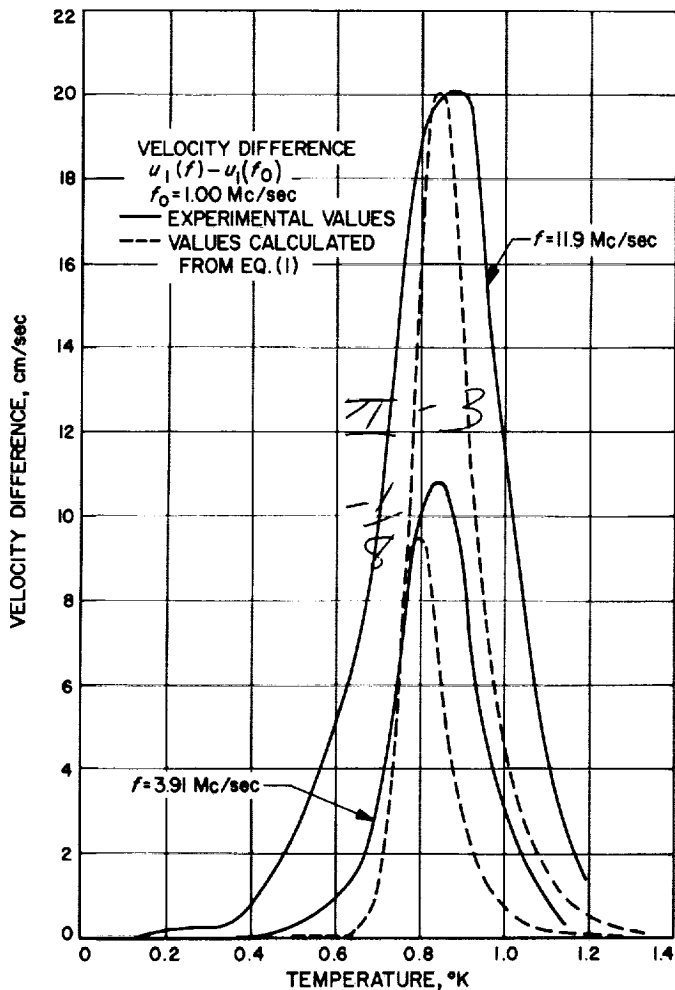
In ordinary fluids, and also in liquid helium II at frequencies not too high and temperatures not too low, the absorption of sound can be described in terms of three phenomenological transport coefficients: (1) attenuation arising from the flow of heat between regions of the sound wave where the temperature is high to regions where it is low is proportional to the thermal conductivity κ; (2) when a fluid is compressed unidirectionally, the local stress is initially anisotropic; energy dissipation resulting from the collisions that restore hydrostatic pressure is proportional to the shear viscosity η; (3) in many liquids and gases, there are degrees of freedom, such as translation and vibration in a gas of polyatomic molecules, between which energy can be exchanged. When such a fluid is compressed without shear, a finite time is required for the equilibrium pressure to develop. Pressure and density changes induced by a sound wave will be slightly out of phase with each other, and some energy will be expended by the wave to heat the fluid. Sound absorption by this mechanism is related to a so-called coefficient of second viscosity ζ.

In liquid helium II near 1°K, the combined sound absorption from thermal conduction and from shear viscosity is an order of magnitude smaller than what is measured. Khalatnikov attributes the bulk of the absorption to second viscosity, and assumes that it is the numbers of phonons and rotons per unit volume, N_{ph} and N_r, that represent the internal degrees of freedom. When the fluid is compressed, the equilibrium number densities are altered: N_{ph} decreases and N_r increases. Changes from the previous equilibrium densities to the new ones are effected by inelastic collisions among the excitations,

Fig. 3. Velocity differences plotted against temperature of liquid helium

principally those in which three phonons merge to form two — the five-phonon process — and in which a phonon and a roton interact to produce two rotons. Following rarefaction, the inverse processes occur. The cross-sections for these higher-order scattering processes will be small in comparison with those for the elastic collisions that promote viscosity and heat conduction. The interpretation of the absorption peak near 1°K is that here the characteristic relaxation times for changes in the number densities N_{ph} and N_r are long enough to be comparable with the periods of the sound waves.

From equations given by Khalatnikov, it is possible to derive the following expression for the wave velocity $u_1(\omega)$ at angular frequency ω:

$$u_1(\omega) = u_1(0) + \omega^2 \frac{(\omega^2 - \omega_1^2)\,\Delta U_{max} + \omega_1^2\,\omega_2\,(\zeta_2/2\rho u_{10})}{(\omega_1^2 - \omega^2)^2 + \omega^2\,\omega_2^2} \tag{1}$$

where ρ is the density and u_{10} the sound velocity at $T = 0°K$. The symbol ΔU_{max} denotes the maximum velocity difference between zero and infinite frequency and ζ_2 the coefficient of second viscosity; these two quantities and the characteristic frequencies ω_1 and ω_2 are related to various thermodynamic parameters for the liquid and to two transport coefficients which Khalatnikov obtained from an analysis of the microscopic collision processes. These expressions are complicated and we omit them here. A description of the manner in which they have been evaluated will be given in a later article.

In Fig. 3, values of the differences $u_1(11.9 \times 10^6) - u_1(1.0 \times 10^6)$ and $u_1(3.91 \times 10^6) - u_1(1.0 \times 10^6)$, computed from Eq. (1), are plotted against the temperature. In certain respects the agreement between these curves and those derived from our measurements is good. The theory appears to account quite well for the height and location of the dispersion maximum, and for an upward shift in its position as the frequency increases from 4 to 12 Mc/sec. There are, however, some quantitative discrepancies. The observed maxima are broader than the predicted ones, and their peak values are located 0.05°K higher. This temperature interval is small, but it encompasses a change in the characteristic frequencies ω_1 and ω_2 by a factor of two. On the high-temperature side of the peaks, the measured velocity differences are considerably greater than the calculated ones, although it is here that errors in the velocity measurement are large, and it may be premature to demand better agreement.

Below 0.7°K, the two sets of curves differ both qualitatively and quantitatively. The calculated velocity differences approach a high-frequency asymptote at low temperatures, and below 0.75°K one would expect such a small change in velocity from 4 to 12 Mc/sec that we could not measure it in our experiment. Our computations show that the maximum dispersion $\Delta U_{max} \simeq 0.05$ cm/sec at 0.5°K, decreasing exponentially at lower temperatures. The measured velocity differences at 4 and 12 Mc/sec are in fact still well separated below 0.5°K, and at this temperature the change in velocity over the frequency interval 4 to 12 Mc/sec is more than 2 cm/sec, 40 times greater than that predicted for the entire range from zero to infinite frequency.

It is at low temperatures that our measurements are the most reliable, and we believe that, small as they are, the discrepancies between the calculated and predicted velocities are significant. Our velocity measurements thus support the conclusion reached previously on the basis of attenuation measurements alone: that the theory of Khanlatnikov provides a good description of sound propagation in the neighborhood of 0.8°K and above, but is inadequate at lower temperatures.

In an earlier report (SPS 37-30, Vol. IV, pp. 64–66), evidence was presented that thermal conduction makes a small contribution to the absorption of sound in liquid helium over the temperature range 0.6 to 0.8°K. The effect of this mechanism on the dispersion will be discussed at another time.

C. The Electro-Optic Effect in BaTiO₃

A. R. Johnston

1. Review of Results

The experimental study of the electro-optic effect in barium titanate has continued since an earlier report of preliminary results (SPS 37-26, Vol. IV, p. 52). This experiment was designed to yield the optical response to an electric field applied along both principal directions in single crystals of $BaTiO_3$. The contribution of the piezoelectrically induced strain to the observed optical response was directly measured. Most measurements of the electro-optic effect in other materials have been made under steady-state, or effectively steady-state, conditions without separating the direct field effect from the strain-coupled part. It is believed that these measurements,

which specifically determined the strain contribution, will provide a sound basis for an understanding of the phenomenon in $BaTiO_3$. At the same time, the data are useful for potential future application of the material as electro-optical modulators. The experimental techniques which were developed for the study of $BaTiO_3$ can be used to obtain similar data for other materials.

Data were obtained both for an [001] applied field, E_c, yielding r_c, and for a [100] field, E_a, yielding r_{42}, over a temperature range covering the tetragonal (ferroelectric) structure in $BaTiO_3$, 5°C to 115°C. Here, the conventional definitions were used; $\Delta(1/n_i^2) = r_{ij}E_j$, where E_j is the jth component of the electric field, n_i is the ith refractive index, $i = 1$ to 6, r_{ij} is the constant which relates the magnitude of the index change Δn_i to the applied field. Nye (Ref. 7) is a good reference for these definitions. For this work, $r_c = r_{33} - (n_a/n_c)^3 r_{13}$ is used to express the total observed response to an [001] field, since r_{33} and r_{13} are not separated. At least three crystals were used for each orientation, in order to avoid problems from domain structure and crystal defects. All samples were flux-grown butterfly wings.[1] Each crystal was poled to a single-domain structure using procedures similar to those described earlier (SPS 37-20, Vol. IV, p. 24). Separate determinations were made first under conditions of constant stress (low frequency) and then at constant strain (high frequency).

The strain contribution to r_c at room temperature was found to be ⅘ of the total low-frequency response, and for r_{42} the strain part was ½ of the total. Thus the strain-coupled response is an important part of the total effect, and should be taken into account if the material is used in modulators operating at high frequency ($\gtrsim 5$ Mc). The half-wave voltage was roughly 400 v at low frequency and 2 kv at zero strain. The optical response per unit of induced polarization, x_{ij}, was found to be constant with temperature in all cases except for E_c at low frequency. The absolute magnitude of x_{ij} was normal, compared to the ~0.05 meters squared per Coulomb (m^2C^{-1}) found in most other materials, ranging from 0.75 m^2C^{-1} to 0.03 m^2C^{-1} in $BaTiO_3$. In addition, the quadratic response above the Curie transition at 120°C was observed, and a large strain contribution was found, even though the structure is centro-symmetric and, therefore, nonpiezoelectric in this state.

The basic polarimeter used, in which a Faraday cell modulator was incorporated to obtain an electronic signal for nulling, originated with Hardy at M.I.T. (Refs. 8

[1] Obtained from Harshaw Chemical Co., Cleveland, Ohio.

and 9). Some modifications were made to the previously described polarimeter for the low-frequency measurements and the apparatus used is described in Ref. 10. The details of the low-frequency measurements on $BaTiO_3$ have been reported recently in Ref. 11 and will not be repeated here.

The strain-free effect was determined by observing the transient optical response to an applied field in the form of a fast-rising step. The response was observed on a nanosecond time scale, and the initial response, before the crystal could develop any strain in response to the suddenly applied field, indicated the magnitude of the electro-optical effect at constant strain. These results have also been recently reported in Ref. 12, so are not described here, but the experimental method will be described below.

A paper which describes the nature of some of the possible devices using electro-optic modulators, and which also relates the results for $BaTiO_3$ mentioned above to the previously published data on the electro-optic effect in other materials, has also been prepared (Ref. 13). An effort will be made through future experiments to supplement this presently available data, which cover about 25 materials. However, only a few basic types are represented, if similarly behaving isomorphs are grouped together. Since no really satisfactory material that is a significant practical improvement on KDP has been found to date, broadening the available experimental basis is an important factor both in seeking better materials empirically, and for obtaining better basic understanding of their behavior.

2. The Experimental Method

As mentioned above, the zero strain condition was obtained by observing the transient response of a single-domain sample to an applied field step. The technique which was developed to do this is new. A sampling oscilloscope provided the required time resolution, and at the same time, the sampling mode made possible the averaging which was necessary in order to recover the desired signal from noise, by using a straightforward phase-coherent detection scheme.

The optics shown in Fig. 4, a conventional polarimeter with a Senarmont compensator and Faraday cell modulator, served for r_c. This configuration is useful for determining changes in retardation measured with respect to fixed axes. Two crystal specimens are introduced only to compensate their large spontaneous birefringence. This
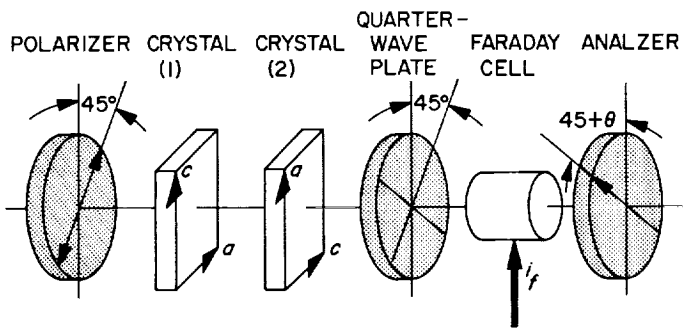
Fig. 4. The optics used for determining r_c

configuration produces linearly polarized light in the region where the Faraday cell is located. The plane of polarization at this point is rotated from its original plane through an angle θ which is proportional to the total retardation δ introduced by the crystal; $\theta = \delta/2$.

In the conventional photoelectric photometer, the Faraday cell is used to wobble the plane of polarization in order to obtain an error signal that can be used for nulling the analyzer prism manually. Indeed, this could also be done in the transient case by sampling the photomultiplier output at a predetermined time after the field step reached the crystal, as mentioned in Ref. 11. The sampled photomultiplier signal could then be filtered, the cutoff frequency of the filter being determined by the noise rejection required. Although possible in principle, this method would be entirely too slow to permit tracing out the time behavior of the transient by repeated measurements with the sampling time varied.

The technique actually used made use of phase coherent sampling of the photomultiplier output in an equivalent manner to reject noise and to provide time resolution, but allowed much faster observation. To do this, the analyzer was rotated through an arbitrary angle θ_0 from the null position observed with no field applied to the crystal, in order to obtain a variation in light intensity at the photomultiplier which is proportional to the change in angular orientation of the plane of polarization. The orientation of the analyzer is not varied. The sampling oscilloscope is then used as the error sensor, and the Faraday cell as the feedback element in a feedback loop. The sampling oscilloscope is triggered alternately, first at the time the field is applied to the crystal, and then midway between pulses. If a difference in light level exists between these two times, then the output of the oscilloscope will change from one level to the other alternately, producing an AC signal. The amplitude of this signal indicates the difference in light level existing at these two times, and therefore also the angular rotation

of the plane of polarization between sampling times. In simplest terms, the oscilloscope output is amplified, and used to drive the Faraday cell. The amplitude of the alternating rotation produced by the Faraday cell is subtracted from that produced by the induced retardation of the crystal, and is automatically adjusted by the feedback loop so that the intensity difference seen by the photomultiplier is zero. This means the rotation produced by the Faraday cell must be equal to that produced by the crystal, and can be quantitatively measured by observing the Faraday cell current.

The light intensity transmitted by the analyzer is

$$I = I_0 \sin^2 \theta; \quad \theta = \theta_0 + \Delta\theta$$

where θ is the instantaneous orientation of the plane of polarization, measured from the null, and θ_0 is the fixed offset mentioned above. Taking the differential of the above expression, for small θ_0

$$\Delta I = -2 I_0 \theta_0 \Delta\theta.$$

ΔI is the signal which is effective in driving the loop, while I is the average DC light level on the photomultiplier. A small change in the orientation of the light incident on the analyzer $\Delta\theta$ therefore produces a proportional change in the light level at the photomultiplier.

A block diagram shown in Fig. 5 illustrates the operation of the feedback loop. The parts within the dotted line are located on the polarimeter, as seen in Fig. 4. The photomultiplier and sampling oscilloscope must be capable of faithfully following the fast changes in light intensity. The pulser is a mercury-switch type, delivering a pulse up to 500 v in amplitude to the crystal, the actual voltage being chosen to produce a convenient rotation, typically 1 to 3 deg.

The timing circuits provide two identical trigger pulses which are used to trigger the oscilloscope. The first trigger pulse occurs as close as possible to the start of the high-voltage pulse. The second trigger pulse occurs at an arbitrary, but fixed, time after the first trigger pulse, such that it lies midway between consecutive high-voltage pulses. The second trigger pulse occurs when there is no voltage applied to the crystal. A flip-flop whose output is used as the reference signal for the demodulator and modulator is also driven by these pulses. The oscilloscope sweep is controlled manually; the timing of the trigger pulse is fixed. Fig. 6 illustrates the timing of these events. The pulser output, at a frequency near 500 pps, determined by the mercury relay, is shown on the first line.
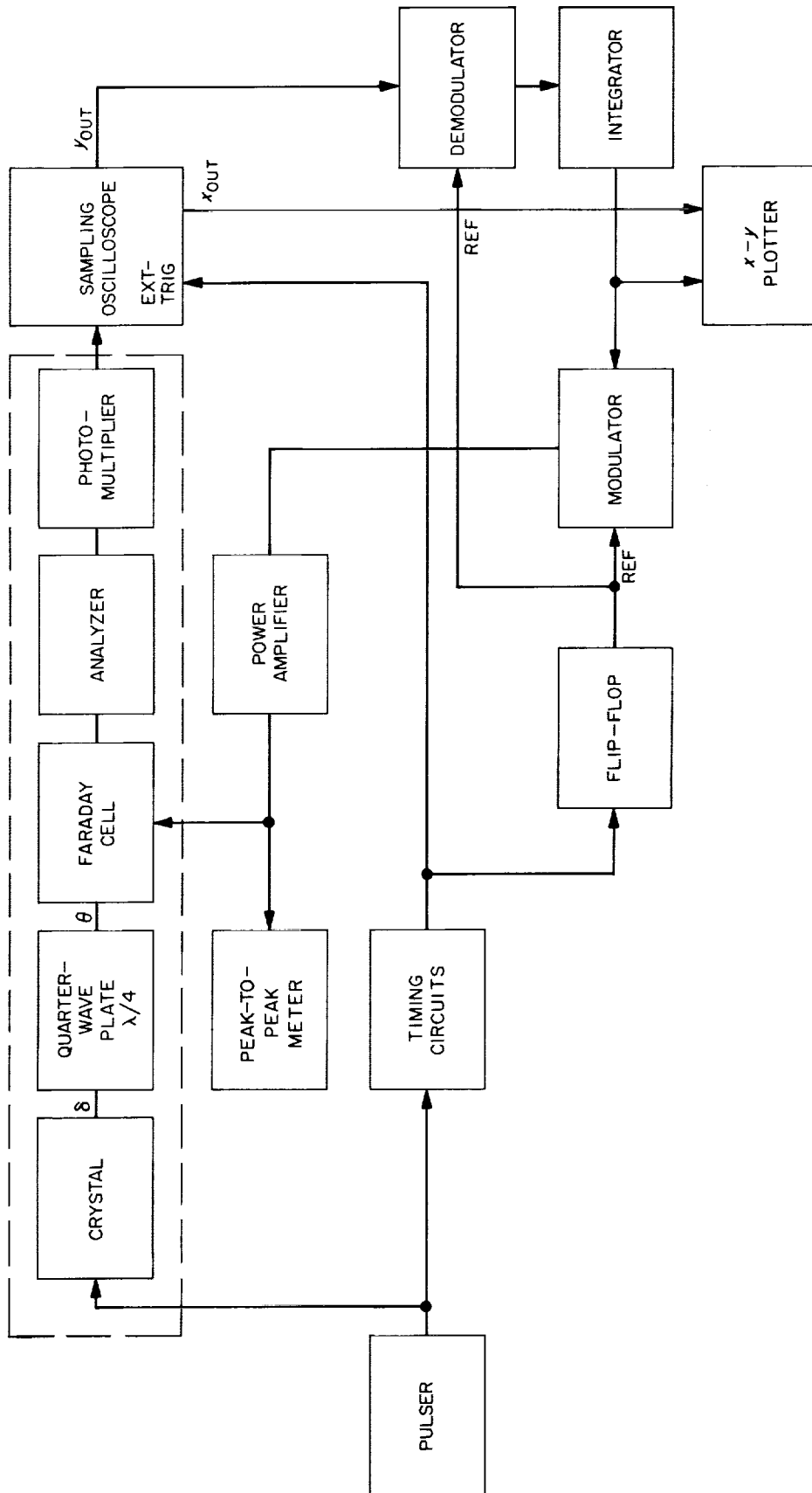
Fig. 5. Block diagram illustrating the operation of the apparatus
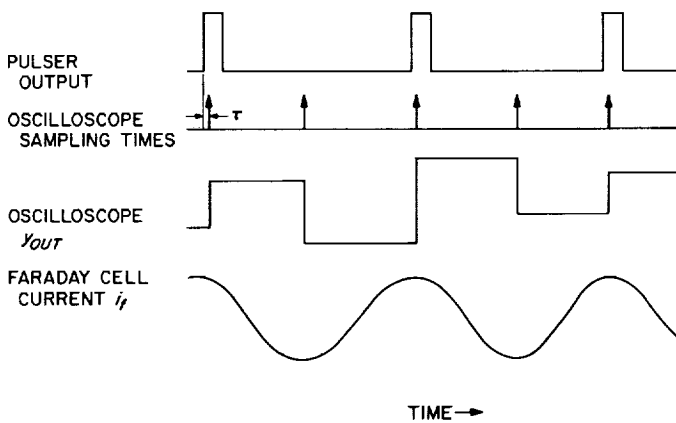
Fig. 6. Diagram showing timing of the events during a typical observation

The width of each pulse is greatly exaggerated in order to be visible in the figure. The actual pulse width was 100 ns. The photomultiplier output is sampled at time τ_s measured from the leading edge of each pulse as shown in the second line. τ_s is controlled manually by the internal sweep circuits of the oscilloscope. The oscilloscope is also triggered midway between the pulses, when there is no voltage applied to the crystal. The oscilloscope thus samples the voltage at the output of the photomultiplier near the time of the first trigger pulse. It then "holds," or provides this fixed voltage as its y-output until the time of the second trigger pulse. The y-output then changes and holds a voltage proportional to the photomultiplier output at that sampling time. The output of the oscilloscope, on the third line, is therefore a noisy audio signal, in phase with the pulser, whose average amplitude is a measure of the difference between the light level existing at time τ_s after the voltage step, and at an intermediate time when the effect of the voltage pulse on the crystal is not present.

The magnitude of the signal part is greatly exaggerated in the figure. Actually, the noise would dominate, even with the loop opened; and in operation the coherent signal is driven to zero, leaving only noise. Conventional demodulation, with an integrator to reject the noise, and to provide a large DC loop gain, and remodulation, are used to obtain the current to drive the Faraday cell, shown in the last line of Fig. 6. The Faraday cell is thus operated at a low frequency, the pulse repetition rate. Maximum deflection of the plane of polarization by the Faraday cell occurs with alternating sign at each sampling time. The magnitude of the alternating current in the Faraday cell is automatically adjusted in response to the large total gain in the feedback loop to yield the same light level at time τ_s during the pulse as at a time

when the crystal is undisturbed, which implies the same angular orientation of the plane of polarization incident on the analyzer. This procedure is similar to the action of the chopper in a DC amplifier, and eliminates drift of the spontaneous retardation as a byproduct.

Quantitative determination of the retardation induced in the crystal is obtained by calibrating the Faraday cell and measuring the peak amplitude of its drive current. The gains in the loop are large enough that the Faraday-cell calibration is the only parameter that need be known in order to measure δ, the retardation.

The average light intensity at the photomultiplier which could conveniently be attained using a 100-watt high-pressure mercury arc and a 30-Å filter centered on the 5461-Å line roughly corresponds to 10 photoelectrons per nanosecond. The sampling gate in the oscilloscope is open for 1 ns, so it is clear that the photon noise is much larger than the electro-optically induced signal, which typically might be 5% of the total photocurrent. An averaging time, defined as the reciprocal of the angular frequency at which the loop has unity gain, of roughly 0.5 min at the 500-pps rate was typically required.

In practice, the sampling time τ_s was swept very slowly over the transient to be observed, by means of the external sweep input of the oscilloscope. The averaging time was adjusted by changing the capacitor in the integrator (Fig. 5) in order to obtain an acceptable noise level. The sweep rate must always be kept sufficiently slow for the feedback loop to be able to follow changes. A compromise must of course be struck between the noise level in the x-y tracing obtained, and the time necessary to obtain it. A typical tracing for a sample of tetragonal BaTiO$_3$ is shown in Fig. 7, on two time scales.

The shape of the transient during the initial 5 to 10 ns of the upper tracing is entirely due to the photomultiplier. The initial deflection indicated by δ_i is taken as a measure of the strain-free response. The electro-optic response itself is thought to be very much faster than this, being determined by the optical lattice modes of the crystal. At present, the response time, T_R on Fig. 7, of 6 ns available from the apparatus is determined only by the photomultiplier. If desired, an improvement in T_R to the order of 0.5 ns should be possible by substituting a Bohm diode for the photomultiplier, provided a large enough sample and appropriate collecting optics were available to make possible the required increase in light intensity. The lower tracing covers a much longer time interval, graphically showing the strain effect. The vertical scale is also
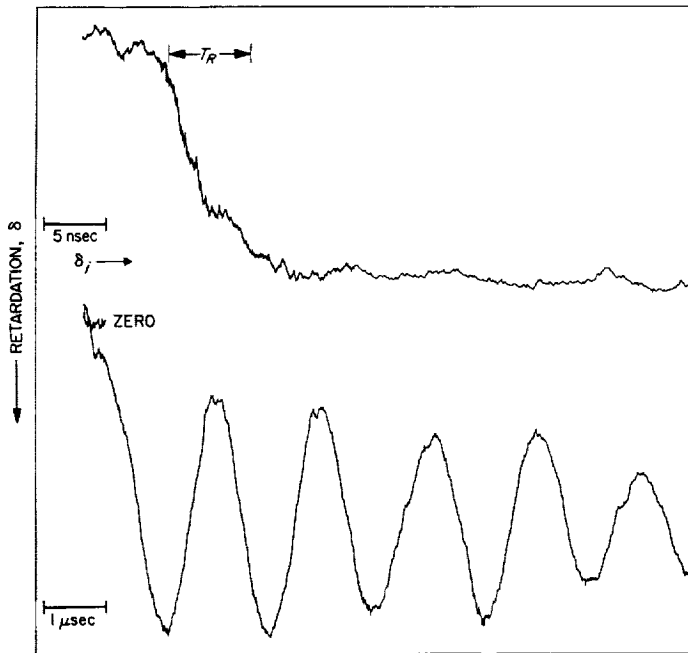
Fig. 7. Typical tracing from x-y plotter showing the induced retardation as a function of time, taken from Ref. 12

compressed, and the slight shoulder at the beginning of the tracing indicates the initial transient shown by the upper tracing. The effect of the mechanical resonance of the crystal is the dominant feature. The observed frequency corresponds with the fundamental stretching mode of the crystal wafer normal to the light path.

The optics for measurement of r_{42} are inherently somewhat different since, here, the direct effect of the applied field is to rotate the optic axes of the crystal. If the technique for observing transient r_c can be thought of as being derived from rotating the analyzer to obtain a null, then the corresponding technique for r_{42} is derived from rotating the sample to achieve a null. This can be done with two opposing Faraday cells as shown in Fig. 8. The
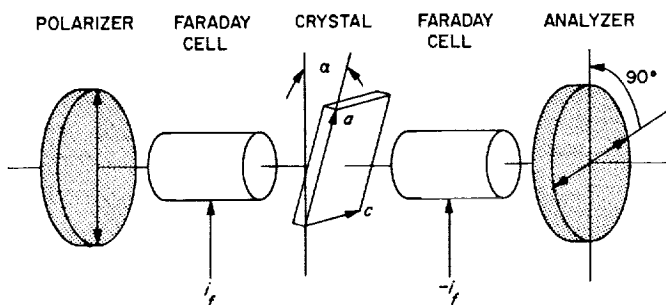
Fig. 8. The optics used for determining r_{42}

two Faraday cells generate an equivalent rotation of the sample placed between them. The peak-to-peak value of the Faraday-cell current was used to measure the quantity r_{42} in a procedure analogous to the one described for r_c.

A paper describing this technique for measuring transient electro-optic effects in detail is planned for future publication.

D. Charge Transport in a Neutral Plasma Such as Exists in an Ideal Thermionic Energy Converter

K. Shimada

1. Introduction

A theoretical analysis of the transport of current-carriers through an electrically neutral plasma was made. The plasma was considered to be weakly ionized and isotropic, and the magnetic field was neglected. Furthermore, the divergence of the net current density was set equal to zero for current continuity. The electric field was determined from the net space-charge density. The analysis indicated that the particle density could be considered to be a superposition of two densities, one which was independent of coordinates and one which depended on the coordinates. The coordinate-independent term determined the current flow through the plasma in conjunction with the mobilities of the charge carriers. The coordinate-dependent term was the result of the ambipolar diffusion of ions and electrons. The distribution of particle density described by this term followed the Boltzmann distribution with an equivalent temperature nearly equal to the electron temperature. However, the coordinate dependence of this term was important only near the plasma edge adjacent to the collector, since it existed only for a few "ambipolar diffusion lengths." Within this thin region, sinks of particles were found.

The physics of cesiated surfaces, which are the most commonly used in thermionic converters, is being extensively studied by many researchers. However, the transport properties of charge carriers in an interelectrode space are not well understood especially in converters which are operated in an ignited mode. Increased efforts in this area are indicated by increased numbers of reports

in recent years. A better understanding of these transport properties could be useful in obtaining larger output voltages in thermionic converters by reducing transport losses. This voltage increase may be achieved if the reduction of plasma losses and the increase of percentage of ionization at larger output voltages can be realized by suitable methods.

As the preliminary phase, the proposed thermionic research task has two-fold goals: (1) an understanding of the physics of carrier transport and (2) a decrease of plasma losses in thermionic converters. Part of the theory derived during the theoretical study of carrier transport is described below.

2. Theory of Carrier Transport

The problem of current conduction through a thermionic diode which is operating in an ignited mode will be divided into three sub-problems, i.e., (1) the emitter sheath problem, (2) the plasma problem, and (3) the collector sheath problem. This division is based upon the different physical principles involved in each region. The emitter not only injects thermionically emitted electrons but also ions which are generated at the hot emitter surface. Consequently there will be a region similar to the Langmuir double-sheath region adjacent to the emitter where the current conduction is achieved free of collisions. The collector collects electrons and ions if the back-emission from the collector can be neglected. The plasma, when volume ionization is taking place, is a region dominated by collisions between neutral particles and charge carriers. Collisions between charge carriers themselves can be neglected in thermionic diodes which have weakly ionized plasmas. The solutions obtained for each region must be matched at the boundaries. The introduction of a pre-sheath between the plasma and the sheath, as a transition region, appears to be inevitable for the boundary conditions to be satisfied. With this in mind, the problem of current conduction through a neutral plasma is studied. The basic assumptions are:

(1) The plasma is electrically neutral, i.e., $n_+ = n_-$.

(2) The plasma is weakly ionized so that the mobility and the diffusion constant can be determined from consideration of elastic collisions between charge carriers and neutrals only.

(3) Electron and ion temperatures are considered constant throughout the plasma.

(4) The plasma is isotropic.

(5) The mobility and the diffusion constants are independent of carrier energies.

(6) The current density is small enough such that the magnetic effect can be neglected.

(7) Current density is continuous, i.e., the sources and sinks of the *net* current density are zero; this means that generation and recombination in the plasma must always occur in pairs.

(8) The electric field is consistent with distribution of the charge density in the plasma.

The last two assumptions are different from assumptions made by other researchers (Refs. 14–17). Typically, the divergence of *each* specie of current carriers is assumed proportional to the electron density, or the divergence of each specie is assumed to be zero. Also, the self-consistency of the electric field is normally revoked, whereas this author assumes the source-sink-less condition for the net electric current which is composed of *both* species of current carriers. In other words, the divergence of *each* specie is equal to one another which is exactly the result of pair generation and pair recombination of particles.

Consider the portion of a plasma bounded by a plane at $x = 0$ and $x = d$ in a one-dimensional model of a large plasma, as shown in Fig. 9.
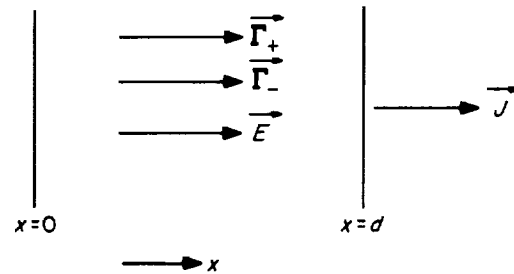
Fig. 9. One-dimensional plasma

Here Γ_+ and Γ_- are the ion and electron flow rates in particles per unit area per second, respectively. $\mathbf{E}$ is the self-consistent electric field satisfying Gauss' theorem. Therefore, we have:

$$\Gamma_+ = n_+\mu_+\mathbf{E} - D_+ \nabla n_+ \tag{1}$$

$$\Gamma_- = -n_-\mu_-\mathbf{E} - D_-\nabla n_- \tag{2}$$

$$\nabla \cdot \mathbf{E} = (1/\epsilon_0) \nabla \cdot \mathbf{D} = (e/\epsilon_0)(n_+ - n_-) \tag{3}$$

(See Table 1 for definitions of terms.) Since a neutral plasma is assumed, we have

$$n_+ = n_- = n \tag{4}$$

Table 1. Definitions of terms

Γ_+ = vector ion flow density, particles/m²-sec

Γ_- = vector electron flow density, particles/m²-sec

μ_+ = ion mobility, m²/sec-volt

μ_- = electron mobility, m²/sec-volt

E = vector electric field intensity, volt/m

n_+ = ion volume density, particles/m³

n_- = electron volume density, particles/m³

D_+ = ion diffusion constant, m²/sec

D_- = electron diffusion constant, m²/sec

e = electron charge, 1.6×10^{-19} coulomb

ϵ_0 = dielectric constant of vacuum, 8.85×10^{-12} farad/m

J = vector electric current density, amp/m²

$V(x)$ = electrostatic potential at x with respect to $x = 0$, volts

k = Boltzmann's constant, 1.38×10^{-23} joules/°K

T_{eq} = equivalent temperature of particles defined by Eq. (43), °K

D_a = ambipolar diffusion constant, m²/sec

μ_a = equivalent mobility defined by Eq. (51), m²/sec-volt

L_a = ambipolar diffusion length defined by Eq. (16), meter

τ_a = ambipolar time constant defined by Eq. (18), sec

E_x = a constant, equal to the electric field existing between $x = 0$ and $x = d$

n_0 = a constant, equal to the net charge in excess of n_1 at $x = 0$

n_1 = a constant, equal to the coordinate independent part of the charge density existing between $x = 0$ and $x = d$

i = unit vector in the x-direction

Thus,

$$\Gamma_+ = n\mu_+ E - D_+ \nabla n \qquad (5)$$

$$\Gamma_- = -n\mu_- E - D_- \nabla n \qquad (6)$$

$$\nabla \cdot E = 0 \qquad (7)$$

Since the current density is continuous, we have

$$\nabla \cdot (\Gamma_+ - \Gamma_-) = 0 \qquad (8)$$

Solving Eqs. (5), (6), (7), and (8) for n (see Sect. 3 for details), we obtain

$$n = n_0 e^{-sx} + n_1 \qquad (9)$$

where

$$s = \frac{\mu_+ + \mu_-}{D_- - D_+} E_x \qquad (10)$$

and n_0, n_1, and E_x are constants to be determined from boundary conditions. The first term in Eq. (9) can be identified as the particle density which is governed by the ambipolar diffusion, and the second term is the particle density which determines the net electric current density J, given by:

$$J = e(\mu_+ + \mu_-)n_1 E_x \qquad (11)$$

Further examination of Eqs. (9) and (10) reveals that the coordinater-dependent part of the carrier density follows the pseudo-Boltzmann distribution described by an equivalent temperature T_{eq}, such that

$$n_0 e^{-sx} = n_0 e^{+\frac{eV(x)}{kT_{eq}}} \qquad (12)$$

where T_{eq} satisfies the Einstein relationship based on the ambipolar diffusion constant D_a and the equivalent mobility μ_a.

$$\frac{e}{kT_{eq}} = \frac{\mu_a}{D_a} \qquad (13)$$

If the current conduction is dominated by the electrons, we have

$$T_{eq} \simeq T_- \qquad (14)$$

and

$$n_0 e^{-sx} = n_0 e^{\frac{eV(x)}{kT_-}} \qquad (15)$$

This is what one expects in the Boltzmann distribution of a single-specie electron system. Furthermore, one sees that the dimension of $1/s$ is a length. Defining the ambipolar diffusion length, L_a, by

$$L_a \overset{\Delta}{=} \frac{1}{s} = \frac{D_- - D_+}{(\mu_+ + \mu_-)E_x}$$

$$= \frac{D_a}{\mu_a E_x} \quad \text{for } E_x > 0. \qquad (16)$$

We arrive at the following for L_a:

$$L_a = \frac{D_a L_a}{\mu_a E_x L_a} = D_a \tau_a \frac{1}{L_a}$$

or

$$L_a = (D_a \tau_a)^{1/2} \qquad (17)$$

where τ_a is defined by

$$\tau_a \overset{\Delta}{=} \frac{L_a}{\mu_a E_x} \qquad (18)$$

Thus, the charge density given in Eq. (9) may also be written as

$$n = n_0 e^{-x/L_a} + n_1 \qquad (19)$$

In a typical plasma where electrons dominate in conduction and in temperature, we have

$$L_a \simeq \frac{D_-}{\mu_- E_x} \qquad (20)$$

Using the Einstein relationship, we obtain

$$L_a \simeq \frac{kT_-}{eE_x}$$

For a plasma in a thermionic converter, let $T_- = 2 \times 10^3$ °K, $E_x = 100$ volt/cm $= 10^4$ volts/meter,

$$L_a \simeq 1.38 \times 10^{-23} \times 2 \times 10^3 / 1.6 \times 10^{-19} \times 10^4$$

$$= 1.72 \times 10^{-5} \text{ meter}$$

$$= 1.72 \times 10^{-2} \text{ mm}$$

$$= 0.68 \text{ mils} \tag{21}$$

The magnitude of the ambipolar diffusion length, thus calculated, is approximately one-third of a typical inter-electrode gap of 2 mils, such as is found in a solar thermionic diode. Substituting the value of L_a obtained in Eq. (21) for $E_x > 0$ into (19), the charge density n becomes:

$$n = n_0 e^{-x/1.72 \times 10^{-2}} + n_1 \tag{22}$$

where x is in millimeters. This is the charge density in the plasma of a thermionic converter of planar geometry. Here $x = 0$ corresponds to the plasma edge near the collector, and $x = d$ corresponds to the plasma edge near the emitter in Fig. 9. In such a case, when $E_x > 0$ and $J > 0$, n_1 must be also positive from Eq. (11). However, n_0 can be either positive or negative as long as n itself in Eq. (22) is positive. If n_0 is positive, n decays monotonically with respect to x from collector to emitter. Examination of the divergence indicates that sinks of particles exist near the collector in such a case. It is shown by

$$\nabla \cdot \Gamma_+ = \nabla \cdot \Gamma_- = -s^2 D_a n_0 e^{-sx} < 0 \tag{23}$$

More careful examination of the boundary conditions should be made together with boundary conditions required by the sheath regions.

3. Derivations

The particle flow rate, the particle densities, and the divergence of particles in a steady-state, electrically neutral, weakly ionized plasma are derived below. Physical interpretations are given in the preceding text.

The equations for the particle flow rates are:

$$\Gamma_+ = \mu_+ n_+ \mathbf{E} - D_+ \nabla n_+ \tag{24}$$

$$\Gamma_- = -\mu_- n_- \mathbf{E} - D_- \nabla n_- . \tag{25}$$

Since the plasma is considered to be electrically neutral, we have

$$n_+ = n_- = n . \tag{26}$$

Thus, from Eqs. (24) and (25),

$$\Gamma_+ = \mu_+ n \mathbf{E} - D_+ \nabla n \tag{27}$$

$$\Gamma_- = -\mu_- n \mathbf{E} - D_- \nabla n . \tag{28}$$

The divergence of the particle flow then becomes:

$$\nabla \cdot \Gamma_+ = \mu_+ \nabla \cdot (n\mathbf{E}) - D_+ \nabla \cdot \nabla n$$

$$= \mu_+ (\mathbf{E} \cdot \nabla n + n \nabla \cdot \mathbf{E}) - D_+ \nabla^2 n . \tag{29}$$

From Gauss' theorem we have,

$$\nabla \cdot \mathbf{E} = (1/\epsilon_0) \nabla \cdot \mathbf{D} = (1/\epsilon_0) e (n_+ - n_-) = 0 \tag{30}$$

Hence, from Eq. (29) we obtain

$$\nabla \cdot \Gamma_+ = \mu_+ (\mathbf{E} \cdot \nabla n) - D_+ \nabla^2 n . \tag{31}$$

Similarly

$$\nabla \cdot \Gamma_- = -\mu_- (\mathbf{E} \cdot \nabla n) - D_- \nabla^2 n . \tag{32}$$

Continuity of the electric current requires the condition:

$$\nabla \cdot J = e \nabla \cdot (\Gamma_+ - \Gamma_-) = 0 . \tag{33}$$

Combining Eqs. (31), (32), and (33), we obtain:

$$(\mathbf{E} \cdot \nabla n) (\mu_+ + \mu_-) - \nabla^2 n (D_+ - D_-) = 0 . \tag{34}$$

For a one-dimensional case, Eq. (34) yields:

$$E_x \frac{dn}{dx} (\mu_+ + \mu_-) - (D_+ - D_-) \frac{d^2 n}{dx^2} = 0 \tag{35}$$

where E_x is the constant electric field in the x-direction which is obtained from Eq. (30). Since (35) is a homogeneous linear differential equation of second order having constant coefficients, the solution can be obtained from its characteristic equation given by:

$$pE_x (\mu_+ + \mu_-) - (D_+ - D_-) p^2 = 0 . \tag{36}$$

Thus,

$$p = 0 \quad \text{or} \quad \frac{\mu_+ + \mu_-}{D_+ - D_-} E_x, \text{ if } D_+ \neq D_- .$$

Therefore, the solution of Eq. (35) is given by

$$n = n_0 \exp \left(\frac{\mu_+ + \mu_-}{D_+ - D_-} E_x x \right) + n_1 \tag{37}$$

or

$$n = n_0 \exp \left(-\frac{\mu_+ + \mu_-}{D_- - D_+} E_x x \right) + n_1 . \tag{38}$$

where n_0, n_1, and E_x must be determined from the boundary conditions.

For simplicity in writing, rewrite Eq. (38) as follows:

$$n = n_0 \exp(-sx) + n_1 \tag{39}$$

where

$$s = \frac{\mu_+ + \mu_-}{D_- - D_+} E_x . \tag{40}$$

The electric potential $V(x)$ at x with respect to the origin is given by

$$V(x) \overset{\Delta}{=} -\int_0^x \mathbf{E} \cdot d\mathbf{x}$$

$$= -E_x x . \tag{41}$$

Combining Eq. (41) with (38) we obtain:

$$n = n_0 \exp\left(\frac{\mu_+ + \mu_-}{D_- - D_+} V(x)\right) + n_1 . \tag{42}$$

Define the equivalent temperature T_{eq} of the particles which participate in the ambipolar diffusion, by the following:

$$\frac{e}{kT_{eq}} \overset{\Delta}{=} \frac{\mu_+ + \mu_-}{D_- - D_+} \tag{43}$$

Then Eq. (42) becomes

$$n = n_0 \exp(eV(x)/kT_{eq}) + n_1 . \tag{44}$$

The first term in Eq. (44) is the part of the particle density which follows a pseudo-Boltzmann distribution.

The particle flow rate can now be expressed in terms of n_0 and n_1. From Eq. (39) into (27), one obtains:

$$\Gamma_+ = \mu_+ n_0 \exp(-sx)\mathbf{E} + \mu_+ n_1 \mathbf{E} + D_+ s\, n_0 \exp(-sx)\, \mathbf{i}$$

$$= n_0 \exp(-sx)[\mu_+ \mathbf{E} + \mathbf{i} D_+ s] + \mu_+ n_1 \mathbf{E}$$

In the one-dimensional case, the above equation reduces to

$$\Gamma_+ = n_0 \exp(-sx)[\mu_+ E_x + D_+ s] + \mu_+ n_1 E_x . \tag{45}$$

Since s satisfies Eq. (40), we obtain

$$\Gamma_+ = n_0 \exp(-sx) \frac{D_- \mu_+ + D_+ \mu_-}{\mu_+ + \mu_-} s + \mu_+ n_1 E_x$$

$$= -\frac{D_- \mu_+ + D_+ \mu_-}{\mu_+ + \mu_-} \frac{dn}{dx} + \mu_+ n_1 E_x$$

$$= -D_a \frac{dn}{dx} + \mu_+ n_1 E_x \tag{46}$$

where D_a is the ambipolar diffusion coefficient defined by

$$D_a \overset{\Delta}{=} \frac{D_- \mu_+ + D_+ \mu_-}{\mu_+ + \mu_-} . \tag{47}$$

In general

$$\Gamma_+ = -D_a \nabla n + \mu_+ n_1 \mathbf{E} \tag{48}$$

Similarly

$$\Gamma_- = -D_a \nabla n - \mu_- n_1 \mathbf{E} . \tag{49}$$

Therefore the electric current density $\mathbf{J}$ is given by

$$\mathbf{J} = e(\Gamma_+ - \Gamma_-)$$

$$= e\, n_1 \mathbf{E}(\mu_+ + \mu_-) . \tag{50}$$

Also from Eqs. (45) and (40) we obtain

$$\Gamma_+ = n_0 \exp(-sx) \frac{\mu_+ D_- + \mu_- D_+}{D_- - D_+} E_x + \mu_+ n_1 E_x$$

$$= n_0 \exp(-sx) \mu_a E_x + \mu_+ n_1 E_x$$

where μ_a is the equivalent mobility defined by

$$\mu_a \overset{\Delta}{=} \frac{\mu_+ D_- + \mu_- D_+}{D_- - D_+} . \tag{51}$$

Combining Eqs. (43), (47), and (51), we obtain the Einstein relationship of particles which are undergoing ambipolar diffusion as follows:

$$\frac{\mu_a}{D_a} = \frac{e}{kT_{eq}} . \tag{52}$$

The divergence of particles can be obtained from Eq. (31) with (39):

$$\nabla \cdot \Gamma_+ = \nabla \cdot \Gamma_- = \mu_+ E_x(-s) n_0 e^{-sx} - D_+ s^2 n_0 e^{-sx}$$

$$= n_0 e^{-sx}(-s)[\mu_+ E_x + D_+ s]$$

$$= -s^2 D_a n_0 e^{-sx} . \tag{53}$$

Eq. (53) shows that there are sinks of particles close to the edge of plasma which is electrically positive with respect to the other edge of plasma.

References

1. Kittel, C., *Elementary Statistical Physics,* Chap. 44, John Wiley and Sons, Inc., New York, 1958.

2. Jeffers, W. A., Jr. and Whitney, W. M., *Physical Review, 139:* A1082, 1965.

3. Khalatnikov, I. M., *Zh. Eksperim. i Teor. Fiz. 20:* 243, 1950.

4. Khalatnikov, I. M., *Zh. Eksperim. i Teor. Fiz. 23:* 21, 1952.

5. Chase, C. E., *Proceedings of the Royal Society* (London), *A220:* 116, 1953.

6. Dransfeld, K., Newell, J. A., and Wilks, J., *Proceedings of the Royal Society* (London), *A243:* 500, 1958.

7. Nye, J. F., *Physical Properties of Crystals,* Clarendon Press, Oxford, 1960.

8. Hardy, A. C., Fopiano, P. J., Trageser, M. B., U.S. Patent No. 2,974,561; 1961.

9. Williamson, S. J., Weingart, J. M., Andrews, R. D., *Journal of the Optical Society of America, 57:* 337, 1964.

10. Weingart, J. M., and Johnston, A. R., *Proceedings of the 1963 Symposium on Ellipsometry,* NBS, Washington, D.C., 1963.

11. Johnston, A. R., and Weingart, J. M., *Journal of the Optical Society of America, 55:* 828, 1965.

12. Johnston, A. R., *Journal of Applied Physics, 7:* 195, 1965.

13. Johnston, A. R., presented at the AGARD Conference on Opto-Electronic Components and Devices, Paris, France, September 6–9, 1965.

14. Hansen, L. K., and Warner, C., "Theory of Cesium Diode Operation in the Arc Mode," to be published in *The Proceedings of the International Thermionic Conference,* London, September 1965.

15. Rasor, N. S., "Analytical Description of Cesium Diode Phenomenology," *Report presented at the 25th Annual Conference, Physical Electronics,* Massachusetts Institute of Technology, Cambridge, Mass., March 1965.

16. Johnson, E. O., "Studies of Externally Heated Hot Cathode Arcs," *R.C.A. Review,* XVI, 498, December 1965.

17. Wilkins, D. R., "The Open Circuit Voltage and Thermionic Converter Diagnostics," *Report of the Thermionic Conversion Specialist Conference,* pp. 275–284, Cleveland, Ohio, October 1964.

ENGINEERING MECHANICS DIVISION

VII. Materials

A. Carbon and Graphite Research

D. B. Fischbach and W. V. Kotlensky

1. Deformation-Enhanced Graphitization of Pyrolytic Carbons, D. B. Fischbach

It has been reported (Refs. 1–3 and SPS 37-33, Vol. IV, pp. 75–78) that high-temperature plastic deformation of pyrolytic carbons significantly increases the degree of graphitization obtained, relative to that obtained by heat treatment alone. The degree of graphitization is defined as $g = (6.88 - c)/0.172$, where c is the unit cell height determined by X-ray diffraction. The deformation may be by tension parallel to the substrate or by compression perpendicular to the substrate.

At JPL, previously reported conclusions on the effect of deformation on graphitization have been based, in part, on comparisons of the g values of the undeformed butt and deformed gage portions of specimens tested in tension (either constant strain-rate or creep). Since graphitization is a thermally activated process, a similar difference in g would result if the temperature of the butt portion were significantly lower than that of the gage section.

Optical pyrometer measurements of the temperature distribution over a portion of the sample length have shown that a definite temperature gradient exists. The specimens are nominally 3 in. long with approximately 1 in. at each end clamped in holders that are attached to pull rods. Only the central 1½ in. of the holder-specimen assembly could be viewed with the pyrometer. The temperature was found to decrease with distance from the center of the viewed area. Extrapolation of observed data to the positions of the butt ends of the sample indicated a temperature difference between center and ends of the order of 100°C for center temperatures ranging from 2400 to 2800°C.

Since the extrapolation of temperature gradient data is very uncertain, an independent estimate of the gradient was obtained by comparison of the observed butt and gage section g values for a specimen heat-treated at 2600–2800°C under zero stress. This was accomplished by mounting two specimens side by side in a modified holder. One specimen supported the weight of the holder and lower load train; the other was cut in two at the center of the gage section so that it supported no load. The observed degree of graphitization g_0 of the butt and gage portions of the cut specimen were

compared with the values g_e expected from studies of the kinetics of graphitization of the same lot of pyrolytic carbon (Sample B in Ref. 4). The results are shown by the open and filled circles in Fig. 1. The g_e values were based on the observed center temperature. The gage g_o values were in reasonable agreement with expectations. The butt g_o values were lower than expected by amounts corresponding to a temperature difference of 100 deg at 2600°C; 70 deg at 2700°C; and 50 deg at 2800°C—in general agreement with the extrapolated pyrometer measurements. It can be concluded that a significant temperature difference exists between the center and butt portions of the sample.

To determine whether g is enhanced by deformation, butt and gage section data for samples deformed 3–24% in tensile creep at 2600–2800°C were compared with

Fig. 1. Comparison of observed (g₀) with expected (g₀)
degree of graphitization of butt and gage portions
of deformed and undeformed tensile creep
specimens of pyrolytic carbon

expected values (also shown in Fig. 1 by the open and filled triangles). Here the g_e values for the butt data were based on corrected butt temperatures estimated from the temperature-difference data given above. The good agreement of the butt g_o and g_e values further confirms the gradient results. However, the deformed gage section g_o values were appreciably higher than expected for all samples deformed more than about 3%. Therefore, deformation does indeed enhance the degree of graphitization. The effect of tensile deformation on g is equivalent to a 250–300°C increase in temperature. A similar enhancement of preferred orientation and apparent crystallite diameter values by deformation was also observed. These results are consistent with those of studies on the kinetics of graphitization without deformation (Ref. 3 and SPS 37-31, Vol. IV, pp. 103–105). In those studies, graphitization was found to be closely associated with X-ray microstructure; and microstructural changes are greatly enhanced by high temperature deformation (Ref. 5 and SPS 37-33, Vol. IV, pp. 75–78). Planned are further studies on the detailed temperature distribution over the specimen length and on the effect of the amount and rate of deformation on the rate of graphitization.

2. Creep Data Analysis, W. V. Kotlensky and D. B. Fischbach

Tensile creep data for both coke-pitch graphite (Ref. 6) and pyrolytic carbon (SPS 37-30, Vol. IV, pp. 71–72) fit the empirical equation

$$\varepsilon = A + Bt^n \qquad (1)$$

where ε is the total creep strain, A is the time-independent (elastic, etc.) component of the strain, t is creep time, n is a constant of the material (0.4–0.6), and B is a material parameter that is a function of the stress and temperature of the test. Creep is well known to be a thermally activated process. Therefore, B in Eq. (1) must contain the Boltzmann factor $K = \exp(-\Delta H/RT)$, where ΔH is the activation energy, T is the absolute temperature, and R is the gas constant. However, before ΔH can be determined from the observed B values, the functional form of B must be known. J. E. Dorn (Refs. 7 and 8) has shown that for metals

$$\varepsilon = f(tK). \qquad (2)$$

Applying Eq. (2) to Eq. (1) (with n and stress constant) it is found that $B = B'K^n$ where B' is independent of temperature. Then ΔH may be determined from the slope of a plot of $(1/n) \ln B$ versus $1/T$. Such a plot for

the pyrolytic carbon data gives $\Delta H \simeq 250$ kcal/mole, the same value as determined by direct application of Eq. (2) [by superimposing data obtained at different temperatures on an $(\varepsilon - A)$ versus tK plot where K is treated as an adjustable parameter].

In Ref. 6, it was incorrectly assumed that $B = B'K$. Therefore, the apparent activation energies determined there, using Eq. (1), are equal to $n\Delta H$ and must be multiplied by approximately 2 $(n \simeq 0.5)$ to get the correct value.

B. Phonon Drag Thermopower in Dilute Copper Alloys

I. Weinberg

Recent theoretical work in the transport properties of solids has resulted in a general expression for the phonon drag thermopower in pure metals (Ref. 9). Using this general expression, Huebener (Ref. 10) has obtained a relation for the change in phonon drag thermopower due to point defects. These results have been applied to the change in phonon drag thermopower caused by lattice vacancies in gold (Ref. 10). If, as in the case of alloys, the solute atom may be treated as a point defect, and elastic scattering of phonons assumed, then the theory developed for lattice vacancies in gold can be applied to the scattering of phonons by solute atoms in dilute alloys. A program of this nature has already been carried out for Cu–Si and Cu–Al alloys (Ref. 11). The current work represents an extension of the previous treatment to dilute alloys of gold and silver, respectively, in copper. At the same time, results and analysis are presented for an additional Cu–Si alloy. The measured change in phonon drag thermopower for the current alloys is shown in Figs. 2 and 3.

Assuming elastic scattering of phonons by impurities, P. G. Klemens (Ref. 12) has shown that τ_i, the relaxation time for phonon impurity scattering, is given by

$$\tau^{-1} = a\omega^4 \qquad (1)$$

where a is the scattering parameter and ω is the phonon frequency. In obtaining the data of Figs. 2 and 3, the primary objective is to obtain, from the experimental
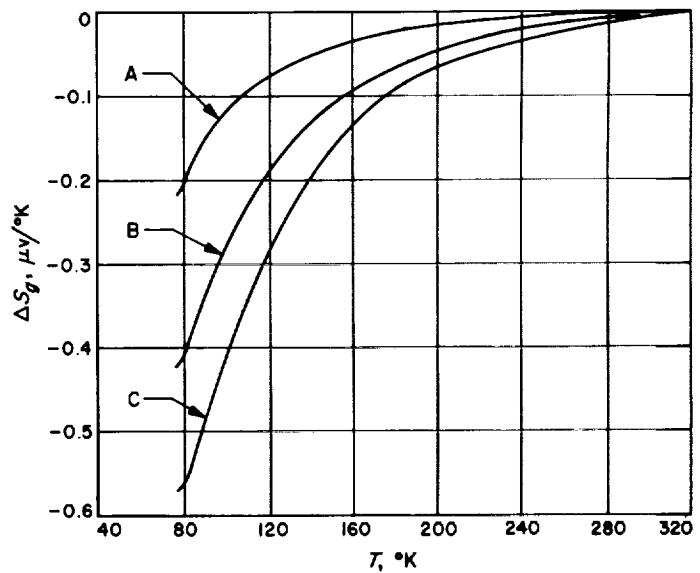
Fig. 2. ΔS_g, the change in phonon drag thermopower for: (A) Cu + 0.03 at. % Au; (B) Cu + 0.09 at. % Ag; and (C) Cu + 0.2 at. % Ag
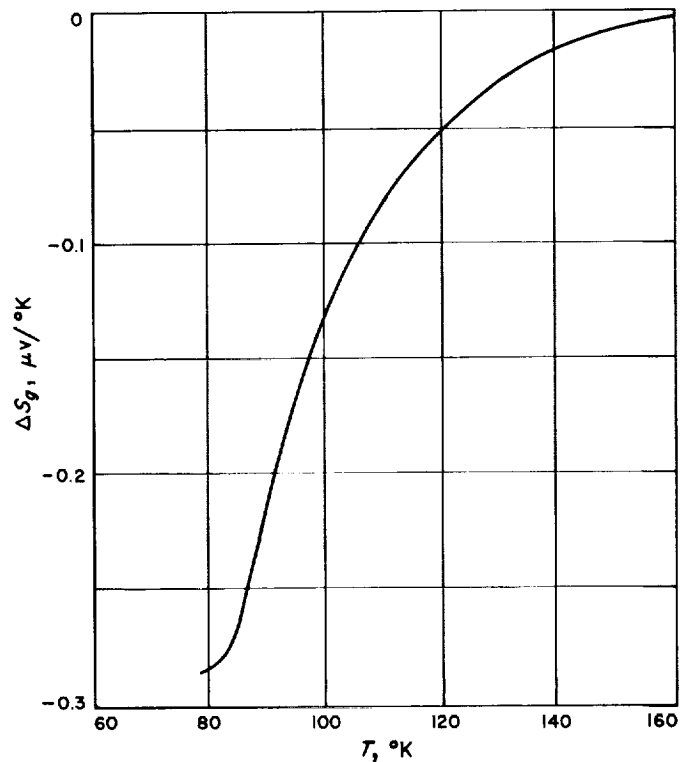
Fig. 3. ΔS_g, the change in phonon drag thermopower for Cu + 0.44 at. % silicon

ΔS_g values, a value for the scattering parameter. Procedures for obtaining a from thermoelectric experiments have been detailed in Ref. 11. Values of a obtained from the present ΔS_g values are shown in Table 1, where they are designated as a_{exp}.

Table 1. Scattering parameters for various copper alloys

Solute	Atomic % solute	a_{exp} [a] (10^{-43} sec^3)	a_m [b] (10^{-43} sec^3)
Au	0.03	1.28	0.93
Ag	0.09	5.62	0.31
Ag	0.2	14.6	0.68
Si	0.44	2.33	0.96

[a] a_{exp} is obtained from the data of Figs. 2 and 3 via a technique detailed in Ref. 11.
[b] a_m is a theoretical value based on mass defect scattering, Eq. (2).

Using the Klemens theory, it is a relatively simple matter to calculate a for the scattering of phonons by a substitutional impurity differing only in mass from the atoms of the host crystal. In addition, this theory contains terms due to changes in force constants and the strain field associated with a point defect. However, the calculation for phonon scattering mechanisms, other than mass difference, are relatively inexact. This is due partly to approximations in the theory and partly, in the case of metals, to the difficulties associated with computing interatomic force constants. Hence only the mass difference calculation is considered in the present case. The scattering parameter so calculated is designated by the symbol a_m. Thus (Ref. 12)

$$a_m = \frac{\Omega f}{4\pi v_s^3}\left[\frac{\Delta M}{M}\right]^2 \qquad (2)$$

where Ω is the atomic volume, f is the fraction of solute atoms present in the metal, and v_s is the phonon velocity. M is the average mass for the atoms of the host crystal and ΔM is the difference in mass between the solute atoms and copper. The values computed for a_m are shown in Table 1.

The Cu–Au results indicate that in this alloy, the mass defect mechanism accounts for most of the phonon scattering. In Cu–Si the results indicate a contribution from lattice distortion which is of the same magnitude as the mass defect term. The Cu–Ag results indicate that phonons are scattered largely through the mechanism of lattice distortion in this alloy. These results are in agreement with the estimated contribution of the various phonon scattering mechanisms based on the range of solid solubilities (Ref. 13) in these alloys.

C. Sterilization of Impact Limiters

E. C. Bernett

Development work to date on Voyager lander-type impact limiters indicates that balsa wood is one of the most promising materials for this application. Since the Voyager lander must be sterilized prior to launch, a program has been started which will evaluate the effects of various sterilization cycles on the energy absorption properties of balsa.

Dry heat exposure at 145°C for 108 hr or 125°C for 498 hr results in decomposition of the wood itself. Weight losses of up to 17% have been noted. The crushing behavior of the exposed wood is very erratic and the samples often split during the compression test, indicating that the transverse strength of the wood is degraded. The specific energy absorption is somewhat reduced by the dry heat exposure. The 498-hr cycle at 125°C was more detrimental than the 108-hr cycle at 145°C. It was observed that heat exposure caused dimensional changes ranging from 0.001 to 0.029 in./in. with the large shrinkage occurring perpendicular to the wood grain. In a thick section such as a full-size limiter, it is very likely that the shrinkage would result in splitting.

Balsa limiter technology developed for Ranger required that the limiter be sealed with an outside cover to maintain constant moisture content in the wood. Heat sterilization runs, as above, were therefore made on balsa samples which were sealed in closely fitting capsules. A major disadvantage to this approach was the high-pressure rises (up to 147 psig) that occurred during the exposure. When the samples were removed from the capsule, large amounts of reaction products escaped and weight losses up to 19.5% were recorded. The specific energy absorption properties of samples that were heat sterilized in the sealed capsules were reduced by about 50%.

Further work on the effects of sterilization on balsa wood is planned. Cycles that include the ethylene oxide decontamination treatment will be checked. Also, various pretreatments that might minimize the detrimental effects of sterilization will be evaluated.

References

1. Kotlensky, W. V., and Martens, H. E., "Mechanical Properties of Pyrolytic Graphite to 2800°C," *Proceedings of the Fifth Conference on Carbon*, Vol. II, pp. 625–638, Pergamon Press, London, 1963.

2. Kotlensky, W. V., and Martens, H. E., "Tensile Properties of Pyrolytic Graphite to 5000°F," *High Temperature Material II* (AIME), Vol. 18, Interscience Publishers, New York, 1963.

3. Bragg, R. H., Crooks, D. D., Fenn, R. W., Jr., and Hammond, M. L., "The Effect of Applied Stress on the Graphitization of Pyrolytic Graphite," *Carbon*, Vol. 1, pp. 171–179, 1964.

4. Fischbach, D. B., "Kinetics of Graphitization I. The High Temperature Structural Transformation in Pyrolytic Carbons," TR 32-532, Jet Propulsion Laboratory, Pasadena, California (in production).

5. Kotlensky, W. V., and Martens, H. E., "Structural Changes Accompanying Deformation in Pyrolytic Graphite," *Journal of the American Ceramic Society*, Vol. 48, pp. 135–138, 1965.

6. Martens, H. E., Button, D. D., Fischbach, D. B., and Jaffe, L. D., "Tensile and Creep Behavior of Graphites Above 3000°F," *Proceedings of the Fourth Conference on Carbon*, pp. 511–529, Pergamon Press, London, 1960.

7. Dorn, J. E., "Some Fundamental Experiments on High Temperature Creep," *Journal of the Mechanics and Physics of Solids*, Vol. 3, pp. 85–116, 1954.

8. Dorn, J. E., "The Spectrum of Activation Energies for Creep," *Creep and Recovery*, pp. 255–283, American Society for Metals, Cleveland, Ohio, 1957.

9. Bailyn, M., "Transport in Metals," *Physics Review*, Vol. 120, p. 381, 1960.

10. Huebener, R., "Thermoelectric Power of Lattice Vacancies in Gold," *Physics Review*, Vol. 135A, p. 1281, 1964.

11. Weinberg, I., "Phonon-Drag Thermopower in Cu-Al and Cu-Si Alloys," *Physics Review*, Vol. 139A, p. 838, 1965.

12. Klemens, P. G., "The Scattering of Low-Frequency Lattice Waves by Static Imperfections," *Proceedings of the Physical Society*, Section A68, p. 1113, London, 1955.

13. Hansen, M., *Constitution of Binary Alloys*, McGraw-Hill Book Co., Inc., New York, 1958.

VIII. Applied Mechanics

A. Equilibrium Radiation From the Shock Layer for Atmospheric Entry to Mars

F. Wolf and J. Spiegel

In the continuing effort to obtain experimental information on equilibrium radiation of shock layers for entry bodies in the atmosphere of Mars, additional data from the JPL and General Electric shock-tube facilities are presented to supplement the previous summary in *SPS 37-33*, Vol. IV, pp. 99–103. The majority of the shock-tube data were obtained from the measured total intensity of radiation in the stagnation area, behind a standing shock of a small hemispherical model (Ref. 1). Only a small number of points were sidewall measurements of radiation behind the moving (incident) shock. Thus, to correlate radiant intensity of the gas against the flight free-stream density, the standing shock measurements are plotted in Figs. 1 and 2 at an equivalent free-stream density that would correspond to equal values of stagnation density and enthalpy. For comparison, approximate total radiative intensity, calculated with

the Ames computer program (Ref. 2) and corrected for a CN_{red} oscillator strength, $f_{CN_{red}} = 0.0072$, is shown with free-flight velocity from 18,000 to 30,000 ft/sec as a parameter. The JPL radiative intensity calculations are being revised and will include a contribution from the CO (4th positive) that was missing from the previous computed radiance program. This transition occurs in the far ultraviolet and has not been observed up to now, because of the higher cut-off wavelength of the quartz windows and lack of vacuum ultraviolet equipment.

The scarcity of data below a free-stream pressure (p_∞) of 1 mm Hg is apparent from both atmospheric models; therefore, further experimental work is planned to obtain additional radiative intensity data in the low density range adjacent to the bulk of the existing points at $p_\infty = 3$ mm Hg and higher. The current G.E. contract provides for the measurement of radiance behind the incident shock at two flight velocities and for free-stream pressures starting at 5 mm Hg and decreasing in steps down to the minimum obtainable, presumably 0.3 mm Hg. This value is due to the limit in the measurable cavity gage output. For these experiments the sidewall technique was preferred, as opposed to the stagnation point
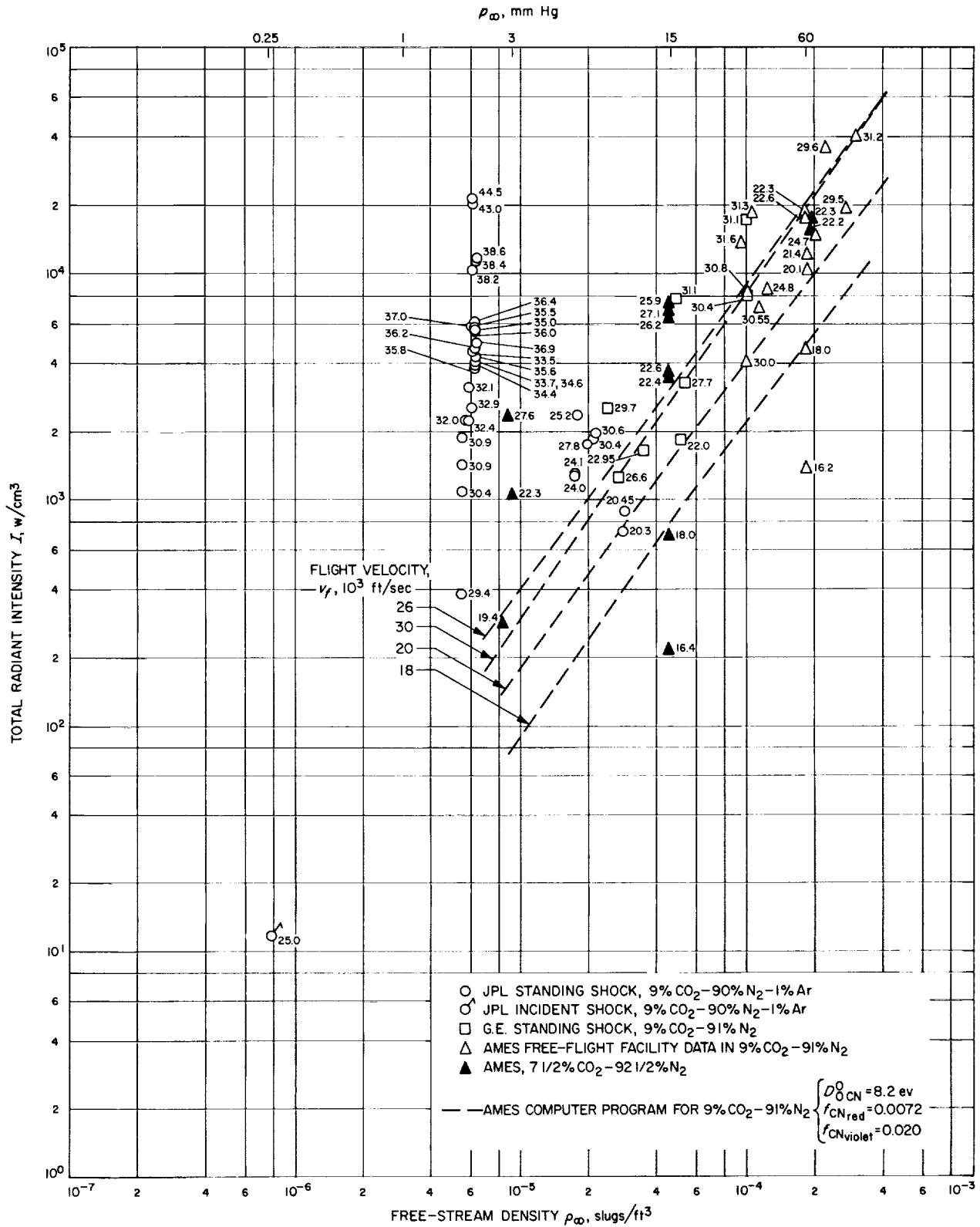
Fig. 1. Free-flight gas radiance of CO_2–N_2–Ar mixtures versus free-stream density
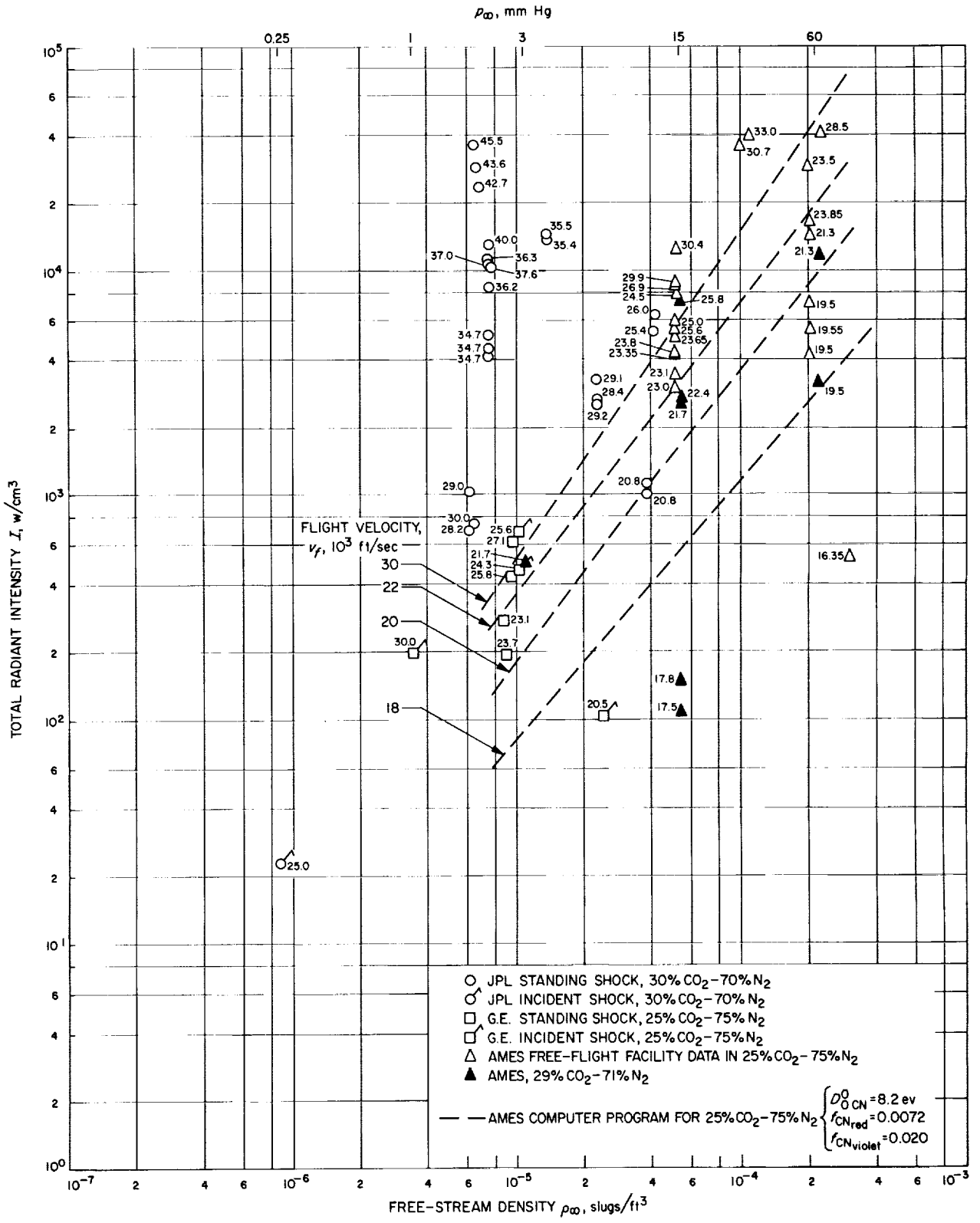
Fig. 2. Free-flight gas radiance of CO_2–N_2 mixtures versus free-stream density

cavity gage, although effects of the shock-tube boundary layer on the total observed radiation are unknown. At the upper limit of the free-stream pressure, the sidewall measurements are affected by self-absorption of the peak intensities and this effect must be accounted for by reduction of the experimental measurements, using analytical methods.

A further extension of radiation measurements into the lower densities would be possible, first by a qualitative spectral survey of the low density gas radiance, then by selecting the major contributors and measuring their intensities with a monochromator and photomultiplier sensors.

A primary requirement of data correlation is the establishment of a local density dependence of gas radiance to be derived from experiment and useful for application to a complete entry trajectory of a capsule. By experimental means this is not yet within the present state-of-the-art,

as the considerable scatter of the data, shown in Figs. 1 and 2, indicates. Consequently, the Ames theoretical density dependence was used to obtain a cross plot of the data by adjustment to three fixed density levels (Figs. 3 and 4), as a function of velocity. The bandwidth of experimental scatter shown seems to be approximately proportional to the local density of experimental points, except perhaps in the high velocity range where continuum radiation predominates (Ames computations account for molecular radiation only, Ref. 2). Thus, there is still some lack of reproducibility and corroboration of experimental results originating in different testing facilities. At present, the best estimate of shock-layer radiance could be based on an average curve in, or (conservatively), the upper bound of the bands covering the data points for a specific free-stream density. The resulting radiance as a function of free-flight velocity would be combined with a density function $n\,(\rho_\infty,\,v_{flight})$ obtained analytically from one of the computer programs calculating a synthetic spectrum of gas radiance.
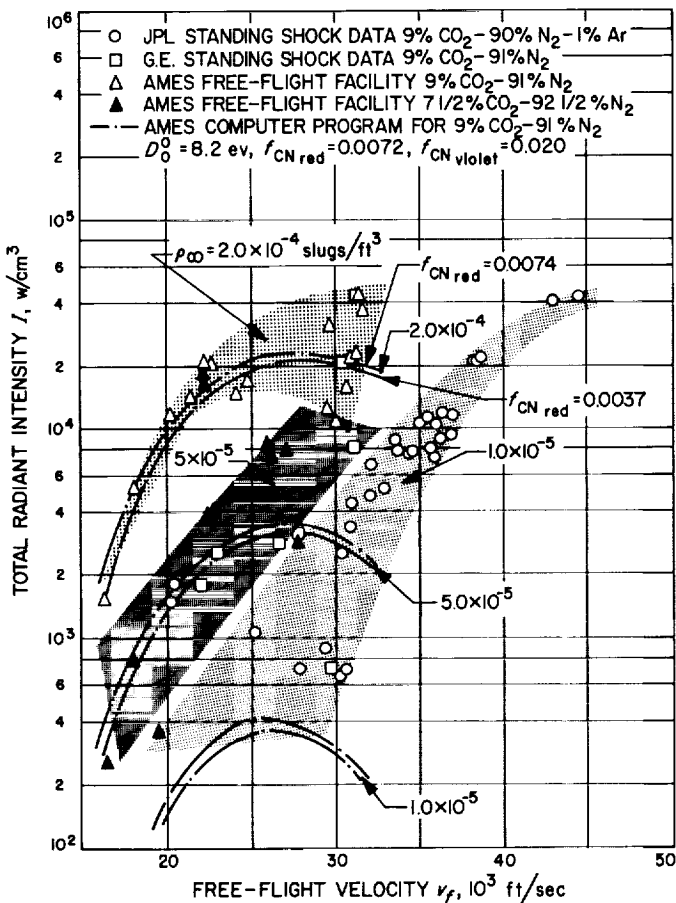
Fig. 3. Gas radiance of CO_2–N_2–Ar mixtures with adjusted nominal density
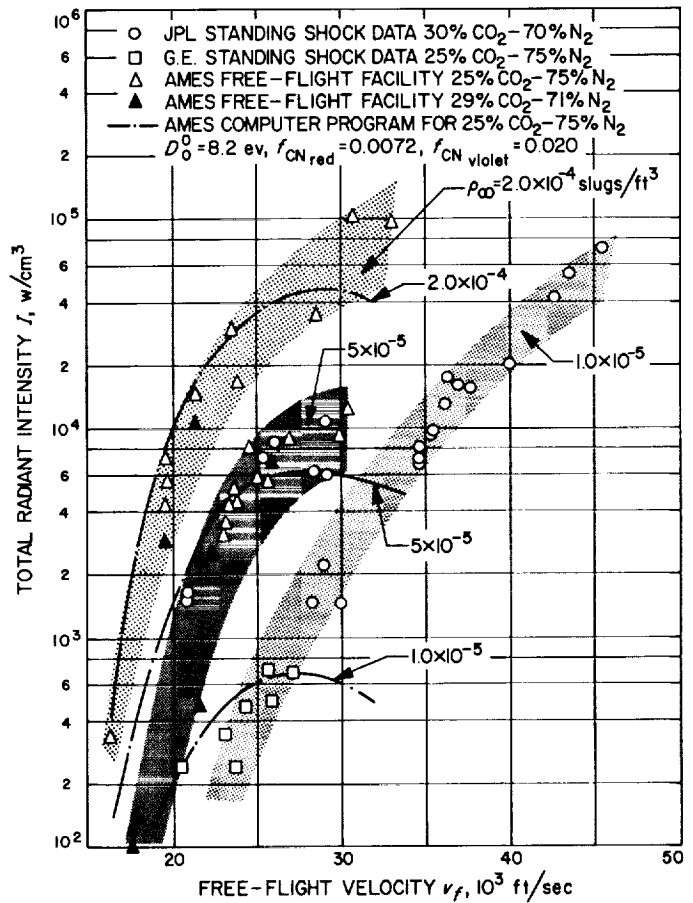
Fig. 4. Gas radiance of CO_2–N_2 mixtures with adjusted nominal density

In *SPS 37-33*, it was indicated that the exponent of density dependence (ρ_∞^n) for the JPL and Ames calculations differed. This can be shown to be related to the different heats of formation of CN used in each calculation. Assuming two model gas mixtures containing C and N in the same proportion, but with two different dissociation energies D_{0CN}^0, we obtain different CN number-densities in these mixtures at elevated temperatures. If the CN radiation forms the major part of the total gas radiance over a certain flight velocity range, the density exponent n of the total radiation per unit volume will follow the CN particle number variations with

$$n_2 - n_1 = \Delta n = \frac{\Delta D_{0CN}^0}{RT_2^2} \rho_\infty \left(\frac{\partial T_2}{\partial \rho_\infty} \right)_{v_f},$$

where n_1 and n_2 correspond to the two model gas mixtures of assumed different dissociation energies D_{0CN1}^0 and D_{0CN2}^0, ρ_∞ is the free-stream density and T_2 the temperature behind the normal shock. Typically, for a mixture of 30% CO_2–70% N_2 the dependence of T_2 on free-stream pressure for the extremes of the Mars entry velocity range is:

v_f, ft/sec	$\rho_\infty = 0.1$,	1.0,	15.0,	mm Hg	Δn_{v_f}
18000	$T_2 = \sim 5200$	5700	6200	°K	0.100
30000	$T_2 = 7000$	8000	9000	°K	0.075

With this temperature free stream pressure relation, the corresponding values of Δn at the $p_\infty = 1.00$ mm Hg level are calculated as shown in the last column above and are in good agreement with the differences between Ames and JPL calculated density exponents indicated in *SPS 37-33*, Vol. IV, p. 102, Fig. 6.

B. A Basic Thermal Analysis System Utilizing Computers

J. A. Hultberg

Computer programs that are used for the analysis of unmanned spacecraft temperature control consider both radiation and conduction heat transfer.

The general heat conduction equation is:

$$\frac{\partial}{\partial x}\left(k_x \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y}\left(k_y \frac{\partial T}{\partial y} \right) + \frac{\partial}{\partial z}\left(k_z \frac{\partial T}{\partial z} \right) + q = c\rho \frac{\partial T}{\partial t}$$

(1)

where

k is the conductivity

T is the temperature

q is the heat input

c is the specific heat of the material

ρ is the density of the material

t is the time

Radiation heat transfer is governed by the Stefan–Boltzmann law. Radiation is a boundary condition for the conduction equation.

For the majority of spacecraft temperature control problems, certain simplifying assumptions are usually made:

(1) Directional conductivity properties are neglected.

(2) Temperature-dependent conduction properties are neglected.

(3) Spectral radiation properties are considered in a solar temperature range and in a spacecraft temperature range. Following the first solar reflection, the properties are represented by one value.

(4) Directional radiation properties are neglected except for radiation on highly reflective surfaces.

(5) Radiation from a surface is assumed to follow Lambert's cosine law.

(6) Grey surfaces are assumed.

(7) Temperature-dependent emissivity is taken into account only for temperature-controlling devices such as louvers.

The discrete form of the combined radiation-conduction coupled equations has the form:

$$\sum_j \alpha_{ij}\sigma T_j^4 + \sum_j \beta_{ij}T_j + \gamma_i + \delta_i \frac{dT_i}{dt} = 0 \qquad (2)$$

where

α_{ij} is the radiation exchange factor between the ith and jth area

β_{ij} is the conductance between the ith and jth volume

γ_i is the energy input to the ith surface or volume

δ_i is the thermal mass of the ith volume

σ is the Stefan–Boltzmann constant

Equations of this type may be solved with existing programs, such as in Refs. 3 and 4.

The calculation of the elements (α_{ij}) of the radiation exchange matrix involves the evaluation of the configuration factor integral. The configuration factor is of the form:

$$F_{ij} = \frac{1}{A_i} \int\int_{A_i} \int\int_{A_j} \frac{\cos\theta_i \cos\theta_j \, dA_j \, dA_i}{\pi d^2} \qquad (3)$$

where

A   is the area

θ is the angle between the outward directed normal and a line joining dA_i and dA_j

d   is the distance between dA_i and dA_j

Integrals of this type may be evaluated for a wide variety of geometries by means of programs in Refs. 5, 6, 7, and 8.

The calculation of the conductances, β_{ij}, for the conduction matrix, as well as the calculation of the power inputs γ_i, and the thermal masses δ_i, is usually performed without any computer program.

Thus it can be seen that the calculations for radiation-conduction coupled temperature control require numerous hand calculations and, essentially, two descriptions of the spacecraft geometry to the computer. The double description of spacecraft geometry to the computer is very time-consuming for the program user. Furthermore, if the number of discrete intervals is changed, the coefficients α_{ij}, β_{ij}, γ_{ij}, and δ_{ij} must be recomputed. It is a tedious task to change the number of discrete intervals.

This suggests that it would be desirable to be able to describe spacecraft geometry to the computer only once. Such a program would be known as a thermal analysis system (TAS) since it would combine features of thermal analysis programs and configuration factor programs. The program being developed combines the solution of the discrete form of the radiation-conduction coupled equations, the configuration factors solution, and the conductance calculation. This program requires only one description of geometry. The Oppenheim network method is currently used for the radiation portion Ref. 9.

Two examples are described to illustrate the versatility of the program.
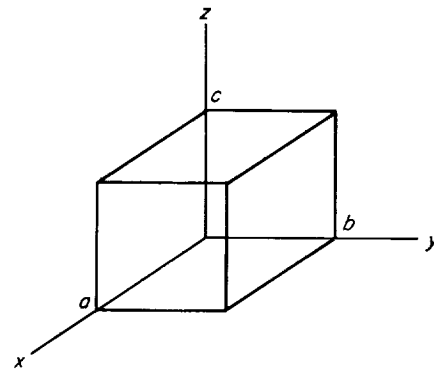
Fig. 5. One-unit geometry

(1) A rectangular parallelepiped (Fig. 5). The geometry description includes lengths a, b, c, the wall thicknesses, and the power inputs. The physical description includes the surface coatings, specific heat, density, and the material conductivity.
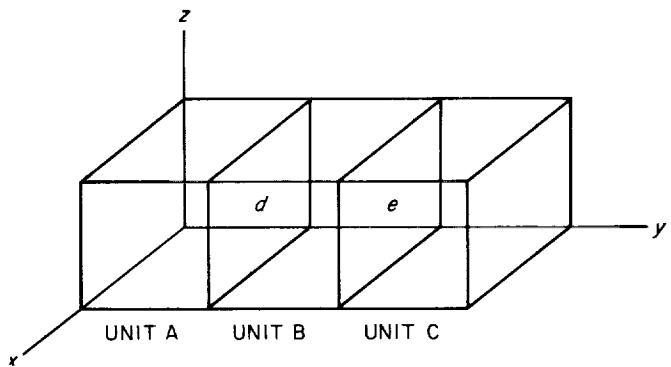
Fig. 6. Three-unit geometry

(2) A simple geometry for three adjacent rectangular parallelepipeds (Fig. 6). This example is composed of a combination of those units in Example (1). It is necessary to give the geometrical and physical information, as in Example (1), as well as further geometric information to indicate that side d of unit A is the same as side d of unit B and side e of unit B is the same as side e of unit C. Such a configuration could be constructed to analyze the model reported in Ref. 10.

The one- and three-unit examples show how the geometry and physical properties may be varied. The wall thickness and coating on a surface may be altered by changing one value on an input card.

Thus, the techniques described here combine several of the existing computing techniques. The major change is in the manner in which the geometric and physical data are transmitted to the computer. The reduction in user's preparation time, together with the increase in computer time, is obvious. The great value of a program of this type is that changes in spacecraft geometry are more easily transmitted to the program.

References

1. Thomas, G. M., and Menard, W. A., "Experimental Measurement of Non-Equilibrium and Equilibrium Radiation from Planetary Atmospheres," (to be published in the *AIAA Journal*).

2. Arnold, J. O., Reis, V. H., and Woodward, H. T., "Theoretical and Experimental Studies of Equilibrium and Non-Equilibrium Radiation to Bodies Entering Postulated Martian and Venusian Atmospheres at High Speeds," presented at the Second AIAA Aerospace Sciences Meeting, New York, January 25–27, 1965.

3. Strong, P. F., and Emslie, A. G., "The Method of Zones for the Calculation at Temperature Distribution," Paper 65-WA/HT-47 at ASME Winter Annual Meeting, November 7–11, 1965, Chicago, Illinois.

4. Sepetoski, W. K., Sox, C. H., and Strong, P. F., "Description of a Transient Thermal Analysis Program for Use With the Method of Zones," Arthur D. Little, Inc., No. C-65670, August 1963.

5. Toups, K. A., "A General Computer Program for the Determination of Radiant-Interchange Configuration and Form Factors, CONFAC I," North American Aviation, S.I.D. 65-1043-1, October 1965.

6. Toups, K. A., "A General Computer Program for the Determination of Radiant-Interchange Configuration and Form Factors, CONFAC II," North American Aviation, S.I.D. 65-1043-2, October 1965.

7. Dummer, R. S., and Breckenridge, W. T., Jr., "Radiation Configuration Factors Program," General Dynamics Astronautics, Space Sciences Laboratory, Report ERR-AN-224.

8. Plamondon, J. A., "Numerical Determination of Radiation Configuration Factors for Some Geometrical Situations," Technical Report No. 32-127, Jet Propulsion Laboratory, Pasadena, California, July 1961.

9. Oppenheim, A. K., "Radiation Analysis by the Network Method," *Transactions of the ASME*, Vol. 78, pp. 725–135, 1956.

10. Fowle, A. A., Gabron, F., and Vickers, J. M. F., "Thermal Scale Modeling of Spacecraft: An Experimental Investigation," presented at the AIAA Space Simulation Testing Conference, November 16–18, 1964, Pasadena, California, AIAA Publication CP-11 (to be published in the *AIAA Journal of Spacecraft and Rockets*).

PROPULSION DIVISION

IX. Research and Advanced Concepts

A. Thermal Radiation From Ionized Argon As Determined by Application of Near-Black Cavities

E. J. Roschke

1. Introduction

Experimental and analytical heat transfer results for ionized argon flow through convergent-divergent nozzles have been presented and discussed in SPS 37-22, -23, and -24, Vol IV. Experimental measurements for total heat flux to the mixing chamber, nozzle, and diffuser walls were presented in the form of axial heat flux distributions which were determined by calorimetric means. Ionization was produced by passing argon through an electric arc upstream of the mixing chamber and nozzle. Experimental data were then incorporated into a laminar boundary-layer analysis that was utilized to predict convective heat transfer in the nozzle. The latter results were compared with test results. Because of the nature of the measurements it was, at that time, not possible to separate the total heat transfer to the walls into radiative and convective components; therefore it became necessary to

neglect the effects of thermal radiation in the theoretical analysis. This problem was noted in SPS 37-22 and plans were outlined to attempt total (rather than spectral) thermal radiation measurement by means of a hohlraum, or black-body cavity.

The principal objective of the proposed thermal radiation measurements is to determine the relative contribution of radiation to the total heat transfer over a range of operating conditions and configurations. If the radiative component were to prove as small as 5% of the total heat transfer, its determination to even moderate accuracy would not be required and rather gross error would be tolerable. A secondary objective of such measurements is to study thermal radiation from ionized argon under the various experimental conditions imposed.

This report will describe the apparatus (particularly the hohlraums) in present usage, and will present some preliminary data obtained from the hohlraums. The advantages and disadvantages of hohlraums are enumerated.

2. Heat Transfer Apparatus

The arrangement of the heat transfer apparatus is shown schematically in Fig. 1. This configuration differs
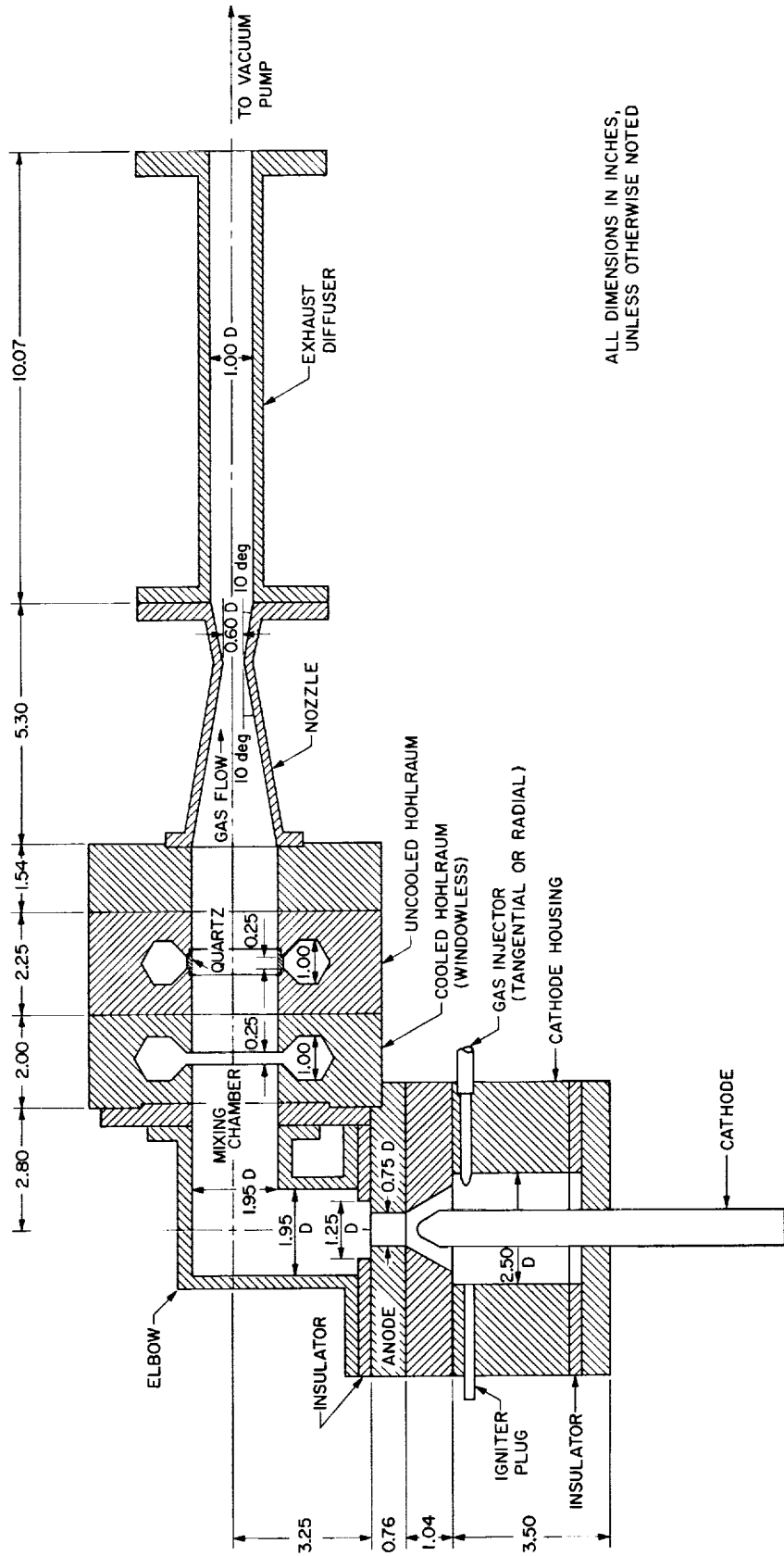
Fig. 1. Ionized gas heat transfer apparatus showing location of hohlraums

from previous ones, chiefly in the spatial position of the cathode and anode with respect to the mixing chamber and nozzle. Former configurations (see previously mentioned SPS references) were axially symmetric throughout, whereas the newer configuration is constructed so that the ionized gas must negotiate a 90-deg bend before entering the mixing chamber. This feature shields the arc from the direct view of the rest of the apparatus and should simplify interpretation of thermal radiation phenomena by minimizing the direct effects of the electric arc. In this work, arc radiation itself is of secondary interest. Fig. 1 does not indicate the many coolant passages employed for calorimetric heat transfer measurements, nor does it show the location of a considerable number of static pressure taps which are used to obtain fluid dynamic data. The present apparatus is actually composed of more individual sections than shown; means are available for inserting sections containing view ports in certain desirable locations.

The hohlraums are located in the mixing chamber approximately midway between the ionized gas inlet and the nozzle inlet. They are of semi-toroidal shape concentric with the axis of the wall of the mixing chamber, which permits the use of larger openings than possible with a more conventional cylindrical cavity utilizing a small, circular opening. Relatively large openings were considered necessary to intercept sufficiently large amounts of thermal radiation for detection purposes. Rather low radiant heat flux values were anticipated. The cavities have polygonal rather than circular cross-sections for reasons which will be enumerated.

Initially two hohlraums were employed, a windowless cavity opening directly to the gas stream and a second cavity isolated from the gas stream by means of an annular or ring-shaped quartz window. These are referred to in Fig. 1 as the cooled and the uncooled hohlraum, respectively. The cooled cavity was surrounded by a number of small coolant passages (not shown in Fig. 1) so that calorimetric heat transfer data could be obtained. This hohlraum is not in general use because the data obtained from it were considered unreliable due to internal gas convection and conduction effects induced by secondary flow within the cavity; it is presently being modified to accept a quartz window similar to that used in the uncooled hohlraum.

An enlarged cross section of the uncooled hohlraum is shown in Fig. 2; a quartz window isolates the cavity from the gas stream, and is partially cooled by adjacent coolant passages both upstream and downstream. Although means
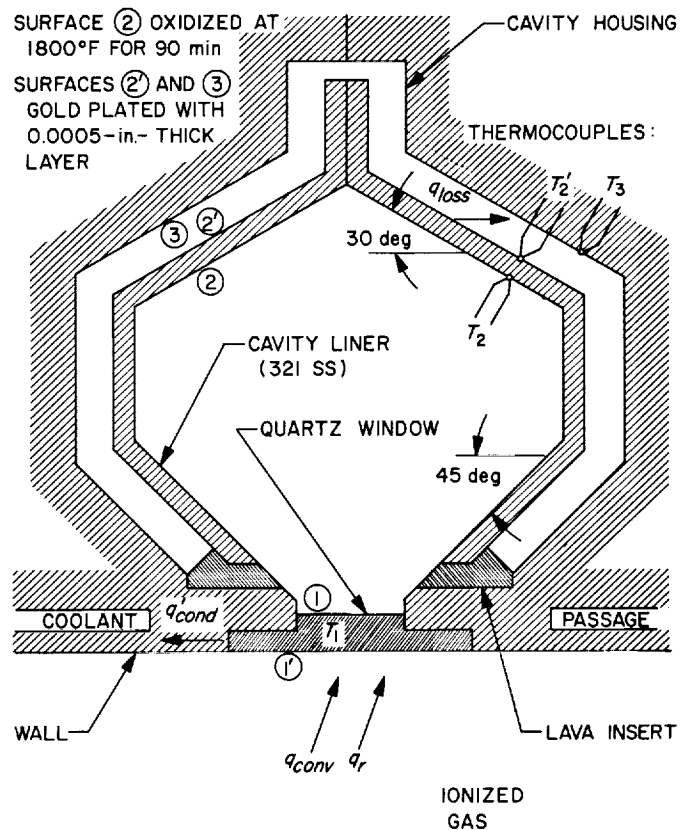
Fig. 2. Details of uncooled hohlraum and terminology for thermal analysis

are available for evacuating the cavity, all tests to date have been obtained with argon in the cavity at a pressure approximately equal to the local wall pressure on the stream side. The cavity has a double-wall construction consisting of a thermally insulated liner placed within the surrounding housing. Cooling of the liner is accomplished only by radiation exchange with the housing; it is desirable to minimize this loss denoted by q_{loss} in Fig. 2. Surface of the three heat transfer surfaces A_2, A_2' and A_3 was intended to maximize the hohlraum performance. Each of the three surface temperatures T_2, T_2', and T_3 are measured at four circumferential positions around the cavity by means of surface thermocouples. The cavity was given a polygonal cross section to facilitate fabrication and to reduce direct reflection loss back through the quartz window. Note that the quartz window area that is subject to gas convection is larger than the cavity neck area (based on minimum neck width) which is defined as the pertinent area transmitting thermal radiation. O-ring seals between the quartz window and the cavity housing (not shown in Fig. 2) serve to seal the cavity environment from the ionized gas. The stainless steel cavity liner is 0.050-in. thick and the liner is separated from the housing

by a nearly uniform 0.090-in. gap. The minimum neck or window width is 0.25 in. and the maximum cavity width is 1.0 in. (Fig. 1).

3. Simplified Thermal Analysis; Effect of Window Excluded

If the quartz window is completely passive, thermally and optically, a thermal balance on the hohlraum cavity is especially simple and the heat transfer terms q_{conv} and q'_{cond} may be eliminated from consideration. At thermal equilibrium, the total radiant energy incident on the cavity opening and actually absorbed is equal to the radiant energy emitted by the cavity plus whatever loss occurs through the cavity wall. For an isothermal cavity, T_2 = constant,

$$\alpha_a A_1 F_{10} q_r = \epsilon_a A_1 \sigma T_2^4 + A'_2 q_{loss}$$

where α_a and ϵ_a represent an apparent absorptivity and emissivity of the cavity opening, respectively. In the first approximation $\alpha_a/\epsilon_a \sim \alpha_2/\epsilon_2$ and $F_{10} \sim 1.0$ so that

$$q_r = \frac{\epsilon_2}{\alpha_2}\sigma T_2^4 + \frac{1}{\alpha_a}\frac{A'_2}{A_1}q_{loss} \tag{1}$$

$$\alpha_a = \alpha_2 \bigg/ \left[\alpha_2 + \frac{A_1}{S}(1 - \alpha_2) \right]. \tag{2}$$

Eq. (2) is strictly applicable only to spherical cavities; its use and some of its limitations are discussed in Ref. 1. An equivalent expression for Eq. (2) is given in Refs. 1, 2, and 3. If the cavity walls are grey, then $\alpha_a = \epsilon_a$ and

$$\epsilon_a = \epsilon_2/[\epsilon_2 + f(1 - \epsilon_2)] \tag{3}$$

where $f = A_1/S$. This expression is plotted as the upper curve in Fig. 3 with f equivalent to that existing in the uncooled hohlraum, and may be compared to an analogous result for a semi-infinite cylindrical groove adapted from Ref. 4. It is likely that ϵ_a for the present hohlraum has a value intermediate between the two results depicted in Fig. 3.

In the present application, measured values of wall temperature were used to calculate q_{loss} from the following expression

$$q_{loss} = \frac{\epsilon\sigma(T_2'^4 - T_3^4)}{1 + \frac{A'_2}{A_3}(1 - \epsilon)} \tag{4}$$
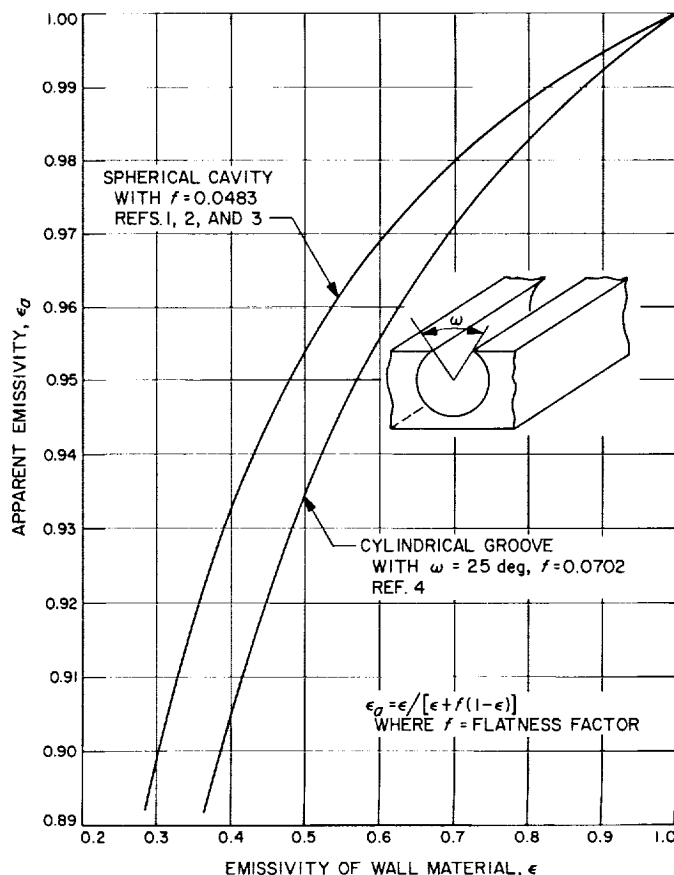
Fig. 3. Apparent emissivity of hohlraum with finite opening, and with grey diffusely reflecting walls

since $\epsilon = \epsilon'_2 \sim \epsilon_3$. Eq. (4) applies to the net radiant heat flux exchanged between infinite parallel plates, concentric spheres, or infinite concentric cylinders and is valid for diffuse or specular radiation. It is considered applicable in the present case since the cavity liner is completely enclosed by the housing.

Eqs. (1), (2), and (4) were utilized to estimate the magnitude of q_r in the present experiments. Fig. 4 indicates the relative effect of α_2/ϵ_2 and small variations in T_2 on the incident radiant heat flux calculated by means of Eq. (1). Typical values of T_2 and q_{loss} were utilized to obtain this result. The normalizing parameter $\bar{q}_r$ represents the value of q_r obtained for grey walls and an average selected temperature $\bar{T}_2$ of the cavity wall. As T_2 is allowed to vary plus or minus 4% from its average value, variations of 16 to 33% are incurred in the computed value of q_r, depending on the value of α_2/ϵ_2. The amount of variation specified for T_2 is typical of the range of variation observed experimentally among the four surface thermocouples attached to Surface 2.
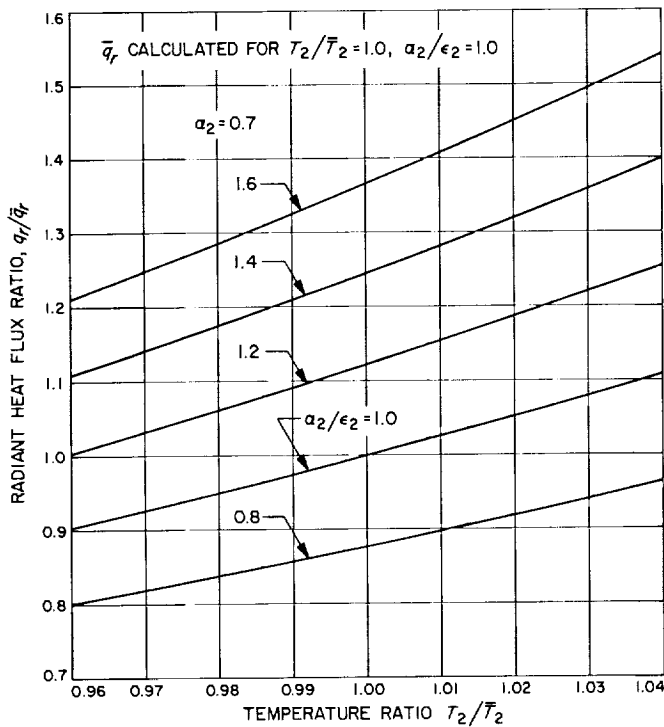
Fig. 4. Effects of non-grey cavity walls and small
variations in cavity wall temperature on radiant
heat flux, calculated neglecting presence of
quartz window

4. Preliminary Experimental Results

Data for seven tests are summarized in Table 1. Total
heat flux to the wall was determined calorimetrically and
results that were obtained just upstream of the uncooled
hohlraum window are listed. Radiant heat flux from the
argon gas as obtained using Eq. (1), and representative
optical properties selected from Refs. 5, 6, and 7, are

Table 1. Some preliminary radiation results based
on uncooled hohlraum measurements

Test No.	Stagnation pressure, psia	q_r [a] Btu/in.^{2} sec	q_{tot} [b] Btu/in.^{2} sec	q_r/q_{tot}
112-3H	4.04	0.00159	0.0730	0.022
112-4H	3.89	0.00196	0.0803	0.024
112-5H	4.00	0.00213	0.0846	0.025
112-6H	2.04	0.00077	0.0392	0.020
112-7H	2.03	0.00089	0.0440	0.020
112-8H	2.07	0.00103	0.0518	0.020
112-11H	1.04	0.000825	0.0439	0.019

[a]Computed disregarding window, using Eq. (1).
[b]Measured just upstream of hohlraum location.

compared to the total heat flux. Thus, only 2 to 2½% of
the total heat transfer was in the form of radiation for
this series of tests. As will be discussed later, these results
are probably too high due to the influence of the quartz
window, which tends to emit, as well as transmit, radi-
ation into the cavity.

Unfortunately, the complete data for these tests were
not yet reduced and analyzed at this writing so such
pertinent and necessary information as the stagnation
temperature and enthalpy of the gas stream and the
degree of ionization were not available. Recent spectro-
scopic measurements made just downstream of the un-
cooled hohlraum, for operating conditions not greatly
different from those of the tests reported here, have indi-
cated that continuum radiation was virtually nonexistent
at that location and that whatever radiation occurred was
line-spectra radiation.

Values of the total heat loss from the cavity $(q_{loss}A_2')$
varied from 28 to 39% of the total incident radiation
(q_rA_1) in this series of tests. Total heat loss tended to
increase with increasing q_r. Relatively large departures
from isothermal cavity conditions occurred, absolute
value of temperature T_2 varied by as much as 5% from one
circumferential location to another. Values of T_3 varied
by more than 6%. Possible reasons for this will be dis-
cussed later. These differences could not be attributed
solely to experimental error since the temperature pattern
on the 12 surface thermocouples was consistent from test
to test. For reasons not yet understood, the 45-deg surface
of the hohlraum nearest to the window consistently at-
tained the highest temperature within the cavity.

5. Discussion

The hohlraum, as a device for measuring total radiation
from a hot gas, is simple in design and construction,
rugged, usable to relatively high temperatures, and free
of complex instrumentation. In principle, all that need be
measured are the various surface temperatures required
in the heat balance so that calibration is not necessary.
However, as used in the present application, the hohlraum
cannot be used for spatial resolution of incident radiant
energy or spectral studies. In addition, it is difficult in
practice to obtain isothermal cavity or window conditions
and the measurement of incident radiant heat flux re-
quires fairly precise knowledge of the optical properties
of participating surfaces. Lack of such knowledge may
result in a requirement for calibration. The operation and
performance of the hohlraum is strongly dependent on
the transmitting capabilities of the quartz window. Since

small surface contaminations of the quartz may markedly affect the transmissivity of the window, periodic transmission checks may be required to isolate this uncertainty.

Experimental thermocouple measurements for T_2, T_2' and T_3 indicate that all three surfaces were nonisothermal but each set of readings showed consistent increases or decreases at a given circumferential location. The true nature of the hohlraum temperature distribution is not known because of an insufficient number of thermocouples. It is possible that reading differences were due to differences in the bonding (spot-weld) of the thermocouples and/or radiation and conduction errors in the leads. However, due to the configuration of the heat transfer apparatus (Fig. 1) several other possibilities exist: (1) radiation emitted by the gas is anisotropic, (2) the quartz window does not view the apparatus symmetrically and, perhaps most significant, (3) rearrangement of the flow after negotiating the 90-deg bend produces circumferential variations in convective heat transfer to the window which produces similar variations in quartz emission into the cavity.

The effect of nonisothermal cavity conditions is difficult to assess, particularly in the present hohlraum which has a rather complicated shape. Very little information on this effect is available in the literature. In Refs. 8 and 9, the effect of temperature gradients along the wall on the apparent emissivity and rate of radiant emission from simple cylindrical cavities has been examined and found to be pronounced. In Ref. 10, the effect was found to be relatively less for V-groove cavities. In general, it appears that temperature gradient effects: (1) are greater for diffuse rather than for specular surfaces, (2) increase with increasing emissivity of the wall material, and (3) produce a relative increase in cavity emission when the wall near the opening is warmer than the cavity base. The information contained in Ref. 11, although not suitable for direct application here, gives valuable insight into the effects of nonisothermal conditions in hohlraums.

Since the quartz window was not directly cooled, it is likely that its effective temperature was at least as high as the cavity temperature T_2. In addition to its role as a transmissive agent for thermal radiation from the gas, the window also absorbs heat by convection on the gas side. This heat is partially released in the form of emission, both to the gas and into the cavity. It is ascertained from Ref. 5 that the total emissivity of quartz is relatively high at moderate temperatures. Thus, the effect of the window is apt to be important, in that values of q_r determined from Eq. (1) do not include quartz emission and are thus an upper bound on true values of q_r.

Thermal radiation incident on the quartz window is not due entirely to gas radiation if the gas is optically thin; in the present case, the gas is probably optically thin since the pressure is not high and the path length is relatively small. The quartz window also receives radiation from itself and from the duct walls, both directly and by reflection. The gas may be construed as being a

---

Definition of symbols

A	surface area
f	flatness factor of cavity
F	configuration or view factor
q	heat flux, thermal energy transfer per unit time and area
S	sum of areas A_1 and A_2
T	temperature, absolute
α	absorptivity
ε	emissivity
σ	Stefan–Boltzmann constant

Subscripts

a	apparent
$cond$	conduction component
$conv$	convection component
$loss$	loss term
r	radiative component
0	fictitious surface (gas)
1	window surface
2	cavity surface of hohlraum (liner)
3	cavity-housing surface

Superscripts

' (prime)	denotes gas-side of window when used with 1
	denotes outside surface of cavity liner when used with 2
	denotes flux per unit area of surface 1 when used with subscript $cond$
— (bar)	denotes average value when used with T
	denotes value obtained using $\overline{T}$ and $\alpha_2 = \varepsilon_2$ when used with q_r

fictitious surface area and all former effects may be lumped in a configuration factor F_{10} from window to gas. It is anticipated that F_{10} approaches unity in the present apparatus.

6. Conclusions

Preliminary experiments using a hohlraum have indicated that the total radiation from ionized argon in the present apparatus cannot exceed approximately 2% of the total heat transfer to the wall, for the experimental conditions imposed. With some refinements in both technique and calculation procedures, it is judged that the hohlraum is adequate for establishing the relative magnitude of the radiative heat transfer component. It is probable that the effect of the quartz window cannot be neglected when determining thermal radiation from hohlraum measurements. This effect is currently under analytical investigation. Symbols used in this Section are defined in the table on p. 92.

B. A Design for an All-Liquid Heat-Rejection Radiator System

J. J. Volkoff

1. Introduction

The objective of this radiator design program is to estimate, from specified system requirements, the minimum weight of the radiator system and the dimensions of the related components comprising the radiator system used for space power plants.

A heat-dissipation system is required for most space power plants. The most effective method for removing waste heat developed by the power plant is by employing a coolant which rejects the absorbed waste heat by conduction to a thermal radiator. The radiator in turn rejects the heat by thermal radiation to the space environment.

A radiator system is composed of supply and return headers and panels. The panel arrangements (Ref. 12) most likely to be considered for a radiator system are shown in Fig. 5. The tapered cylindrical headers are arranged and sized so that they will distribute coolant

PANEL	PARAMETERS		
	m_1	m_2	m_3
ONE	I	I	I
TWO	2	0.5	I
FOUR	I	0.5	0.5
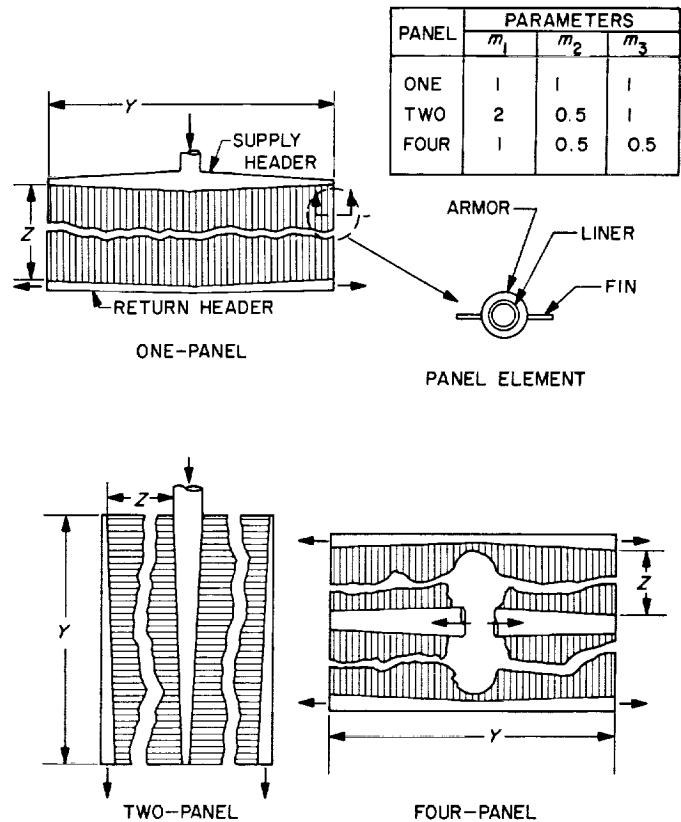
Fig. 5. Panel configurations

uniformly to the array of finned-tube elements comprising the panel section. The cylindrical tubes are of uniform diameter. The finned-tube element shown in Fig. 5 consists of a tube liner, if necessary, to be compatible with highly corrosive materials such as liquid metals, an armor enveloping the tube to resist meteoroid impact damage, and adjacent rectangular fins.

The three design constraints that the radiator system must satisfy are total heat rate to be rejected, fluid flow requirements, and protection requirement to resist meteoroid impact damage. However, there may be size constraints imposed upon the heat-rejection system because of spacecraft system integration requirements. The heat-rejection system must then be divided into a number of radiator systems, each responsible for a designated fraction C_{13} of the total thermal energy Q_{rej} to be rejected by the combined radiator systems. The thermal energy Q_1 rejected by a radiator system can then be expressed by

$$Q_1 = C_{13} \cdot Q_{rej} = NZ\sigma\epsilon_t\theta_t^4 \left[F_{t\pi}(d_i + 2\Delta_a + 2\Delta_c) \right.$$

$$\left. + 4F_f L\eta \right] + \sum_{k=1}^{n} (F_h A_h \sigma \epsilon_h \theta_h^4)_k \qquad (1)$$

where the first term represents the heat rejected by the finned-tube elements and the second term represents the heat rejected by the headers. The parameter N is the number of tubes, Z is the length of each tube, σ is the Stefan–Boltzmann constant, ε_t is the net hemispherical emittance coefficient for the finned-tube element, θ_t is the mean surface temperature of the tube, F_t, F_f and F_h are geometric view factors, d_i is the inside tube diameter, Δ_a and Δ_c are wall thicknesses for the armor and tube liner respectively, L is the fin length, η is the heat transfer effectiveness of the fin, and A_h is the outer surface area of the header. The summation in the second term is carried out over the total number of headers used for the radiator system.

The pressure drop ΔP_t of the fluid flow in the tubes can be expressed by a form of the Darcy equation [Eq. (1)] as:

$$\Delta P_t = \left[0.83 + \frac{(0.184)\,Z}{d_i Re^{0.2}} \right] \cdot \frac{\rho_l V_t^2}{2g} \qquad (2)$$

where Re is the mean Reynolds number of the fluid, ρ_l is the fluid density, V_t is the fluid velocity, and g is a gravitational constant.

The armor requirement is defined as follows (Ref. 13):

$$\Delta_a = \mathcal{R}\,(0.562)\,(10^{-5})\,K_8 K_{10}\left[\frac{A_v \tau}{C_{13}\,(-\log_e P_{(0)})} \right]^{0.248} \qquad (3)$$

where the vulnerable area A_v includes the headers as well as the tubes, and the protection requirement ratio $\mathcal{R}$ defines the armor ratio required for the specified interplanetary mission. The parameter K_8 is a factor which normalizes the considered target material properties with respect to the meteoroid-particle material properties and is shown by Fig. 6. The factor K_{10} is a constant, equal to 1.82, that converts the units of the parameters expressed in Eq. (3) from the metric system to the British system.

It is evident from Eqs. (1) and (3) that the number of variables is large and that they are interrelated. For example, the heat rejection is directly proportioned to the number of tubes [Eq. (1)], the fluid velocity is also a function of the number of tubes, [Eq. (2)], and the vulnerable area is not only a function of the number of tubes, but includes the headers which are also a function of the number of tubes, [Eq. (3)]. This describes calculations which indeed are tedious. Therefore, JPL Computer
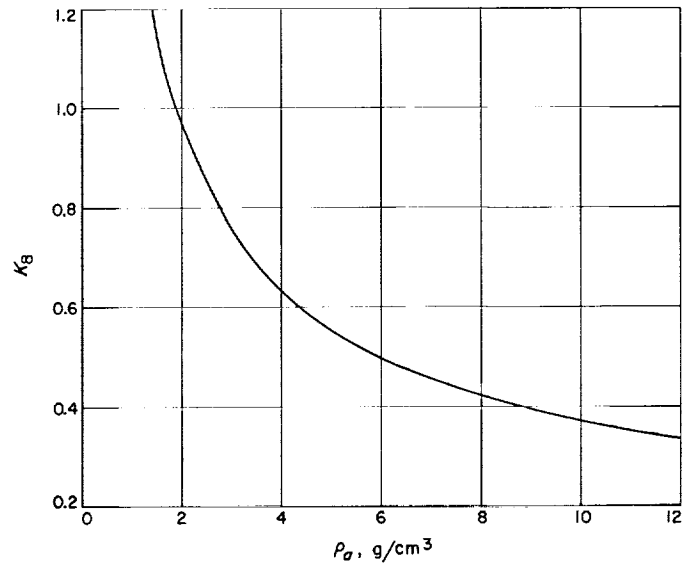
Fig. 6. Factor K_8 vs target material density

Program No. M091 has been developed to design the minimum weight radiation system.

2. Program Description

The inputs to the program are read in. First, the design constraints: these are the total heat rate to be rejected, the design pressure drop, the total coolant weight flow rate, the fraction of the total heat to be rejected by the considered radiator system, the probability of no meteoroidal impact damage, the mission time, and the surface temperature of the tube and supply and return headers. Then the material properties are read into the program: the thermal conductivity of the armor and fin, the density of the armor, fin, liner, fluid, and the viscosity of the fluid. The variable design inputs are the fluid velocity at entrance to the supply header, the panel configuration, the inside diameter of the tube, and the tube diameter increment required for the program in order to calculate the results for four sets of tube diameters for one run.

From Eq. (2), the length of a tube in the panel is given by

$$Z = \left[\Delta P_t - \frac{(0.83)\,\rho_l}{2g}\,V_t^2 \right] \cdot \left(\frac{g d_i Re^{0.2}}{(0.092)\,\rho_l V_t^2} \right) \qquad (4)$$

and the total number of tubes N for the radiation system

$$N = \frac{4 \dot{W} C_{13}}{\rho_l \pi d_i^2 V_t} \qquad (5)$$

where $\dot{W}$ is the weight flow rate of the fluid.

The length of the header Y is expressed as

$$Y = Nm_2 (d_i + 2L + 2\Delta_a + 2\Delta_c) \qquad (6)$$

where m_2 is a constant defined by the panel configuration as shown in Fig. 5. It can be seen that the length Y is directly proportional to the finned-tube geometry. The geometry of the finned-tube element of the panel may be designed so that either a minimum weight of the element configuration results, or so that a minimum weight of the radiator system results. The minimum weight finned-tube design is defined as that geometry whose heat rejection rate per unit element weight is maximum for a given tube diameter, specified tube temperature, and materials. The minimum weight radiator design incorporates an optimization between the headers and the panel such that there results a maximum heat rejection per radiator system weight. This design suggests that the weight of the resulting finned-tube element may be slightly greater than that for the minimum weight finned-tube design. However, when the header weights are combined with the panel weights, the resulting weight for the minimum weight radiator design is less than that weight for the minimum weight finned-tube design. The results obtained from these two designs for a selected radiator system are given comparatively in Fig. 7, which shows a weight difference of about 0.5% between each design. The primary difference between the two designs is their respective number of iteration procedures. The

minimum finned-tube weight design is selected for this program because its results are reasonably minimum and it contains iteration procedures which require less program time, by at least one order of magnitude, than for those iteration procedures contained in the minimum weight radiator design.

The variables of the finned-tube elements, as described in the parentheses of Eq. (6), are found from Eq. (3) and a subroutine which defines fin-length for the minimum weight design of the finned-tube element. This fin-length L, derived by maximizing the heat rejected for total element weight, is obtained from:

$$\frac{8\rho_f \sigma \epsilon \theta_a^3}{k_f} L^3 + \frac{3\pi \rho_f \sigma \epsilon F_t \theta_a^3 (d_i + 2\Delta_c + \Delta_a)}{k_f F_f \eta_f} L^2$$
$$- \epsilon N_c \left[\rho_c \pi \Delta_c (d_i + \Delta_c) + \rho_a \pi \Delta_a \right.$$
$$\times \left. (d_i + 2\Delta_c + \Delta_a) \right] = 0. \qquad (7)$$

The root thickness of the fin is

$$t = \frac{2\sigma \epsilon \theta_a^3 L^2}{k_f} \qquad (8)$$

where ρ_f, ρ_c and ρ_a are the density of the fin, liner, and armor respectively, θ_a is the mean surface temperature of the armor, and k_f is the thermal conductivity of the fin.

The vulnerable area of the tubes is (Ref. 13):

$$A_{v_t} = NZ \left[\pi (d_i + 2\Delta_c + 2\Delta_a) - 2t \right]. \qquad (9)$$

The total vulnerable area used in Eq. (3) is:

$$A_v = A_{v_t} + A_{h_t} \qquad (10)$$

where A_{h_t} is the combined vulnerable area of the headers.

The surface areas of the supply and return headers are found, respectively, by the equations

Fig. 7. Comparison of radiator heat rejection per unit weight obtained from the minimum weight finned-tube design and the minimum weight radiator design

$$A_{h_s} = \pi Y \left[\left(\frac{D_{s_i} + D_{s_o}}{2} \right) + 2\Delta_c + 2\Delta_a \right] \qquad (11)$$

$$A_{h_r} = \pi Y \left[\left(\frac{D_{r_i} + D_{r_o}}{2} \right) + 2\Delta_c + 2\Delta_a \right] \qquad (12)$$

$$A_{h_t} = A_{h_s} + \frac{1}{m_2} A_{h_r} \qquad (13)$$

where

$$D_{s_i} = \left(\frac{2m_1 C_{13}\dot{W}}{\pi \rho_l V_h}\right)^{1/2}$$ (14)

$$D_{s_o} = 1.5\left(\frac{2m_1 C_{13}\dot{W}}{\pi(n-1)\rho_l V_h}\right)^{1/2}$$ (15)

$$D_{r_i} = 1.5\left(\frac{2C_{13}\dot{W}m_3}{\pi(n-1)\rho_l V_h}\right)^{1/2}$$ (16)

$$D_{r_o} = \left(\frac{2C_{13}\dot{W}m_3}{\pi \rho_l V_h}\right)^{1/2}.$$ (17)

The generalized flow chart for the program is shown by Fig. 8. After the input data are read into the program, the tube length and number of tubes for a specified tube diameter are calculated from Eqs. (4) and (5) for an initial arbitrary tube velocity V_2. With these results, the program calls the subroutines containing Eqs. (3), (7)–(17) and calculates the header length Y by Eq. (6). Due to the interrelationship of the different variables shown in these equations, iteration is used until convergence is obtained at the program node a.

At this point, the radiator system has a geometry for the panel and headers, and satisfies the fluid flow and the armor protection requirements. The program proceeds to compute for this geometry the heat rejection rate given as follows.

The heat rejected by the panel, the supply header, and the return header are, respectively:

$$Q_r = NZ\sigma\epsilon_t\theta_t^4\left[F_{t\pi}(d_i + 2\Delta_a + 2\Delta_c) + 4F_fL\eta\right]$$ (18)

$$Q_{h_s} = \epsilon_{h_s}\sigma A_{h_s}F_{h_s}\theta_3^4$$ (19)

$$Q_{h_r} = \frac{1}{m_2}\epsilon_{h_r}\sigma A_{h_r}F_{h_r}\theta_4^4.$$ (20)

The total heat rate rejected Q_C for the geometry is:

$$Q_C = Q_r + Q_{h_s} + Q_{h_r}.$$ (21)

If Q_C is not equal to the required heat rejection rate Q_1 within a specified percentage (i.e. 0.3%) at program node b, the program then selects a second arbitrary tube velocity V_3 and iterates a second set of results which specify Z, N, and geometry. The selection of the velocity
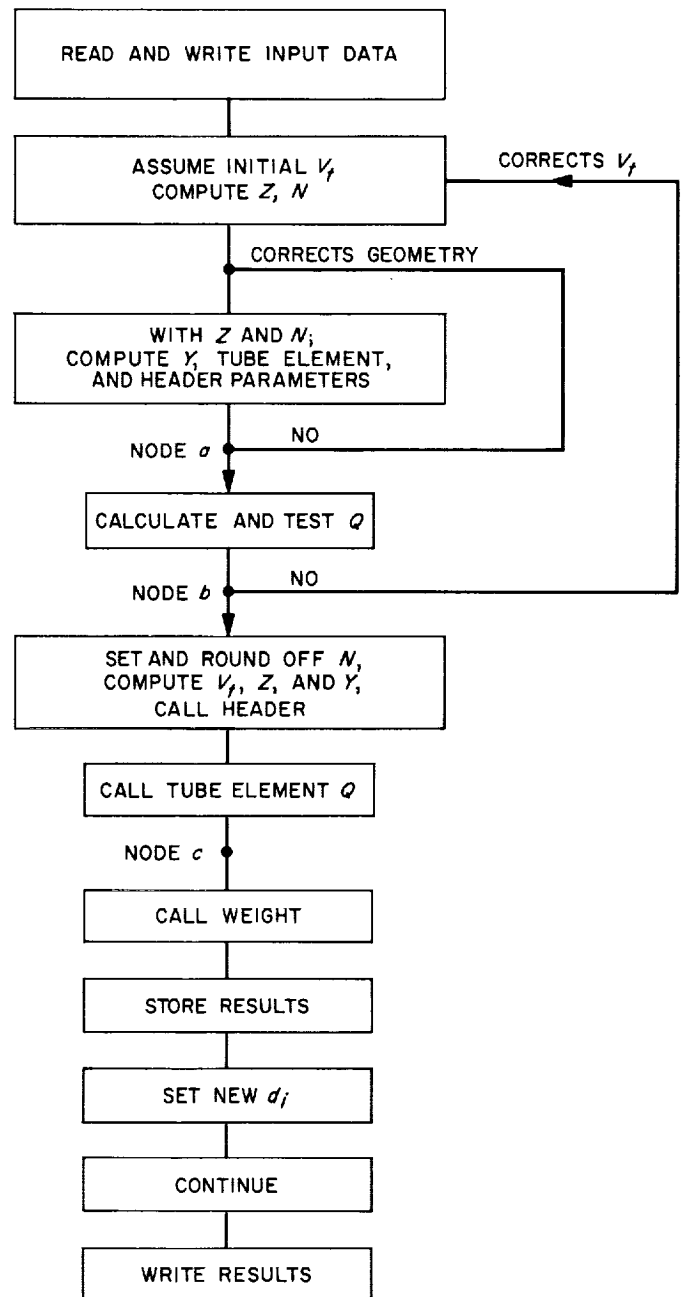
Fig. 8. Program flow chart

V_t corresponding to Q_1 is obtained by a logarithmic interpolation as follows:

$$\log V_t = \log V_2 + \frac{\log Q_1 - \log Q_2}{\log Q_3 - \log Q_2}(\log V_3 - \log V_2).$$ (22)

Iteration proceeds until convergence at node point b occurs, thereby satisfying all the radiation system requirements. The program then rounds off N to the nearest

whole number of tubes. A corrected tube velocity, Z, radiator geometry, and Q_c are then recomputed, based on the realistic value of N. At this node point c, the weight of the radiator components can then be calculated as follows.

The total liner weight for the headers is:

$$W_{hdr_c} = \frac{\pi Y}{2} \rho_c \Delta_c \left[D_{s_i} + D_{s_o} + \frac{1}{m_2} (D_{r_i} + D_{r_o}) \right].$$

(23)

The total armor weight enveloping the headers is:

$$W_{hdr_a} = \pi Y \rho_f \Delta_a \left[\frac{D_{s_i} + D_{s_o}}{2} + \left(1 + \frac{1}{m_2}\right)(2\Delta_c + \Delta_a) \right.$$

$$\left. + \frac{1}{2m_2}(D_{r_i} + D_{r_o}) \right].$$

(24)

The panel weight consisting of the total finned-tube elements is:

$$W_r = NZ \{ 2\rho_f t L + \pi [\rho_c \Delta_c (d_i + \Delta_c) + \rho_a \Delta_a$$

$$\times (d_i + 2\Delta_c + \Delta_a)] \}.$$

(25)

After the weights are calculated, the program then stores all the results and continues to the next tube diameter that is specified. Four sets of results corresponding to four different tube diameters described by the inputs to the program are written.

3. Program Output

The program output describes the inside tube diameter, fin length, fin thickness, armor thickness, tube liner thickness, number of tubes, diameter of the supply and return headers (both at their respective inlet and outlet ends), header length, tube length, length to diameter ratio of the tubes, mass flow rate in the tubes, in-tube flow area, mean liquid velocity in the tubes, planform area of the panel, panel aspect ratio (i.e. Z/m_2Y), panel specific heat rejection rate (Btu/hr lb), heat rejection rate by the radiation system, combined weight of the headers, panel weight, total radiator system weight, total weight of the liner, total weight of the armor, and total weight of the fins.

The symbols used in this Section are defined in the table on this page.

Definition of symbols

A	surface area, ft²
C	constant
D	radiator header inside diameter, ft
d	tube diameter, ft
F	view factor, sterad
g	gravitational constant, ft/sec²
L	fin width, ft
m	constant designating panel arrangement—Fig. 1
N	number of tubes
$P_{(0)}$	probability of no catastrophic impacts by meteoroidal particles
Q	heat transfer rate, Btu/hr
$\mathcal{R}$	protection requirement ratio for an interplanetary mission relative to near Earth[13]
τ	mission time, sec
Re	Reynolds number
V	flow velocity, ft/sec
Y	header length, ft
Z	length of tubes, ft
ε	surface net hemispherical emittance
η	heat transfer effectiveness of the fin
Δ	thickness, ft
σ	Stefan-Boltzmann constant, 0.1712×10^{-8} Btu/hr ft² °R
ρ	density, lb/ft³
θ	temperature, °R
t	fin thickness, ft
ΔP	pressure drop, lb/ft²
$\dot{W}$	coolant weight flow, lb/sec

Subscripts

f	fin
h	header, return or supply
i	inside, inlet
c	liner
t	tubes
v	vulnerable
rej	reject
o	outlet
a	armor
l	liquid

C. Liquid MHD Power Conversion

D. Elliott, D. Cerini, L. Hays, and E. Weinberg

The long lifetimes required of electric-propulsion powerplants makes cycles without rotating components desirable. Such a cycle under investigation at JPL is the liquid metal magnetohydrodynamic (MHD) system shown schematically in Fig. 9. In this cycle a fluid, such as cesium, circulates in the vapor loop and causes a liquid metal, such as lithium, to recirculate through an MHD generator in the liquid loop. The cesium leaves the radiator (or radiator-loop condenser) as condensate, flows through an electromagnetic pump and regenerative heat exchanger to the nozzle, vaporizes on contact with the lithium, atomizes and accelerates the lithium in the nozzle, separates from the lithium in the separator, and returns to the radiator through the regenerative heat exchanger. The lithium leaves the separator at high velocity (typically 500 ft/sec), decelerates through the production of electric power in the MHD generator, and leaves the generator with sufficient velocity (typically 300 ft/sec) to return through a diffuser to the reactor (or reactor-loop heat exchanger) where the lithium is reheated.

The program being followed is to investigate the hydraulic and electrical problems of the system in full scale [50–300 kw (e)] cold tests with NaK simulating the lithium and nitrogen gas simulating the cesium vapor, to study the materials problems in small-scale loop tests with 2000°F cesium and lithium, and then to combine the results in a 2000°F prototype ground-test system.

1. 5-kw Generator

Prior to fabrication of a full-scale NaK-nitrogen system, the basic problems of ac MHD generators are being investigated using an ac MHD generator of 5-kw nominal output power which has been fabricated for testing with pressure-fed NaK. Although limited by its small size to no more than 30–40% efficiency, the 5-kw generator incorporates all of the features believed necessary for 50–70% efficiency at the 50–300 kw level, i.e., electrical length of only one wavelength to minimize winding losses; length-to-gap ratio below 20 to minimize friction loss; rms magnetic field greater than 5000 gauss to make the electric retarding force large, as compared with the friction force; compensating poles at inlet and exit to cancel the pulsating currents produced by a traveling magnetic field in a short channel; and conductivity-reducing devices, such as axial vanes, in the compensating-pole regions to minimize eddy-current losses under the
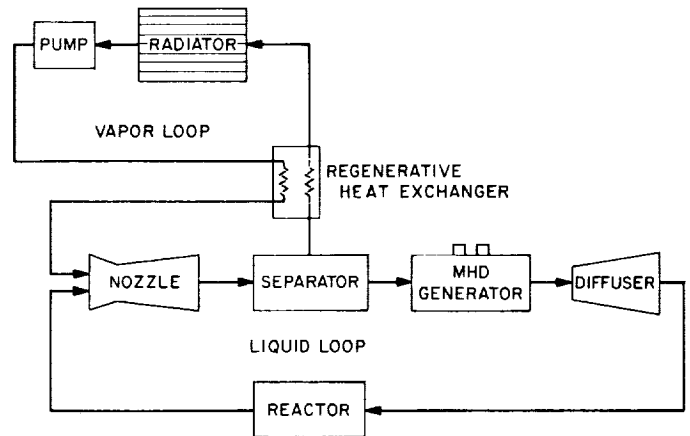
Fig. 9. Liquid MHD power conversion cycle

compensating poles. A discussion of these considerations and of the 5-kw generator design may be found in *SPS 37-33*, Vol. IV, pp. 142–150.

Fig. 10 shows the 5-kw ac generator prior to winding the stators, and Fig. 11 shows the flow channel with the upper stator block removed. The stator blocks are clamped to copper side strips to form the flow channel and then clamped between a nylon inlet nozzle and nylon exit block. The stator blocks are held together by the two halves of the micarta housing, and the nylon blocks are clamped to the stator blocks by the inlet and outlet pipe flanges. During operation, a flow of mineral oil through the housing cools the windings and removes NaK leakage.

Fig. 12 shows the stator blocks before winding. The blocks are bonded from 8-mil Hiperco-50 (49% Fe, 49% Co, 2% V) chosen for high saturation flux and low core loss, and the slots are formed by electrical discharge machining. The five inner slots on the upper stator and the six inner slots on the lower stator carry the traveling-wave windings, and the large end-slots carry the compensating-pole coils, together with the return sides of the traveling-wave windings. The different width at top and bottom of each inner slot approximates the optimized slot shape to be discussed later.

The wound stator blocks are shown in Fig. 13. The conductors are 0.130- × 0.068-in. rectangular film-coated magnet wire with 2-mil mylar separators and slot liners, impregnated with varnish and then baked. Each of the upper slots contains 12 conductors of one phase, and the lower slots contain 6 conductors of each of two phases. The combination provides a 5/6-pitch winding with two slots per pole per phase or, in effect, a six-phase winding.

PRESSURE MEASUREMENT LINES

INLET PIPE

UPPER STATOR BLOCK

EXIT PIPE

MICARTA HOUSING

0 2 4 6 8 10
INCHES
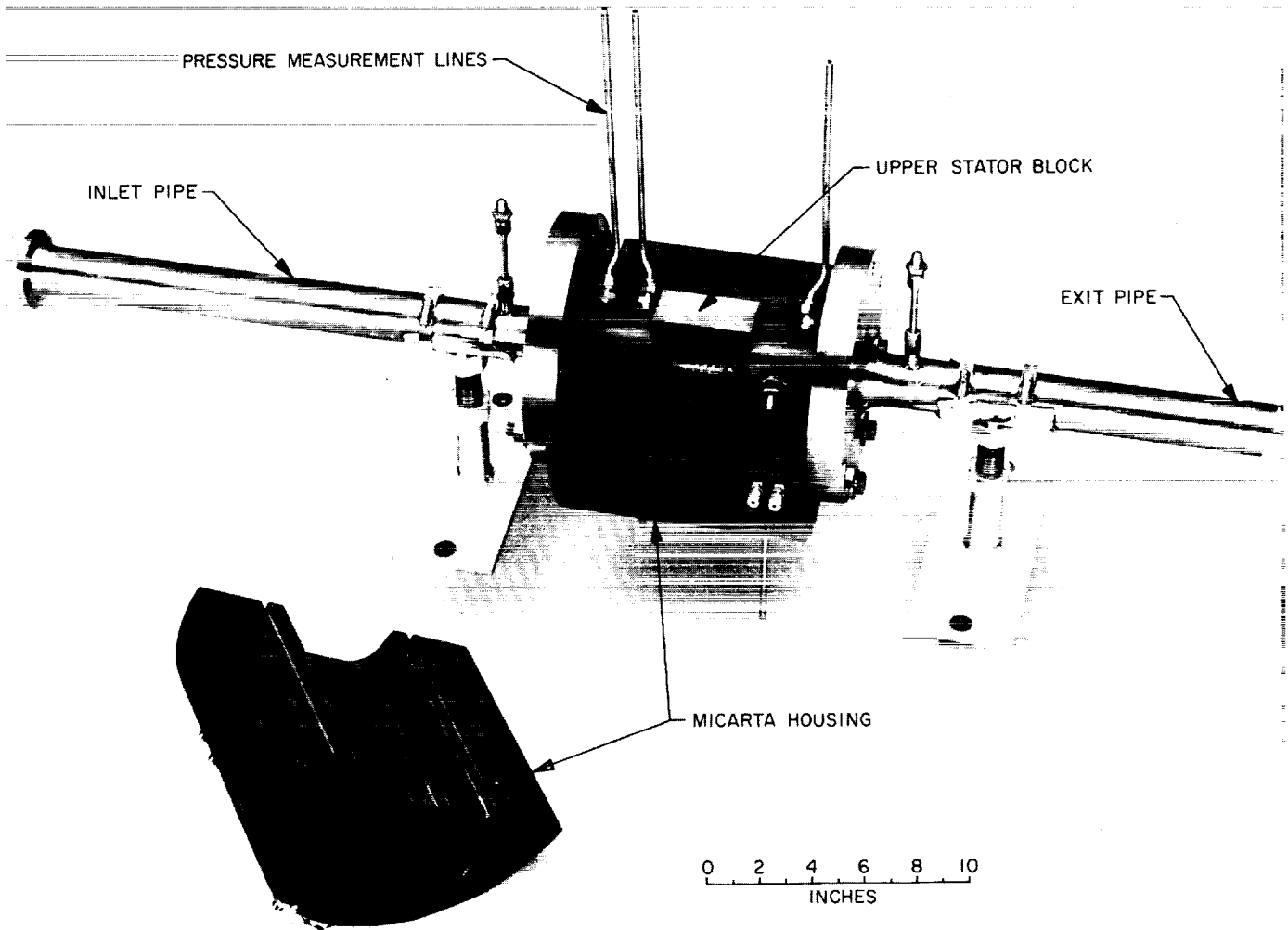
Fig. 10. Experimental 5-kw ac MHD generator

The generator has been assembled, leak-checked, and the magnetic field profile determined using the test arrangement shown in Fig. 14. The generator was mounted vertically, the inlet pipe and flange were removed, the housing and channel were filled with mineral oil for cooling, and current was supplied to the windings from dc power supplies. The magnetic field along the channel was measured by moving a Hall-effect probe down the channel and plotting the output versus distance on an X-Y plotter. The resulting field profiles are presented in Fig. 15 for various phase angles at an equivalent rms current of 90 amp/phase. The wave shape is a satisfactory approximation to a traveling sinusoidal wave. However it was found that, due to an error, the Hiperco-50 laminations had not been properly heat-treated and the desired 9000-gauss peak field cannot be reached. A second set of stator blocks is being fabricated while tests proceed at reduced performance with the initial assembly.

The remaining tests before running with NaK are ac empty-channel measurements to set the excitation capacitance values and to measure the ac winding and core losses.

2. 50-kw Generator

For the NaK-nitrogen system the ac generator will have 50 kw nominal output power. The design parameters tentatively chosen are: length, 4.3 in.; width, 6.06 in.; inlet gap, 0.400 in.; exit gap, 0.515 in.; fluid inlet velocity, 270 ft/sec; fluid exit velocity, 210 ft/sec; rms field, 6500 gauss; and frequency, 575 cps. To provide decreasing
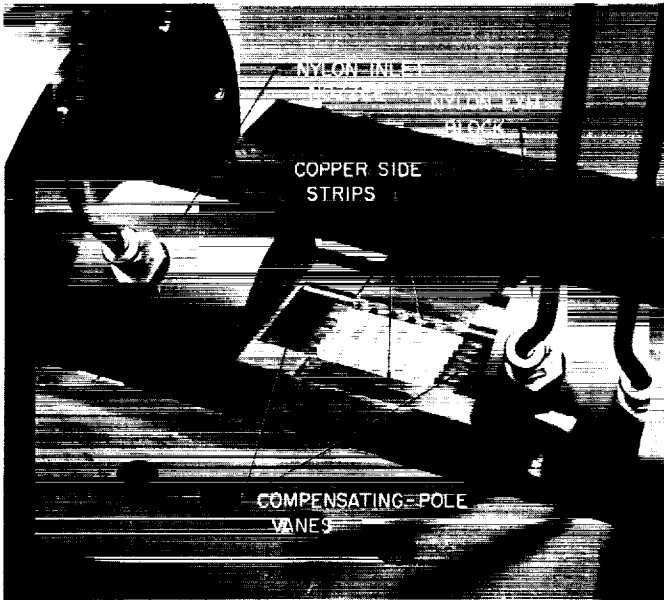
Fig. 11. AC generator flow channel with upper stator block removed
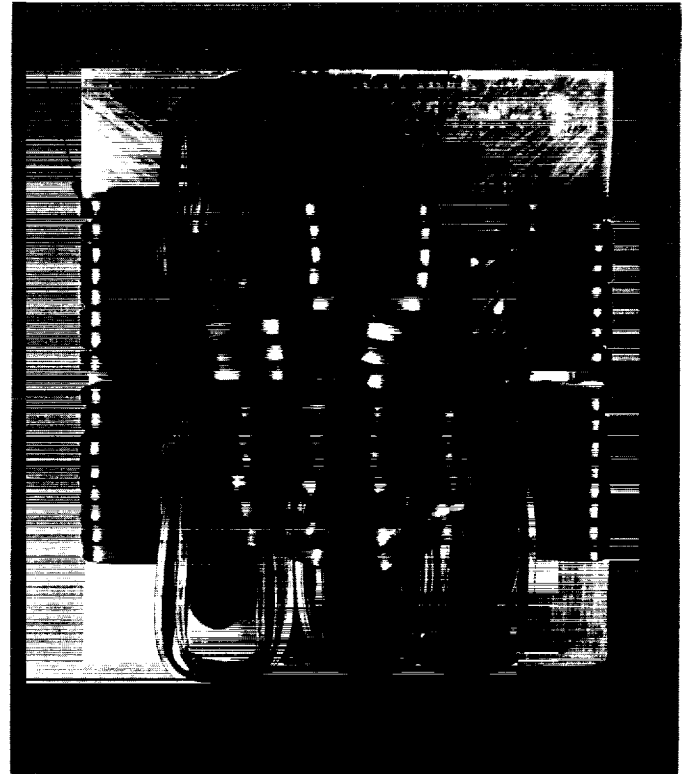
Fig. 13. AC generator stator blocks after winding
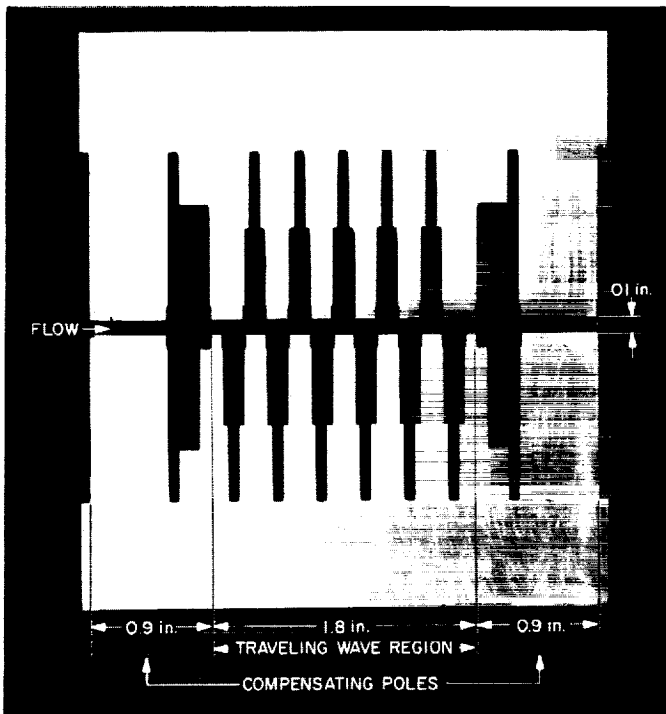
Fig. 12. AC generator stator blocks before winding

Fig. 14. Test arrangement for measuring ac generator magnetic field with dc excitation
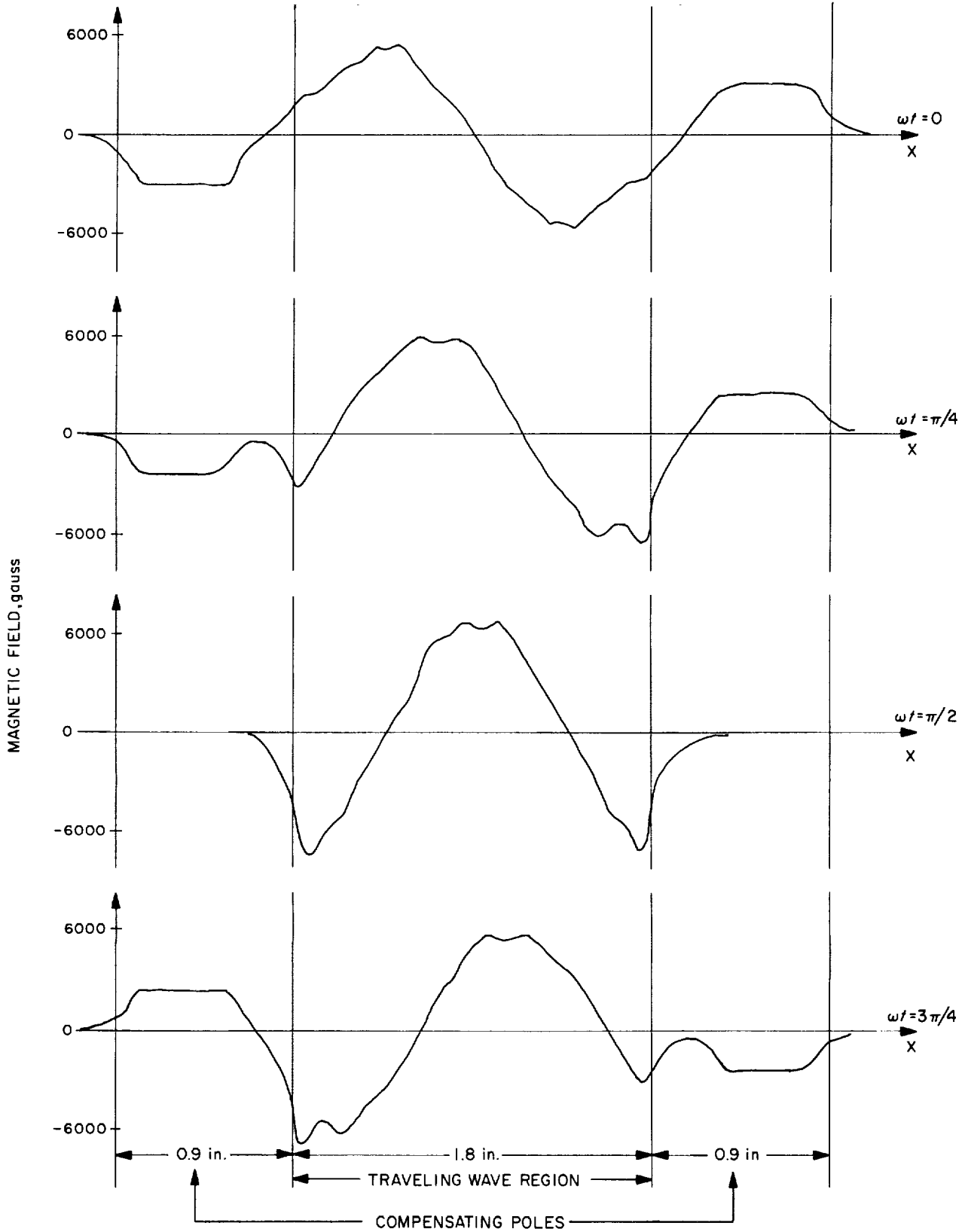
Fig. 15. Measured magnetic field profiles in ac generator at various phase angles for 90-amp rms phase current

wave velocity with distance to match the decreasing fluid velocity, the slot pitch decreases from inlet to exit. An iron mockup of the stator blocks was tested with dc excitation and the variable pitch was found to provide the desired variation in wave speed.

To minimize winding losses, the 50-kw generator employs tapered slots to optimize the distribution between iron and copper. Fig. 16 illustrates the problem. Slots of pitch p are to be contoured so that all of the iron in the tooth between them becomes saturated simultaneously when the peak gap field B_m passes the tooth. Any iron that remains unsaturated could be replaced by copper with reduction in winding loss; hence, complete tooth saturation is the optimum situation.

Saturation at the tooth tip results from making the slots wide enough at the gap so that

$$pB_m = (p - w_1) B_{sat}$$

or

$$w_1 = p\left(1 - \frac{B_m}{B_{sat}}\right) \tag{1}$$

where B_{sat} is the saturation flux density of the iron.

It then remains to taper the slots to provide increasing tooth width to carry the slot flux. As indicated in Fig. 16, the tooth can be considered as divided into a center zone carrying the gap flux, and into two outer zones carrying the slot fluxes; the slot contour that saturates the outer zones is the desired one. Consider, first, the empty-channel

Fig. 16. Stator geometry for optimization of slot shape

case. When B_m is centered over the tooth, the slots carry equal and opposite currents I and $-I$, the slot fluxes are equal, and the three zones are divided by straight lines perpendicular to the gap as shown. Consequently, each increment of slot depth dy carries a flux $B\,dy$ which is exactly equal to the flux $B_{sat}\,dw/2$ in the corresponding increment of iron width. Thus, the slope of the slot must be

$$\frac{dw}{dy} = w' = \frac{2B}{B_{sat}}. \tag{2}$$

The flux density crossing the slot at y is equal to μ_0/w times the enclosed current. Thus,

$$B = \frac{\mu_0 i}{w} \int_0^y w\,dy \tag{3}$$

where i is the current density.

Combining Eqs. (2) and (3), the relationship between w, B_{sat}, and i is

$$w\,w' = \frac{2\mu_0 i}{B_{sat}} \int_0^y w\,dy. \tag{4}$$

Differentiating Eq. (4) yields the differential equation for w

$$w\,w'' + w'^2 = \frac{2\mu_0 i\,w}{B_{sat}}. \tag{5}$$

Eq. (1) provides the boundary condition at $y = d$. The boundary condition at $y = 0$ is given by Eq. (4) as

$$w'(0) = 0 \tag{6}$$

since w_2 is, in general, not zero.

The slot current is

$$I = i \int_0^d w\,dy. \tag{7}$$

The power dissipated per unit length of slot is I^2 times the resistance per unit length $\mathcal{R}/A$, where $\mathcal{R}$ is the copper resistivity and A is the copper cross section. But A is equal to I/i; hence, the power dissipated in the winding per unit slot length is

$$P'_w = iI\,\mathcal{R} \tag{8}$$

and the minimum loss for a given slot current occurs when the current density is minimum, corresponding to increasing the slot depth d until the slot comes to a sharp point ($w_2 = 0$).

In general, with current in the fluid, the symmetrical situation of Fig. 16 does not occur, but the slot fluxes obtained from the above treatment can be used to construct a phasor diagram for the tooth flux to determine the equivalent current I for which the slots must be designed.

Fig. 17 presents the slot current I, for maximum-depth (sharp-pointed) slots, as a function of current density for slot pitch $p = 1.8$ cm, peak gap flux density $B_m = 8500$ gauss, and saturation flux density $B_{sat} = 20,000$ gauss, conditions of interest for the 50-kw generator. The equivalent slot current required is $I = 2600$ amp, requiring a current density of approximately 1000 amp/cm^2 and a slot depth of approximately 6 cm. By comparison, rectangular slots would require approximately 2300 amp/cm^2 and would use more than twice the power of the tapered
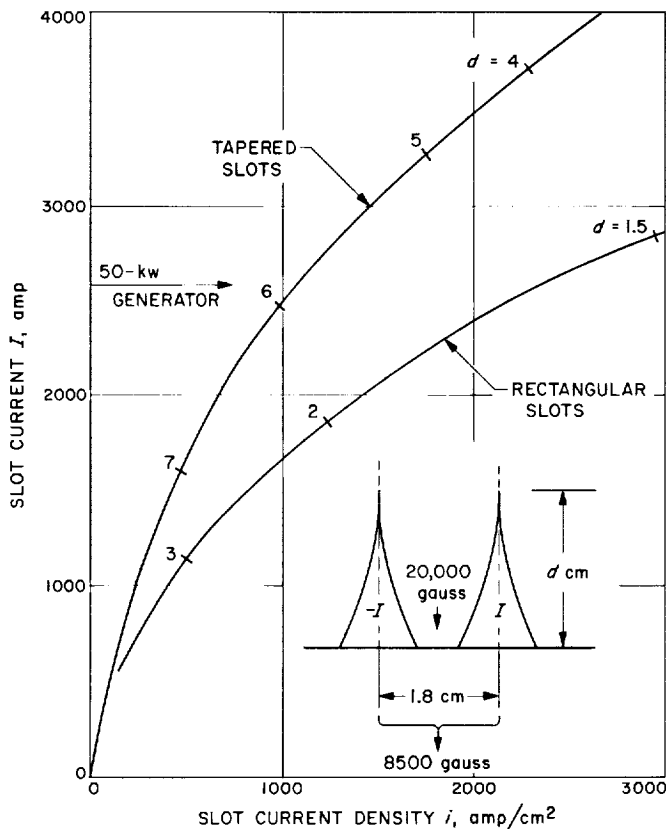
Fig. 17. Variation of slot current with current density for optimum tapered slots and rectangular slots

slots. The actual slot shape chosen for the 50-kw generator is $d = 4$ cm, $w_1 = 1.0$ cm, and $w_2 = 0.25$ cm, since this shape requires negligibly greater current density than the sharp-pointed slot and simplifies fabrication.

Analysis of the 50-kw generator is continuing. Design of the separator and diffuser is nearly completed, and fabrication of the latter components will begin in December.

3. Lithium Heater Development

Development of a 150-v heater for lithium at temperatures to 2000°F at high power levels (5 Mw) has continued. The purpose is to develop a heater design for the 2000°F prototype system and also to provide a heater for immediate use in the 100-kw (t) cesium-lithium erosion loop. Two conduction heaters with Cb-1% Zr sheaths, beryllia insulation, and tantalum center conductors were tested at temperatures to 2300°F. A heater with Cb-1% Zr sheath, alumina insulation and tantalum center conductor had previously been operated in 2080°F lithium for 500 hr (SPS 37-32, Vol. IV, pp. 134-141). These preliminary tests were made to establish feasibility and to provide data on the electrical resistivity and breakdown field strength for beryllia and alumina in the swaged form at elevated temperatures.

The first beryllia heater tested was described in SPS 37-35, Vol. IV, pp. 165-169. Cb-1% Zr blocks were pressed onto the heater to provide increased radiant heat transfer area. However, nonuniform contact between the sheath and blocks resulted in local high temperatures (>2500°F) and resultant failure of the sheath.

The second beryllia heater rejected the generated heat by direct radiation from the sheath to the water-cooled vacuum chamber wall. Fig. 18 shows this heater after 50 hr of operation at sheath temperatures greater than 2000°F. A total of ten cycles from ambient to 2000°F was conducted, in addition to the steady state tests, with no deleterious effects on the beryllia.

The results of electrical tests performed on this heater while at temperature are shown in Fig. 19. The resistivity was determined for the temperature range of 1500 to 2300°F by applying a potential from the sheath to the center conductor. A range of resistivities from 5×10^8 to 8×10^4 $\Omega \cdot$cm was found. The values fall between those for the compressed powder form (Ref. 14) and the solid fired form (Ref. 15).

The potential applied across the beryllia was varied from 50 to 200 v with no apparent effect on resistivity.
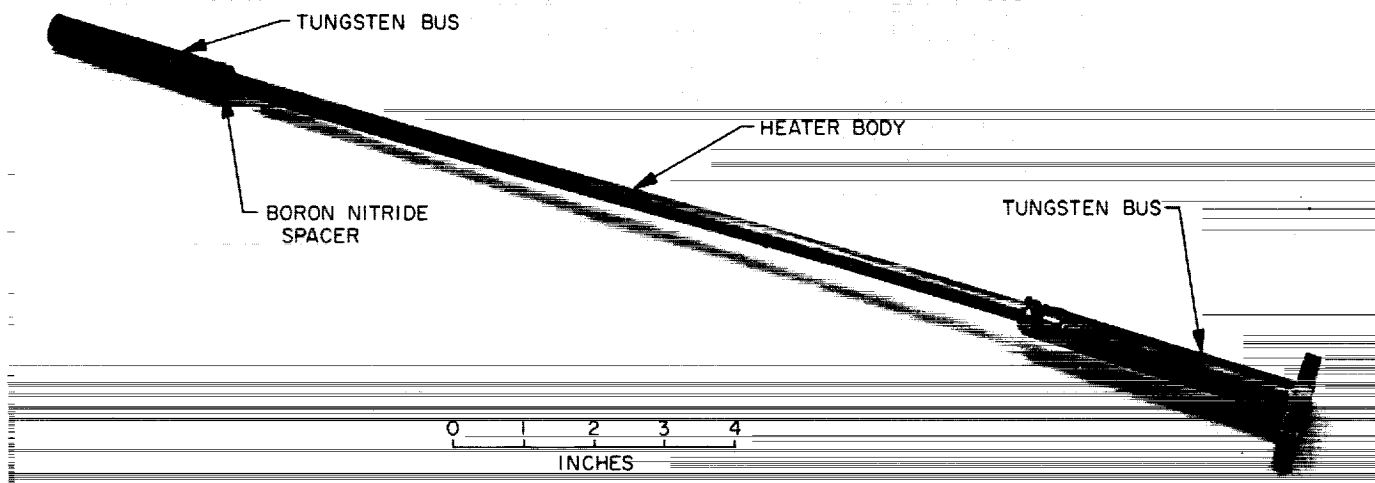
Fig. 18. Experimental conduction heater using beryllia insulation
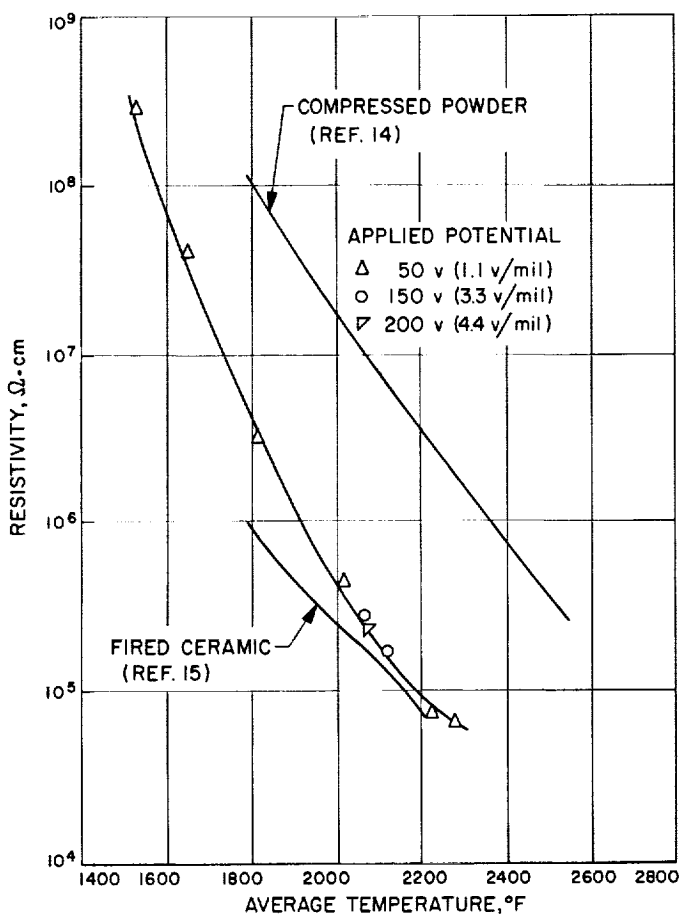
Fig. 19. Measured resistivity of beryllia insulation in the experimental heater and comparison with values for compressed powder and fired ceramic

The maximum field strength applied was 4.4 v/mil at a temperature of approximately 2100°F. This value did not initiate breakdown of the ceramic and is higher than the value of 3 v/mil assumed in preliminary design of the 5-Mw heater.

Fig. 20 shows the lithium heater for the 100-kw erosion loop, prior to welding. This component has four beryllia-insulated heating elements in parallel. The heating elements (as in the test units described above) are geometrically similar to the design for the 5-Mw heater. Further single-element tests, together with the results from the 100-kw erosion loop heater, should provide sufficient information for the final design of the 5-Mw unit.

4. Cycle Analysis

Further cycle performance calculations have been made for the single- and two-component separator cycles and for the jet condenser cycle in order to provide a basis for choice of the final cycle and working fluids for a prototype system.

To compare the performance of the different cycles, the cycle efficiency, prime (isothermal) radiator area, and specific weight were individually optimized for each cycle at maximum cycle temperatures of 1800, 2000, and 2200°F for a number of working fluids. The assumptions and calculation procedures were described in SPS 37-34, Vol. IV, pp. 163–170. Cesium-lithium and potassium-lithium were investigated as working fluid combinations for the two-component separator cycle. Potassium
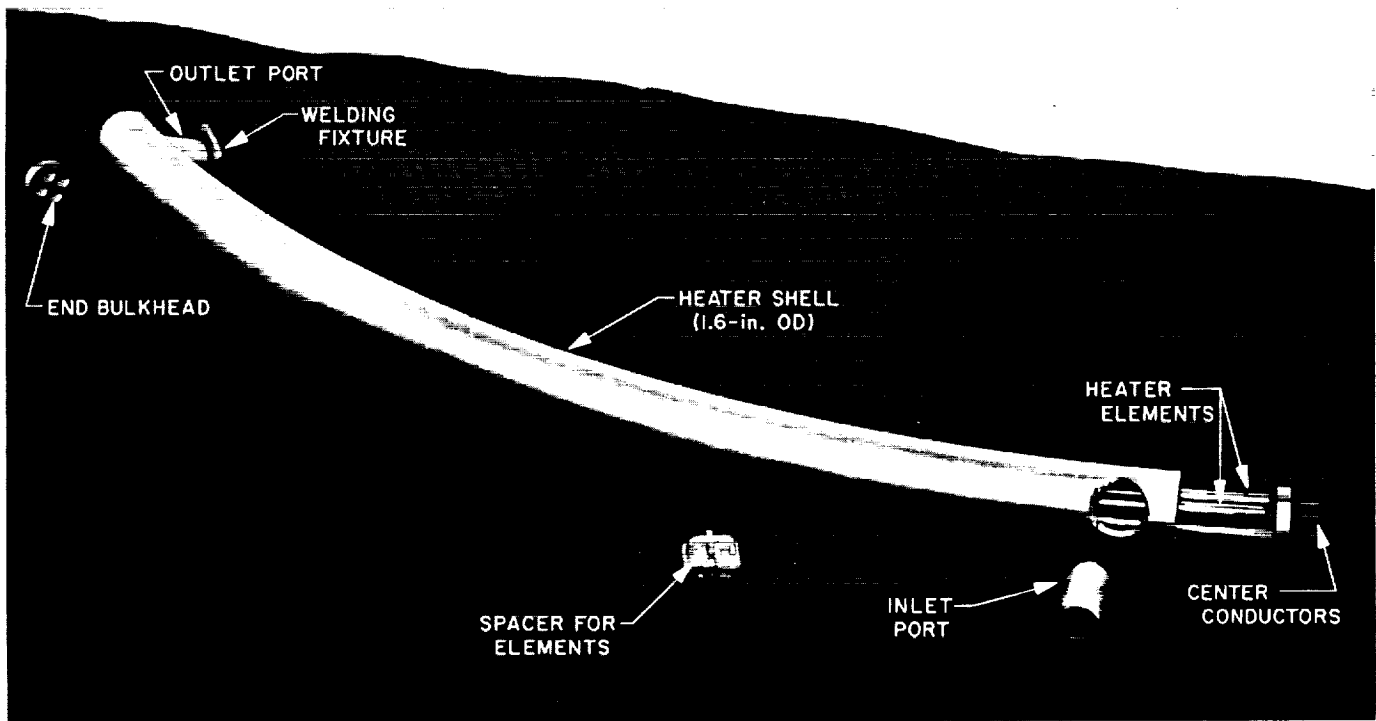
Fig. 20. Lithium heater for the 100-kw erosion loop

and cesium were the fluids considered for the single-component separator and jet condenser cycles.

Fig. 21 illustrates the trade-off between cycle efficiency and radiator area. The highest efficiencies are theoretically attainable with the jet condenser cycles, but at the expense of radiator areas which are larger than those for the separator cycles. For example, at a maximum cycle temperature of 2000°F, a theoretically attainable efficiency for the cesium jet condenser cycle is 11%, versus 7.5% maximum for the cesium-lithium separator cycle and 4% maximum for the cesium separator cycle. At these efficiencies, however, the prime radiator area is 6.0 ft²/kw for the jet condenser cycle versus 2.6 ft²/kw for the two-component separator cycle and 4.0 ft²/kw for the single-component separator cycle. The minimum prime radiator area for the jet condenser cycle is approximately 3.4 ft²/kw, which occurs at a cycle efficiency of 6%. (The separator cycles exhibit efficiency peaks because separator area and friction increase with decreasing nozzle exit pressure. No such exit-pressure penalty was assumed for the condenser cycles.)

Similar trends were found at maximum cycle temperatures of 1800 and 2200°F. At 2200°F, however, the performance of the two-component separator cycle, relative

to the others, is less favorable because of increased solubility of cesium in lithium. The minimum radiator areas at this temperature are 2.1 ft²/kw for the two-component separator cycle versus 2.3 ft²/kw for the best single-component (cesium) separator cycle and 2.4 ft²/kw for the jet condenser cycle. The maximum temperature at which the potassium-lithium combination is feasible is 1800°F due to the high solubility of potassium in lithium; the maximum efficiency at 1800°F is 1.8%.

The influence of maximum cycle temperature on minimum prime radiator area is shown in Fig. 22. This temperature has a marked influence because of the fourth-power dependence of radiator area. For the cesium-lithium cycle the minimum area decreases from 3.0 ft²/kw at 1800°F to 2.1 ft²/kw at 2200°F. The area for the single-component (cesium) separator cycle decreases from 4.7 ft²/kw at 1800°F to 2.4 ft²/kw at 2200°F. The reduction in radiator area is less pronounced for the two-component system than for the single-component system due to the increased solubility of cesium in lithium which lowers the efficiency and results in an increase in the total heat rejected.

Minimum specific weight, from the simple formula discussed in SPS 37-34, is plotted versus maximum cycle
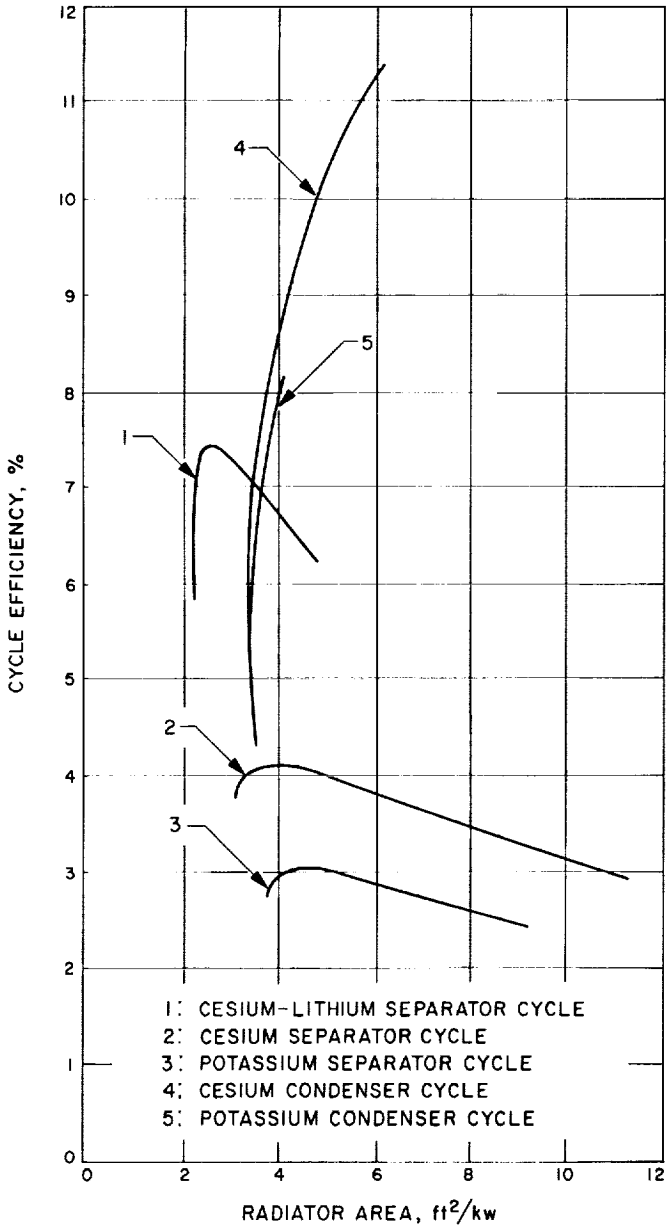
Fig. 21. Variation of cycle efficiency with specific prime radiator area for liquid MHD systems at 2000°F maximum cycle temperature

Fig. 22. Effect of maximum cycle temperature on minimum radiator area for liquid MHD cycles
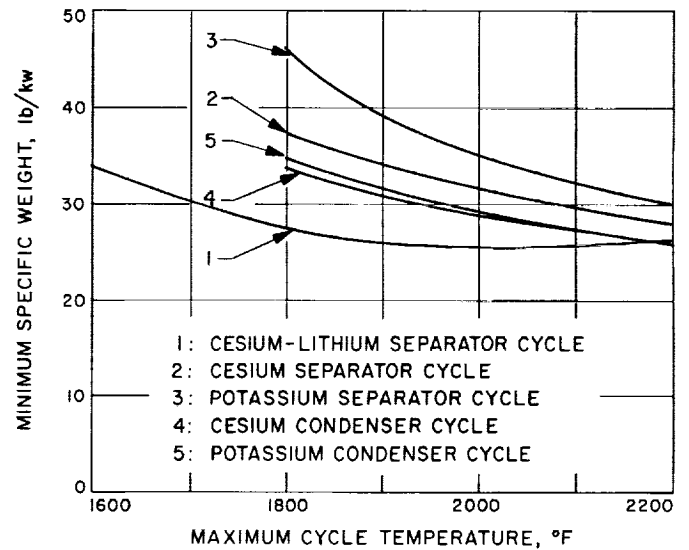
Fig. 23. Effect of maximum cycle temperature on specific weight of liquid MHD cycles

temperature in Fig. 23, showing the same trend as the minimum radiator area. The two-component separator system has the lowest weight at the temperatures considered.

The effect of maximum cycle temperature on the maximum efficiencies for the three separator cycles is shown in Fig. 24. The efficiencies of the single-component separator cycles increase with the maximum cycle tem-

perature, but the efficiency of the two-component cesium-lithium cycle eventually decreases with temperature due to the solubility effect.
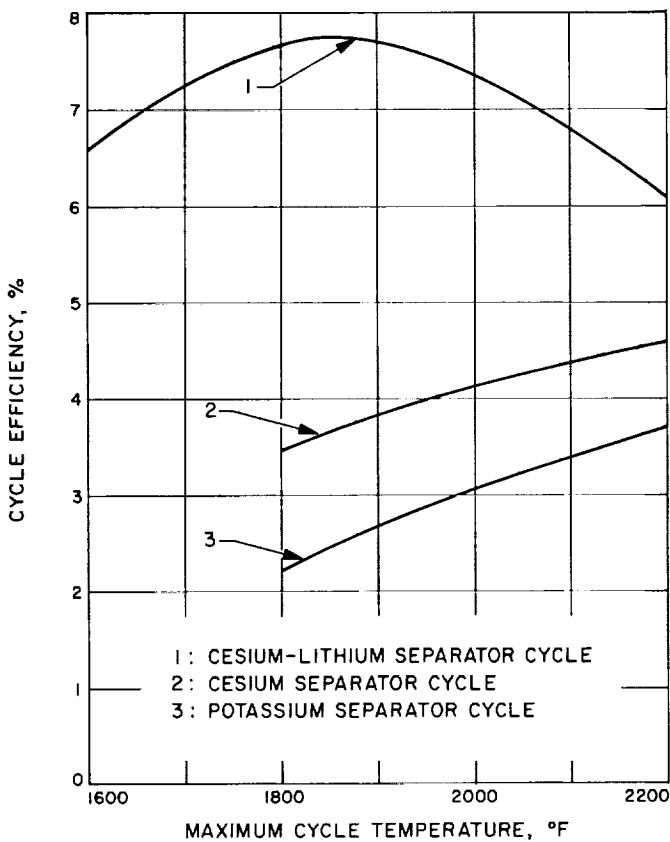
Fig. 24. Effect of maximum cycle temperature on efficiency of liquid MHD separator cycles

On the basis of these calculations, it appears that the cesium-lithium separator cycle is the best choice for space applications and that low area and weight are attainable even at 1800°F reactor temperature.

D. Plasma Investigation in a Reversed Current Electron Bombardment Ion Engine

T. D. Masek

1. Introduction

For several years, the electron bombardment ion engine has been studied as a possible source of primary propulsion for advanced spacecraft (Refs. 16–18). A laboratory program by the Application Studies Group has accompanied this effort to determine systems requirements and

to make basic studies of the plasma in the engine (Refs. 19 and 20; SPS 37-32, Vol. IV, p. 141 and SPS 37-33, Vol. IV, p. 150). The work reported here is an extension of previous plasma investigations of a different engine configuration.

The electron bombardment engine is normally operated with a cathode in the center of the engine chamber, and an annular anode. The plasma ion density distribution, which determines the ion beam distribution from the engine, has been found to be highly nonuniform with a maximum at the center and an order of magnitude drop across the chamber (Refs. 19 and 20). As indicated in SPS 37-32, this is thought to be due to the plasma formation mechanism and the radially increasing volume that the plasma must fill. Ionization takes place mainly near the cathode since electrons originating at the cathode rapidly lose energy in collisions with plasma electrons. The ions and electrons formed in ionization diffuse across the chamber, but the large volume that the plasma must fill requires the ion or electron density to decrease radially. The mercury plasma has approximately 10% "primary" electrons throughout which have not lost much of their initial energy but the major portion of the ionization still appears to take place near the cathode. A possible method of improving the ion density distribution, by reversing the position of the cathode and anode, is presented here.

A more uniform distribution seems possible with the reversed configuration when the following points are considered. If the ionization takes place near the cathode, ions and electrons will be diffusing into a decreasing volume. If no other processes occur, the ion density should not be significantly different than before. However, electrons and ions also diffuse axially so that many produced at the periphery will travel axially, rather than radially, tending to reduce the buildup at the center. This favorable condition does not arise in the normal configuration because many ions and electrons originating at the center and drifting axially are not available to diffuse radially.

The reason for not using the reversed configuration before is probably due to the feeling that cathode heating power would be excessive. A method for allowing low cathode temperature is described in the next section.

2. Reversed Current Engine

The terminology "reversed current" results from the reversed direction of the arc current flowing from cathode to anode.

FEED LINE

MAIN HOUSING

CATHODE
HEATER LEADS

PROPELLANT FEED

RADIATION SHIELD
0.002 in. SHIM STOCK
(5 LAYERS)

MOLYBDENUM
CATHODE

ANODE ADAPTER
(BORON NITRIDE)

SHADOW SHIELD

ANODE

SEAL, COPPER
(SQUARE)

PROPELLANT
FEED

ALUMINUM
GRIDS

GRID HOLDER

INSULATOR
(96% ALUMINA)

FEED LINE

COVER RING

CATHODE HEATER
(TANTALUM SHEATH,
TUNGSTEN—RHENIUM CONDUCTORS)

SHADOW SHIELD
(BORON NITRIDE)

Fig. 25. Reverse current electron bombardment ion engine

The engine used in this investigation (shown schematically in Fig. 25) presently uses cesium as a propellant, but the use of mercury seems feasible. The engine's primary elements—the cathode, anode, housing, and grid structure—will be described.

a. Cathode. The cathode, 4.0 in. in diameter and 1.8 in. wide, was fabricated from a 0.005-in. thick molybdenum sheet drilled with a series of 0.030 holes spaced approximately 0.25 in. in centers. The sheet was spot welded to stainless steel rings to form a box structure. Four ⅛-in. stainless tubes were welded into the rear ring to supply propellant through the cathode. The cathode heater consists of two separate windings of tantalum shielded tungsten-5% rhenium, tungsten-26% rhenium thermocouple wire. Similar wires from each winding were connected in parallel, providing two windings of two resistances since the 5% and 26% rhenium wires have different resistivities. Using equal numbers of turns, the cathode heater power can be varied without changing the magnetic field since the field changes with current, and resistance heating changes with current squared. A sheet of stainless steel shim stock was spot welded on the outside of the rings to form an annular cavity. Finally, five layers of dimpled 0.002-in.-thick shim stock were attached for thermal shielding.

Cesium is fed through the cathode to take advantage of its work-function reducing properties (Refs. 21–23). Fig. 26 is a plot of work function versus degree of cesium surface coverage (1.0 represents a monolayer 1-atom thick) and shows the large change of work function with coverage (Ref. 22). Electron emission from a surface is given by the Richardson–Dushman equation (Ref. 24)

$$j = AT^2 \exp\left(- \frac{e\phi}{kT}\right) \qquad (1)$$

where

$A = 120$ amp cm$^{-2} \cdot$ °K^{-2}

$T = $ surface temperature, °K

$e = $ electronic charge, coulombs

$\phi = $ work function, v

$k = $ Boltzmann constant, 1.38×10^{-16} j $\cdot$ °K^{-1}

Experimental data for electron emission from molybdenum is shown in Fig. 27 and Ref. 25. This plot, in principle, represents Eq. (1) calculated with ϕ a function of temperature. The straight lines represent constant work function. The numbers on each S-curve represent
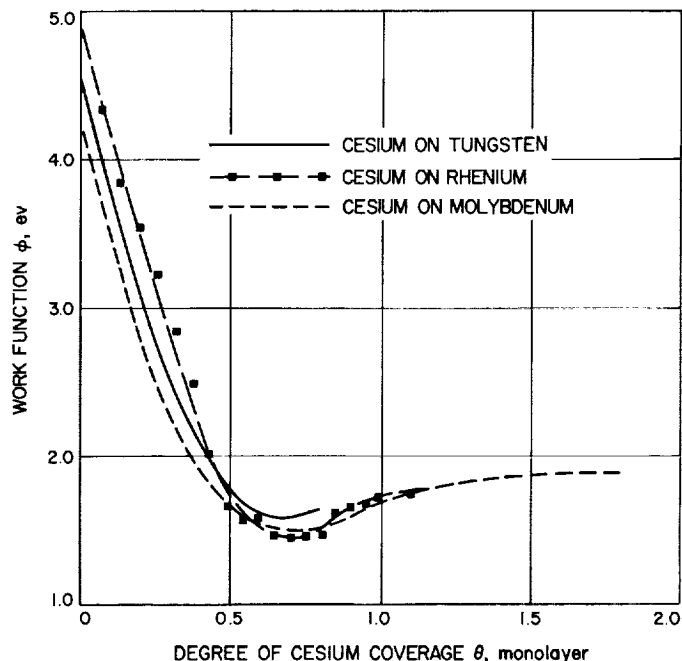
Fig. 26. Work function variation as a function of surface coverage of cesium

atom arrival rates at the surface, supplied from a cesium source at the temperature indicated. The curves indicate that the emission depends on the coating rate and surface temperature. The dotted curve is data from Ref. 26. It can be seen that the minimum work function reported varies from 1.5 in Fig. 26 to 1.65 in Fig. 27 (dotted curve). The dashed lines are extrapolations of the solid curves. To obtain 10.0 amp total from an area 150 cm², the cathode temperature must be approximately 875°K with a work function of 1.6 v. An atom arrival rate of about 3×10^{18} is also required. A calculation of the atom density near the cathode indicated about 10^{12} cm^{-3} would be present. Calculations also show that approximately 10^{16} atoms/sec/cm² strike the cathode. It therefore seems unlikely that 10^{18} atoms could be striking the cathode. Ref. 25 indicates that the electron emission S-curves can be shifted upward by surface contamination and field emission by two orders of magnitude and that the minimum work function can be reduced to less than 1.5. It seems reasonable to explain the measured emission by this curve shift.

A cathode temperature-power calibration was performed in a bell jar to find heating power requirements. Fig. 28 shows calibration data for thermally shielded and unshielded tests. This calibration is an upper limit because the cathode is actually more shielded in the engine, it is bombarded by ions from the plasma, and
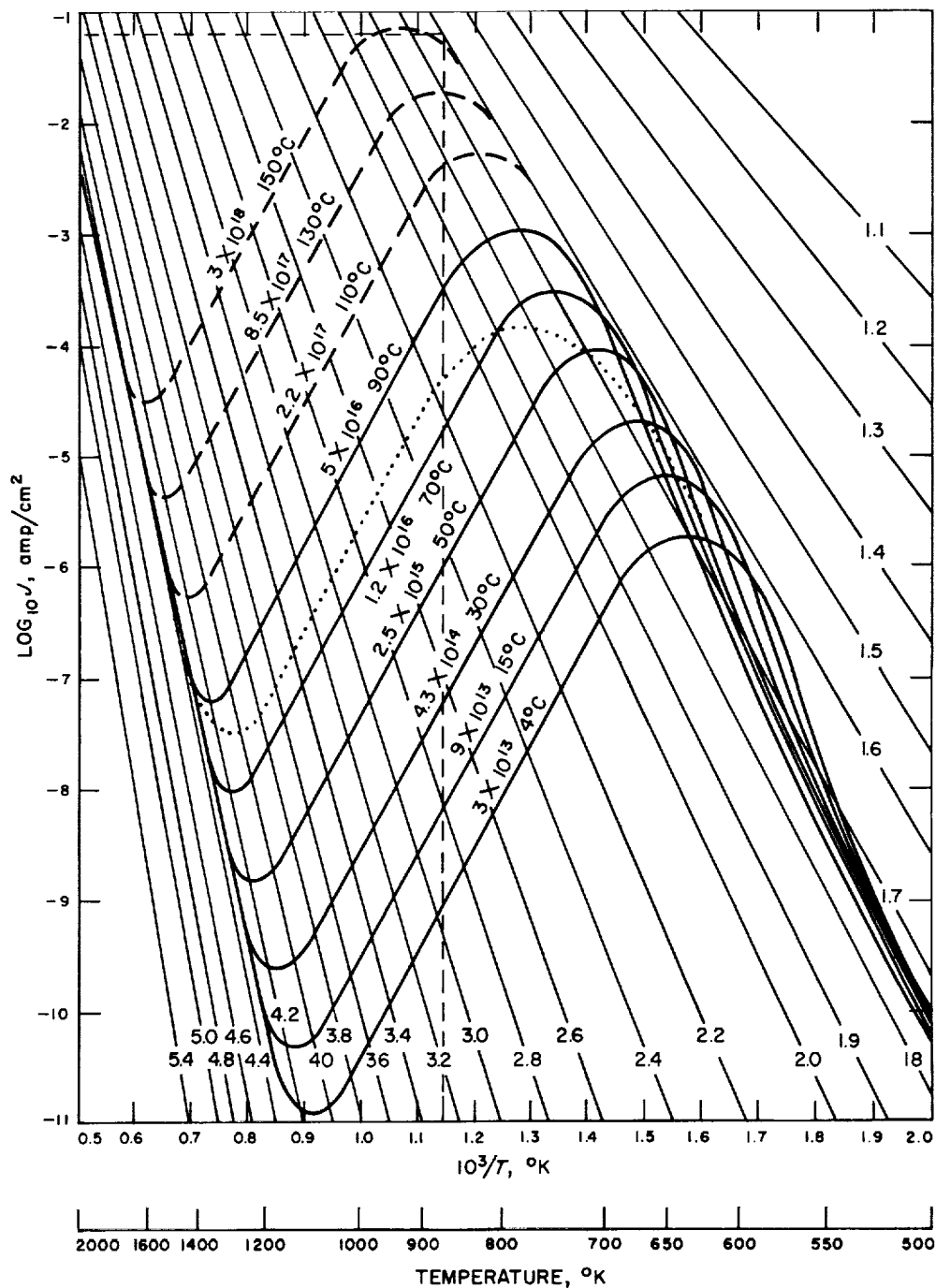
Fig. 27. Electron emission from cesiated molybdenum
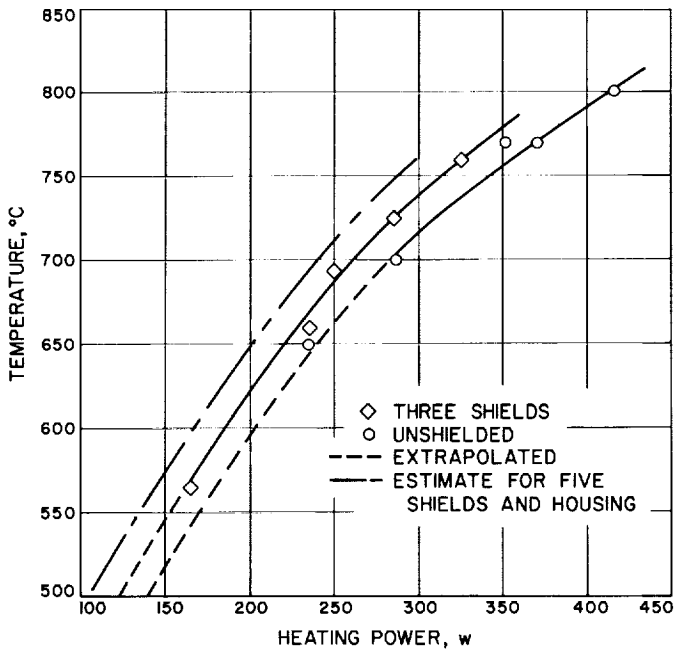
Fig. 28. Reverse current engine cathode calibration

Fig. 29. Comparative tests with various electrode materials

it is heated by radiation from the plasma. The large cathode area, low emission density, and low temperature combined with the relatively low arc voltage, which gives a low ion bombardment damage rate, should give this cathode a long life expectancy.

b. Anode. The anode was fabricated of molybdenum because of the possibility of high temperature operation due to heating by the arc current. It is 0.40 in. in diameter and approximately 2 in. long. A boron nitride support is used to electrically insulate the anode from the cathode and a stainless steel "shadow shield" is provided to reduce cesium coating on the boron nitride.

c. Engine housing. The housing, fabricated from stainless steel, has a removable back plate and an insulator support ring. The cathode and anode attach to the back plate and the grid support insulators are threaded into the support ring. The housing is supported by high voltage insulators to isolate the engine from the mounting structure.

d. Grid structure. The screen grid, attached to the housing, provides a boundary for the plasma while the accelerating grid, supported on insulators, accelerates and focuses ions from the engine. Both grids were made of 6061-T6 aluminum, based on previous experiments at JPL with cesium and work at Electro-Optical Systems (SPS 37-32 and Ref. 27). Fig. 29 (from Ref. 27) shows a plot of drain current from the accelerating grid and

represents relative electron emission rates from various grid materials. The electron current is a result of thermionic emission from the cesium coated grid surface which is heated by ion interception and radiation. The curves only represent relative current because the data also includes ion interception current.

The assembled engine is shown in Figs. 30–32. An overall view of the engine (Fig. 30) shows the grids and insulators. Fig. 31 is a side view showing the mounting structure, cesium boiler, and wiring. The internal portions of the engine may be seen in Fig. 32. In this view, the cathode, anode, feed lines, and rear shielding can be
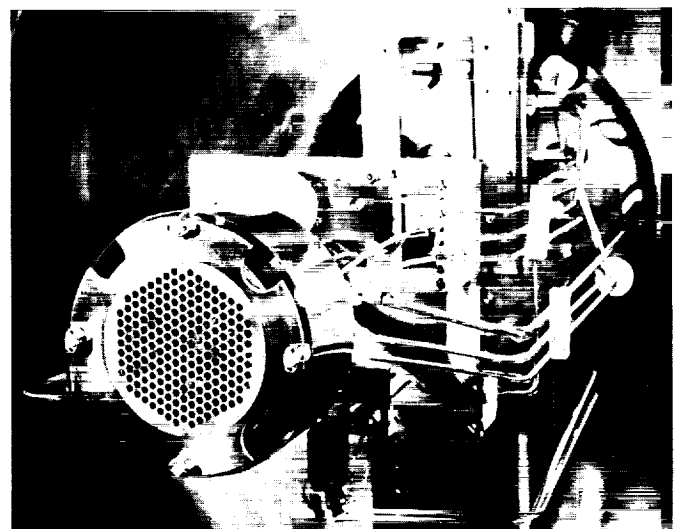
Fig. 30. Reverse current engine experimental setup, front view
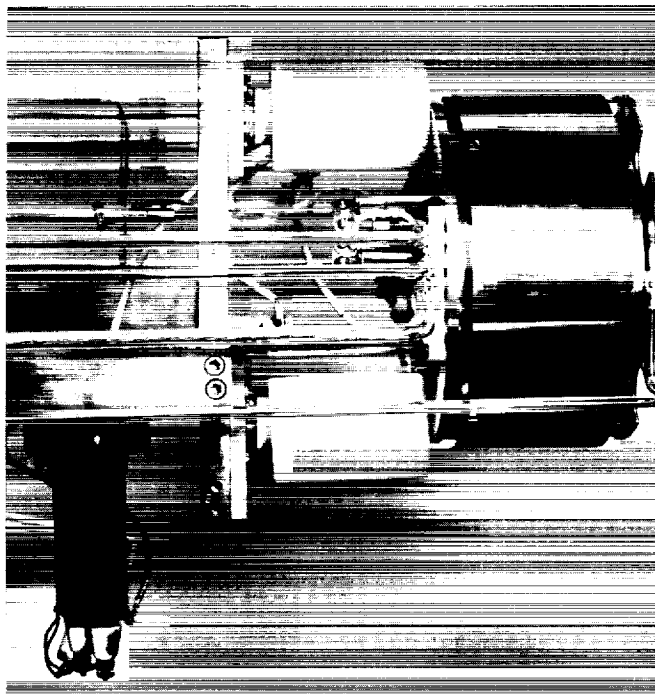
Fig. 31. Reverse current engine experimental setup,
side view

Fig. 32. Reverse current engine experimental setup,
arc chamber view

observed. The glass probe in Fig. 32 is a Langmuir probe, used in making plasma measurements.

Fig. 33 is an electrical wiring schematic that includes all wiring except that for the boiler temperature controller system and thermocouples. Included in the diagram is the circuitry for Langmuir probe measurements.

3. Engine Operation

The engine previously described operated efficiently in the initial tests which totaled approximately 35 hr. The cathode operated as expected but with less heating power than anticipated. The cathode heaters appeared to have successfully performed the double roles of heater and magnetic field coils.

The drain currents in the accelerating grid circuit were extremely high but after disassembling the engine it was found that the grid was incorrectly assembled. The spacing between the screen and accelerating grids was intended to be 0.075 in., but was actually approximately 0.137 in. This incorrect spacing led to high ion interception heating of the accelerating grid, causing electron emission to the screen grid which became extremely hot at one time and warped such that the spacing increased.

Table 2. Reverse current engine operating
conditions, Run 401

	1	2	3	4	5	6
Anode voltage, kv	2.60	3.2	3.2	3.0	3.2	3.1
Bias voltage, kv	1.65	2.0	2.0	1.5	2.0	2.05
Beam current, ma	155	170	192	158	195	100
Interception, ma	12	15	15.5	8.7	17	8.0
Arc current, amp	9.85	11.70	11.75	9.2	13.0	4.7
Arc voltage, v	8.5	8.7	9.4	10.4	8.1	14.0
Cathode A current, amp	1.15	1.0	0.3	1.0	0.9	1.0
Cathode A voltage, v	16.5	11.0	3.0	11.5	10.0	11.0
Cathode B current, amp	1.0	1.5	0.4	1.15	1.5	1.5
Cathode B voltage, v	7.0	10.5	3.0	8.0	10.5	10.0
Anode power, w	403	545	615	473	625	310
Bias power, w	20	30	31	13	34	16
Arc power, w	83	102	110	95	105	66
Cathode power, w	26	26	2	21	25	26
Total power, w	532	704	758	604	789	418
Specific impulse,[a] sec	6220	6900	6900	6680	6900	6800
Thrust, lb × 10⁻³	2.96	3.6	4.05	3.24	4.12	2.08
Power to thrust ratio, kw/lb	179	195	187	186	191	200
Power efficiency, %	75.6	77.5	81.2	78.8	79.4	74.1

[a] Uncorrected for propellant utilization.
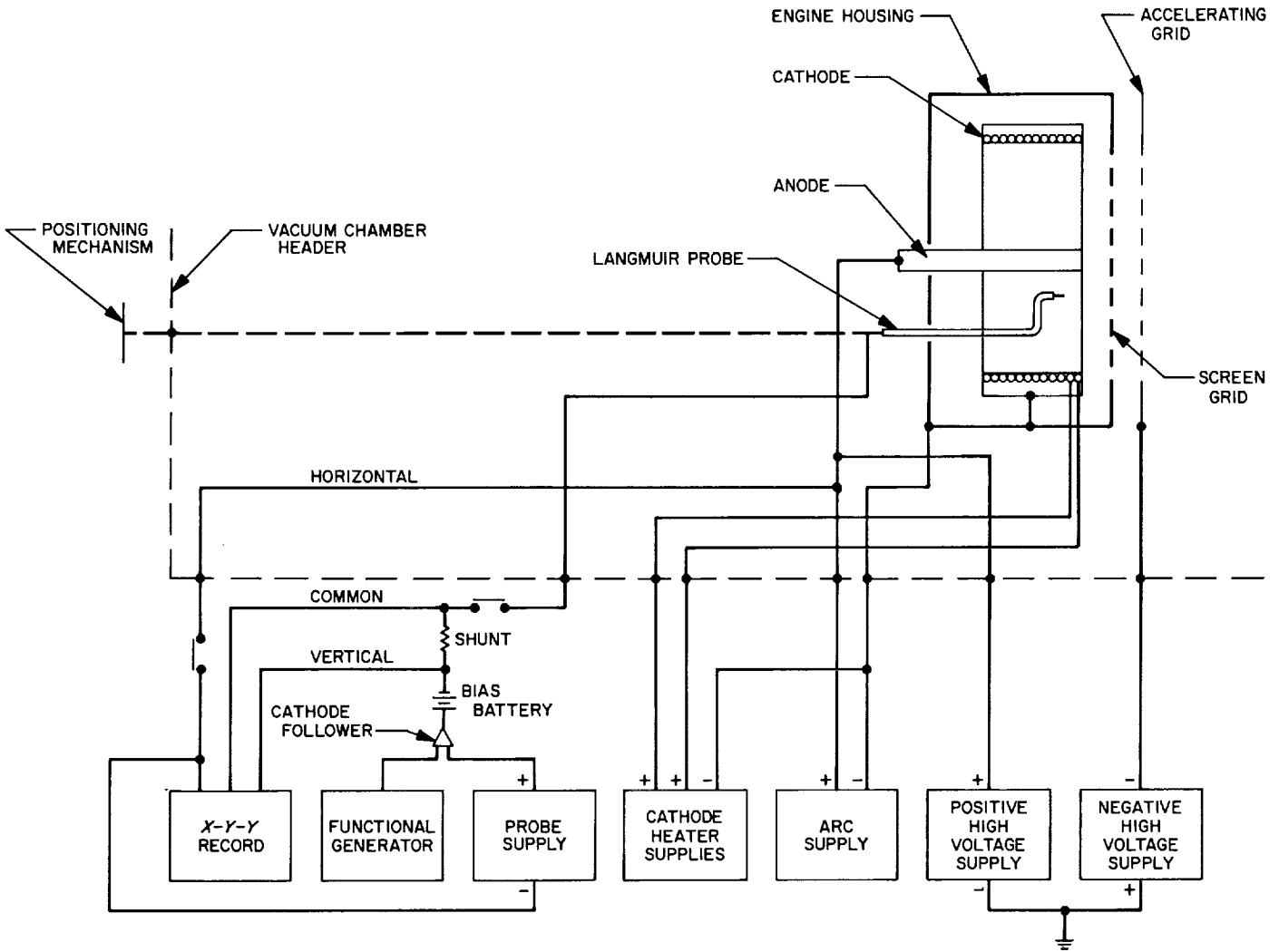
Fig. 33. Electrical wiring schematic diagram

The drain current during most of the test was 5 to 10% of the beam current. With proper grid spacing, this problem should be rectified in the next tests.

Engine performance for the initial tests is shown in Table 2. The anode and bias (accelerating grid) voltages shown are higher than would normally be necessary because of the large grid spacing. (The current drawn from the engine is inversely proportional to the spacing, squared.) The first three columns of data are the engine operating conditions for probe measurements presented in the next section.

Thermocouple measurements were made on the engine and feed system during operation, using a unique isolation system described in Ref. 28. A schematic of the sys-

tem is shown in Fig. 34 and a typical set of operating temperatures is given in Table 3.

Table 3. Typical reverse current engine temperature measurements

Location	Temperature, °F
Engine housing	500
Engine rear plate	575
Propellant feed line	575
Boiler outlet flange	550
Boiler temperature control	550–600

4. Plasma Measurements

Probe measurements of the plasma were made with a movable Langmuir probe similar to the one in Fig. 35.
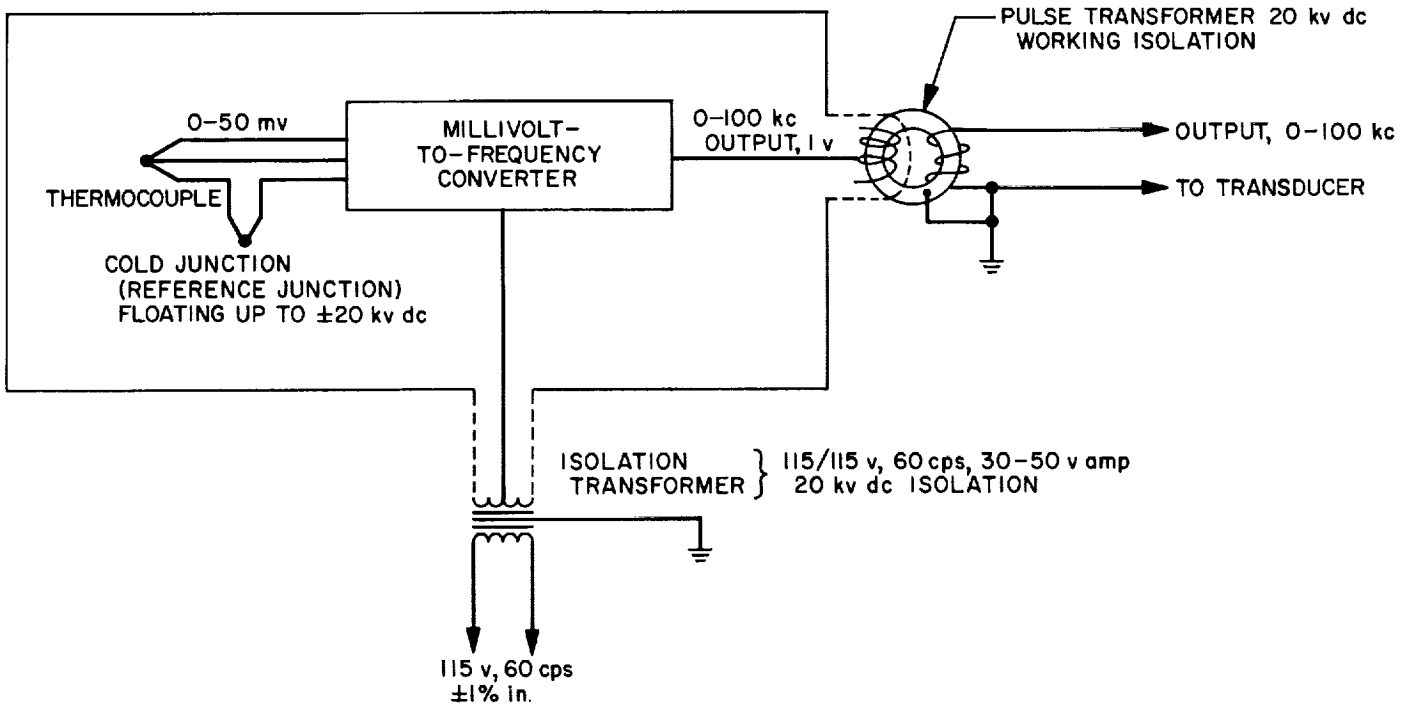
Fig. 34. High voltage thermocouple isolation system
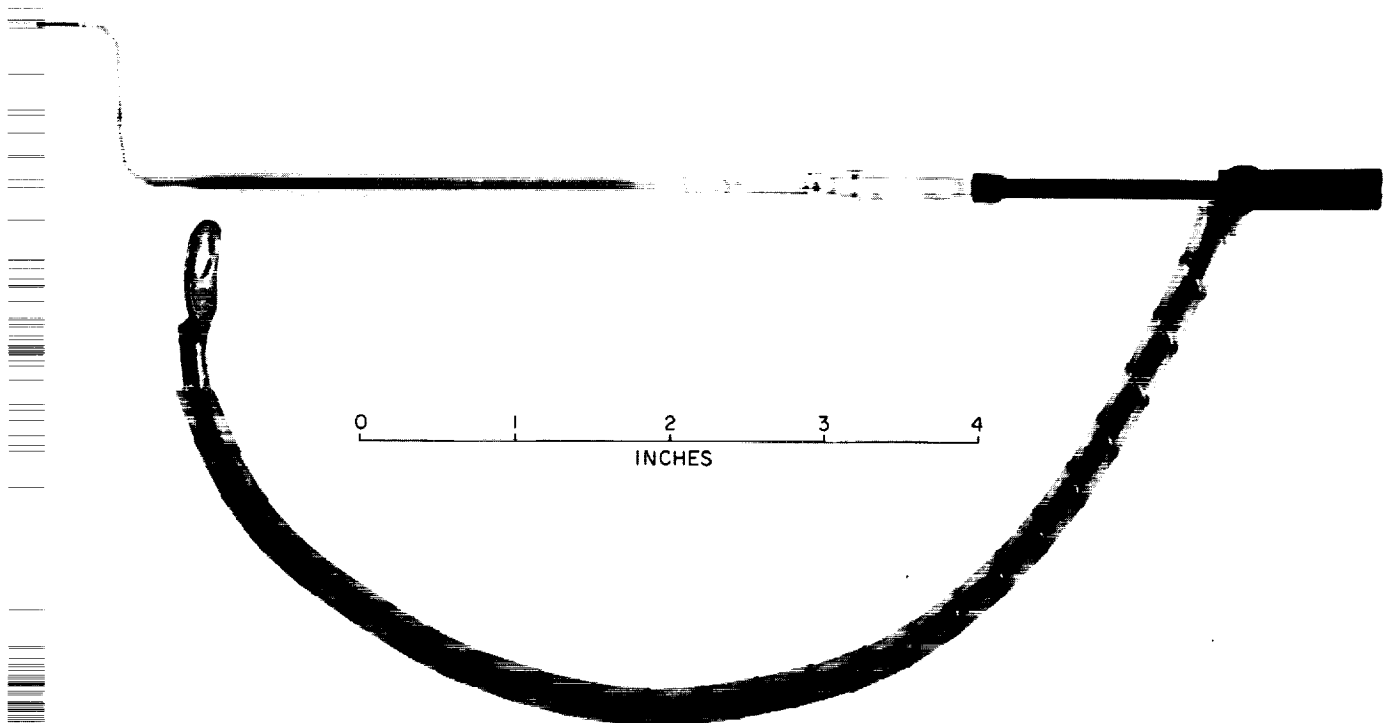
Fig. 35. Typical Langmuir probe

The theory of the Langmuir probe has been outlined in other work and will not be presented here. Fig. 36 is a schematic diagram showing the possible positions of the probe.

Normalized plots of ion density for the entire engine with three different operating conditions are shown in Fig. 37 (a–c). As previously indicated, the operating conditions for these measurements are given in Table 2.
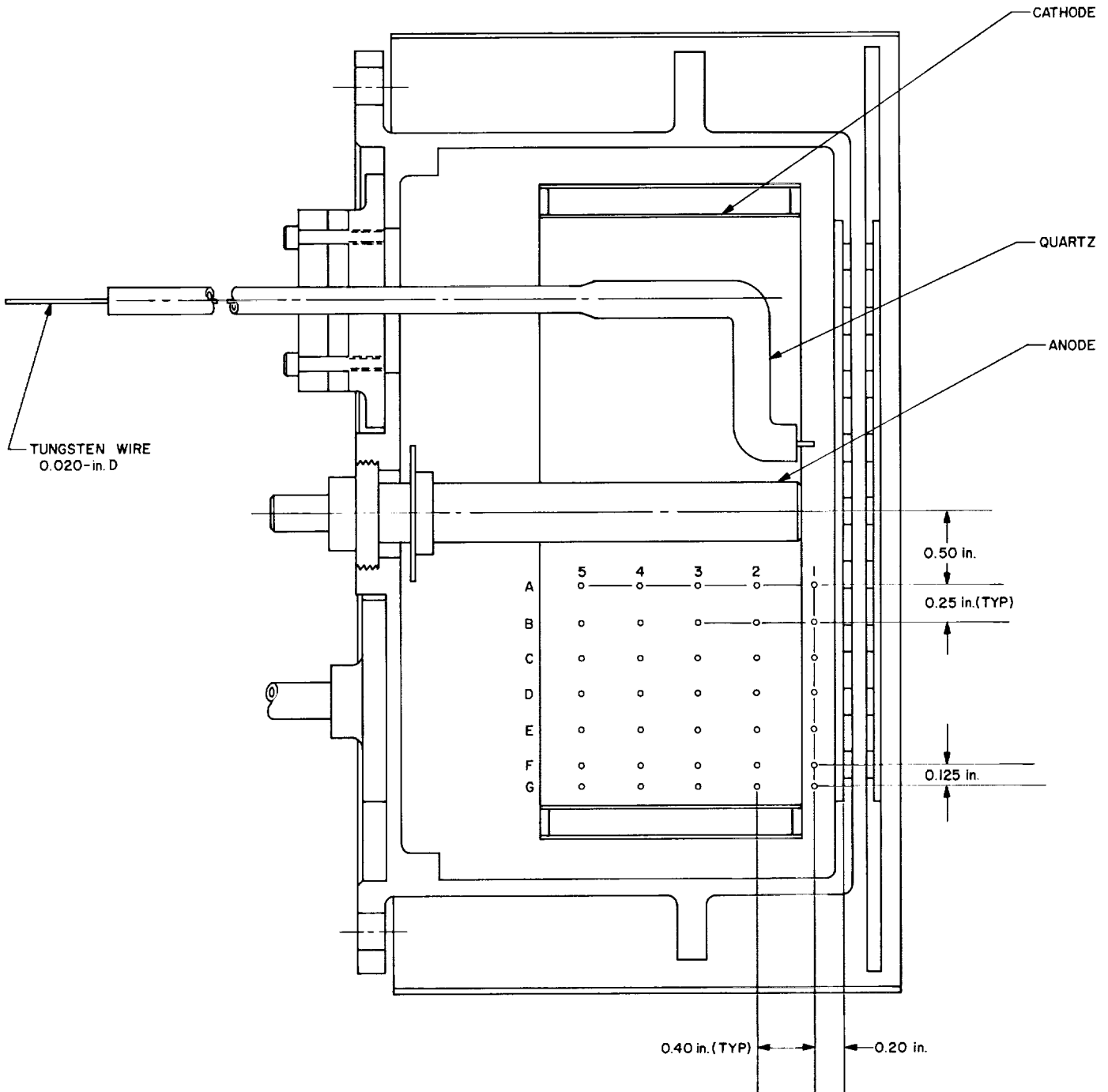
Fig. 36. Langmuir probe positions
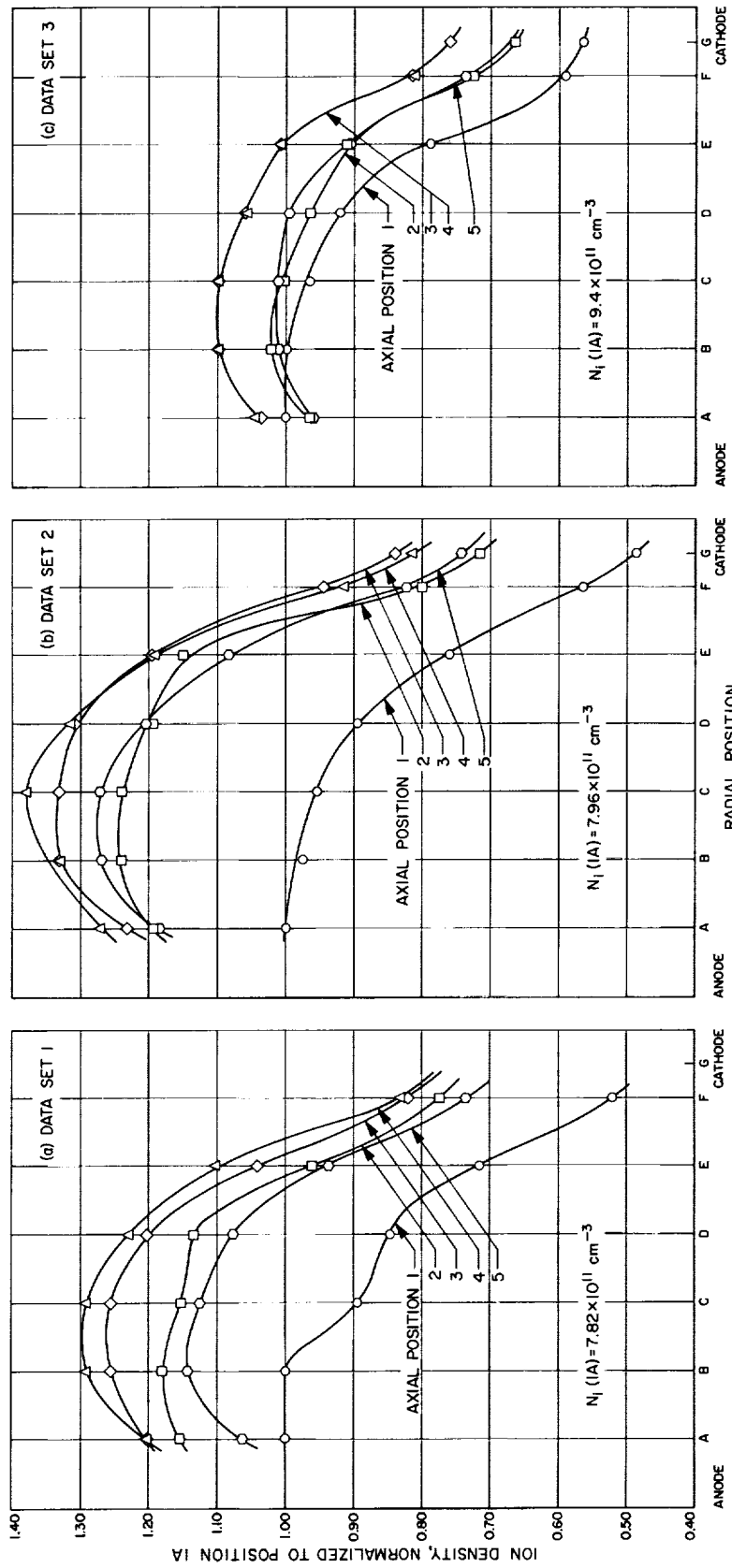
Fig. 37. Reverse current engine ion density distribution

Typical normalized ion densities measured in conventional mercury and cesium engines (Ref. 20 and *SPS 37-32*) are compared with these in Fig. 38 for the grid position. It can be seen that the ion density in the reverse current engine is much more uniform than in the other two engines. The electron energies and plasma potentials for the three operating points are plotted in Figs. 39 and 40. Typical electron temperatures and plasma potentials are given in *SPS 37-32* for the conventional engines. The electron energies are more uniform than in other engines tested, probably because of the higher plasma density in the present tests, allowing electrons to equilibrate energies quickly. The plasma potentials, with respect to the anode

potential, are remarkably well represented by smooth curves. In general, the single Langmuir probe is not accurate for this measurement so these data should be used with great care.

5. Conclusions

The primary conclusion to be drawn from the previous discussion is that the reversed cathode-anode configuration improved the ion density uniformity by approximately a factor of five (ratio of the ion density at the cathode to that at the anode) over conventional engines. It is likely that a larger engine of this type would show additional uniformity improvement.

In addition, these tests indicate that the cathode power requirements are not too severe (less than 5% of the total power) and may be further reduced at higher plasma densities. It was also shown that the cathode could be fabricated simply and relatively inexpensively.

Additional plans for this engine include zero gravity feed system tests and Faraday probe measurements in the beam. A separate magnetic field coil will also be added to attempt to optimize the magnetic field in the plasma.

Fig. 38. Ion density distribution comparison for three engines

E. Static Pressure Drop Along the Wall of a Constant-Diameter Duct Which Contains a Decaying Swirling Flow of Argon

P. F. Massier

1. Introduction

Most of the information that exists in the literature which pertains to skin friction and pressure drop for swirling flows applies to fluids flowing through constant-diameter tubes that contain swirl generators such as twisted tapes (Refs. 29–33). In general, investigators of these flows have been interested in the increased heat-transfer rate attainable when swirl is introduced into the flow and the pressure head required to maintain the swirl. The amount of swirl introduced by twisted tapes has been such that the tangential component of velocity has ranged from values approximately equal to the axial

Fig. 39. Reverse current engine electron energy
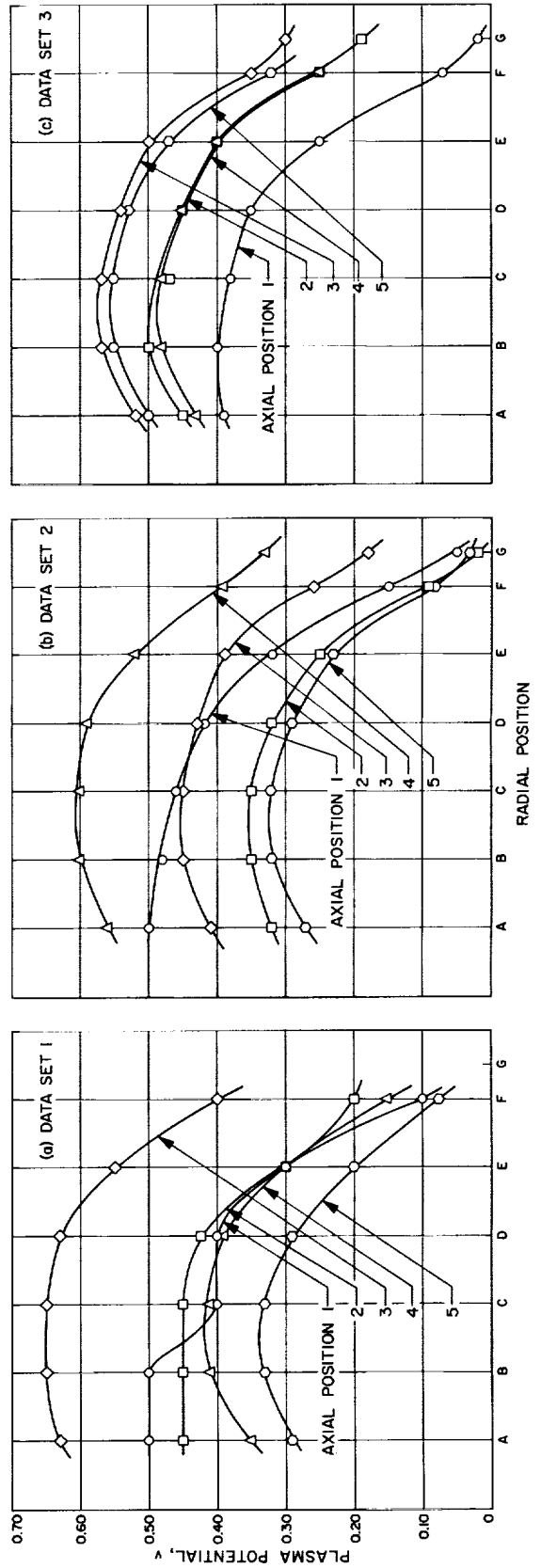
Fig. 40. Reverse current engine plasma potential

component of velocity, downward to values much less than the axial velocity.

This report pertains to swirling flows that differ from the twisted tape studies in several respects. First, the magnitude of the tangential velocity near the cylindrical wall was much greater than the magnitude of the axial velocity. Second, the pipe did not contain swirl generators; instead, the fluid was introduced tangentially at one axial location through four holes equally spaced circumferentially and the swirl was allowed to decay as the fluid flowed along the pipe. Since there were no swirl generators, the tangential velocity distribution outside the boundary layer approximated that of solid body rotation in the vicinity of the centerline and of a potential vortex between the solid body rotation and boundary layer regions.

The fluid that was used to obtain the experimental data was argon at approximately room temperature; hence, the flow was externally adiabatic. The stagnation pressures ranged from 1.0 to 40 psia and the ratio of tangential velocity near the wall to axial velocity ranged from approximately 6.5 to 10.4. Turbulence measurements were not made; however, the magnitudes of the Reynolds numbers indicate that the flow outside the boundary layer

was probably turbulent. Swirling flows of this type may be encountered in the inlet and mixing-chamber sections of electric arc jets although, in regions of these devices where the gas is at a high temperature, the flow would generally not be externally adiabatic and it may be laminar.

The test apparatus is shown in Fig. 41 and the analysis used to determine the flow field outside the boundary layer from the measurements is given in SPS 37-33, Vol. IV, pp. 133–141. Additional results of the flow field are presented in SPS 37-34, Vol. IV, pp. 149–157.

Other investigations of decaying swirling flows have been made. Nissan and Bresan (Ref. 34) obtained experimental values of pressure drop as well as velocity distribution along vertically mounted constant-diameter tubes in which water was injected tangentially at the bottom. By use of the tangential momentum equation, Talbot (Ref. 35) derived an approximate solution for the tangential velocity as a function of the radius and the axial coordinate for the decay of an incompressible laminar flow with swirl. Kreith and Sonju (Ref. 36) obtained a series solution of the tangential momentum equation for the radial distribution of the tangential velocity along a decaying swirling flow that is turbulent.
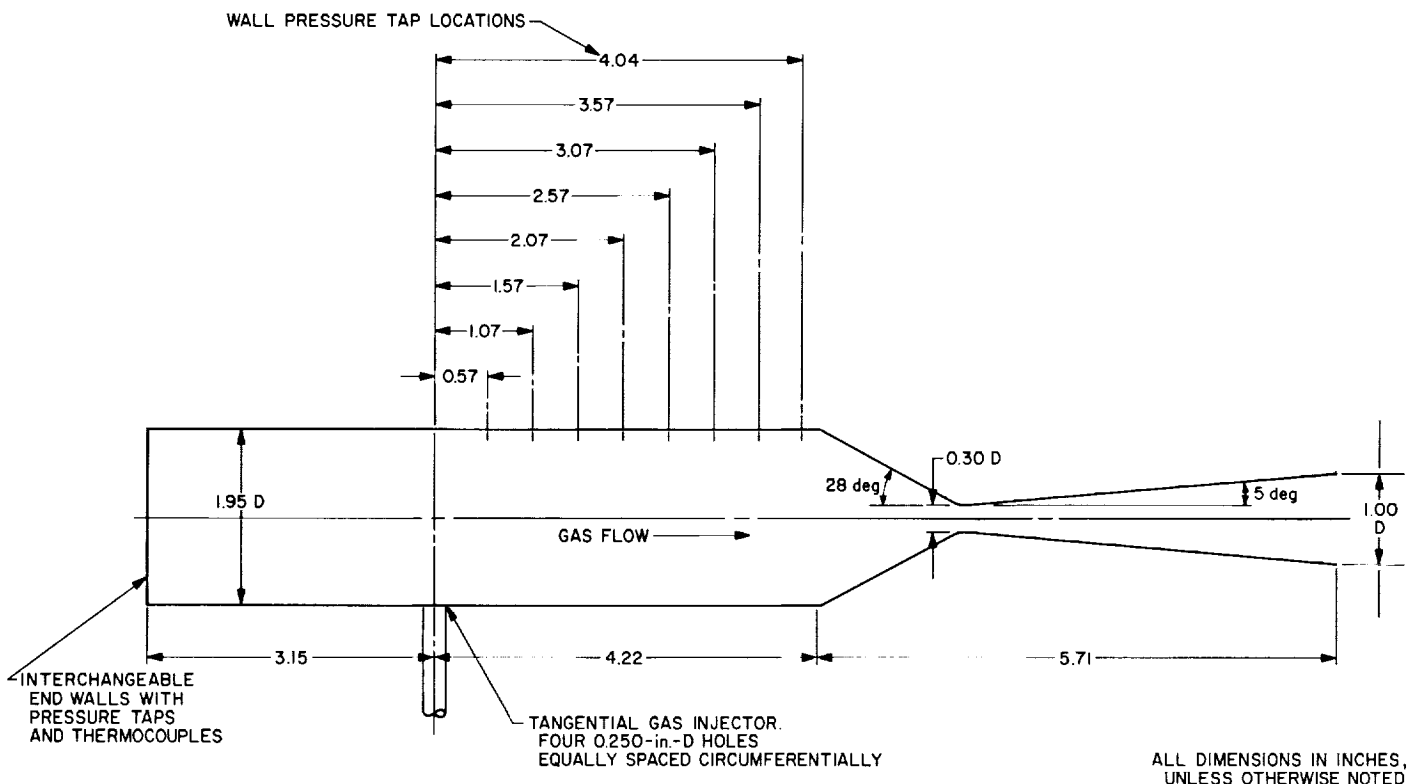
Fig. 41. Swirl flow test apparatus

To obtain numerical results, the initial velocity distribution must be known, as well as knowledge of the eddy diffusivity. The present discussion pertains to pressure drop resulting from the turbulent swirling flow of a compressible fluid and consequently differs from the studies of Refs. 34–36.

2. Data Analysis

A definition of the friction factor for fully developed nonswirling flow in a constant-diameter duct, as given in Ref. 37, is:

$$\Delta p = f\left(\frac{l}{D}\right)\frac{\rho w^2}{2g}. \qquad (1)$$

The quantity l represents the length of the duct which, for nonswirling flow, is the distance travelled by the fluid as it flows through the duct. For swirling flow, however, the distance travelled by the fluid is greater than the duct length. Therefore, to determine a friction factor for swirling flow that would be consistent with Eq. (1) the velocity should be a mean resultant fluid velocity in the region in which the velocity distribution is a consequence of the wall shear. Also, the length should be the actual distance travelled along the direction of this mean resultant velocity. Furthermore, the pressure drop must be that brought about only as a consequence of the wall shear and not by the fluid acceleration or by the fluid shear that may result from the existence of a tangential velocity distribution resembling that of a potential vortex.

In the present investigation, the flow was not fully developed because the influence of the wall shear had not penetrated to the centerline of the duct. For two of the tests performed, the radial distribution of the tangential velocity component in the constant-diameter duct upstream of the nozzle is shown in Fig. 42. These velocities were determined, as discussed in SPS 37-33 and -34, from temperature and static pressure measurements along the end wall shown in Fig. 41. The axial component of the velocity obtained from the measured mass flow rate was assumed to be uniform. The radial distributions of the swirl angles are shown in Fig. 43. The swirl angle β is defined as the

Fig. 42. Radial distributions of tangential and axial components of velocity upstream of the nozzle
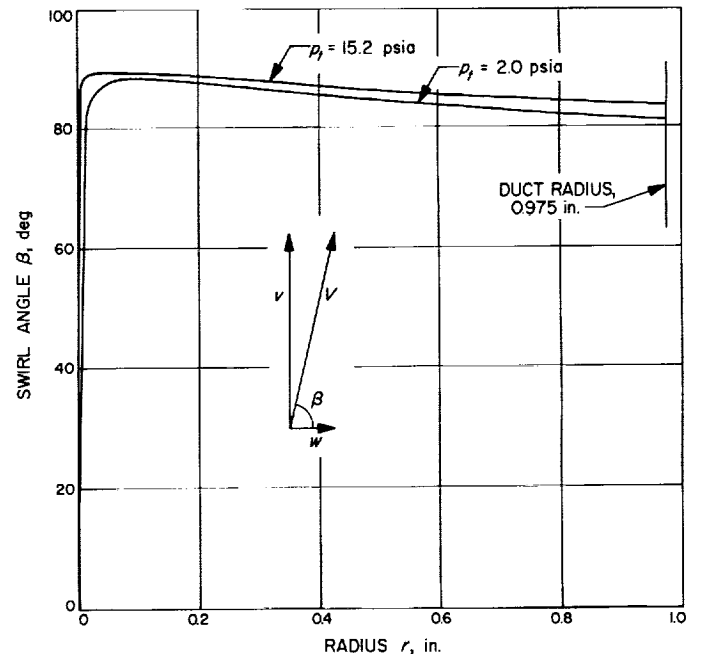
Fig. 43. Radial distribution of swirl angle upstream of the nozzle

angle between the resultant velocity and the axial component of velocity as shown in this figure. The velocities shown in Fig. 42 are located in the main stream outside both the end wall and the cylindrical wall boundary layers. These velocity distributions resulted from the radially inward flow of tangential momentum introduced at the injector, and from the associated shear stresses in the main stream that retarded the tangential motion of the fluid. The shear stresses in the main stream prevented the fluid from acquiring a potential vortex distribution all the way to the centerline. Presumably, however, the shear stresses at the wall as the fluid progressed along the duct had no influence on these velocity distributions in the main stream once the tangential velocity distributions were established. The shear stresses in the main stream account for a portion of the static pressure drop that occurs along a duct which contains a swirling flow of this type. Hence, the values of pressure drop obtained in these experiments are a consequence of at least three effects: (1) the shear stresses in the boundary layer adjacent to the cylindrical wall, (2) the shear stresses in the main stream outside the boundary layer and, (3) the change in velocity of the fluid in the main stream that results from a growing boundary layer. The configuration of the nozzle attached to the exit end of the duct also influenced the flow field in the duct and may in turn have effected the pressure drop.

In these experiments, insufficient instrumentation was provided to isolate the effect of each influencing factor on the combined pressure drop; consequently, results of only the combined pressure drop are shown. Eq. (1) is modified by substituting the resultant velocity V_w at the main stream edge of the boundary layer for the axial velocity and by substituting L, the distance travelled by particles moving at V_w, in place of the duct length. The resulting definition of the pressure-drop parameter Φ is:

$$\Phi \equiv \frac{\Delta p}{\left(\dfrac{L}{D}\right)\dfrac{\rho V_w^2}{2g}} \cdot \qquad (2)$$

The distance travelled by the swirling fluid at velocity V_w can be expressed in terms of the corresponding swirl angle β_w and the axial distance of travel.

If the boundary layer is thin, the distance travelled in one revolution at velocity V_w is

$$S = \frac{\pi D}{\sin \beta_w} \cdot \qquad (3)$$

Hence the axial distance travelled in one revolution is

$$z = S \cos \beta_w = \frac{\pi D}{\tan \beta_w} \cdot \qquad (4)$$

Consequently, if l, which is the length of the duct downstream of the injector, is the total axial distance travelled, and n is the number of revolutions then the total distance travelled is

$$L = nS = S \frac{l}{z}$$

or

$$L = \frac{l}{\cos \beta_w} \cdot \qquad (5)$$

Eqs. (2) and (5) may be combined to form the following equation which can be used to obtain experimental values of the pressure-drop parameter Φ:

$$\Phi = \frac{\Delta p}{\dfrac{l}{D \cos \beta_w} \dfrac{\rho V_w^2}{2g}} \qquad (6)$$

In general, V_w is a function of L since the swirl decays; but in this instance since the maximum axial length between the farthest upstream and downstream cylindrical wall pressure taps was only about two duct diameters, V_w was taken to be a constant as determined by the end-wall pressure taps.

3. Results

The resultant velocity V_w, and the swirl angle β_w, at the main-stream edge of the cylindrical-wall boundary layer are shown in Figs. 44 and 45, respectively, as a function of stagnation pressure. These values were introduced into Eq. (6) to evaluate the pressure-drop parameter. Reynolds numbers based on this resultant velocity and the duct diameter are shown in Fig. 46; the Reynolds numbers were not used but are merely shown for reference.

Static pressures were measured at eight locations between the injector and nozzle inlet as shown in Fig. 41; hence, it was possible to obtain seven values of the pressure-drop parameter along the duct for each test condition. Since for each test condition the boundary layer was developing along the duct, the physical dimension selected for the Reynolds number to establish its relationship with the pressure-drop parameter was the length of travel of the fluid instead of the duct diameter.
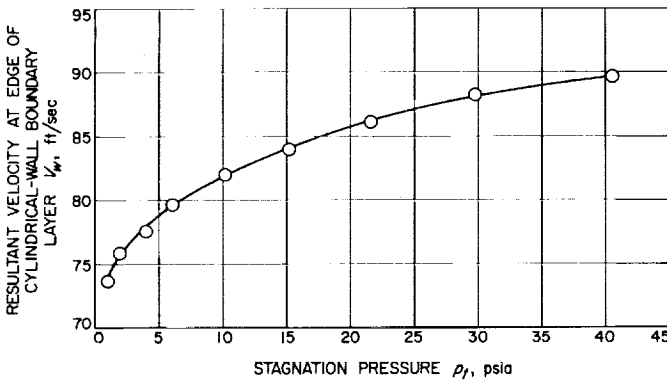
Fig. 44. Resultant velocity at main-stream edge of the boundary layer versus stagnation pressure
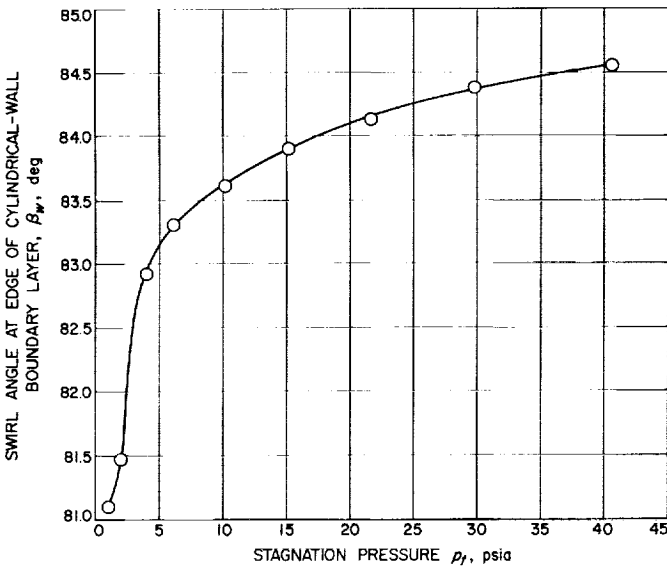
Fig. 45. Swirl angle at main-stream edge of the boundary layer versus stagnation pressure

Fig. 46. Reynolds number based on diameter versus stagnation pressure

the average value between the two upstream pressure taps. This value of Φ is also the highest value attained at the given pressure. Successively larger values of the Reynolds number and, except in a few instances, lower values of Φ occur at adjacent downstream pressure taps. All values of Φ shown represent averages between the farthest upstream pressure tap and the pressure tap at which the Reynolds number was evaluated. It is apparent that the trends of the data are similar for nearly all tests. The deviations at the lower pressures are attributed to scatter rather than conclusive evidence of decreasing and increasing values of Φ along the duct. The pressures could be read within about 0.5 mm oil since manometers were used. At the lowest stagnation pressure of 1 psia the minimum Δp was 1.5 mm oil, hence the error for the data point at the lowest Reynolds number could be quite large. At the highest pressure of 40 psia, however, the largest Δp was 68.5 mm oil, hence the accuracy at the largest Reynolds number, based on reading error, is very good. A probable reason why the data do not overlap at different stagnation pressures is that the influence of the wall shear had not completely penetrated the flow field.

This length of travel consisted of the distance from the injector centerline to the pressure tap centerline. Experimental results of the pressure-drop parameter versus the Reynolds number are shown in Fig. 47. For each test condition (stagnation pressure, p_t) the data points represent the evaluation of Φ by Eq. (6) using the wall pressure difference between the first pressure tap downstream of the injector and each of the successive pressure taps, as well as the corresponding axial spacing for l. Other variables in Eq. (6) were constant for each test condition, but changed as the stagnation pressure was changed.

For each set of data at any given stagnation pressure the value of Φ at the lowest Reynolds number represents

These results were obtained as a consequence of a current investigation of heat transfer from ionized gases using the same components of the test apparatus. The cold flow tests were made to aid in the understanding of the fluid dynamics of swirling flows which are sometimes used to attain good performance of electric arc jets. These swirling flows have a significant influence on the convective heat transfer. The pressure-drop results represent only a small portion of the information obtained from these tests; hence, measurements of sufficient detail to establish

TEST No.	p_f, psia	T_f, °F
112-19C	1.0	73
112-20C	2.0	76
112-21C	4.0	70
112-22C	6.1	70
112-23C	10.2	68
112-24C	15.2	69
112-25C	21.6	59
112-26C	29.8	59
112-27C	40.6	56

PRESSURE-DROP PARAMETER, Φ

REYNOLDS NUMBER BASED ON TOTAL DISTANCE TRAVELED, $(\rho V_w L)/(\mu)$
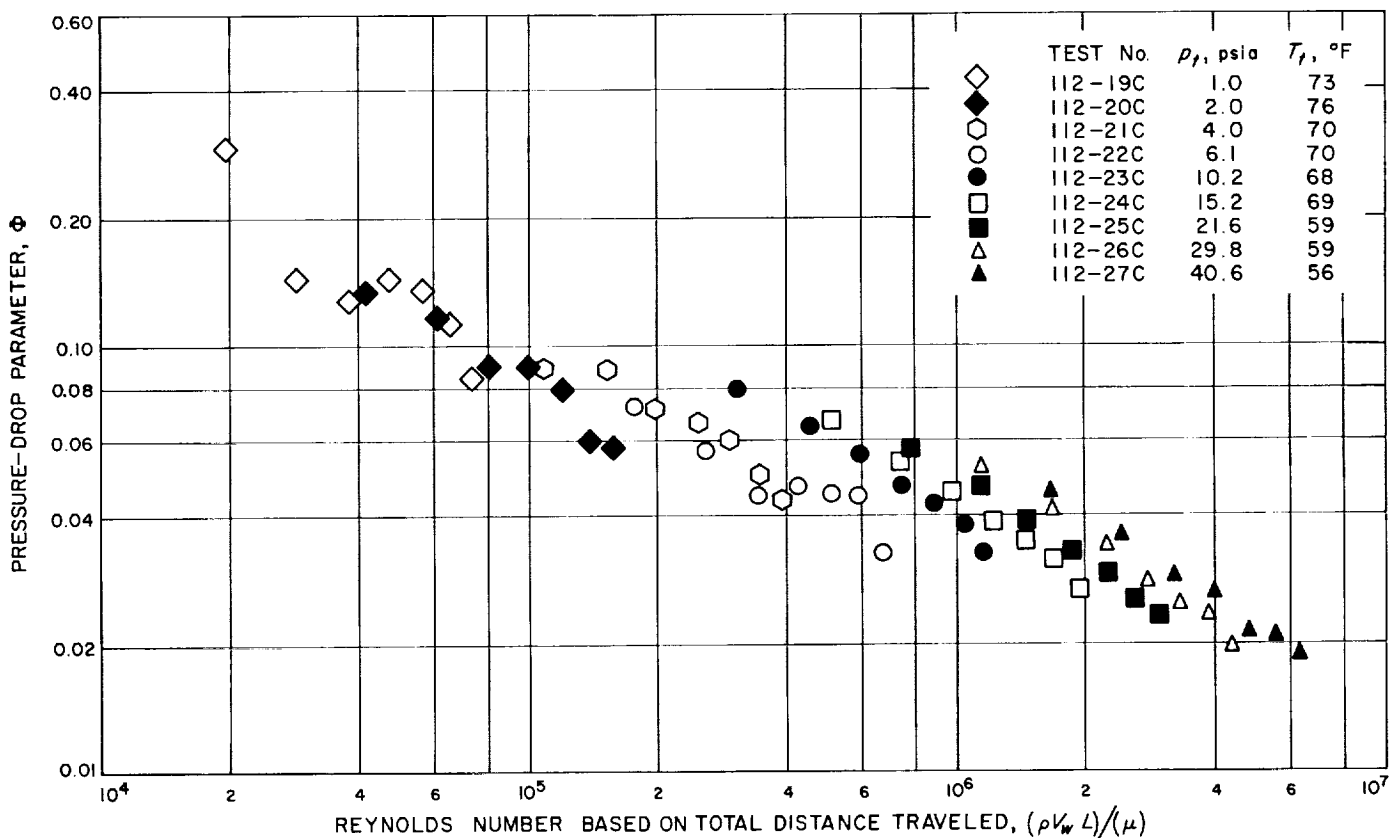
Fig. 47. Static pressure-drop parameter versus Reynolds number based on length of fluid travel

the skin friction in the mixing chamber were not warranted since this would require a major effort in itself. It would be useful to at least compare the relationship between the pressure-drop parameter for flows with and without swirl; however, this has not been done.

4. Summary and Conclusions

Experimental results were obtained for a pressure-drop parameter Φ, as a function of Reynolds number (based on distance of fluid travel) for the decay of a swirling compressible fluid in a constant-diameter duct. This parameter was defined in terms of the distance of fluid travel, the resultant velocity, and the swirl angle, all of which were evaluated at the main-stream edge of the cylindrical-wall boundary layer. The swirl angle and resultant velocity were evaluated by measurements along the end wall and were taken to be constant at a given stagnation pressure, even though the swirl was decaying as a consequence of the shear stresses. This was considered to be justifiable because the duct was comparatively short. The swirling fluid investigated was presumably not fully developed because the influence of the wall shear was

believed not to have penetrated to the centerline of the duct. It was found that within the Reynolds number range of the tests, which was between 2×10^4 and 6×10^6, the relationship between Φ and Re_L followed similar trends along the duct for each value of stagnation pressure (mass flow rate). At the lower Reynolds numbers, the pressure-drop parameter was higher.

It was concluded that more measurements than the wall static pressure were needed to evaluate the friction factor because the pressure drop was a consequence of at least three effects: (1) the skin friction along the cylindrical wall, (2) the shear stresses within the main stream of the fluid resulting from the velocity distribution of the vortex and, (3) the change in velocity of the fluid in the main stream resulting from a developing boundary layer. An additional effect was probably the influence of the configuration of the nozzle which was attached to the exit end of the duct.

The symbols used in this Section are defined in the table on p. 124.

Definition of symbols

D	diameter	v	tangential velocity
f	friction factor	V	resultant velocity
g	gravitational constant	w	axial velocity
l	total axial distance travelled by fluid particles	z	axial distance that a fluid particle travels in one revolution
L	total distance travelled by fluid particles		
n	number of revolutions	β	swirl angle
p	static pressure	ρ	static density
r	radius	Φ	pressure-drop parameter defined by Eq. (2)
Re_D	Reynolds number based on diameter	*Subscripts*	
Re_L	Reynolds number based on distance of fluid travel	t	stagnation conditions
		w	conditions at main-stream edge of cylindrical-wall boundary layer
S	axial distance travelled by fluid particles		

References

1. Sparrow, E. M., "Radiant Emission, Absorption and Transmission Characteristics of Cavities and Passages," Symposium on Thermal Radiation of Solids, San Francisco, California, March 1964. Proceedings published as NASA SP-55, pp. 103–115, edited by S. Katzoff, 1965.

2. Sparrow, E. M., and Jonsson, V. K., "Absorption and Emission Characteristics of Diffuse Spherical Enclosures," Transactions of ASME, Series C, *Journal of Heat Transfer*, Vol. 84, No. 2, pp. 188–189, May 1962.

3. Williams, C. S., "Discussion of the Theories of Cavity-Type Sources of Radiant Energy," *Journal of the Optical Society of America*, Vol. 51, No. 5, pp. 564–571, May 1961.

4. Treuenfels, E. W., "Emissivity of Isothermal Cavities," *Journal of the Optical Society of America*, Vol. 53, No. 10, pp. 1162–1171, October 1963.

5. Gubareff, G. G., Janssen, J. E., and Torborg, R. H., "Thermal Radiation Properties Survey," 2nd Edition, Honeywell Publication, Honeywell Research Center, Minneapolis, 1960.

6. Bevans, J. T., Gier, J. T., and Dunkle, R. V., "Comparison of Total Emittances with Values Computed from Spectral Measurements," Transactions of the ASME, Vol. 80, No. 11, pp. 1405–1414, (also discussion, pp. 1415–1416), October 1958.

7. Ohlsen, P. E., and Etemad, G. A., "Spectral and Total Radiation Data of Various Aircraft Materials," North American Aviation Report No. NA57-330, Engineering Department, July 1957.

References (Cont'd)

8. Sparrow, E. M., "Radiant Emission Characteristics of Nonisothermal Cylindrical Cavities," *Applied Optics*, Vol. 4, No. 1, pp. 41–43, January 1965.

9. Kelly, F. J., "On Kirchhoff's Law and Its Generalized Application to Absorption and Emission of Cavities," AIAA Paper No. 65-135, AIAA 2nd Aerospace Sciences Meeting, New York, January 1965.

10. O'Brien, P. F., and Heckert, B. J., "Effective Emittance of a Blackbody Cavity at Nonuniform Temperature," *Illuminating Engineering Society*, Preprint No. 4, National Technical Conference, Miami Beach, Florida, September 1964.

11. Streed, E. R., McKellar, L. A., Rolling, R., Jr., and Smith, C. A., "Errors Associated with Hohlraum Radiation Characteristics Determinations," Symposium on Measurement of Thermal Radiation Properties of Solids, Dayton, Ohio, September 1962. Proceedings published as NASA SP-31, pp. 237–252, edited by J. C. Richmond, 1963.

12. Krebs, R. P., Haller, H. C., and Aver, B. M., "Analysis and Design Procedures for a Flat, Direct-Condensing, Central Finned-Tube Radiator," Lewis Research Center, Cleveland, Ohio, NASA TN D-2474, September 1965.

13. Volkoff, J. J., "Protection Requirements for the Resistance of Meteoroid Penetration Damage to Interplanetary Spacecraft Systems," Technical Report No. 32-410, Jet Propulsion Laboratory, Pasadena, California, July 1, 1964.

14. Ryshkewitch, E., *Oxide Ceramics*, Academic Press, New York, 1960.

15. Bacon, J., "The Evaluation of Materials for the Application to Magnetohydrodynamic Power Generation," Report No. A-210173-1, Pratt & Whitney Aircraft Division, East Hartford, Connecticut, 1962.

16. Kaufman, H. R., "An Ion Rocket with an Electron-Bombardment Ion Source," NASA Technical Note D-585, January 1961.

17. Kaufman, H. R., and Reader, P. D., "Experimental Performance of Ion Rockets Employing Electron Bombardment Ion Sources," ARS Electrostatic Propulsion Conference, ARS Preprint 1374–1360, Monterey, California, November 1960.

18. Mickelsen, W. R., and Kaufman, H. R., "Electrostatic Thrustors for Space Propulsion, Present and Future," British Interplanetary Society Symposium, London, England, October 1963.

19. Strickfaden, W. B., and Geiler, K. L., "Probe Measurements in an Operating Electron Bombardment Engine," *AIAA Journal*, Vol. 1, p. 1815, 1963.

20. Kerrisk, D. J., and Masek, T. D., "Effects of Plasma Non-Uniformity on Grid Erosion in an Electron Bombardment Ion Engine," *AIAA Journal*, Vol. 3, p. 1060, 1965.

21. Taylor, J. B., and Langmuir, I., "The Evaporation of Atoms, Ions and Electrons from Caesium Films on Tungsten," *Physics Review*, Vol. 44, p. 423, 1933.

22. Charbonnier, F. M., Swanson, L. W., Cooper, E. C., and Strayer, R. W., "Investigation of the Migration, Desorption and Voltage Breakdown Properties of Cesium Films on Refractory Electrodes," AIAA Electric Propulsion Conference, Preprint 63020, Colorado Springs, Colorado, 1963.

References (Cont'd)

23. Rump, B. S., and Gehman, B. L., "Work Function Measurements of Nickel, Molybdenum, and Tungsten in a Cesium-Hydrogen Atmosphere," *Journal of Applied Physics*, Vol. 36, p. 2347, 1965.

24. Nottingham, W. B., "Thermionic Emission," Technical Report 321, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1956.

25. Wilson, R. G., "Electrode Surface Physics Research," Summary Report, NASA Contract NAS 3-5249, Hughes Research Laboratories, 1965.

26. Aamondt, R. L., Brown, L. J., and Nichols, B. D., "Thermionic Emission from Molybdenum in Vapors of Cesium and Cesium Fluoride," *Journal of Applied Physics*, Vol. 33, p. 2080, 1962.

27. Speiser, R. C., Sohl, G., and Reid, G. C., "Ion Rocket System Research and Development," NASA Contract NAS 305250, Electro-Optical Systems, Inc., Quarterly Report, Pasadena, California, December 1964.

28. Cremer, Hans H., "Transducer Operation at Kilovolt Level," JPL Industrial Invention Flash Sheet, Serial Number 554, 1964.

29. Gambill, W. R., and Bundy, R. D., "An Evaluation of the Present Status of Swirl-Flow Heat Transfer," ASME Paper No. 62-HT-2, Presented at ASME-AIChE Heat Transfer Conference, Houston, Texas, August 5–8, 1962.

30. Kreith, F., and Margolis, D., "Heat Transfer and Friction in Turbulent Vortex Flow," *Applied Scientific Research*, Section A, Vol. 8, pp. 457–473, 1959.

31. Koch, R., "Pressure Loss and Heat Transfer for Turbulent Flow," VDI Forschungsheft, Vol. 24, Series B, No. 469, pp. 1–44, 1958.

32. Smithberg, E., and Landis, F., "Friction and Forced-Convection Heat-Transfer Characteristics in Tubes with Twisted Tape Swirl Generators," ASME Transactions, *Journal of Heat Transfer*, Vol. 86, Series C, No. 1, pp. 39–49, February 1964.

33. Gambill, W. R., Bundy, R. D., and Wansbrough, R. W., "Heat Transfer, Burnout, and Pressure Drop for Water in Swirl Flow Through Tubes with Internal Twisted Tapes," Chemical Engineering Progress Symposium Series, No. 32, Vol. 57, pp. 127–137, 1961.

34. Nissan, A. H., and Bresan, V. P., "Swirling Flow in Cylinders," *American Institute of Chemical Engineers Journal*, Vol. 7, No. 4, pp. 543–547, December 1961.

35. Talbot, L., "Laminar Swirling Pipe Flow," *Journal of Applied Mechanics*, Vol. 21, pp. 1–7, March 1954.

36. Kreith, F., and Sonju, O. K., "The Decay of a Turbulent Swirl in a Pipe," *Journal of Fluid Mechanics*, Vol. 22, Part 2, pp. 257–271, June 1965.

37. Schlichting, H., *Boundary Layer Theory*, Fourth Edition, p. 503, McGraw-Hill Book Company, Inc., New York, 1960.

X. Polymer Research

A. Nonrandomness in Base Sequences of DNA's

J. Moacanin and R. Simha[1]

1. Introduction

One of the central problems in molecular biology is the determination of the arrangement of base units in natural and synthetic nucleic acids and the consideration of the possibilities for replication mechanisms and structure-function relationships this may suggest. The problem of base sequences acquires increased significance in the light of the recent successes in the elucidation of the protein code. From a more general viewpoint of polymer chemistry, nucleic acids provide an interesting case of a quaternary copolymer. The development of a generalized mathematical framework that allows the analysis of correlations which may exist in the arrangement of monomeric units provides a powerful tool for the elucidation of problems such as copolymerization mechanisms, energies of interaction or catalyst specificity. This is of importance to our work on electrical conductivity in polymeric systems. Certain vinyl aromatic/hydrocarbon copolymers are currently being synthesized, and if it develops that short-run sequences are of great importance to conductivity, this new avenue of approach is now available for studying the sequence arrangement.

Experimental analyses of sequences in a series of natural and synthetic DNA's have led investigators to the conclusion that the arrangement of the four bases in the chain departs significantly from a random distribution (Refs. 1–5). A statistical analysis by Simha and Zimmerman (Ref. 6) showed that sequence data for a copolymer of adenine and uracil (Ref. 6) as well as for calf thymus DNA (Ref. 7) must be interpreted on the basis of higher than nearest neighbor effects. That is, the frequency of occurrence of a specific nearest neighbor configuration cannot be predicted from the composition alone and, in addition, involves a specific influence of, at least, the unit preceding the pair in the chain.

The relative frequency of various base arrangements along the chain of some specified overall compositions can be expressed in terms of conditional probabilities which describe departures from randomness. Thus, for example, if only nearest neighbor interactions are operative, the frequency of triplets can be computed from those of the doublet. But disagreement between the computed and observed triplet frequencies would indicate the presence of second nearest neighbor interactions, in which case triplet frequencies would have to be used to compute longer runs.

In this paper, experimental data on pyrimidine runs for the following animal, plant, and bacterial DNA's have been examined for evidence of higher than nearest neighbor effects: calf thymus (Refs. 2, 8, 9, 10), herring

testis (Refs. 8, 9), human spleen (Ref. 11), salmon, *e. esculentus* (sea urchin), *e. coli, p. aeruginosa, a. faecalis, m. lysodeikticus* (Ref. 8), wheat germ, and rye germ (Ref. 10).

We consider in more detail Petersen's recent results on calf thymus DNA (Ref. 9), which are more extensive than those (Ref. 2) analyzed previously (Ref. 6). In support of the generality of the conclusions reached, some of the results on the other DNA's will be discussed.

2. Discussion

The experimental methods of analysis (Ref. 2) yield frequencies of pyrimidine runs C^iT^j, which are given in

Table 1. Comparison of measured sequence data on calf thymus DNA (Ref. 9) and computed estimates assuming nearest-neighbor and random effects

| Sequence | Moles of pyrimidine/100 g atoms of DNA P | | | |
| | Measured | Computed assuming nearest neighbor effect | | Computed assuming random effect |
		(From Column 1ª)	(From Table 2ᵇ)	
C	3.92		4.25	5.35
T	6.23		5.88	7.08
C²	1.99		2.11	2.29
CT	2.96		2.70	3.03
TC	2.23		2.70	3.03
T²	2.62		3.62	4.01
C³	0.76	0.76	0.79	0.74
C²T	2.73	2.77	2.95	2.92
CT²	2.64	2.90	3.78	3.86
T³	1.28	0.83	1.67	1.70
C⁴	0.31	0.26	0.26	0.21
C³T	1.39	1.21	1.29	1.11
C²T²	1.99	1.91	2.43	2.20
CT³	1.53	1.17	2.08	1.94
T⁴	0.55	0.23	0.68	0.64
C⁵		0.08	0.08	0.06
C⁴T	0.39	0.47	0.50	0.37
C³T²	1.06	1.08	1.23	0.98
C²T³	1.27	0.94	1.56	1.30
CT⁴	0.82	0.40	1.00	0.86
T⁵	0.19	0.06	0.26	0.23

ªComputed using the first six measured frequencies: C, T, C², CT, TC, and T². For example, $C^2T = 3(\bar{N}_{CCT} + \bar{N}_{CTC} + \bar{N}_{TCC})$, where $\bar{N}_{CCT} = (C^2/2)(CT/2)/(C)$, and similarly for $\bar{N}_{CTC}$ and $\bar{N}_{TCC}$ (Ref. 7).

ᵇComputed using data from Table 2 following a procedure analogous to computations from Col. 1, except now, $\bar{N}_{CCT} = (N_{AC} + N_{GC}) <p_{CC}> <p_{CT}>$ $(<p_{TA}> + <p_{TG}>)$, and similarly for $\bar{N}_{CTC}$ and $\bar{N}_{TCC}$ (Ref. 7).

Table 1. These represent sums of frequencies for fixed numbers i and j of C and T nucleotide units, respectively. Thus, for example, C^2T is the sum of frequencies of CCT, CTC, and TCC. Each sequence is initiated and terminated by a purine unit. Hence, the tabulated numbers refer to structures of the general type—$pu\ C^iT^j\ pu$—summed over all i, j, G and A combinations. Data on nearest neighbor frequencies are taken from Josse, et al. (Ref. 4) and shown in Table 2.

The extent of departures from randomness is given by the deviations from unity of the ρ's in Table 2 (Ref. 7); also $AT \neq TA$, $TG \neq GT$, etc., make the absence of randomness self-evident. The nonrandomness of sequences was pointed out previously (Refs. 1–4). Turning our attention to pyrimidine runs (Table 2), we note the differences between measured frequencies and those computed assuming randomness (Ref. 7); again, for purine flanked doublets, $CT \neq TC$ is to be noted. On the assumption of nearest neighbor effects, but absence of

Table 2. Nearest-neighbor frequencies N_{JK} (Ref. 4), correlation coefficient ρ_{JK} and configurational probabilities $< p_{JK} >$ for native calf thymus DNA (Ref. 7)

Nearest neighbors	N_{JK}, mole %	ρ_{JK}	$< p_{JK} >$
AA	8.9	1.088	0.311
AT	7.3	0.902	0.255
AG	7.2	1.137	0.252
AC	5.2	0.838	0.182
TA	5.3	0.655	0.187
TT	8.7	1.086	0.307
TG	7.6	1.255	0.269
TC	6.7	1.091	0.237
GA	6.4	1.046	0.299
GT	5.6	0.925	0.262
GG	5.0	1.092	0.234
GC	4.4	0.948	0.206
CA	8.0	1.289	0.369
CT	6.7	1.091	0.309
CG	1.6	0.345	0.074
CC	5.4	1.147	0.249

Nucleotide composition: 28.6 mole % A; 28.3 mole % T; 21.4 mole % G; 21.7 mole % C.

penultimacy, the averaged $<p_{JK}>$ values (Table 2) reduce to p_{JK} and permit C^iT^j values to be estimated (Table 1). Similarly, longer runs were computed using frequencies of the purine flanked singlet and doublet C and T runs. The lack of agreement between the two columns of computed values is apparent. If the assumption of no penultimacy were correct, the two methods of computation should yield the same results. Moreover, neither of the two sets of frequency estimates seem to improve the random model. Thus, these observations offer strong indirect evidence for the presence of penultimate effects. Qualitatively similar conclusions are reached from the examination of data on other DNA's.

Additional insight can be gained by considering pure T or C runs, because for these the averaging effects caused by the summation over the various isomers are eliminated. Thus, the ratio $[T^j/j]/[T^{j-1}/(j-1)]$ yields p_{TT} for $j \geq 2$ and p_{TTT} for $j \geq 3$ in absence or presence of penultimate effects, respectively (Ref. 7). Table 3 lists these ratios for a variety of DNA's. Petersen's data on calf thymus show that the ratio is nearly independent of j for $j \geq 3$, giving convincing evidence for penultimacy. These are probably the most reliable results, since

Table 3. Comparison of ratios of T frequencies

DNA source	$\dfrac{T^2/2}{T}$	$\dfrac{T^3/3}{T^2/2}$	$\dfrac{T^4/4}{T^3/3}$	$\dfrac{T^5/5}{T^4/4}$	Ref.
Calf thymus	0.21	0.33	0.33	0.30	9
	0.21	0.33	0.35	0.40	6
	0.20	0.31	0.35	0.29	8
	0.11	0.46	0.37	–	10
Human spleen	0.24	0.41	0.44	0.48	11
Herring testis	0.19	0.31	0.28	–	8
	0.19	0.38	0.26	0.25	9
Salmon	0.24	0.25	0.35	–	8
E. esculentus (sea urchin)	0.26	0.33	0.31	0.26	8
E. coli	0.27	0.23	0.25	0.37	8
A. faecalis	0.14	0.22	0.17	–	8
P. aeruginosa	0.19	–	–	–	8
Wheat germ	0.21	0.37	0.43	–	10
Rye germ	0.23	0.31	0.49	–	10
Average	0.18	0.32	0.34	0.34	

separate experiments using higher DNA loadings were carried out for the determination of the longer runs, such as T^4 or T^5. The reasonable agreement with results from other experiments on calf thymus further supports these conclusions. Inspection of the values for the other DNA's shows that, in spite of considerable scatter, the same trend is followed, i.e., a lower value for $j = 2$ is followed by higher and more or less constant values for $j \geq 3$. For example, the average values for all the data in the table are 0.18, 0.32, and 0.34 for $j = 2$, 3 and 4, respectively. Data are available also for C runs, but these are less extensive because of the generally low C content and, hence, less conclusive.

Burton, et al. (Ref. 12) observed the constancy of the ratios $[T^n/n]/[T^{n-3}/(n-3)]$ and considered these as giving the frequency of randomly distributed nonoverlapping T-triplets. We note, however, that with our definitions

$$\frac{T^n/n}{T^{n-3}/(n-3)} = \frac{T^n/n}{T^{n-1}/(n-1)} \times \frac{T^{n-1}/(n-1)}{T^{n-2}/(n-2)}$$

$$\times \frac{T^{n-2}/(n-2)}{T^{n-3}/(n-3)} = p_{TTT}^3 \tag{1}$$

Hence, their numerical result follows from our calculations, whereas their model does not imply constant P_{TTT}, as we appear to find.

It should be noted that the comparisons of C^iT^j's are not completely independent, since a given set was calculated using the same nearest neighbor correlation coefficients. Certain pyrimidine frequencies, however, can be examined for consistency in a more direct way. For example, on the assumption of penultimacy, but no higher effects, one can arrive at the following relationship

$$CT^4/5 = \frac{T^2/2}{T}\left\{\left[\left(\frac{T^3/3}{T^2/2}\right)\bigg/\left(\frac{T^2/2}{T}\right) + \lambda\right]\right.$$
$$\left.\times \left(CT^3/4\right) - \lambda\frac{T^3/3}{T^2/2}\left(CT^2/3\right)\right\} \tag{2}$$

where $\lambda \equiv \bar{\Theta}_{CTT}\,\bar{\Theta}_{TTC}/(\bar{\Theta}_{CTT} + \bar{\Theta}_{TTC} - \bar{\Theta}_{TTT})$; Θ's are correlation coefficients for purine-flanked triplets which reduce to unity in absence of penultimacy. All the other terms are experimentally available quantities. For the random model the relationship reduces to

$$CT^4/5 = f_T\left\{2(CT^3/4) - f_T(CT^2/3)\right\} \tag{3}$$

where f_T is the fraction of T-nucleotide bases. For the nearest-neighbor case the equation becomes

$$CT^4/5 = p_{TT} \left\{ 2(CT^3/4) - p_{TT}(CT^2/3) \right\} \qquad (4)$$

where $p_{TT} = \; <p_{TT}> \;$ or $(T^2/2)/T$ when nearest neighbor effects are computed from Table 2 or 1, respectively. Using data on calf thymus (Table 1) and herring testis (Ref. 9) to compute the respective CT^4 values, the following results were obtained:

	Calf thymus	Herring testis
Random [Eq. (4)]	0.73	0.53
Nearest neighbor [Eq. (5)]:		
$(T^2/2)/T$	0.61	0.44
$<p_{TT}>$	0.76	—
Penultimate [Eq. (2) and assume $\lambda = 1$]	0.73	0.60
Observed	0.82	0.59

These results, of course, neither indicate a random base pairing, nor a nearest-neighbor effect, since the other respective conditions are not satisfied; e.g., $f_T \neq (T^2/2)/T \neq \; <p_{TT}>$. To obtain agreement with the observed value for calf thymus we require $\lambda = 2$. From Table 2 one finds that $\bar{\Theta}_{TTT} = 1.55$, which leads to the reasonable condition $\bar{\Theta}_{CTT} \; \bar{\Theta}_{TTC}/(\bar{\Theta}_{CTT} + \bar{\Theta}_{TTC} - 1.55) = 2$; note that if $\bar{\Theta}_{CTT} = \bar{\Theta}_{TTC} = 1$, $\lambda = 2.2$. Similarly, for herring testis agreement with experiment is obtained for $\lambda = 2.84$. Again, one finds that $\bar{\Theta}_{TTT} = 1.94$, and, hence, $\bar{\Theta}_{CTT} \; \bar{\Theta}_{TTC}/(\bar{\Theta}_{CTT} + \bar{\Theta}_{TTC} - 1.94) = 2.84$.

In conclusion, the unequivocal evidence for non-randomness along with the failure of nearest-neighbor effects alone to account for experimental results, give strong evidence for the existence of at least penultimate effects. Furthermore, without violating any of the conditions for penultimacy and using ad hoc, but acceptable, numerical values for λ (i.e., a function of $\bar{\Theta}_{CTT}$ and $\bar{\Theta}_{TTC}$) agreement with experiment can be obtained. Data on triplet frequencies are needed to ascertain the extent of effects higher than penultimacy. We wish to point out, however, that separation of isomers for the triplet pyrimidine runs, C^2T and CT^2 would yield estimates of the $\bar{\Theta}_{IJK}$'s. The latter approach should be feasible with the current separation techniques.

B. Hall Mobility and Electron Spin Resonance of Poly-N-Vinylcarbazole-Iodine Complex

A. M. Hermann and A. Rembaum

1. Introduction

The mechanism of electronic conduction in monomeric or polymeric charge transfer complexes is at present not well understood. Although these materials are known to be characterized by relatively high conductivity and high free-spin concentration as evidenced by electron spin resonance studies, no clear relationship between conductivity and paramagnetic properties has been established.

The purpose of this report is to show that the number of electrical carriers in the poly-N-vinylcarbazole (PVCA)-iodine complex constitutes only a fraction of the total number of free spins, and this may well be a general phenomenon applicable to most charge-transfer systems.

2. Experimental

Reagents and complex preparation. The reagents, their treatment, and preparation of the PVCA-iodine complex is reported in SPS 37-35, Vol. IV, pp. 115–118.

Hall effect. The development of an AC technique for the measurement of the Hall effect in high resistivity semiconductors of low mobility has permitted the determination of the Hall mobility in poly-N-vinylcarbazole complexed with iodine in various proportions.

The method is described in detail elsewhere (Ref. 13). The electric field was applied to the sample at 13⅓ cps, the sample was rotated in a static magnetic field at 20 cps, and the Hall signal was preamplified by field-effect transistors mounted on the rotating shaft near the sample and was synchronously detected at 33⅓ cps.

Resistivity apparatus. Measurements of resistivity were carried out as described in SPS 37-35, Vol. IV.

Free spin concentration measurements. The variation of the free spin concentration for various compositions of PVCA complexed with iodine in the solid state was measured at room temperature *in vacuo* with a Varian

V-4502 EPR spectrometer using 100-kc modulation. The first derivative representations were numerically integrated by the method of first moments, and absolute intensity calibration was made by comparison to the 0.1% pitch in KCl standard. Relative intensities are considerably more accurate than absolute intensities. No evidence of saturation was observed at room temperature for microwave power levels employed.

3. Results

Hall effect. It is convenient to present the data in terms of the simplest single carrier model. One defines the Hall mobility μ_H as

$$\mu_H = \frac{E_H}{E_A} \cdot \frac{1}{B}, \qquad (1)$$

where E_H is the Hall electric field, E_A is the applied electric field, and B is the magnetic field strength. For the electric field in v/cm, and the magnetic induction in webers/cm^2, the units of the Hall mobility are cm^2/v-sec.

Table 4. Hall mobility in PVCA iodine complex at room temperature

Iodine, wt %	Hall mobility, cm^2/v-sec	Sign of Hall coefficient
34	0.5	Negative
40	<0.2	
56	0.3	Negative
77	0.7	Negative
100 (Single crystal, AC plane)	2.9	Positive

Table 4 shows the variation of the Hall mobility of poly-N-vinylcarbazole complexed with iodine in various proportions. The magnitude of the Hall mobility and sign of the Hall coefficient (product of mobility and resistivity) are in agreement with the results of the drift mobility determinations for single crystals of pure iodine (Refs. 13, 14). The variation of the Hall voltage with both magnetic and applied voltage for two iodine concentrations is shown in Figs. 1 and 2. These variations are linear within the precision of the experiment.

Resistivity. Fig. 3 displays the variation of resistivity with reciprocal temperature for the complex at three iodine concentrations, as reported in SPS 37-35. The linearity of these curves shows that the complex obeys the
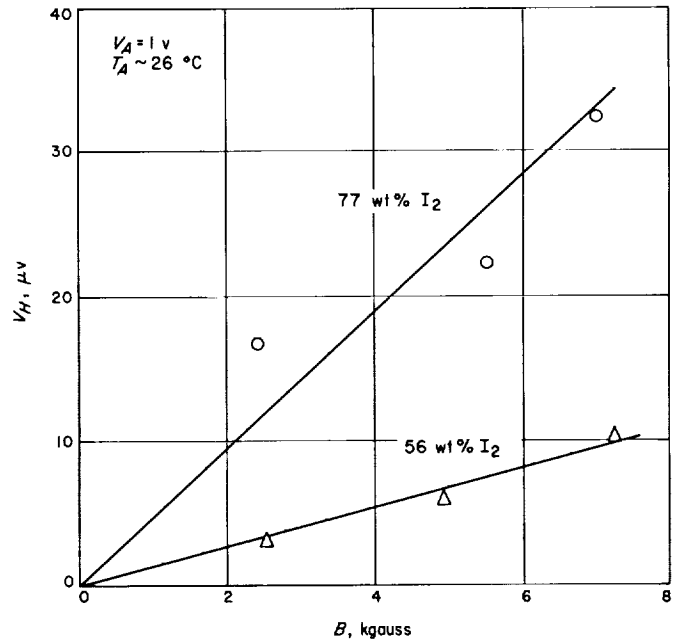
Fig. 1. Variation of Hall voltage with magnetic field strength for PVCA iodine complex at two iodine concentrations
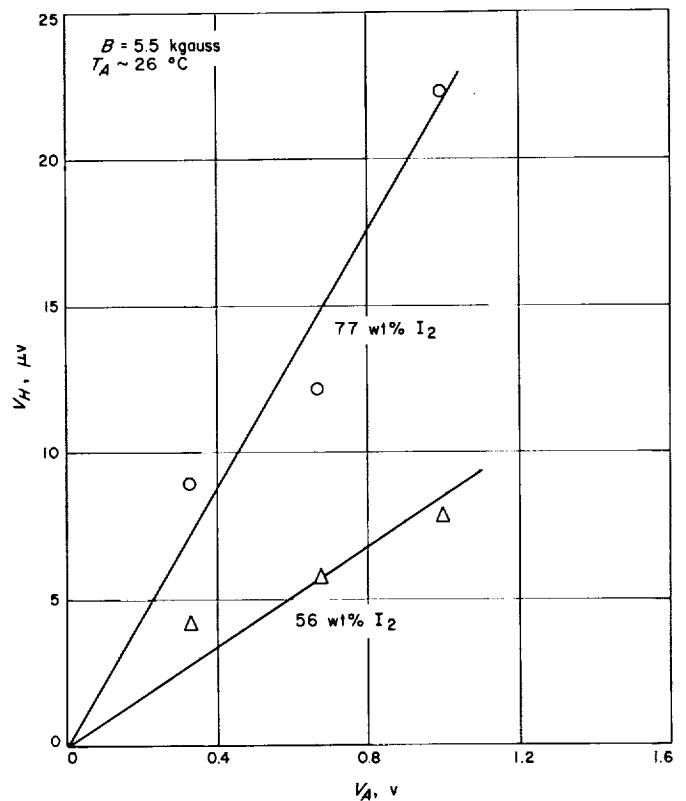
Fig. 2. Variation of Hall voltage with applied voltage for PVCA iodine complex at two iodine concentrations
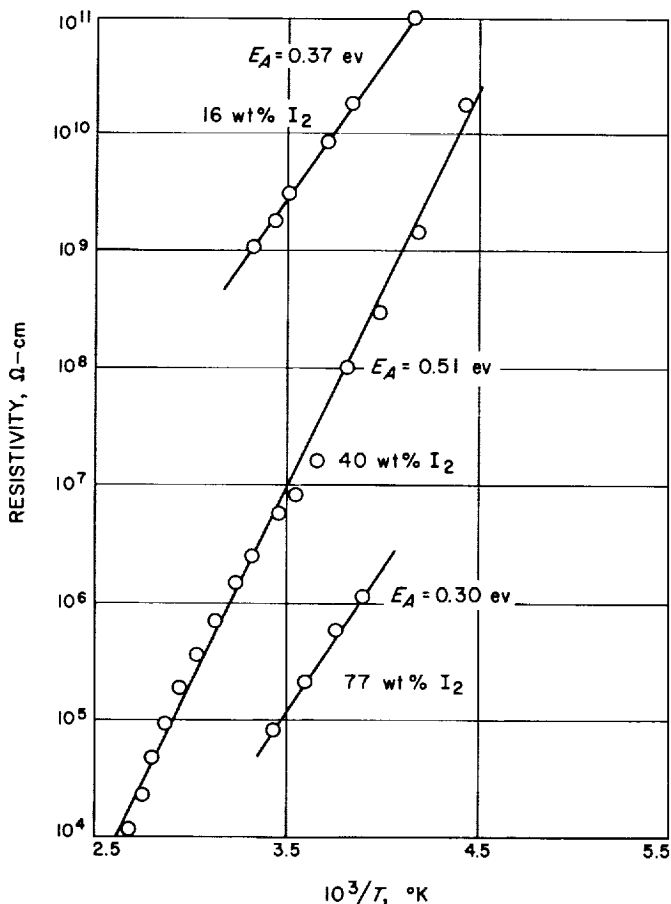
Fig. 3. Variation of resistivity with reciprocal temperature for PVCA iodine complex at several iodine concentrations

IODINE COMPOSITION, wt % (AT ROOM TEMPERATURE)

Fig. 4. Variation of conductivity, density of carriers, and density of spins with iodine content for PVCA iodine complex

Arrhenius relationship, with activation energies E_A (determined by the method of least squares) as shown in the figure.

Fig. 4 shows the variation of conductivity of the complex with iodine content at room temperature. The concentration of carriers is related to the Hall mobility and the conductivity by

$$\sigma = ne\,\mu_H \tag{3}$$

where σ is the conductivity, n is the number of carriers per cm^3, e is the electronic charge, and μ_H is the Hall mobility. Fig. 4 also shows the variation of the concentration of carriers computed from Eq. (3) with iodine content at room temperature.

Electron Spin Resonance. Also shown in Fig. 4 is the variation of the density of free spins with iodine content at room temperature. Although there is some resemblance between the shapes of the spin and carrier curves, the

concentration of free spins is in all cases more than four orders of magnitude larger than the concentration of carriers.

The line shapes of the EPR spectra varied with iodine content; the line widths (separation of peaks of the derivative) were of the order of 8 gauss. The Lande g-factor in the 34% iodine pellet was found to be 2.0041 by comparison with diphenylpicryl hydrazyl (DPPH).

4. Discussion of Results

Although a number of investigations of EPR properties of organic materials have been carried out recently (Refs. 15–21), the relationship between paramagnetic and elec-

trical properties is still not understood. The EPR signals in pyrolyzed materials and conjugated polymers seem to have little correlation with electrical properties (Ref. 18), but spin densities of perylene and pyrene complexed with iodine have the same temperature dependence as conductivities (Refs. 20, 21). It should be pointed out, however, that comparison of temperature dependencies of conductivity and free-spin creation in charge-transfer complexes is insufficient of itself to identify spins as carriers, since the mobility can be temperature-dependent (Ref. 22).

In attempts to identify spins as carriers, workers have been unable to separate quantitatively concentration of carriers from their mobilities. The measurement of the Hall mobility in the PVCA-iodine charge-transfer complex now permits direct computation of the concentration of carriers; comparison can then be made with the density of free spins determined by EPR.

Study of Fig. 4 immediately suggests the following question regarding the PVCA-iodine complex: Since the number of spins always exceeds the number of carriers by four orders of magnitude or more, does the number of carriers represent some portion of the spins which are in an excited state, or do the spins arise as part of a collectivized ground state of molecules, as suggested by Semenov (Ref. 23)? The finite thermal activation energy for spin creation in most materials suggests that free spins are in an excited state (Ref. 17). If one assumes that a portion of the spins do contribute to conductivity, and that the spin-lattice relaxation time for conducting spins is the same as for ground-state spins, one may calculate the carrier mobility from the spin-lattice relaxation time, according to the work of Elliot (Ref. 24). From band theory calculations, including spin-orbit coupling, he finds that for relaxation by lattice scattering in semiconductors the spin-lattice relaxation time T_1 is given by

$$T_1 \sim \frac{T_r}{(g - g_0)^2} , \qquad (4)$$

where T_r is the carrier relaxation time, and $(g - g_0)$ is the difference of the Lande g-factor from that of a free electron. Since we found no evidence for saturation at room temperature, we assume that $T_1 \sim T_2$, the spin-spin relaxation time. T_1, for a line width of 8 gauss, is 8.2×10^{-9} sec.

The mobility can now be calculated from

$$\mu = \frac{e}{m^*} T_r = \frac{e}{m^*} (g - g_0)^2 T_2 \qquad (5)$$

where m^* is the effective mass of the carriers, and e is the electronic charge. For the 34% I, 66% PVCA pellet, the g-value is 2.0041. Using 2.0023 for g_0 and the free electron mass for the effective mass, one finds

$$\mu \sim 45 \text{ cm}^2/\text{v-sec} \qquad (6)$$

The observed Hall mobility for this sample is 0.5 cm²/v-sec.

One would guess, however, that the effective mass in a low mobility semiconductor would be much larger than the free electron mass. If, in fact, the effective mass were 100 times larger than the free electron mass, agreement between experiment and theory could be reached.

Some electrical properties of DPPH have recently been investigated (Ref. 15). If one uses a linewidth of 1.9 gauss (an average of that reported by several authors (Refs. 25–30) and a g-value of 2.0039 (Ref. 31), one finds from Eq. (5) that

$$\mu \sim 150 \text{ cm}^2/\text{v-sec} \qquad (7)$$

where the free electron mass has been used in Eq. (5). Masuda and Yamaguchi (Ref. 15), however, failed to detect a Hall voltage in DPPH. The discrepancy again may be due to a large carrier mass.

A similar calculation applied to the perylene-iodine complex, with values of g and T_2 determined by Singer and Kommandeur (Ref. 21), yields a mobility of 10 cm²/v-sec. The Hall mobility in this complex, however, was found to be below 0.01 cm²/v-sec (Ref. 20). It is unlikely that a large effective mass can account for this discrepancy; it appears, therefore, that either the spin-lattice relaxation times for conducting and ground state spins are not the same, or that band theory calculations are not applicable.

5. Summary

In an attempt to correlate unpaired spins and charge carriers in poly-N-vinylcarbazole complexed with iodine, one finds that the density of carriers deduced from Hall data is at least four orders of magnitude smaller than the concentration of free spins. Until more mobility data become available, it is impossible to determine whether or not this is a general phenomenon. If it can be shown that band theory can be legitimately applied to these systems, and if separate determination of effective carrier mass can be made, it may be possible to predict approximate mobilities from EPR measurements, according to the work of Elliot (Ref. 24).
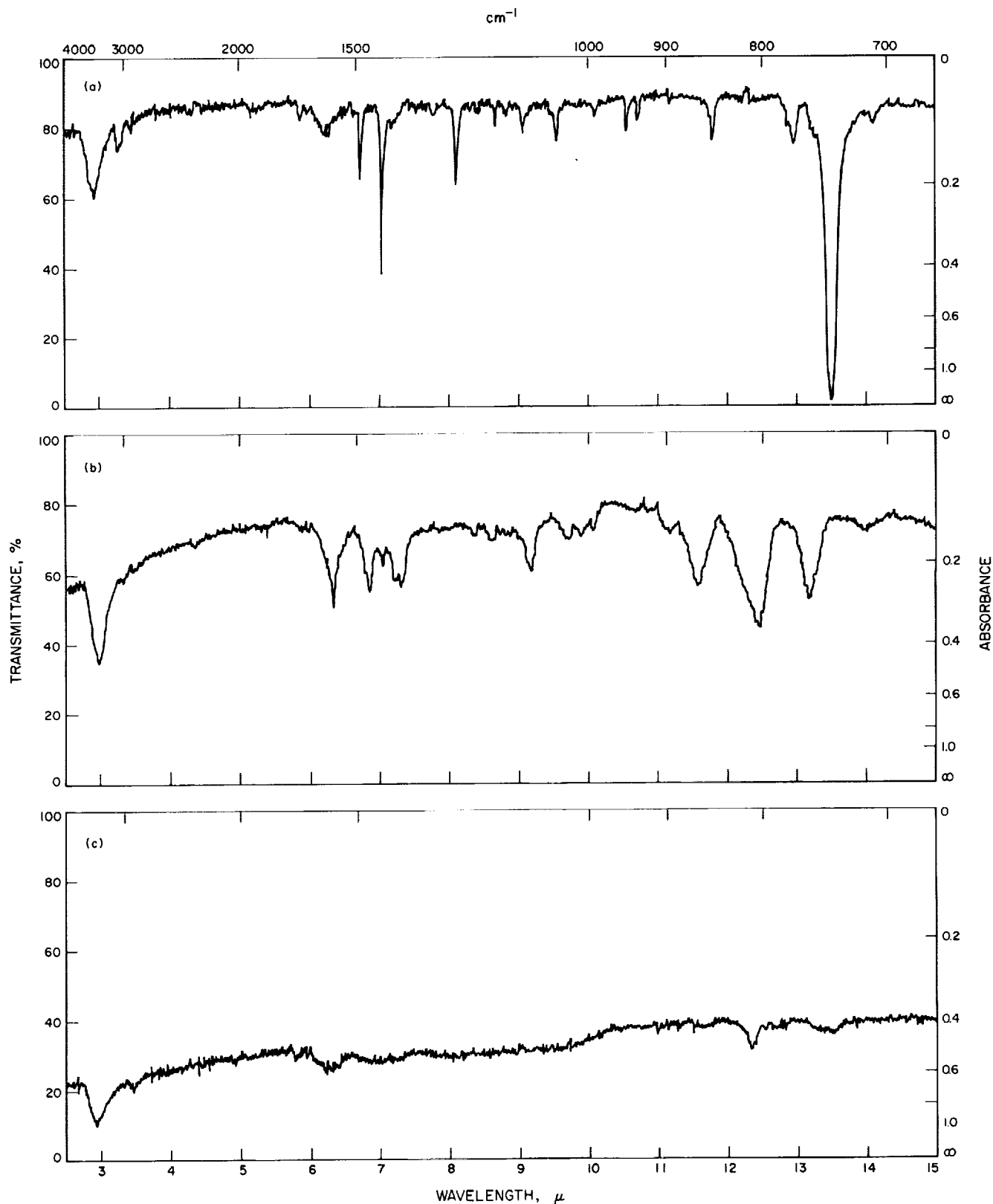
Fig. 5. Infrared spectra (KBr Disks) (a) Triphenylene, (b) Poly(triphenylene), (c) Poly(triphenylene) heated to $\simeq$ 600°C (at about 10^{-4} torr). Weak absorption, $\simeq$ 3 μ, is due to water inherent in the KBr disks.

C. Thermally Stable High Polymers

G. K. Ostrum, D. D. Lawson, R. F. Landel, and J. D. Ingham

The synthesis of poly(benzenes) has received considerable attention due to the desirable thermal stability of these polymers (*SPS 37-30*, Vol. IV, pp. 96, 97). It has been pointed out that chain branching could occur during the polymerization of benzene with Lewis acid catalyst-oxidant (Ref. 32), but the presence or absence of such branching has not yet been conclusively shown, due to the extreme insolubility of the poly(benzenes). Whenever such branching occurs, cyclization to produce a triphenylene nucleus in the polymer chain could occur by the same oxidative cationic polymerization mechanism proposed for chain growth (Reactions I-V, p. 136). So, as part of our study of the preparation and characterization of high-temperature polymers using fused salts as reaction media, we have prepared a poly(triphenylene) using

I
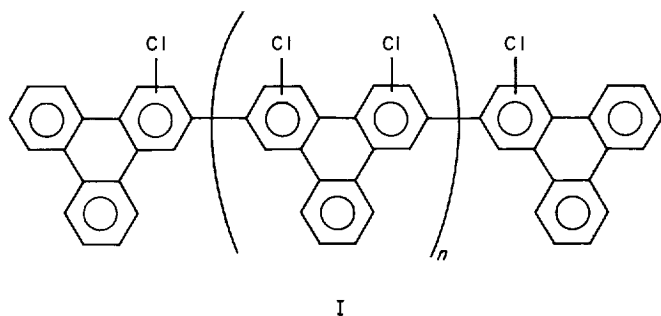
Fig. 6. Proposed structure for the poly(triphenylene)

a eutectic mixture of 60 wt% ferric chloride, 26 wt% sodium chloride, and 14 wt% potassium chloride (Ref. 33), both as "solvent" and source of Lewis acid catalyst-oxidant.

Triphenylene was polymerized at 220 $\pm 10°$C until hydrogen chloride evolution ceased. The tan polymer was readily formed and isolated in 43% yield; nonpolymeric products were not analyzed. Elemental analysis data (C, 76.45; H, 2.96; Cl, 20.90) were analyzed by a computer program (Ref. 34), and results indicate that the repeating triphenylene unit can be 3, 11, 14, 16, 17 ... (or multiples thereof) with the lower values intuitively more likely. The C/(H + Cl) ratio (1.805) indicates that the polymer is completely aromatic (theoretical value, 1.800). The usually intense triphenylene absorption peak at 738 cm^{-1} (Fig. 5a) was not apparent in the infrared spectrum (Fig. 5b) of the polymer. When a sample of the polymer was heated to $\simeq 600°$C at about 10^{-4} torr and the residue examined, there was a significant change in the infrared spectrum (Fig. 5c), indicating very extensive degradation of the polymer. Fig. 5c agrees closely with the known infrared spectrum of carbon black (Ref. 4). The polymer is insoluble in the usual organic solvents, but is soluble to the extent of $\simeq 60\%$ in refluxing phenanthrene (bp 340°C). These data are consistent with the proposed Structure I (Fig. 6) for this poly(triphenylene).

Differential thermal analysis showed that an exothermic reaction occurred at 457°C when the poly(triphenylene) was heated under nitrogen at a programmed rate of 10°C/min (Fig. 7). A subsequent weight loss also was
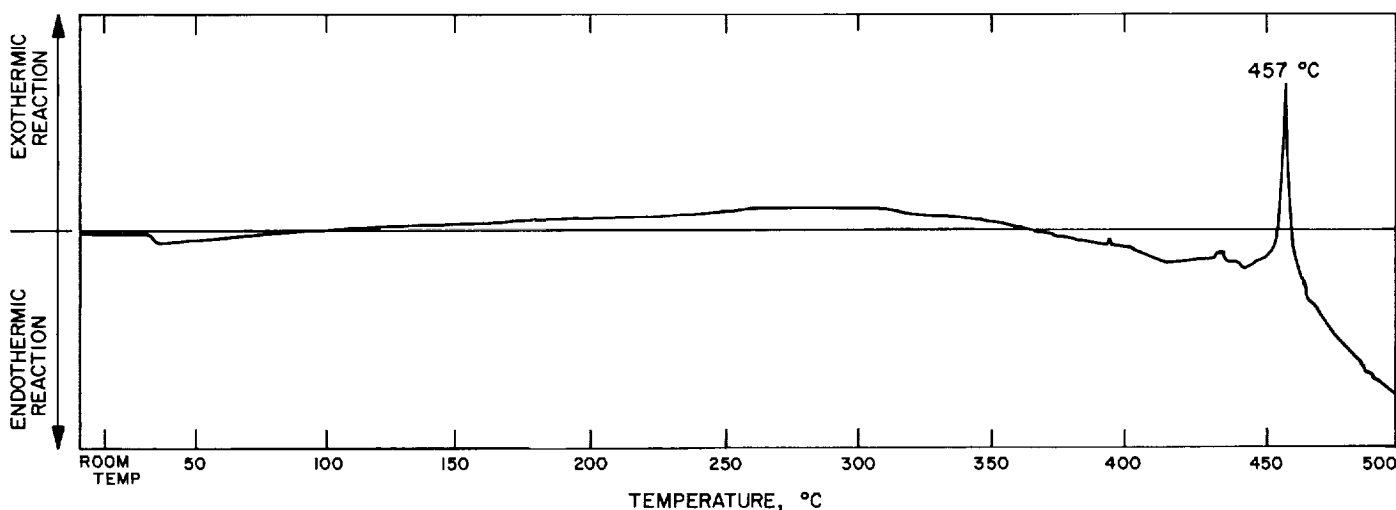
Fig. 7. Differential thermal analysis thermogram for poly(triphenylene)

(I)

(II)

(III)

(IV)

REACTION CONTINUES AS IN I AND/OR III   (V)
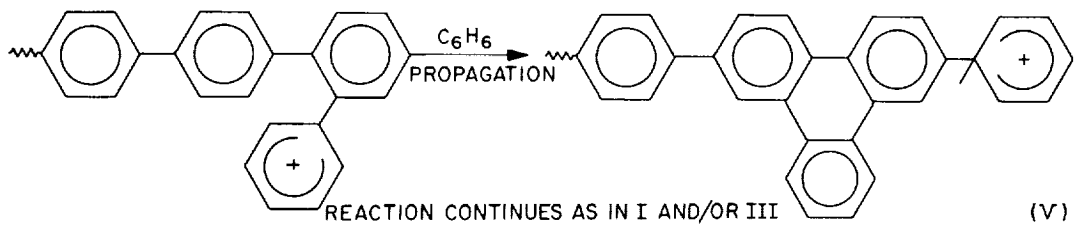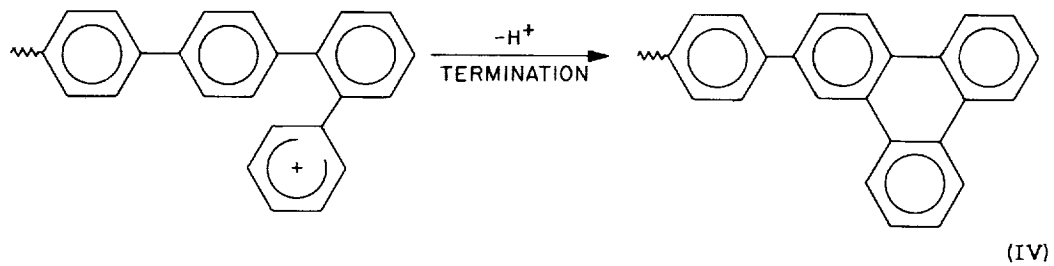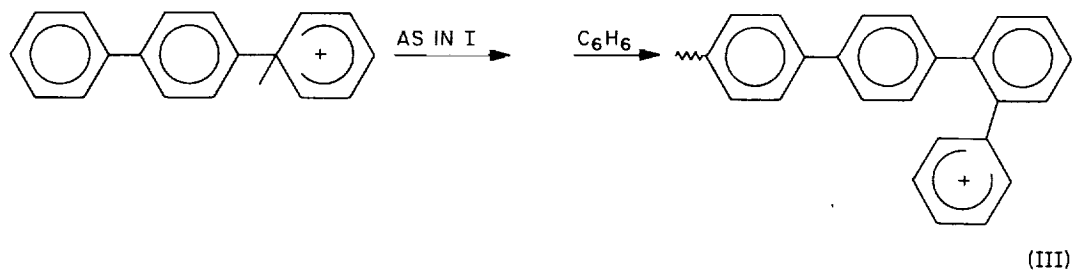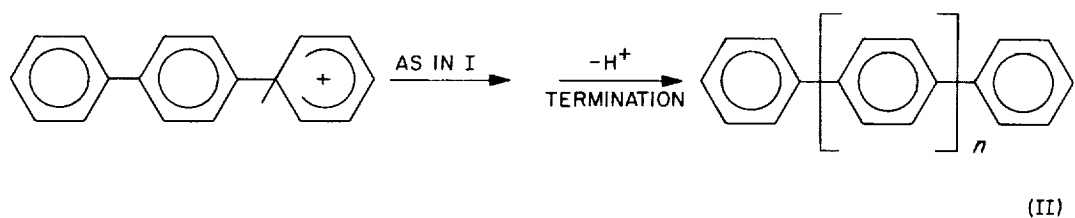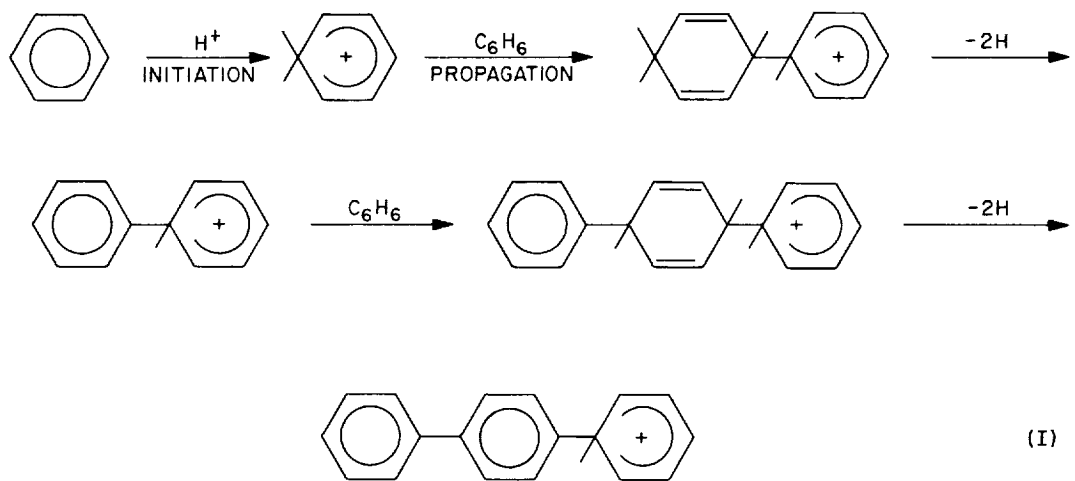
indicated. It is postulated that this exothermic reaction is due to the formation of more highly condensed carbo-cyclic ring systems within the polymer. Further experiments to test this postulate are in progress.

A preliminary determination of thermal gravimetric behavior was also carried out. An 11-mg sample of the polymer was heated in a vacuum system ($<10^{-4}$ torr), and its weight monitored with a Cahn electrobalance, while the furnace temperature was programmed from 200 to 620°C ($\simeq 2$°C/min) and held at the higher temperature overnight. At $\simeq 460$°C the integral weight loss was equivalent to the initial total hydrogen and chlorine content. This temperature also corresponded to a minimum volatilization rate, indicating a change from the end of one type of degradation process to a second type. The latter apparently is characteristic of the polymeric structure obtained at the DTA exotherm. The total weight loss was $\simeq 34\%$. At this stage, the volatilization rate was nearly immeasurably low, and the residue consisted of glistening black carbon particles. Further studies are in progress to determine the volatile products evolved by mass spectrometry, and further characterize the polymer structure during the degradation processes.

D. Rupture of Amorphous Polymers

R. F. Landel and R. F. Fedors

Several facets of the rupture behavior of amorphous, unfilled elastomers have been discussed in SPS 37-18, 37-22, and 37-30, Vol. IV and in *Quarterly Summary Report* 38-14. Since these reports cover a span of three years, the following recapitulation is presented in order to summarize and unify the work in a single SPS.

Rupture of elastomers depends simultaneously on many parameters and, therefore, a comprehensive understanding of fracture requires knowledge of not only their separate effects but also of their interactive effects. Thus, studies have been made on the influence of time (Ref. 36), temperature (Ref. 36), degree of crosslinking (Refs. 37, 38), details of chain structure (Refs. 38, 39) and concentration of diluent (Ref. 40), but seldom has a sufficiently wide range and combination of such variables been covered in any given experimental program so as

to provide an indication of the interactions. In order to synthesize these experimental results into a more coherent whole, fracture in viscoelastic bodies should not be treated as a separate phenomenon, but rather as one facet of the problem of describing their stress-strain-time (σ, ϵ, t) properties (Ref. 41). (Only tensile properties are considered here.) If these properties are represented by a three-dimensional surface, which we call a *physical property surface*, then fracture represents some limiting value or discontinuity on this surface, or boundary to it (Ref. 41). Fig. 8 depicts such a surface for a gum Viton B elastomer.

The projections of this boundary to the σ, t or ϵ, t planes depict the time dependence of fracture, while the projection to the σ, ϵ plane is independent of the time scale. This latter projection, known as the Smith failure envelope, is of great importance because of this independence. Moreover, the failure envelope appears to be independent of the path, so that the same envelope is generated in stress relaxation (Refs. 41–43), creep (Ref. 43), or constant strain rate (Refs. 41–43) experiments. As such it serves as a very useful failure criterion.

On the surface itself, the curves AB and AC represent the paths traced out by a creep (constant stress) experiment and a stress-relaxation experiment, respectively. Note that the starting point A is the same for both. The boundary of the surface at large ϵ values represents the occurrence of rupture; the projection of this boundary curve to the σ, ϵ plane generates the time-independent failure envelope. Shown as the dashed lines are the projections of the creep and relaxation experiments. Rupture occurs when the paths intersect the failure envelope at points B and C.

Thus, in mathematical terms, a useful description of fracture behavior requires at least two functions which can be solved simultaneously to yield values of σ, ϵ and t at break. A possible pair of such functions, for example, would be the set comprised of one expression for the property surface and one expression relating breaking strain ϵ_b to the time-to-break t_b.

There is ample experimental evidence to suggest that over much of the property surface, strain and time are factorable (Ref. 44) so that the equation for the surface may be written as:

$$\sigma = E(t) f(\epsilon, n) \qquad (1)$$

where $E(t)$ is the time-dependent, strain-independent modulus. At long times, E approaches equilibrium and becomes time-independent; here kinetic theory predicts
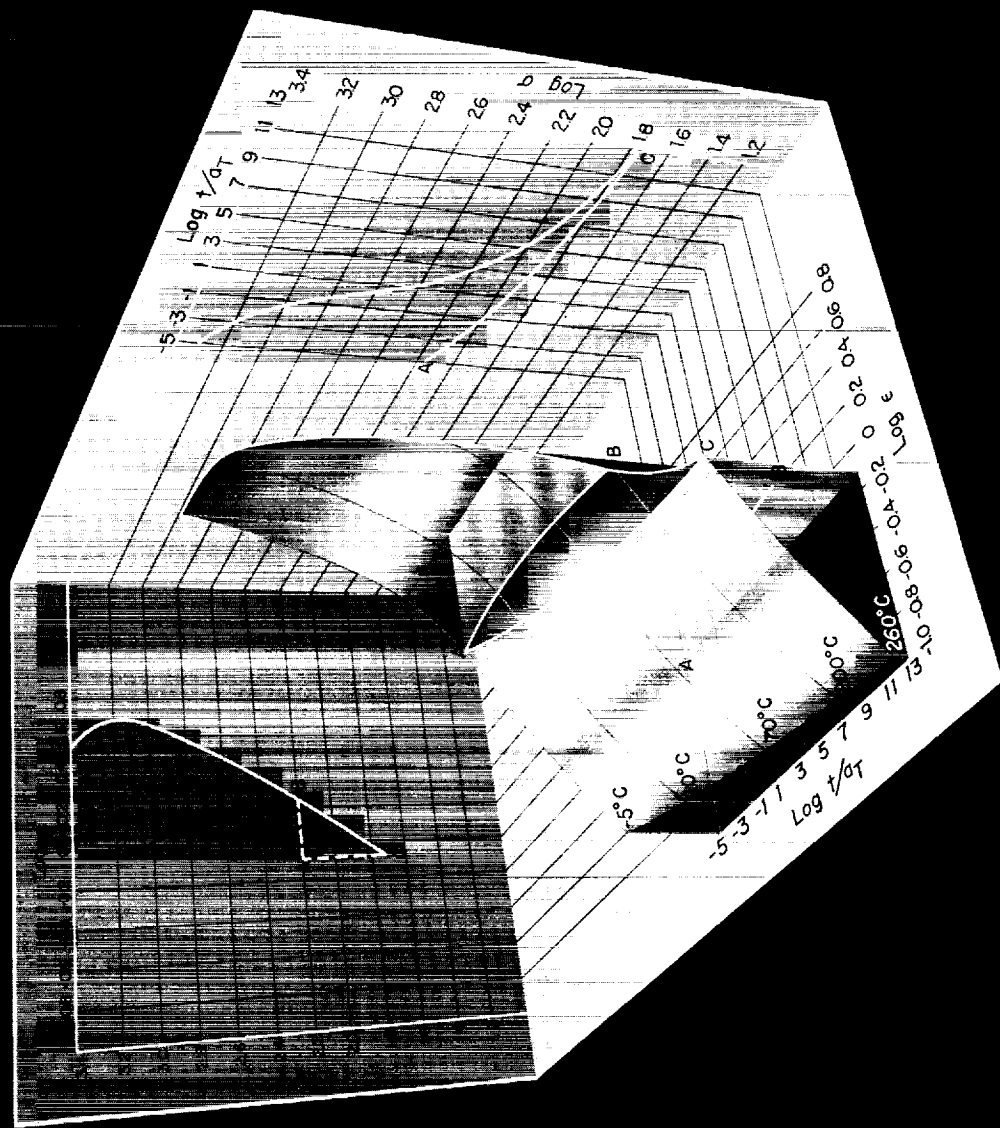
Fig. 8. Stress-strain-time surface for a Viton B elastomer (Ref. 41)

$E_e = 3\nu_e RT$. The function $f(\epsilon, n)$ is generally a non-linear function of strain and also depends on a parameter n which is a measure of chain flexibility. A possible form for the function $f(\epsilon, n)$ which fits data very well has been proposed by Treloar (Ref. 45). It is based on an extension of the kinetic theory of rubberlike elasticity which takes into account the finite extensibility of a polymer chain

$$f(\epsilon, n) = \frac{n^{1/2}}{9}\left[\mathcal{L}^{-1}\left(\frac{\lambda}{n^{1/2}}\right) - \frac{1}{\lambda^{3/2}}\,\mathcal{L}^{-1}\left(\frac{1}{\lambda^{1/2}n^{1/2}}\right)\right]$$

(2)

where $\lambda = \epsilon + 1$ and where $\mathcal{L}^{-1}$ is the inverse Langevin function, e.g., if

$$\frac{\lambda}{n^{1/2}} = \coth\beta - \frac{1}{\beta} = \mathcal{L}(\beta)$$

then

$$\mathcal{L}^{-1}\left(\frac{\lambda}{n^{1/2}}\right) = \beta$$

In Eq. (2), it is easy to show that $f(\epsilon, n) \to \infty$ as $\lambda \to n^{1/2}$ and thus, the magnitude of n provides an upper limit to the maximum value which λ can attain. Thus, if Eq. (2) is valid for values at break, then

$$(\lambda_b)_{max} = n^{1/2}$$

(3)

where $(\lambda_b)_{max}$ is the maximum value of λ_b as obtained from the upturn in the failure envelope at the point where $d\sigma_b/d\epsilon_b = \infty$. Eq (3) can be recast into the more convenient form which takes account of the presence of soluble rubber and free chain ends

$$(\lambda_b)_{max} = \left[\frac{g}{\nu_e N\left(1 + \dfrac{\rho g}{\nu_e \overline{M}_n}\right)}\right]^{1/2}$$

(4)

where g is the gel fraction, N is the molar volume of statistical units, ρ is the density, and $\overline{M}_n$ is the number average molecular weight of the primary molecules. When $\overline{M}_n$ is large, the factor in parenthesis approaches unity and $(\lambda_b)_{max}$ becomes inversely proportional to $\nu_e^{1/2}$. There is experimental data (Ref. 38) to support this type of dependence on ν_e.

In addition, when $\lambda \ll n^{1/2}$, Eq. (2) becomes essentially independent of n, and hence the effect of n on f becomes apparent primarily when $\lambda \to n^{1/2}$. To indicate how well Eq. (2) applies to break data, Fig. 9 shows rupture data obtained with SBR gum at the indicated rates and temperatures. For the comparison it was assumed that $E = 3\nu_e RT$ and that $f(\epsilon, n)$ is given by Eq. (2). The values of n and ν_e required for the curve fit were 50 and
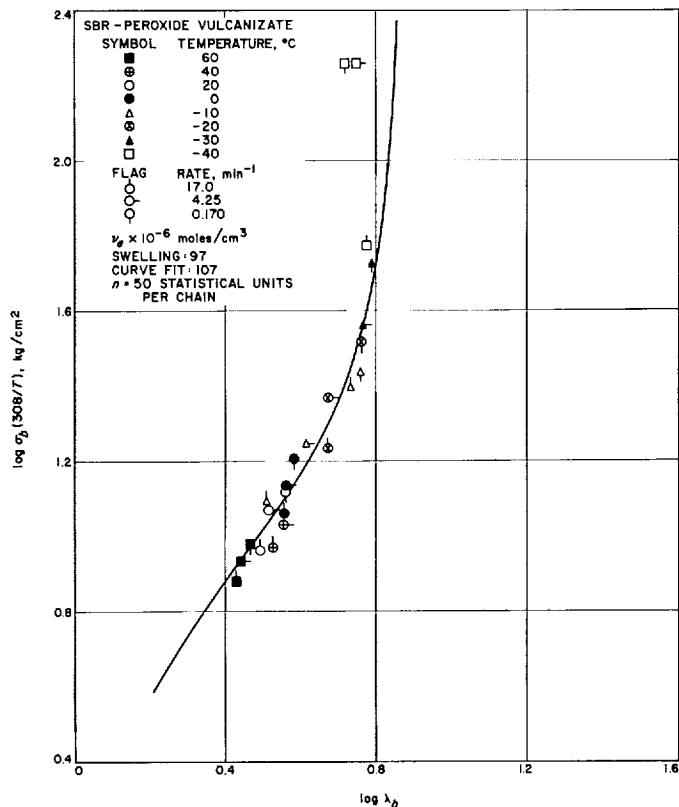
Fig. 9. Comparison of the experimental break data for an SBR gum elastomer with the prediction of Eq. (1)

107×10^{-6} moles/cm^3, respectively. The experimental value of ν_e calculated from equilibrium swelling in benzene is 97×10^{-6} moles/cm^3.

Experimentally it has been observed (Ref. 41) that in the ϵ_b, t_b plane, ϵ_b depends on both t_b and n, i.e.,

$$\epsilon_b = g(t_b, n)$$

(5)

Hence, Eqs. (1) and (5) considered together define a space curve on the property surface which defines fracture for all experimental conditions for which Eqs. (1), (2) and (5) remain valid.

Differentiation and rearrangement of Eqs. (1) and (5) yield an equation which defines the shape and location of the failure envelope in the σ_b, ϵ_b plane:

$$\frac{d\sigma_b}{d\epsilon_b} = f\frac{\left(\dfrac{dE}{dt_b}\right)}{\left(\dfrac{\partial g}{\partial t_b}\right)} + \left[E\frac{\partial f}{\partial n} - f\left(\dfrac{\partial g}{\partial n}\right)\frac{\left(\dfrac{dE}{dt_b}\right)}{\left(\dfrac{\partial g}{\partial t_b}\right)}\right]\frac{dn}{d\epsilon_b} + E\left(\dfrac{\partial f}{\partial\epsilon_b}\right)$$

(6)

Eq. (6) predicts that the shape of the failure envelope depends on many variables simultaneously in a rather complicated fashion. It is interesting to note that since the failure envelope is independent of time, the left-hand side of Eq. (6) can only be a function of ϵ_b and n. This implies that functional relations must exist between the quantities in Eq. (6) which are time-dependent. If the expressions for E, g and f were known, the explicit occurrence of time in Eqs. (1) and (5) could be eliminated to yield the mathematical expression for the envelope.

Fortunately, over a rather wide range of temperature, $E(t)$ is only a very slowly varying function of time which can thus be taken as constant and set equal to its kinetic theory value of $3\nu_e RT$. Under these conditions, Eq. (6) becomes

$$\frac{d\sigma_b}{d\epsilon_b} = 3\nu_e RT\left[\left(\frac{\partial f}{\partial n}\right)\left(\frac{dn}{d\epsilon_b}\right) + \frac{\partial f}{\partial \epsilon_b}\right] \qquad (7)$$

Thus for those conditions under which E can be taken as constant, Eq. (7) predicts that the shape of the envelope depends only on f and on its variation with both ϵ_b and n and on the variation of n with respect to ϵ_b. Furthermore, the explicit dependence of the envelope on ν_e can be removed by normalizing σ_b to unit ν_e, i.e.,

$$\frac{d\left(\frac{\sigma_b T_o}{\nu_e T}\right)}{d\epsilon_b} = 3R\,T_o\left[\left(\frac{\partial f}{\partial n}\right)\left(\frac{dn}{d\epsilon_b}\right) + \frac{\partial f}{\partial \epsilon_b}\right] \qquad (8)$$

where T_o is an arbitrary reference temperature.
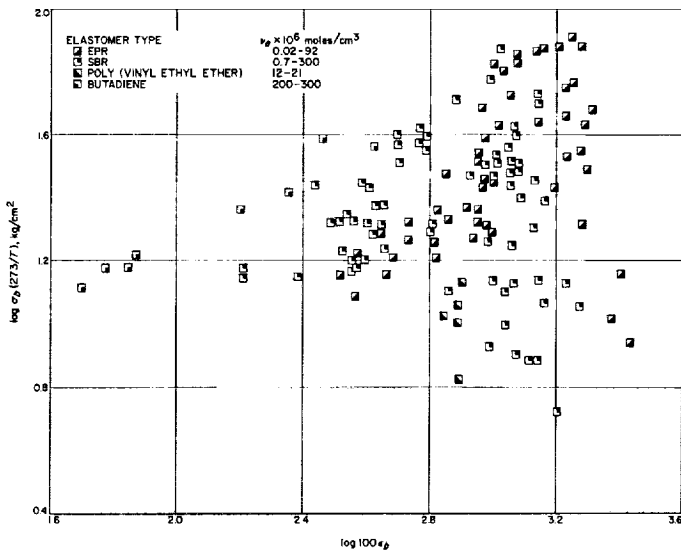
Fig. 10a. Plot of break data obtained at constant strain rate and temperature as a function of crosslink density for four types of gum elastomers
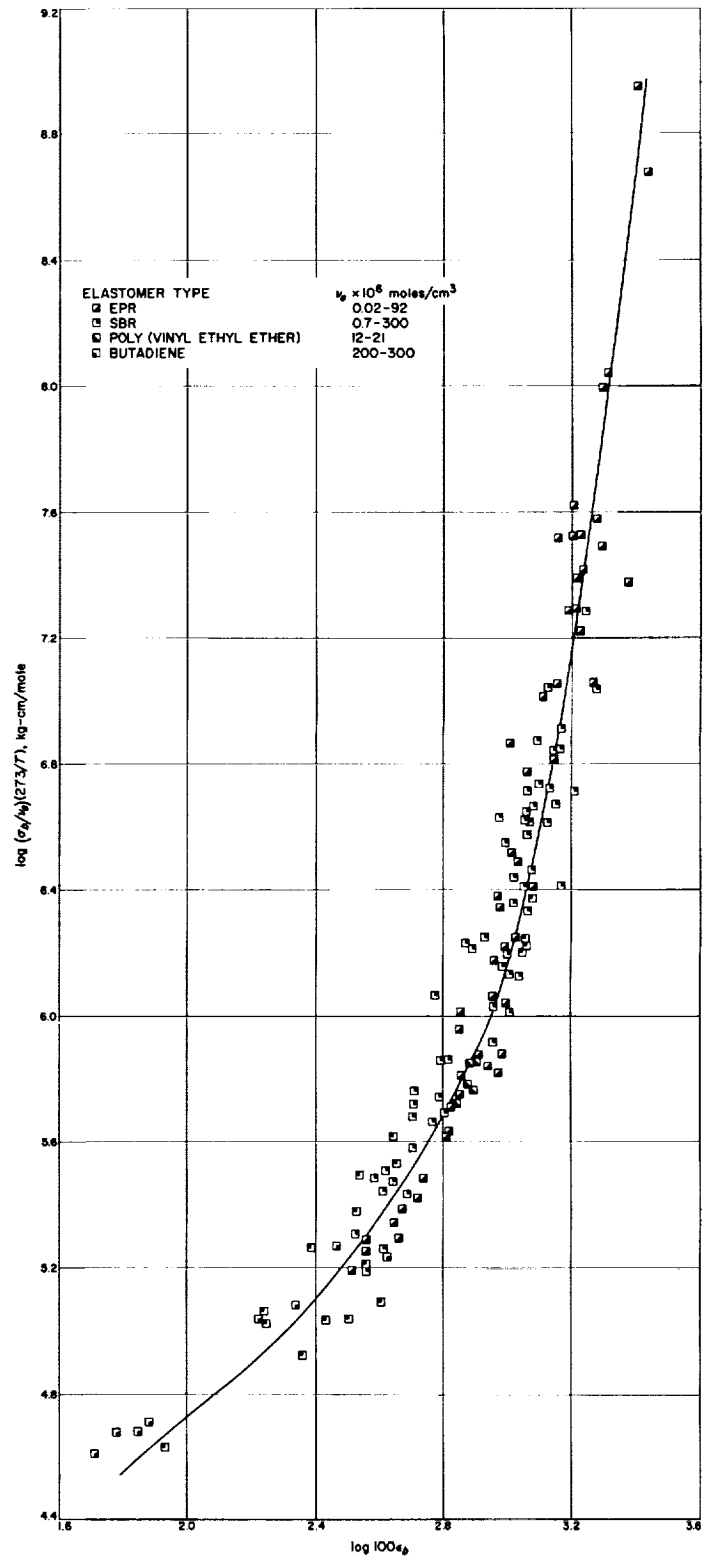
Fig. 10b. Data of Fig. 10a replotted with σ_b reduced to unit $\nu_e T$ as suggested by Eqs. (8) and (10)

In addition, since for small values of the ratio ϵ_b/n, f is essentially independent of n, then Eq. (8) becomes

$$\frac{d\left(\frac{\sigma_b T_0}{v_e T}\right)}{d\epsilon_b} = 3R\, T_o \left[\frac{\partial f}{\partial \epsilon_b}\right] \quad (9)$$

which can be integrated directly to yield

$$\frac{\sigma_b T_o}{v_e T} = 3R\, T_o\, f(\epsilon) \quad (10)$$

It has been shown previously (Refs. 38, 41), that when the breaking stress is normalized to unit v_e, then at temperatures sufficiently above the glass temperature, the reduced stress is a unique function of the breaking strain,

independent of the chemical structure of the chain (epoxy, silicone, butadiene, butyl, fluorocarbon, styrene-butadiene), such that data for some fifteen elastomers can be reduced to a single master curve as predicted by Eq. (10).

In Fig. 10a, the rupture data, obtained at a single rate and temperature, are shown for four elastomers which vary in both chemical structure and crosslink density (Ref. 46)[1,2]. When these data are reduced to unit crosslink density, as suggested by Eqs. (8) or (10), the data super-

[1]Dudek, T., unpublished results.
[2]van der Hoff, B. M. E., unpublished results.
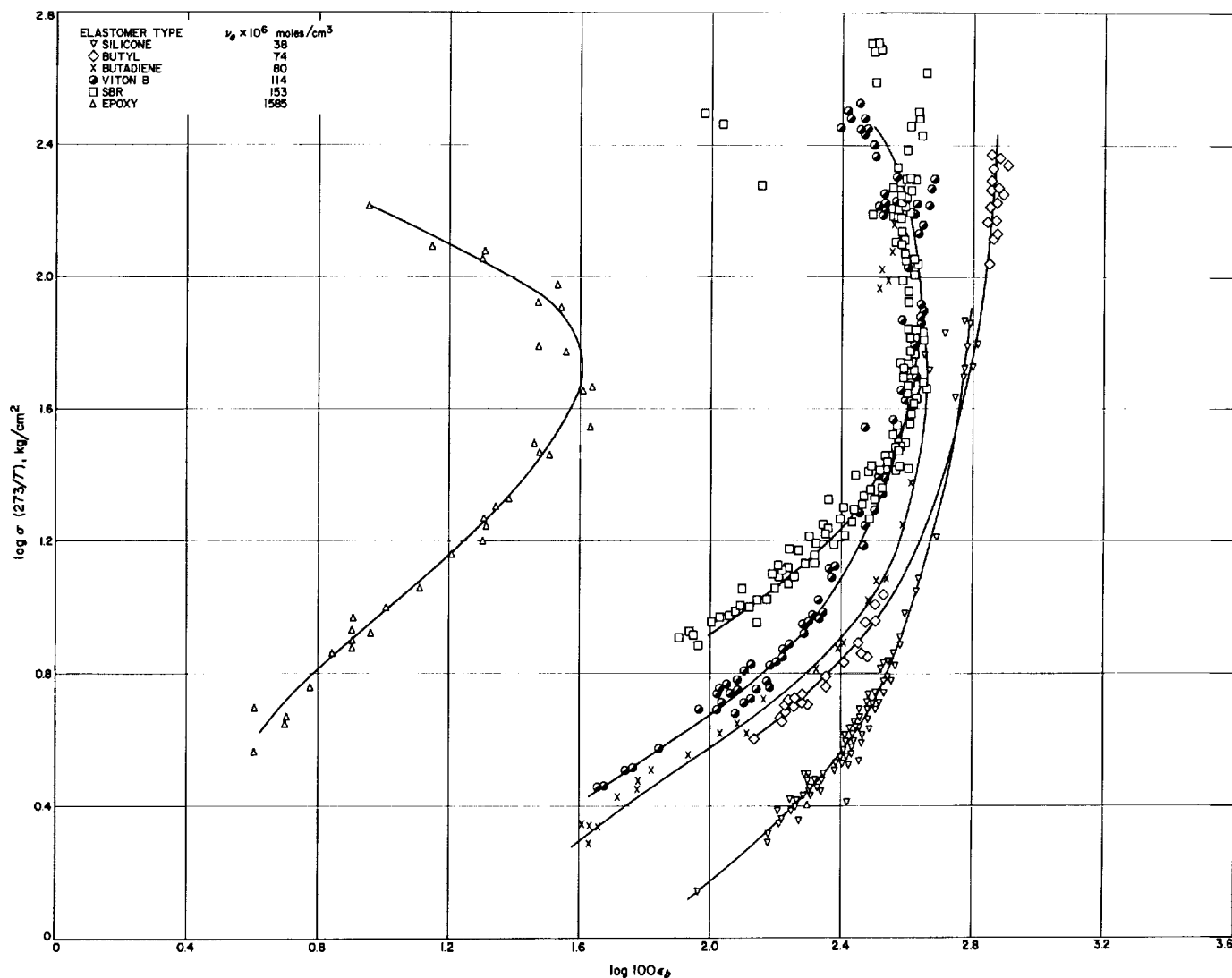
Fig. 11a. Failure envelope data obtained at constant crosslink density as a function of strain rate and temperature for 6 elastomers differing in chemical structure

pose to a single curve to within the experimental uncertainty in the reported ν_e values. The line shown in Fig. 3b represents the average behavior for this set of data.

As a more clear-cut demonstration of the independence of the reduced master envelope on chemical structure, Fig. 11a shows the failure envelopes for five types of
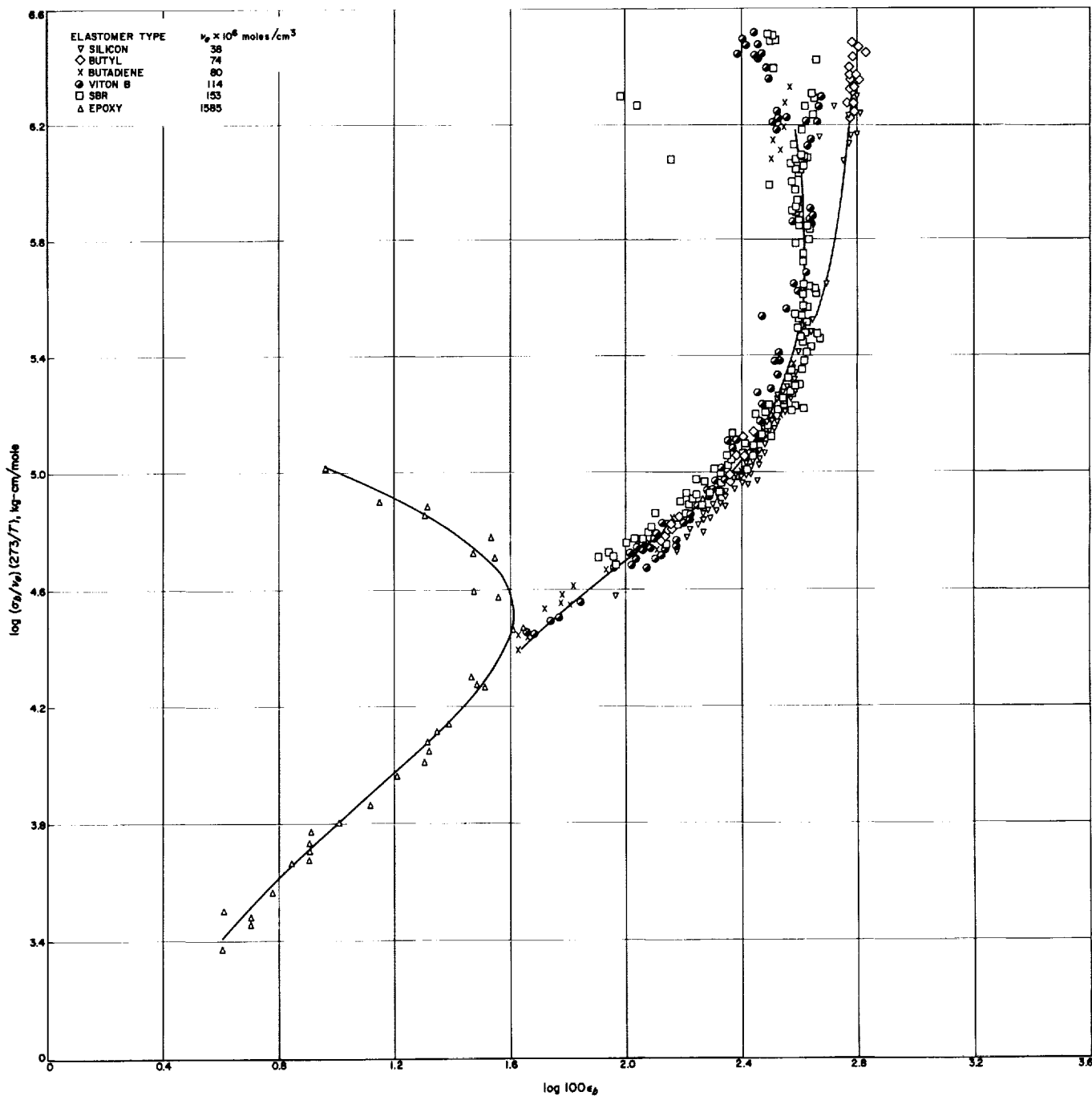
Fig. 11b. Data of Fig. 11a replotted with σ_b reduced to unit $\nu_e T$ as suggested by Eqs. (8) and (10)

rubber, each at a constant crosslink density. These envelopes were obtained by making tests at several strain rates and temperatures (Refs. 38, 39, 47, 48). As the temperature is lowered or the strain rate is increased, the data points move counterclockwise around the envelope. When these envelopes are reduced, as in Fig. 11b, their high-temperature portions superpose as expected, even though ν_e varies by a factor of 40.

It might be argued that, since not only ν_e by T_g and type of backbone are being changed, the reduction is not real but only an artifact stemming from the simultaneous change in several variables. Hence we offer the evidence based on a single rubber, Viton A-HV, as measured by Smith (Ref. 37). Fig. 12a is a plot of his tabulated rupture data, the crosslink densities being given in the key. Fig. 12b shows the reduced failure envelope and it can be
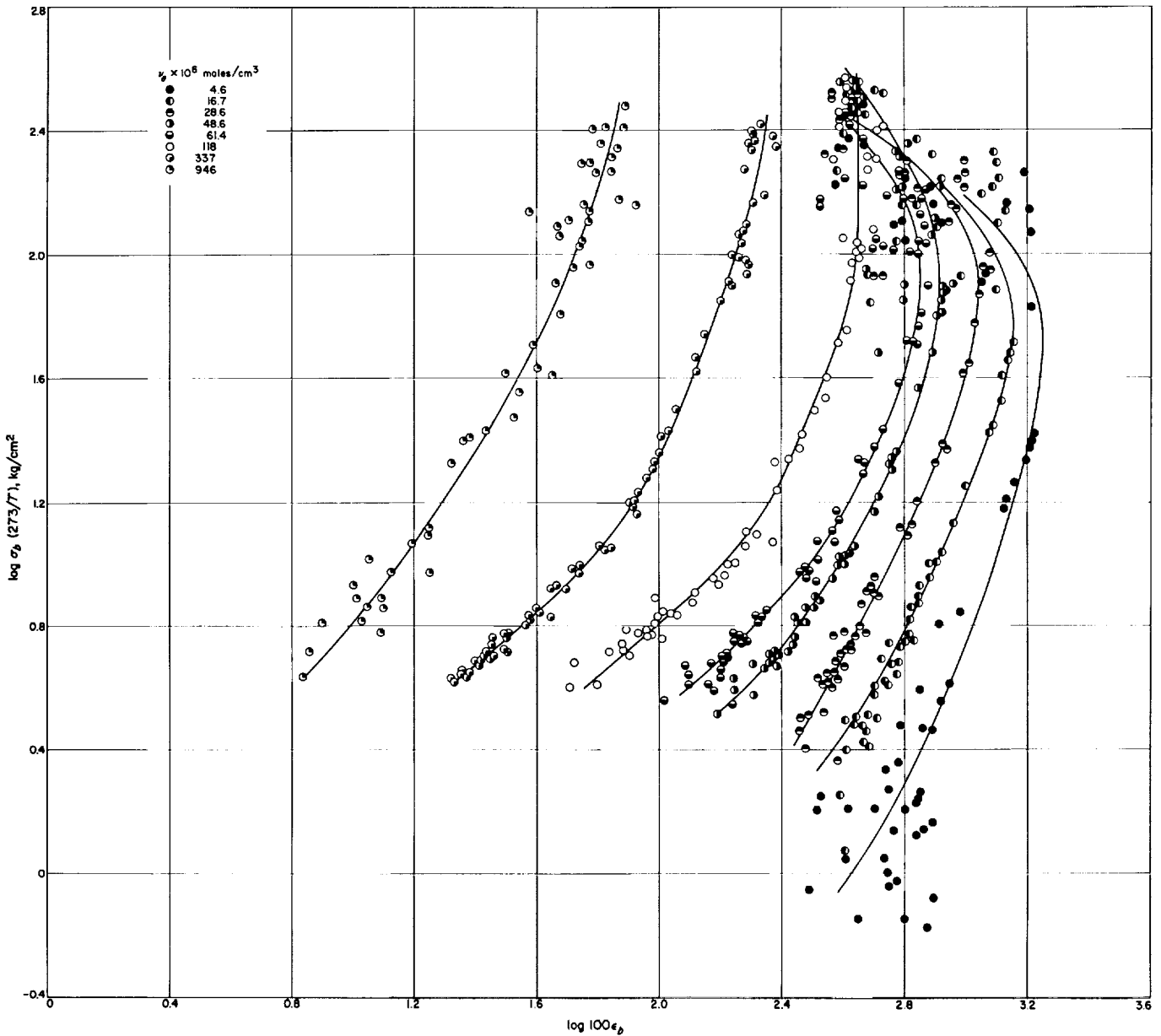
Fig. 12a. Plot of failure envelope data obtained as a function of rate, temperature and crosslink density for a Viton A-HV elastomer
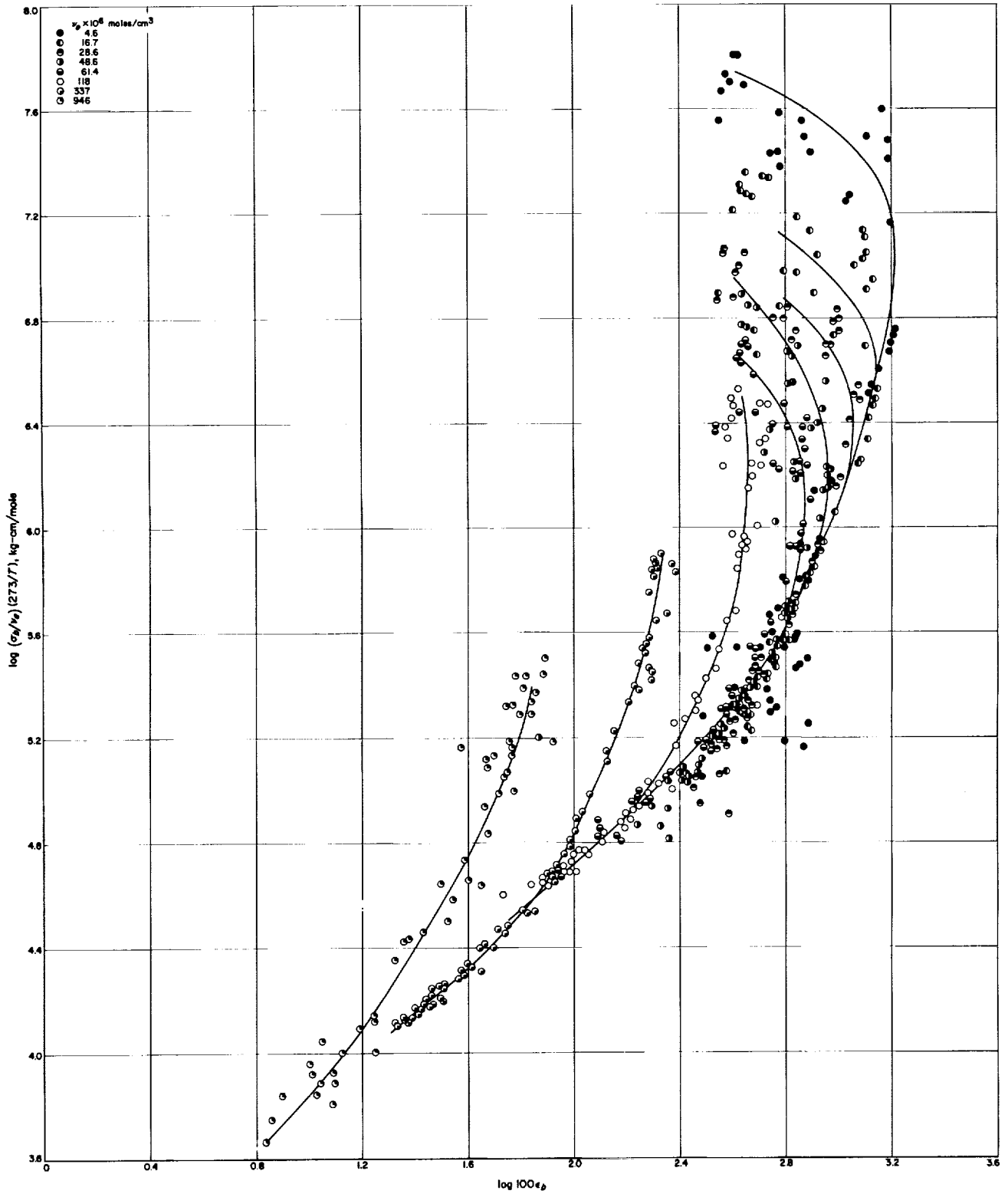
Fig. 12b. Data of Fig. 12a replotted with σ_b reduced to unit $\nu_e T$ as suggested by Eqs. (8) and (10)
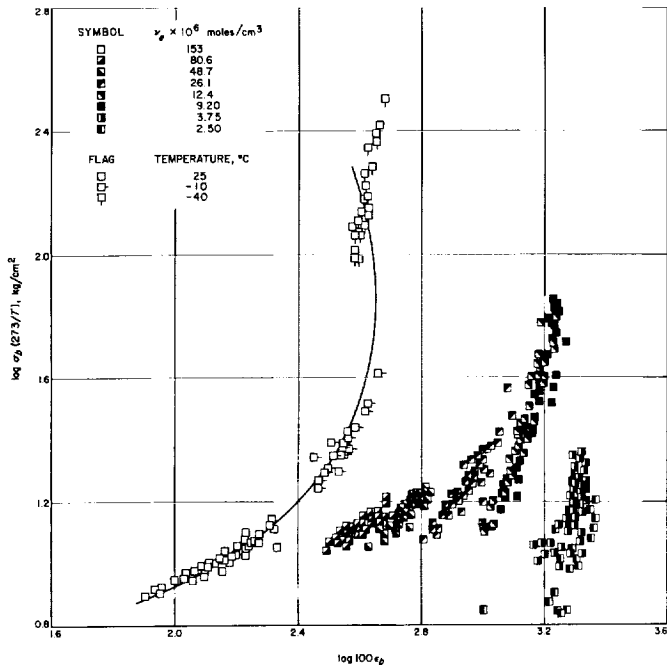
Fig. 13a. Plot of failure data obtained as a function of temperature and sample size for an SBR gum elastomer

Fig. 13b. Data of Fig. 13a with σ_b reduced to unit $\nu_e T$ as suggested by Eqs. (8) and (10)

seen that the high-temperature portions have all merged into a common response curve, independent of ν_e.

The master envelope or common response curve is independent of statistical variations in rupture properties. Hence, statistical studies of the breaking properties as measured for a sample of a given crosslink density at a given strain rate and temperature permit the delineation of portions of the failure envelope for that sample. This is illustrated in Fig. 13a, where the unfilled squares show portions of the envelope for an SBR rubber as measured by the statistical variability of 23 specimens of constant ν_e at the three indicated temperatures (Ref. 38) compared to the envelope formed by varying the strain rate and temperature (the solid line). The filled squares denote the segments formed at 25°C by break data obtained from testing 46 specimens of an SBR gum at constant rate and temperature but varying ν_e values (Ref. 49). In Fig. 13b, these data have been reduced to unit ν_e and compared with the reduced master curve (solid line) obtained earlier. Again, the reduction principle holds at high temperatures and hence is unaffected by statistical variability.

Fig. 14 shows the reduced failure envelopes for all the systems discussed here. The high-temperature portions of all envelopes have been brought together into a master curve, which is essentially the same one shown earlier. Thus the reduced variable concept as proposed is valid for wide ranges in polymer-type T_g, crosslink density, test rate, test temperature, and statistical fluctuations in break data.

Therefore, it is concluded that at sufficiently high temperatures compared to the glass temperature, both experimental evidence (Fig. 14) and Eq. (10) predict that failure envelopes obtained from samples differing in ν_e will superpose to a common response curve $f(\epsilon)$ independent of chemical structure of the polymer when σ_b is normalized to unit $\nu_e T$. At lower temperatures, experiment and Eq. (8) indicate that in such a normalized plot, individual failure envelopes will diverge from the common response curve $f(\epsilon)$, due primarily to the influence of the chain flexibility parameter n. As the temperature is lowered still further, Eq. (6) shows that the shapes of individual envelopes may vary if the time dependences of E and/or g differ. This effect is presumably the factor which produces the difference in shapes between the Viton elastomers and the others shown in Fig. 14. In addition, Eq. (4) relates the maximum value which λ_b can attain to other readily measured parameters, notably ν_e. Thus the knowledge of the parameters, ν_e and n, is sufficient to predict to a good approximation the shape and location of the failure envelope up to the region of $(\lambda_b)_{max}$.
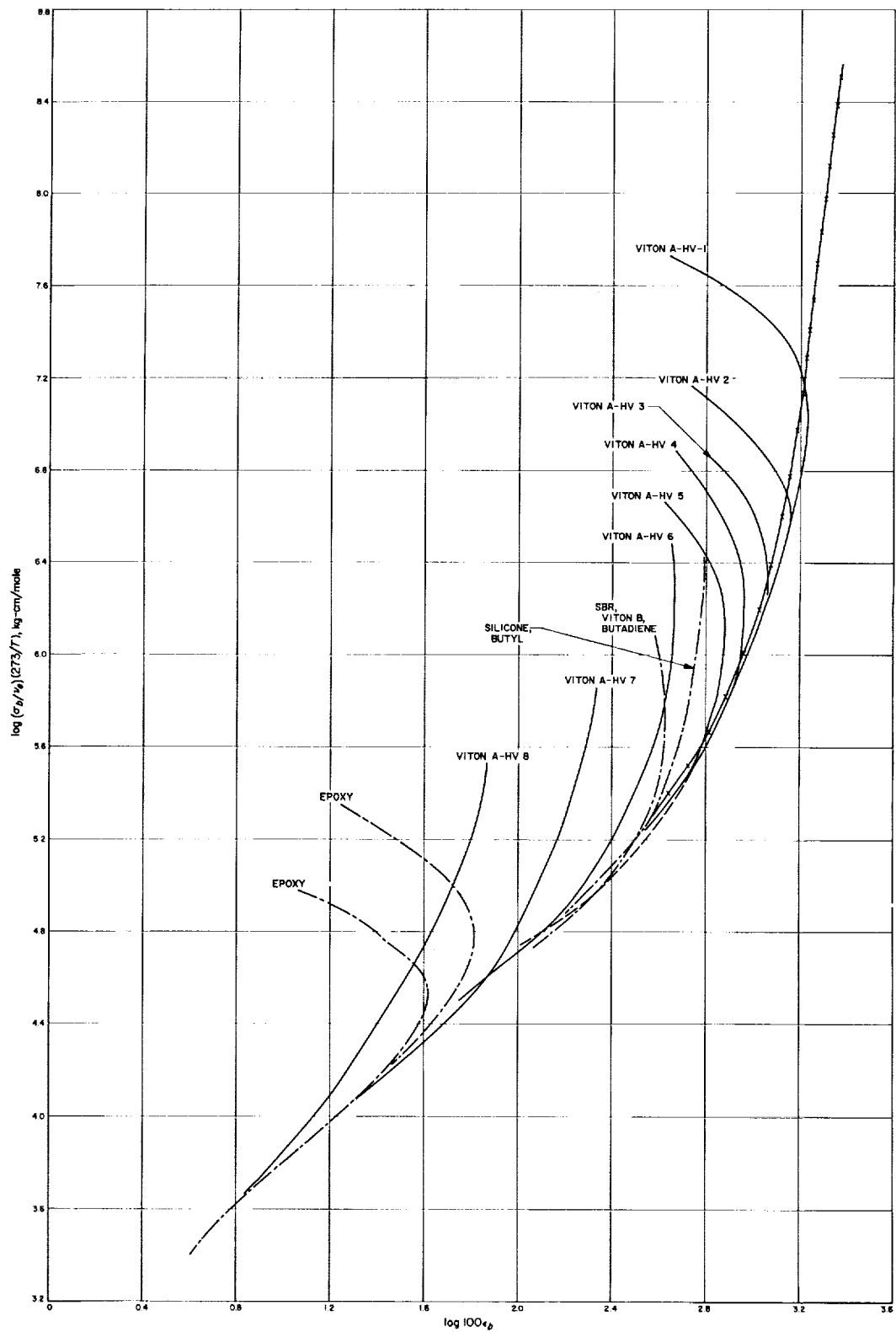
Fig. 14. Data of Figs. 9 to 13b with σ_b reduced to unit $\nu_e T$

E. Structural Integrity of Solid Retrorockets, II

A. San Miguel and E. N. Duran

A study program has been initiated (Ref. 50) to evaluate state-of-the-art experimental techniques and theoretical concepts pertinent to the prediction of the structural integrity of solid retrorockets. The *Surveyor* spacecraft retrorocket was chosen as being representative of the future design trends for retrorockets. The major objective of this program is to establish a reference basis from which to predict the structural integrity reliability aspects of future retrorockets.

The results and status of the four major tasks (Ref. 50) which comprise this program follow:

1. Task 1 Progress

Four ¼-scale rocket cases have been hydrotested to 600 psi. The technique of drilling a 0.046-D hole, and reinforcing the hole with a 0.100-OD × 0.015-in. doubler, has thus been shown to be structurally feasible. Fifty such reinforced holes are illustrated in Fig. 15. These holes are used to extricate guide wires, etc., during the propellant cure cycle. A strain rosette, bonded to the
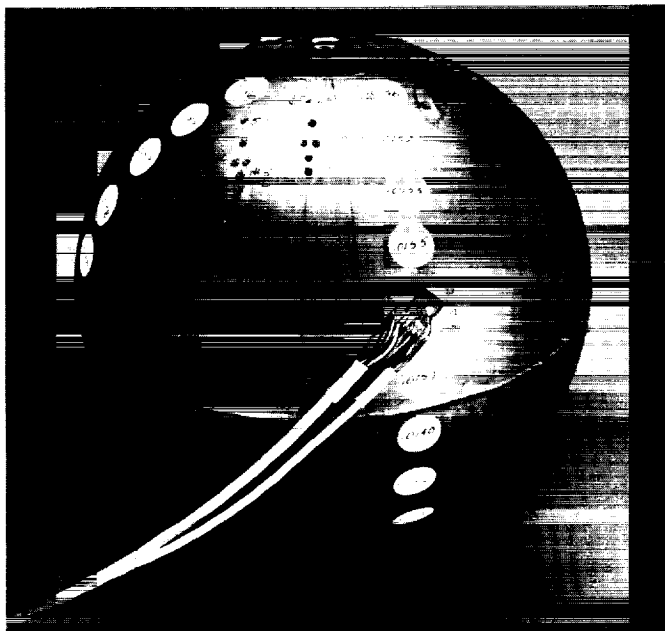
Fig. 15. *Surveyor* retrorocket casing with reinforced access holes (¼-scale)

exterior surface of the case, is employed to estimate average case strains due to pressurization. The locations of the embedded miniature stress rosettes (Ref. 51) are shown in Fig. 16.

The case is made of 410 stainless steel. The uniaxial tensile strength of sample coupons, after heat treatment, indicated excellent isotropic properties and an ultimate strength of 189K psi.

Liner is to be made from Gen-Gard V-45 (butadiene-acrylonitrile rubber with filler). After degreasing and treating the areas to be case bonded, the unvulcanized rubber will be dissolved in methyl ethyl ketone and painted on the inner surface of the case. The liner will then be vulcanized at 300°F (149°C) for 1 hr.

One case is to be mounted with miniature stress rosettes and the second case is mounted with thermocouples in the same locations (Fig. 16).

2. Task 2 Progress

A three-dimensional viscoelastic stress analysis (Ref. 50) is currently being performed by Douglas Aircraft Company under Contract (NAS 7-392). Pertinent physical properties have been measured for a JPL processed TPH 3062 formulation in Task 3 and supplied to Douglas in order to perform their rigorous three-dimensional stress analysis. The solutions (stress fields) from this theoretical analysis will be available in February.

3. Task 3 Progress

The measurement of pertinent physical properties via state-of-the-art techniques (Ref. 52) has been completed for a JPL-processed TPH 3062 propellant. This propellant is an aluminized, ammonium perchlorate–filled carboxy-terminated polybutadiene, and its mechanical properties are different from those of *Surveyor* TPH 3062. *Surveyor* TPH 3062 will soon be characterized in a similar fashion by examining batch check samples supplied by Thiokol.

The basic data obtained are as follows: (1) broad-spectrum tensile properties, (2) stress relaxation modulus, (3) creep compliance, (4) bulk modulus, (5) specific heat, (6) thermal conductivity, (7) coefficient of thermal expansion, (8) glass temperature, (9) density, (10) cure shrinkage, and (11) moisture absorption.

The above data were obtained from specimens with geometries as shown in Fig. 17.

GAGE No.	LOCATION		
	x	y	z
1	1.474	0	4.132
2	4.360	0	-0.500
3	3.928	1.892	-0.500
4	2.314	1.114	-3.138
5	2.088	0	0
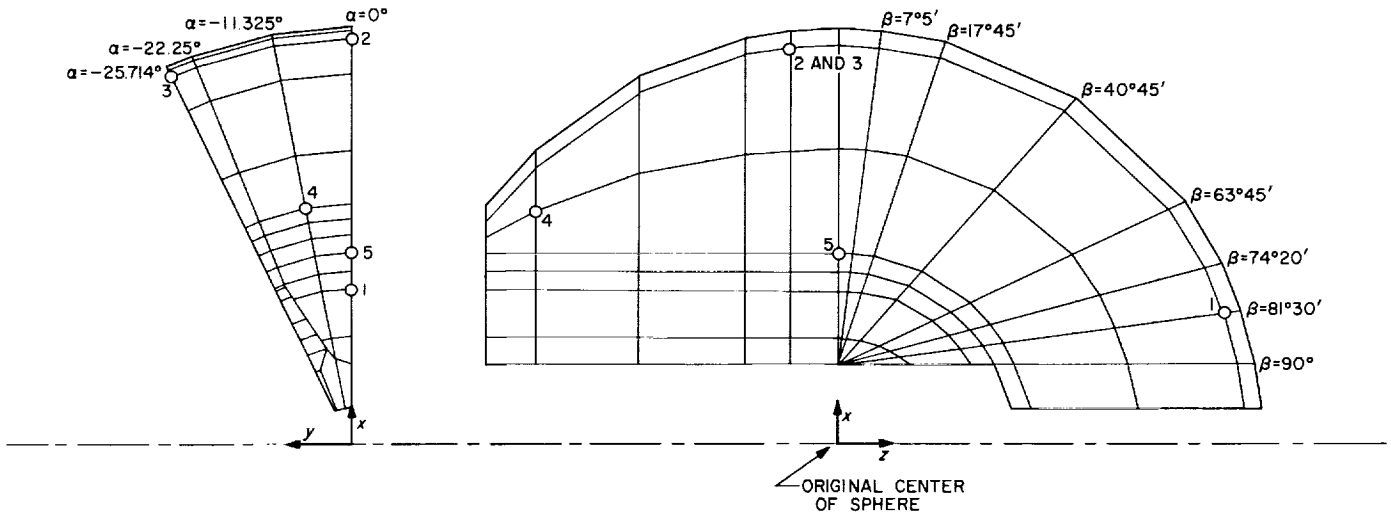
Fig. 16. Locations of the embedded miniature stress rosettes and thermocouples
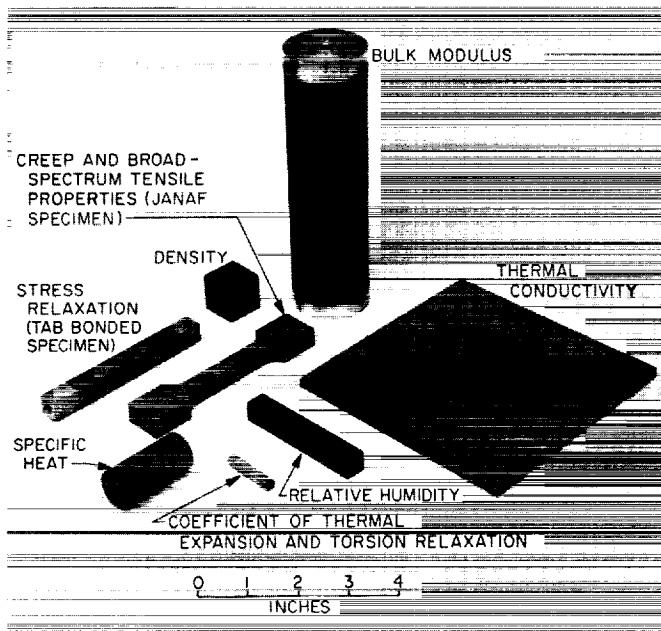
Fig. 17. Test specimens of JPL-processed TPH 3062 propellant

a. Broad spectrum tensile properties. The broad-spectrum tensile properties consist of characterizing, in terms of linear viscoelasticity theory, the uniaxial stress-strain response as a function of temperature and time (Ref. 52). These data were obtained from a JANAF tensile bar configuration. The crosshead speeds were 0.002, 0.2, 2 and 20 in./min. The temperature range was between -70 and $180°F$ (-56 and $82°C$). The strains both lateral and axial, were measured with a 70-mm camera. The strain rates were obtained from the strain-time plots. The time-temperature shift factor a_T was approximated from (Ref. 52).

$$\log a_T = -\frac{8.86\,(T - T_s)}{101.8 + (T - T_s)} \qquad (1)$$

T_s was taken as $262°K$ ($-11°C$). The reduced strain rates ($\dot{\epsilon} a_T$) experimentally obtained lie between 10^{-7} and 10^{+7} (minutes).

A reduced master true stress-strain curve for the JPL propellant is shown in Fig. 18, plotted in double logarithm coordinates, i.e., the logarithms (to the base 10) of
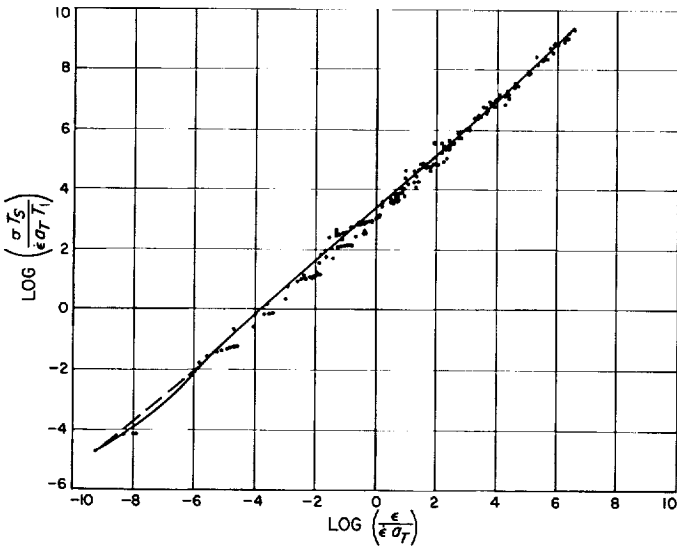
Fig. 18. Reduced master stress-strain curve for JPL-processed TPH 3062 propellant

the values are plotted on ordinary graph paper, rather than the values being plotted on logarithm paper as usually seen. This curve is obtained by multiplying the stress ordinate by T_s/T ($a_T \dot{\epsilon}$) and the strain abscissa by $1/(\dot{\epsilon} a_T)$ for each individual uniaxial stress-strain curve (46 curves).

If the propellant behaved as a quasilinear viscoelastic material, as inherently assumed in this reduction procedure, the superposition of the individual reduced stress-strain curves would produce a continuous reduced master stress-strain curve.

The master stress relaxation modulus (Fig. 19) can be obtained from the reduced master stress-strain curve (Fig. 18), as indicated in Ref. 52.
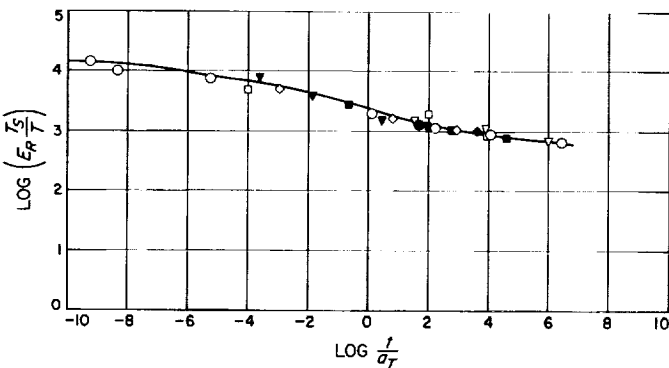
Fig. 19. Master stress relaxation modulus for JPL-processed TPH 3062 propellant

Other broad spectrum data that have been obtained for failure analysis are plots of strain at maximum stress versus the $\log_{10} a_T \dot{\epsilon}$, strain at rupture versus $\log_{10}$, a_T, $\dot{\epsilon}$, $\log T_s/T$ times maximum stress versus $\log_{10}$, $a_T \dot{\epsilon}$, and $\log_{10} T_s/T$ times initial tangent modulus versus $\log_{10} a_T \dot{\epsilon}$. These data, aside from failure evaluation, serve to characterize the failure capabilities of a given propellant system.

b. Stress relaxation modulus. The testing apparatus used to obtain stress relaxation data is shown in Fig. 20. This apparatus consists of a load cell mounted to the base plate of a frame. The cell is surrounded with a water-cooled jacket to maintain the cell's temperature when the apparatus is placed in an oven. Tab-ended specimens are used. The specimen is rapidly loaded (faster than 15 in./sec) to a predetermined length by means of a pneumatic, solenoid-actuated piston.

Fig. 21 illustrates the decay of the relaxation modulus with respect to time for three predetermined strains and three temperatures. Data are listed starting for a time
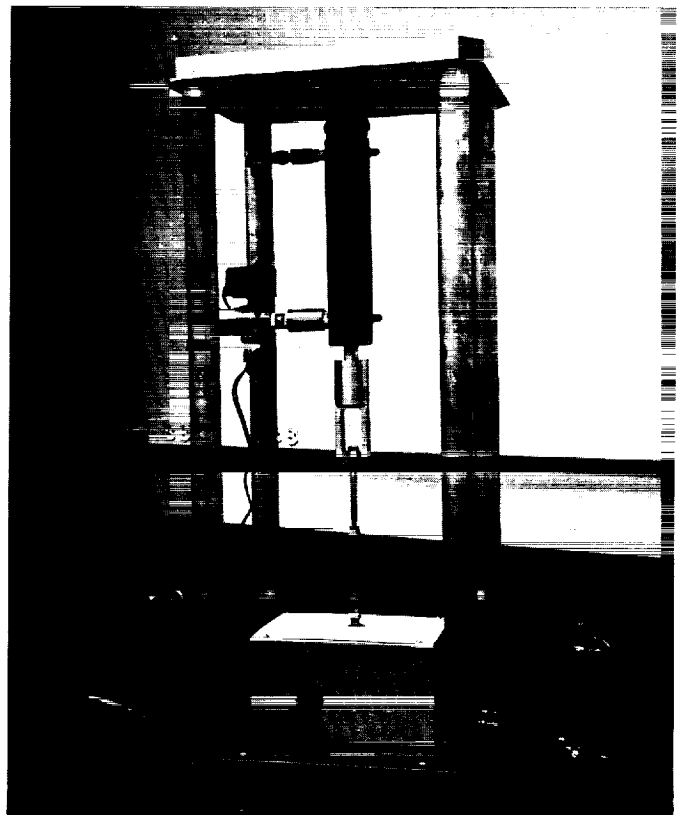
Fig. 20. Testing apparatus used to obtain stress relaxation data
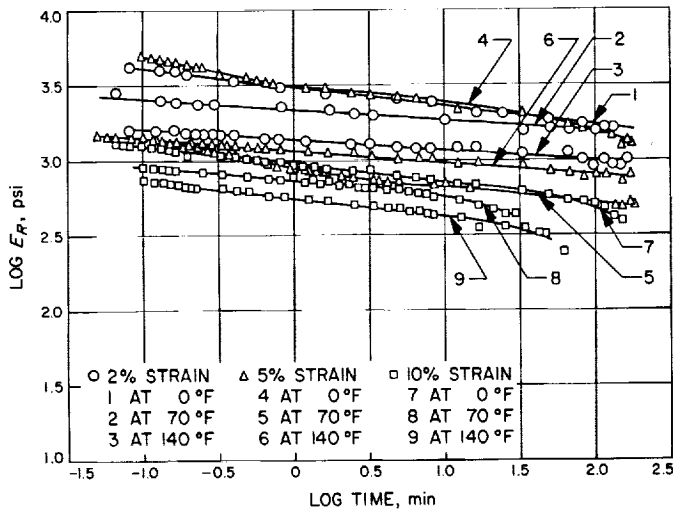
Fig. 21. Stress relaxation modulus as a function of time (uniaxial stress relaxation)

equal to ten times the rise time. Nonlinear effects are evident.

c. Creep compliance. The creep experiment consists of applying a constant load to a JANAF tensile specimen and observing the change of strain with time for various temperatures. Tests were performed at three temperatures (0, 70, 140°F) (−18, 21 and 60°C) and four loads (1.65, 2.66, 5.24, and 10 lb). An assumption was made that the stress relaxation modulus was equal to the reciprocal of the creep compliance. Hence, Fig. 22 was generated.
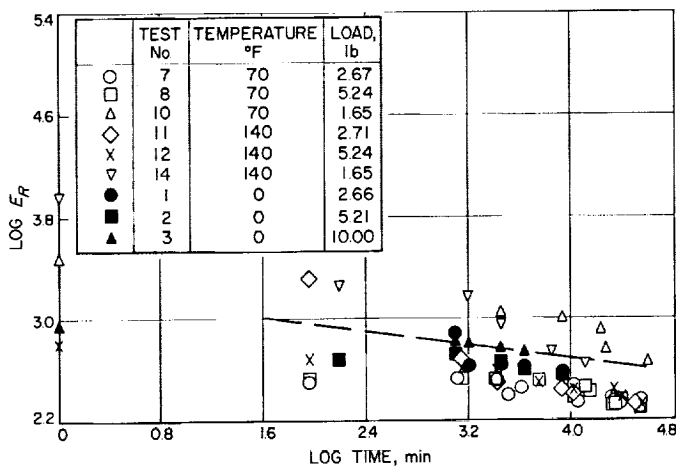
Fig. 22. Stress relaxation modulus as a function of time (uniaxial creep)

Discrepancy among a, b, and c. The discrepancy between a stress relaxation modulus obtained from constant strain rate, relaxation, and creep data is shown in Fig. 23. If the constant strain rate data are taken as the reference, one might infer that the master stress relaxation modulus approximates the actual viscoelastic response of solid propellant within a half a decade. To what extent this discrepancy is due to experimental accuracy remains to be seen. Whether this approximation of the true behavior of solid propellant can be used to obtain engineering stress analysis solutions remains to be seen. In fairness to the linear viscoelastic theory it should be noted that some of the mathematical requirements could not be experimentally duplicated. Note that the stress relaxation and creep data were reduced in terms of engineering stress because of the difficulty of measuring lateral strain, for small values of longitudinal strain, while the constant strain rate data were calculated in terms of true stress. On the other hand, it is experimentally recognized that the solid propellant is not linearly viscoelastic.

A second method was employed to obtain correlation support. Data were obtained from a torsion apparatus (Ref. 53) which allows a measurement of a shear relaxation modulus as a function of time and temperature. This experiment has the undesirable feature that creep and relaxation effects are not separated out of the response. However, the deviation from a true relaxation modulus
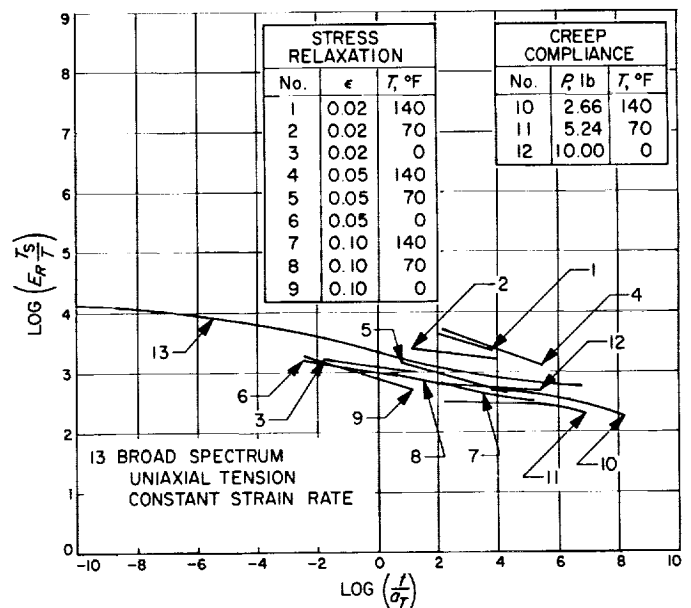
Fig. 23. Stress relaxation modulus obtained by three different tests

may be no greater than that obtained from the constant strain rate experiment. In any event, the familiar 10-sec "shear" relaxation modulus (equal to ⅓ the stress relaxation modulus for small strains and a Poisson's ratio of ½) is shown in Fig. 24 as a function of temperature. Also shown is the predicted shear relaxation modulus employing the information from Fig. 19. The correlation is considered *good* if consideration is given to the experimental difficulties in calculating the modulus from Fig. 18.

d. Bulk modulus. The bulk modulus is, loosely speaking, the ratio of the applied pressure to the relative decrease in volume. If the solid propellant were linearly viscoelastic, it would be desirable to see if the shear modulus is considerably greater than the bulk modulus (e.g., two decades). If this were true, simple extension data would give the same information as simple shear data, e.g., as in Fig. 24.

Fig. 25 illustrates a pressure chamber with a floating linear expansion measuring transducer as described in Ref. 52. A rough measurement for initial "bulk modulus" (600 psi rise in about ½ sec) was about 0.5×10^6 psi. This approximate value was obtained for various pressure rates (200 to 600 psi/sec) and temperatures (140 to 70°F) (60 to 21°C). The "bulk modulus" measure was time-dependent. The significant observation is that the bulk modulus is much greater than the shear modulus.

e. Specific heat. A calorimeter (Ref. 54) was used to measure the mean specific heat of the JPL-processed TPH 3062 propellant (Fig. 26). The calorimetric fluid was distilled water. The temperature range investigated was between 50 and 160°F (10 and 71°C). The values shown in Fig. 26 are comparable to the various values reported (Ref. 55) by six different laboratories for specimens from the same batch of a Thiokol-processed, *Surveyor* TPH-3062.
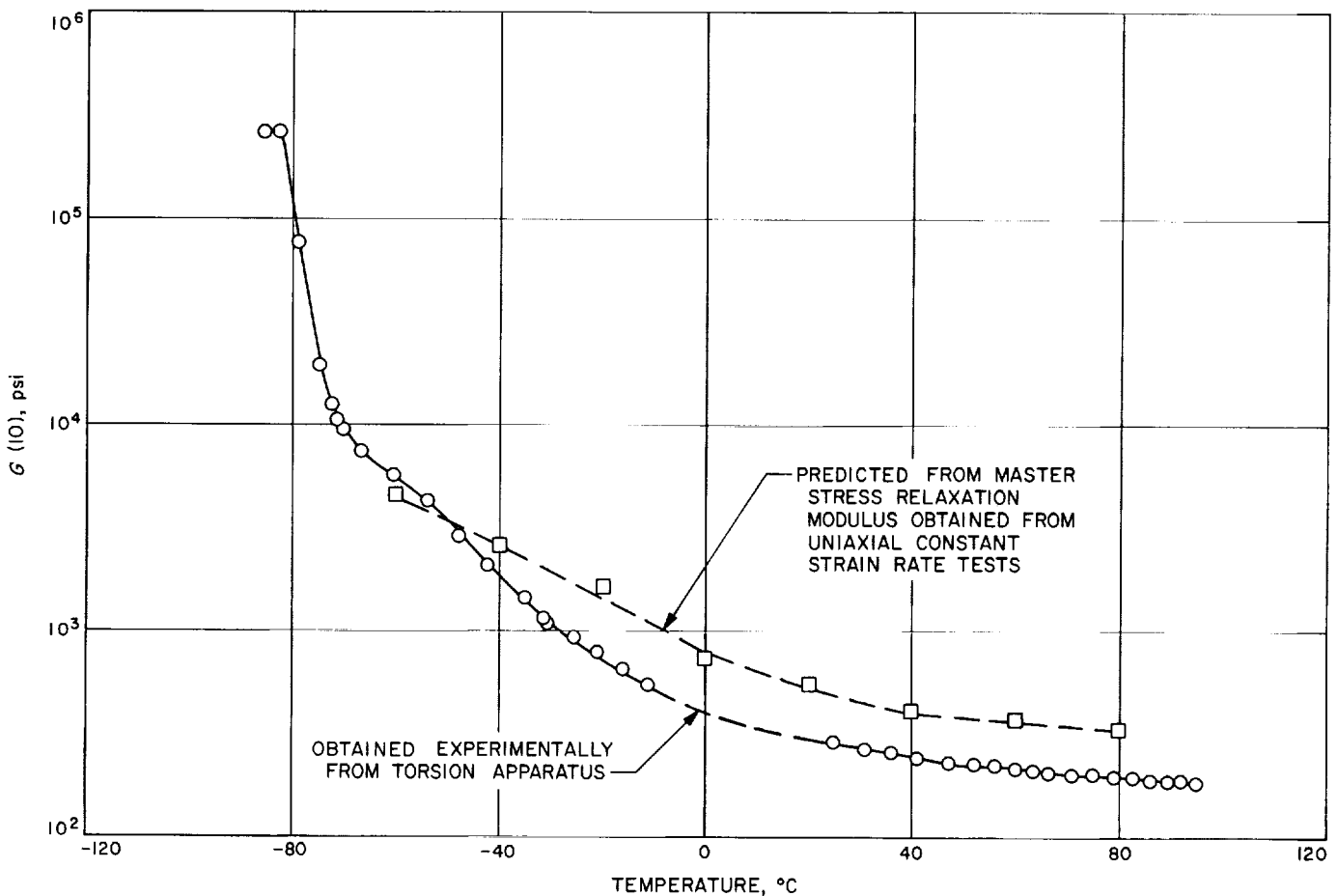
Fig. 24. Shear relaxation modulus (10 sec) as a function of temperature
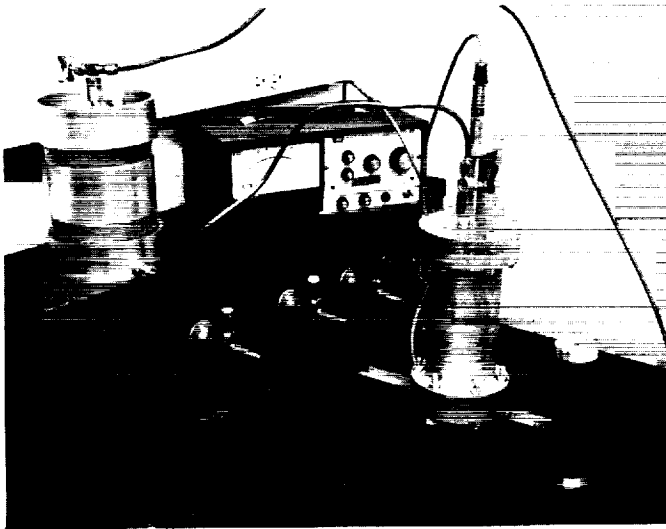
Fig. 25. Bulk modulus measuring apparatus
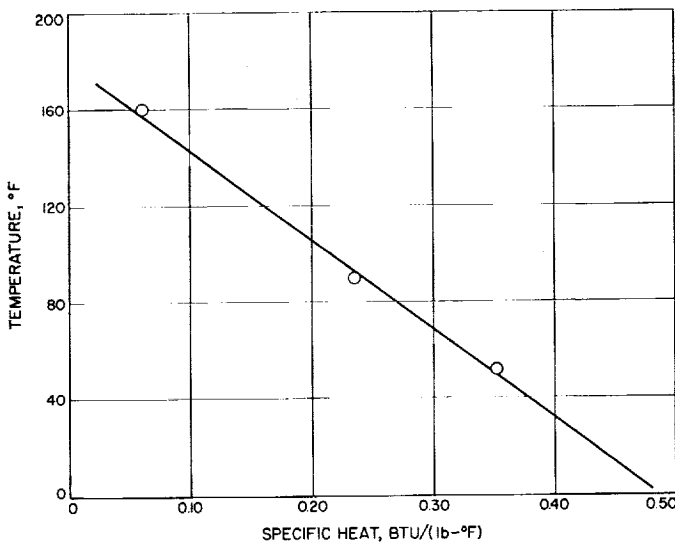
Fig. 26. Specific heat of JPL-processed TPH 3062 propellant as a function of temperature

f. Thermal conductivity. A Cenco-Fitch thermal conductivity apparatus was used to measure the thermal conductivity of sheets of propellant (¼ × 6 × 6 in.). The temperature range measured was from 40 to 212°F (4 to 100°C). The experimental results of thermal conductivity as a function of temperature are shown in Fig. 27. These results also compare favorably with the spectrum of values reported by six different laboratories for specimens from the same batch of a Thiokol-processed *Surveyor* TPH-3062 (Ref. 55).
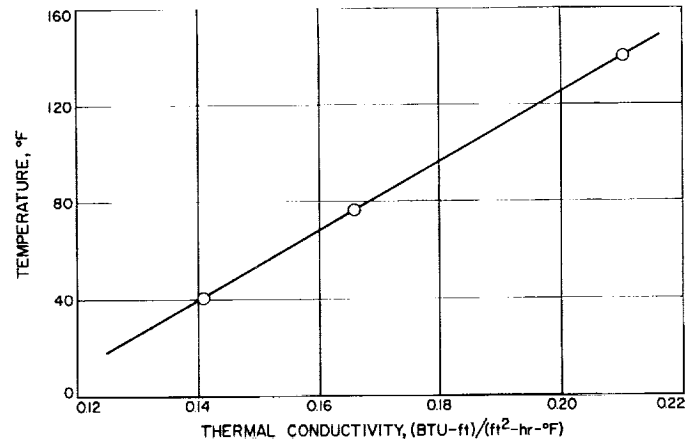
Fig. 27. Thermal conductivity of JPL-processed HT 3062 propellant as a function of temperature

g. Coefficient of thermal expansion. A volumetric dilatometer was used to measure the propellant volume change between the temperature range of −22 to 122°F (−30 to 50°C). The volume coefficient of thermal expansion is 20.9×10^{-5} in.^{3}/(in.^{3}-°F); i.e., a linear coefficient of thermal expansion of 6.9×10^{-5} in./in.-°F).

h. Glassy temperature. Propellant samples (0.064 × 0.243 × 1.312 in.) were tested with a Gehman Torsional Tester (Ref. 53). The glassy temperature transition was observed at −118°F (−83°C) (Fig. 24).

i. Density. The density of three 1-in. cube samples of propellant was measured by three methods. The first method consisted of differential weighing in Dow 200 silicone oil. The second method[1] consisted of direct weighing using measured linear dimensions. The third method employs a soil density apparatus (Ref. 56). All three methods measured a density of 0.063 lb/in^3.

j. Cure shrinkage. The change in volume of solid propellant from the time of casting to reasonable cure is defined as the cure shrinkage. A pycnometer was used to follow the volume change. The confining liquid was distilled water. The resolution of the pycnometer was about ±0.5%. There was no appreciable volume change in 180 hr.

k. Moisture absorption. Twelve specimens (½-in. square by 3¼-in. long) were weighed and placed into cans or bottles containing saturated salts at 40, 60, and 80°F (4, 15, and 27°C). The salts used for each temperature were sodium bromide (60, 60 and 59% RH), sodium iodine

["Standard Operating Procedures – Density of Propellants," JPL Sect. 381, October, 1963.]

(42, 41 and 38% RH), calcium bromide (21, 20%), and potassium acetate (21% RH at 80°F) (27°C) (Ref. 57). Each sample was reweighed each fifth day and compared to the initial weighing. After thirty days there was less than 1% change in weight in any sample.

4. Task 4 Progress

The experimental progress to obtain stress, strain and temperature data from the ¼-scale *Surveyor* (see Task 1) is proceeding in a satisfactory manner.

F. Polymer Permeation Constants From Diffusion Data on Polymeric Foams

E. F. Cuddihy and J. Moacanin

The influence on the outgassing rates of closed-celled, rigid foams from such factors as foam density, polymer type, and temperature have been reported in *SPS 37-34* and *37-35*, Vol. IV. In general, it was found experimentally that: (1) for a particular foam material the outgassing rate was proportional to the inverse of the foam density, (2) the outgassing rate substantially increased with increasing temperature, and (3) the outgassing rate was proportional to the permeation constant of the polymer material comprising the foam.

There was also developed a mathematical relationship between the diffusion constant of the foam and the foam density and the permeation constant of the bulk polymer. This equation is

$$D = P_e \left(\frac{RT}{M}\right)\left(\frac{\rho_0}{\rho}\right)$$
$$\left[\frac{1}{(1 - \rho/\rho_0)^{1/3}} + 1 + (1 - \rho/\rho_0)^{1/3}\right] \quad (1)$$

where

D = diffusion constant

P_e = permeation constant

R = gas constant

M = molecular weight of gas

T = absolute temperature

ρ_0 = density of bulk polymer

ρ = density of foam

This equation can be simplified by noting that the bracketed term varies slowly between 3 and 3.05 as the ratio ρ/ρ_0 varies between 0 and 0.5. Since most foams of practical importance have a ρ/ρ_0 ratio considerably less than 0.5, the term can be considered identically equal to 3, which results in the following equation:

$$D = 3P_e \left(\frac{RT}{M}\right)(\rho_0/\rho) \quad (2)$$

In its simplest terms, this equation predicts for a given foam material that the diffusion constant would be proportional to the inverse of the foam density, a prediction which was supported by experimental evidence.

Additionally, this same equation can be used to calculate the value of the permeation constant of the bulk polymer by employing the experimentally obtained values of the diffusion constant and the foam density. The numerical value so derived could be compared with literature values as a further test of the validity of the equation.

For rigid polyurethane foams filled with CO_2 gas, diffusion constants of 1.24 and 1.76×10^{-6} cm²/sec have been found (*SPS 37-34*, Vol. IV) for foam densities of 6.4 and 4.5 lb/ft³, respectively. Using these numbers, a value for the permeation constant of 5.1×10^{-9} (cc STP)mm/sec/cm²/cm Hg was calculated from Eq. (2). The polyurethane foam used in the study was purchased from the Emerson and Cummings Corp., and is advertised to have excellent high-temperature properties. This fact would imply the possibility of a high aromatic content in the polyurethane structure. Also, the material is a rigid, highly crosslinked plastic.

No data were found in the literature for polyurethane materials having these particular properties and, in fact, the few materials reported in the literature were all flexible. However, it is still possible to arrive at some meaningful conclusions.

Permeation constants of 40×10^{-9}, 18×10^{-9}, 14×10^{-9}, 7.2×10^{-9} (cc STP)mm/sec/cm²/cm Hg were reported, respectively, for the polyurethane materials designated as Adiprene-L (Ref. 58), Vulcaprene (Ref. 59), Estane (Ref. 1), and Urethane PC-6 (Ref. 58). The first three materials are rubbery, whereas Urethane PC-6 (which is used as a coating material) would be more nearly described as having the properties of a stiff but flexible leather. This property can be achieved by increasing the crosslinking density of polyurethanes. It has been reported (Ref. 60) that permeation constants are inversely proportional to crosslinking density, and it is therefore not

surprising that Urethane PC-6 would have the lowest value. However, the polyurethane comprising the foam of this study is rigid, and therefore even more cross-linked; it is observed that its permeation constant is both lower than, and in the same order of magnitude as, that for the Urethane PC-6.

These considerations certainly suggest that the calculated permeation constant for this present material is not unreasonable and is in the proper direction and magnitude when its value is compared with literature values. This result definitely adds credence to the validity of the expression derived to relate the diffusion constant of a foam to the foam's density and the permeation constant of the bulk polymer.

[a]UNSATURATION CAN APPEAR BOTH ALONG THE BACKBONE AND IN A PENDANT GROUP. THE RELATIVE AMOUNTS CAN VARY, DEPENDING UPON THE METHOD OF POLYMERIZATION.

Fig. 28. Chemical structure of PBAA, MAPO, and the crosslinked elastomer

G. Development of Sterilizable Elastomeric Foams

E. F. Cuddihy and J. Moacanin

1. Introduction

A study was made of the feasibility of preparing elastomeric foams which are stable at high temperatures ($\simeq 150°$C), in support of the effort to develop sterilizable solid propellants for *Voyager* capsule applications. It is anticipated that elastomeric foams could be used as propellant liner materials capable of relieving thermal stresses during sterilization heat cycles.

The procedure for preparing a usable elastomeric foam liner is a four-part problem consisting of: (1) the choice of an elastomer, (2) the technique of preparing a foam from this material, (3) determining the optimum cure time and formulations parameters to achieve the best properties of the elastomeric material, and (4) a study of the effects on the elastomeric foam of a high-temperature heat sterilization.

2. Liner Material

The elastomers used for this study were prepared from a copolymer of butadiene and acrylic acid (PBAA) by crosslinking the pendant carboxyl groups with MAPO, a commercial tri-imine (Fig. 28). The PBAA, purchased from the American Synthetics Corp., is a black viscous liquid which is used as is. The PBAA employed in this study (Lot 607) contained 1.06 wt% of the antioxidant

phenyl-β-naphthylamine (PBNA) by JPL analysis. MAPO (Lot 62), purchased from the Interchemical Corp., Carlstadt, New Jersey, is a straw-colored low-viscosity liquid which was also used as is, and was considered for formulation purposes as 100% pure.

The crosslinking occurs by reaction between the carboxyl groups on the PBAA and the imine groups of the MAPO. Therefore, only a single formulation parameter is needed and is defined as the ratio of the equivalents of imine groups to the equivalents of carboxyl groups. The equivalent weight of the MAPO is 71.67 and that of the PBAA used in this study is 1562.7 (determined by acid analysis).

3. Foam Preparation

There are three methods available for generating a gas within a viscous polymer to produce a foam. One is the generation of a gas from the curing reaction (or by reacting from another special agent with the chemicals used in the curing reaction to generate a gas). Another is to froth a viscous polymer liquid by blowing into it a gas under pressure. Finally, a third method, and the one chosen here, is to disperse into the viscous polymer system a solid (called blowing agents) which at elevated temperatures will decompose to yield a gas. Blowing agents having a wide range of decomposition temperatures are commercially available, and five such solids are described in Table 5.

Table 5. Structure and decomposition data on commercial organic foaming agents

Chemical description	Structure	Trade Name	Decomposition range in plastics, °C	Gas yield, ml (STP)/g	Manufacturer
Azobisformamide	H_2N—CO—N=N—CO—NH$_2$	Kempore	160-200	220	National Polychemicals, Inc.
Azobisisobutyronitrile	NC—C(CH$_3$)(CH$_3$)—N=N—C(CH$_3$)(CH$_3$)—CN	VAZO	90-115	130	DuPont
Diazoaminobenzene	⟨ ⟩—NH—N=N—⟨ ⟩	DAB	95-100	115	National Aniline Div., Allied Chemical Corp.
N,N'-Dimethyl-N,N'-dinitrosoterephthalamide	H_3—C—N(NO)—OC—⟨ ⟩—CO—N(NO)—CH$_3$	Nitrosan	90-105	126	DuPont
N,N'-Dinitrosopentamethylenetetramine	(ON—N, CH$_2$)(CH$_2$, N—NO) ring structure with CH$_2$, N, CH$_2$	Opex 40	130-190	220	National Polychemicals, Inc.

To produce a foam, one has to determine the optimum temperature at which the decomposition of the blowing agent will occur simultaneously with the gelation of the prepolymer. If the temperature is not high enough, then the gas may escape before the polymer gels sufficiently. If, on the other hand, the temperature is too high, the polymer may gel before an adequate foam structure can be formed.

To find the optimum conditions for the PBAA-MAPO system, first a series of mixtures was prepared, using for the formulation parameter a ratio of 1.00:1, and 2 g of blowing agent per 100 g of PBAA; each of the five blowing agents listed in Table 5 were used. Separate specimens prepared from each mixture were then heated at fixed temperatures between 100 and 140°C until either the decomposition was complete or gelation had occurred. By this procedure, it was found that the best foam structure was obtained at 135°C, using VAZO blowing agent. This foam is shown on the left hand side of Fig. 29. Also, adequate foam structures were obtained for these conditions when the formulation parameter was increased from a ratio of 1.00:1 to 1.15:1.

With Kempore and Opex 40 the polymer gelled before they decomposed sufficiently, whereas DAB dissolved in the polymer system. This seemed to result in a poor gas yield, and Nitrosan decomposed before the system gelled.

It should be pointed out that the optimum mechanical properties of the elastomer are not achieved at the gel point, and further curing is ordinarily required. Since the optimum properties of the foam can be achieved only when the optimum properties of the elastomer are achieved, it becomes necessary to investigate the properties of the PBAA-MAPO elastomer system.

4. Elastomer Properties

A cure temperature of 135°C was not used, as this study was initiated prior to the foam investigation. For the initial study, four preparations of PBAA-MAPO, having the formulation parameters of 1.00:1, 1.05:1, 1.10:1, and 1.15:1, were used. Curing was carried out to a maximum of 24 days at temperatures of 40, 70, 100, and 146°C. The extent of cure was followed by means of a Shore A Durometer. The results are shown in Tables 6 and 7.

The rate of the curing increased with increasing temperature. At 40°C cure was not completed in 24 days, whereas at 146°C maximum cure was achieved in about 1 hr. However, the highest Shore A hardness (S_A), was obtained at 70°C, with slightly lower values observed at 100°C, and significantly reduced values at 146°C. Furthermore, with increasing time, S_A remained nearly constant at 70°C, but exhibited significant decreases at both 100 and 146°C. In terms of the formulation parameters, maximum properties were achieved at the higher ratios of 1.10:1 and 1.15:1 (high MAPO content).

Undercured or optimally cured elastomers were characterized by a clear golden-colored appearance, whereas

BEFORE STERILIZATION   AFTER STERILIZATION

0   1   2   3   4
INCHES
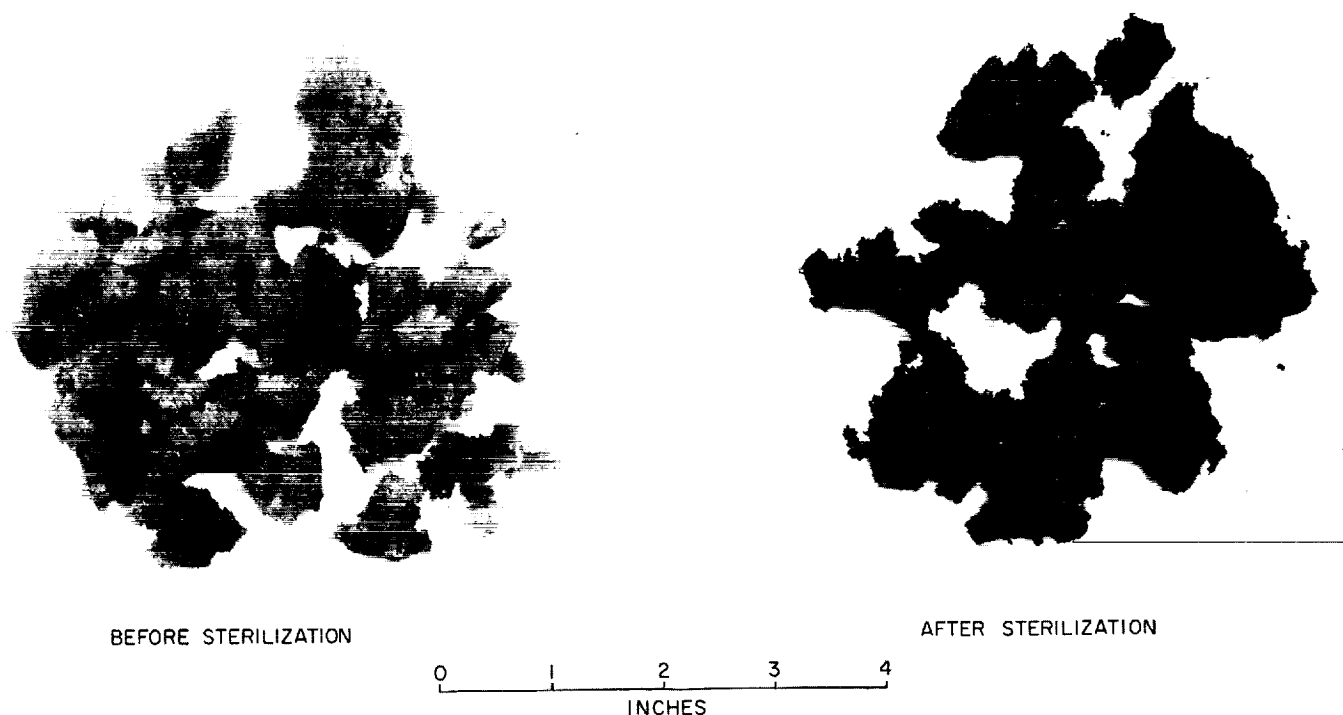
Fig. 29. Foam structure prepared at 135°C from a 1.00:1 PBAA-MAPO elastomer and VAZO blowing agent, before and after a three-cycle sterilization

the overcured materials exhibited darkening, which increased with increasing postcuring. In fact, the color appearance of the elastomers can be used as a criterion of the state of cure. Materials which were being cured at 146°C (295°F) started to darken almost from the beginning. This suggests extensive degradation reactions, which may explain the low S_A values observed for the elastomers cured at this temperature. It is encouraging to recall that the foams prepared at 135°C over a 20-min period were golden-colored in appearance.

Fig. 30 shows the modulus versus temperature curves for the 1.10:1 sample, cured 184 hr at 70°C, which had the highest S_A. The modulus curve was obtained using a Gehman torsion tester (SPS 37-18). The curve marked "A" is that for the first test, and because the sample darkened after heating to 200°C, it was retested, resulting in Curve C.

These curves show that the glass transition temperature T_g of the elastomer is near −90°C, and that no marked change in T_g occurred upon heating the sample to 200°C. Furthermore, for both tests, the modulus values were nearly identical over the temperature interval from 0 to 200°C. Modulus curves obtained for other overcured

(darkened samples) and undercured materials exhibited approximately the same behavior.

5. Sterilization

Currently, it is proposed that space vehicles be subjected to heat sterilization composed of three consecutive 36-hr exposures at 295°F. The PBAA-MAPO elastomer system was selected for consideration because of its asserted capability of withstanding these conditions. Selected test specimens of materials characterized in the top half of Table 6 were subjected to the three-cycle sterilization, and S_A values were measured after each cycle. It was noted that the surface of the samples darkened, whereas below the surface, the elastomers retained their characteristic golden color. This phenomenon is illustrated in Fig. 31, which shows the cross-section of three samples. Sample A was unsterilized, and Samples B and C were subjected to one and three cycles, respectively.

S_A measurements were taken after each cycle only on the exposed darkened surface. After the third cycle, the samples were cut and S_A values taken on the clear cross-sectional surface. These results are tabulated in the lower half of Table 6.

Table 6. Effect on the Shore A hardness of formulation parameter, cure time,
cure temperature and sterilization for a PBAA-MAPO elastomer

Sample number	Cure time, hr	Formulation parameters											
		1.00:1			1:05:1			1.10:1			1.15:1		
		Cure temperature,											
		40°C	70°C	100°C	40°C	70°C	100°C	40°C	70°C	100°C	40°C	70°C	100°C
		Shore A hardness											
1	16	0	0	14	0	0	15	0	0	16	0	0	15
2	40	0	8	15	0	8	15	0	8	19	0	10	15
3	64	0	14	14	0	14	14	0	15	18	0	15	17
4	136	0	18	12	0	18	14	0	21	15	0	21	11
5	184	2	19	11	4	20	9	5	22	9	4	22	14
6	232	5	18	9	6	18	9	8	22	8	8	21	11
7	304	9	19	7	12	21	9	12	21	8	12	21	9
8	376	10	17	7	11	14	5	12	20	9	13	19	6
9	472	10	18	3	14	16	2	13	18	7	13	19	5
10	568	15	18	2	18	18	2	18	18	3	17	18	5
Sample number	Cycle number	Heat sterilization[a]											
1	1			8			2			3			6
	2			11			6			7			5
	3			14			14			23			18
	Inside surface			11			6			12			11
2	1			11			6			2			3
	2			14			14			8			12
	3			16			15			10			11
	Inside surface			11			9			6			8
4	1		8			6			5			5	
	2		9			8			7			8	
	3		16			19			21			22	
	Inside surface		8			7			7			9	
5	1		11			9			6			9	
	2		13			11			12			18	
	3		17			13			13			16	
	Inside surface		10			11			8			6	
10	1	8	9		6	8		5	2		4	4	
	2	12	10		10	9		8	5		8	5	
	3	16	16		20	18		21	13		26	21	
	Inside surface	11	5		13	10		8	7		9	6	

[a] Samples subjected to three consecutive cycles of 40 hr at 295°F under dry N_2. S_A values were obtained on the exposed surfaces after each cycle. After the third cycle, the samples were cut and an S_A measurement made on the inside material.
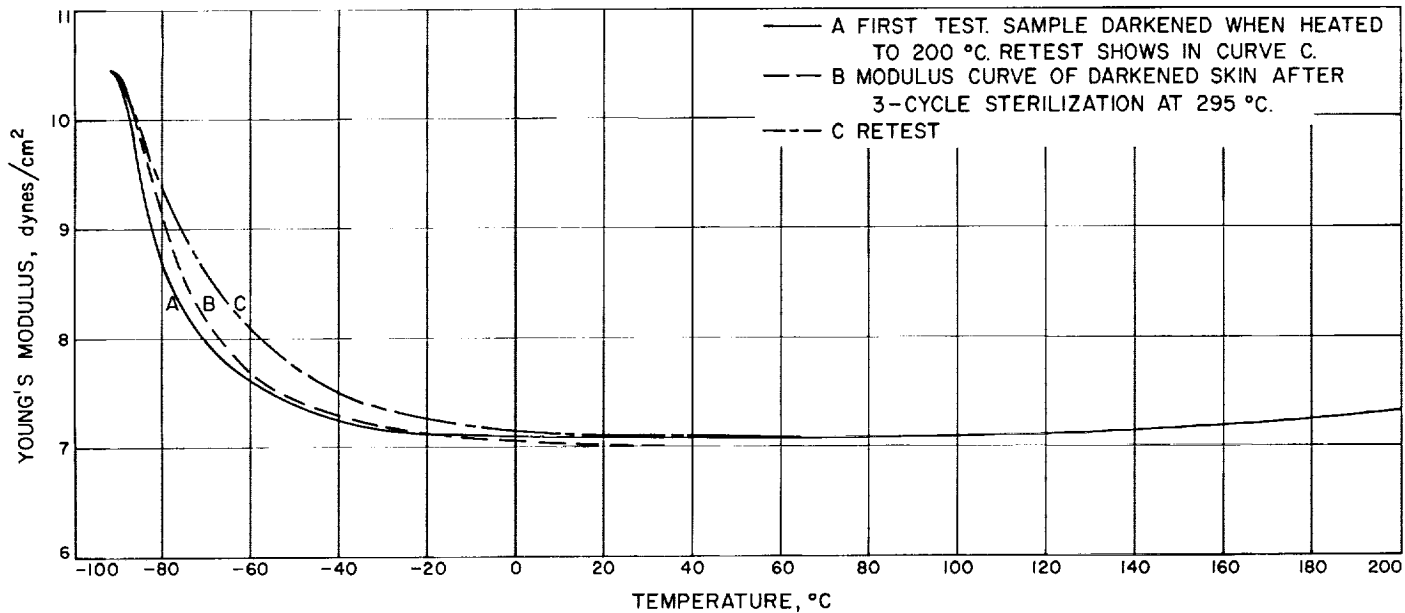
Fig. 30. Modulus curves of a 1.10:1 PBAA-MAPO elastomer cured 184 hr at 70°C

Table 7. Effect of cure at 295°F on the Shore A
hardness of PBAA-MAPO elastomers

Formulation parameters					
Sample number	Cure time, min	1.00:1	1.05:1	1.10:1	1.15:1
1	15	0	0	0	0
2	30	0	0	2	5
3	45	0	6	8	7
4	60	1	8	8	6
5	180	0	7	8	5
6	300	0	4	4	2
7	420	0	3	4	2
8	510	0	4	4	2

For all of the materials, the S_A of the darkened surface first decreased, and then increased with subsequent cycles, showing a skin-hardening effect. In all cases the S_A values for the interior were lower than the initial value as well as that for the darkened surface.

These results also demonstrate that generally better properties were retained after sterilization for those elastomers prepared from lower MAPO concentrations, and that the more undercured materials have exhibited a greater increase and an ultimately higher S_A value for the surface. For more extensively cured materials, the skin-hardening effect was less pronounced.
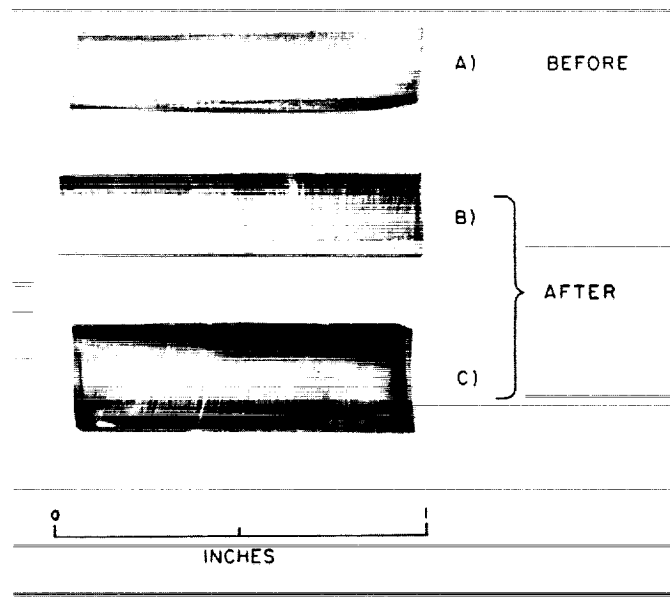
Fig. 31. Surface darkening of cured elastomer resulting from exposure to high-temperature sterilization

Since the foam consists of a maze of thin membranes, it is reasonable to expect that during sterilization these membranes will undergo extensive darkening and skin hardening. The extensive darkening of a foam after a three-cycle sterilization is clearly shown on the right

side of Fig. 29. It is to be noted that this material retained its foam structure after sterilization.

To obtain some idea of the properties of the sterilized elastomer which comprises the foam, the darkened skin section of a solid sterilized elastomer was removed and its modulus curves obtained. The results are shown in Fig. 30, Curve B, where it is seen that the low-temperature properties are basically unchanged, and that it has a lower rubbery modulus at room temperature which is in agreement with its reduced S_A value.

6. Summary and Future Direction

It has been found that elastomers prepared from PBAA and MAPO could withstand the high-temperature sterilization treatment with a reasonable retention of properties. Generally, higher properties are retained as the concentration of MAPO is lowered. During sterilization, these materials are characterized by a hardening and darkening of the exposed surface which gradually advances inward with increasing exposure times. Based on modulus measurements, the glass transition temperature of the materials is near $-90°C$ and appears to be relatively insensitive to the influence of sterilization. Furthermore, no basic differences in modulus behavior were noted, except that which could be expected from a decreasing S_A hardness.

Excellent foam structures were prepared from these elastomers by employing the blowing agent VAZO. Preparative conditions involved exposure of the system for 20 to 30 min at $135°C$, during which the VAZO decomposed and the elastomer gelled sufficiently to top the developed foam structure. No loss of foam structure was noted during sterilization, but the foam became black.

An undesirable feature of the PBAA-MAPO elastomers is their poor tear properties. Since the properties of a foam cannot be better than those of the starting elastomer, the emphasis of future development will be directed toward improved tear properties.

References

1. Jones, A. S., Stacey, M., and Watson, B. E., *Journal of the Chemical Society*, Vol. 2454, 1957.

2. Burton, K., and Petersen, G. B., *Biochemical Journal*, Vol. 75, p. 17, 1960.

3. Shapiro, H. S., and Chargaff, E., *Biochimica et Biophysica Acta*, Vol. 26, p. 608, 1957; Vol. 39, p. 62, 1960.

4. Josse, J., Kaiser, A. D., and Kornberg, A., *Journal of Biological Chemistry*, Vol. 236, p. 864, 1961.

5. Swartz, M. N., Trautner, T .A., and Kornberg, A., *Journal of Biological Chemistry*, Vol. 237, p. 1961, 1962.

6. Simha, R., and Zimmerman, J. M., *Journal of Polymer Science*, Vol. 42, p. 309, 1960; Vol. 51, p. S39, 1961.

7. Simha, R., and Zimmerman, J. M., *Journal of Theoretical Biology*, Vol. 2, p. 87, 1962.

8. Burton, K., *Biochemical Journal*, Vol. 77, p. 547, 1960.

9. Petersen, G. B., *Biochemical Journal*, Vol. 87, p. 495, 1963.

10. Spencer, J. H., and Chargaff, E., *Biochimica et Biophysica Acta*, Vol. 68, p. 19, 1963.

11. Shapiro, H. S., and Chargaff, E., *Biochimica et Biophysica Acta*, Vol. 76, p. 1, 1963.

References (Cont'd)

12. Burton, K., Lunt, M. R., Petersen, G. B., and Siebke, J. C., *Symposium on Quantitative Biology,* Vol. 28, p. 27, 1963.

13. Hermann, A. M., and Ham, J. S., *Review of Scientific Instruments,* Vol. 36, p. 1553, 1965.

14. Many, A., Simhony, M., Weisz, S. Z., and Teucher, Y., *Journal of Chemical Physics,* Vol. 25, p. 721, 1964.

15. Masuda, K., and Yamaguchi, J., *Journal of the Physical Society, Japan,* Vol. 20, p. 1340, 1965.

16. Grishina, A. D., and Vannikov, A. V., *Doklady Akademii Nauk, USSR,* Vol. 156, p. 647, 1964.

17. Pohl, H. A., and Chartoff, R. P., *Journal of Polymer Science, Part A.,* Vol. 2, p. 2787, 1964.

18. Nechtschein, M., *Journal of Polymer Science, Part C.,* Vol. 4, p. 1367, 1963.

19. Ottenberg, A., Hoffman, C. J., and Osiecki, J., *Journal of Chemical Physics,* Vol. 38, p. 1898, 1963.

20. Kommandeur, J., and Hall, F. R., *Journal of Chemical Physics,* Vol. 34, p. 129, 1961.

21. Singer, L. S., and Kommandeur, J., *Journal of Chemical Physics,* Vol. 34, p. 133, 1961.

22. Tobin, M. C., and Spitzer, D. P., *Journal of Chemical Physics,* Vol. 42, p. 3652, 1965.

23. Semenov, N. N., *Journal of Polymer Science,* Vol. 55, p. 563, 1961.

24. Elliot, R. J., *Physical Review,* Vol. 96, p. 266, 1954.

25. Singer, L. S., Kickuchi, C., *Journal Chemical Physics,* Vol. 23, p. 1738, 1955.

26. Swarup, P., and Misra, B. N., *Zeitschrift für Physik,* Vol. 159, p. 384, 1960.

27. Weidner, R. T., and Whitmer, C. A., *Physical Review,* Vol. 91, p. 1279, 1953.

28. Goldsborough, J. P., Mandell, M., and Pake, G. E., *Physical Review Letters,* Vol. 4, p. 13, 1960.

29. Roest, R., and Poulis, N. J., *Physica,* Vol. 25, p. 1253, 1959.

30. van Gerven, L. A., van Itterbeck, and de Wolf, E., *Journal de Physique et le Radium,* Vol. 18, p. 417, 1957.

31. van Roggen, A., van Roggen, L., and Gordy, W., *Physical Review,* Vol. 105, p. 50, 1957.

32. Kovacic, P., and Wu, C., *Journal of Polymer Science,* Vol. 47, pp. 45–54, 1960.

33. Packman, D. I., *Chemical Communications* (London), p. 207, 1965.

34. Usher, D. A., Gougoutas, J. Z., and Woodward, R. B., *Analytical Chemistry,* Vol. 37, pp. 330–332, 1965.

References (Cont'd)

35. Plyler, E. K., and Ball, J. J., *Journal of the Optical Society of America,* Vol. 38, p. 988, 1948.

36. Smith, T. L., *Journal of Polymer Science,* Vol. 32, p. 99, 1958.

37. Smith, T. L., *Proceedings of the Fourth International Congress on Rheology,* Part 2, E. H. Lee, Ed., Interscience Publishers, New York, N.Y., 1965, p. 525.

38. Landel, R. F., and Fedors, R. F., Ibid, p. 543.

39. Smith, T. L., *Journal of Applied Physics,* Vol. 35, p. 27, 1964.

40. Epstein, L. M., and Smith, R. P., *Transactions of the Society of Rheology,* Vol. 2, p. 219, 1958.

41. Landel, R. F., and Fedors, R. F., "Rupture of Amorphous, Unfilled Polymer," Chap. III B in *Fracture Processes in Polymeric Solids, Phenomena and Theory,* B. Rosen, Ed., Interscience Publishers, New York, N.Y., 1964.

42. Smith, T. L., *Journal of Polymer Science,* Vol. 1, p. 3597, 1963.

43. Halpin, J. C., *Journal of Applied Physics,* Vol. 35, p. 3133, 1964.

44. Landel, R. F., and Stedry, P., *Journal of Applied Physics,* Vol. 31, p. 1885, 1960.

45. Treloar, L. R. G., *The Physics of Rubber Elasticity,* Chap. 6, Clarendon Press, Oxford University, London, 1958.

46. Lal, J., and McGrath, J. E., *Rubber Chemistry and Technology,* Vol. 36, p. 1159, 1963.

47. Knauss, W., "Rupture Phenomena in Viscoelastic Materials," Doctoral Thesis, California Institute of Technology, 1964.

48. Kaelble, D., *Journal of Polymer Science,* Vol. IX, p. 1213, 1965.

49. Fedors, R. F., and Landel, R. F., *Transactions of the Society of Rheology,* Vol. 9, p. 195, 1965.

50. San Miguel, A., "Structural Integrity of Solid Retrorockets," SPS No. 37-35, Vol. IV, pp. 105–115, Jet Propulsion Laboratory, Pasadena, California, October 31, 1965.

51. San Miguel, A., Silver, R. H., and Duran, E. N., "On the Practical Application of a Miniature Stress Rosette to Solid Propellant Grain Design," Bulletin of the Fourth Meeting of ICRPG Working Group on Mechanical Behavior, October 15, 1965, SPIA, Johns Hopkins University, Applied Physics Laboratory, Silver Spring, Maryland.

52. ICRPG Solid Propellant Behavior Manual, Johns Hopkins University, Applied Physics Laboratory, Silver Spring, Maryland, Section 4, September, 1963.

53. "Standard Method of Measuring Low-Temperature Stiffening of Rubber and Rubber-like Materials by Means of a Torsional Wire Apparatus," ASTM Document 1053-61, American Society for Testing and Materials, Philadelphia, Pa., 1961.

54. "Standard Method of Test for Measuring Specific Heat of Thermal Insulation," ASTM Designation C351-61, American Society for Testing and Materials, Philadelphia, Pa., 1961.

55. *Surveyor Spacecraft Systems, Bimonthly Progress Summary,* SS-120, SSD 4365R, Hughes Aircraft Co., June–August, 1964.

References (Cont'd)

56. "Standard Method of Test for Shrinkage Factors of Soils," ASTM Document 427-61, American Society for Testing and Materials, Philadelphia, Pa., 1961.

57. Hedlin, C. P., and Trofimerkoff, F. N., "Relative Humidities Over Saturated Solutions of Nine Salts in the Temperature Range From 0° to 90°C," *Humidity and Moisture,* Vol. 3: Fundamentals and Standards, Wexler, A., and Wildback, W. A., Editors, Reinhold Corp., New York, N.Y., 1965, pp. 519–520.

58. Major, C. J., and Kammermeyer, K., *Modern Plastics,* Vol. 39, No. 11, p. 135, 1962.

59. van Amerongen, G. J., *Journal of Polymer Science,* Vol. 5, p. 307, 1950.

60. Rogers, C. E., "Permeability and Chemical Resistance," *Engineering Design for Plastics,* sponsored by the Society of Plastics Engineers, Inc., Eric Baer, Ed., Reinhold Publishing Corp., New York, N.Y., 1964.

XI. Liquid Propulsion

A. Resonant Combustion

R. M. Clayton and J. G. Sotter

1. Experimental Program

Typical chamber boundary pressure distributions associated with the fully developed oscillatory mode of combustion exhibited by two of the 20,000 lbf thrust engine configurations of interest to this program (Ref. 1) have been determined experimentally. Both engine configurations utilized the same 11-in.-D chamber/nozzle hardware and a steady-state chamber pressure of 300 psia; hence, the only significant differences between the configurations were the injectors and propellants used.

Results obtained for the engine configuration using Rocket Motor Injector Research (RMIR) Injector 5 with standard *Corporal* propellants are described in detail in Ref. 2 and are not further discussed here, except for comparison with the results obtained for the second engine. The latter engine, which incorporated RMIR Injector 7 and the N_2O_4 + 50/50 fuel (UDMH/N_2H_4) propellant system, is of primary interest here.

Injector 7 consists of 47 identical unlike-impinging doublets designed to satisfy the mixing uniformity criterion (Ref. 3) for N_2O_4 + N_2H_4 at the peak theoretical c^* mixture ratio of 1.20. The arrangement of the elements in the flat face plate is shown photographically in Fig. 1. This arrangement provides nearly uniform axial mass distribution throughout the major portion of the combustion volume, but does allow some local nonuniformity in the proximity of the chamber wall. Complete design specifications for the injector assembly are described in Ref. 4.

For the purpose of these experiments, 50/50 fuel was substituted for the N_2H_4, in order to circumvent the characteristically noisy combustion exhibited by the injector when operated with N_2H_4. This substitution requires operation of the injector at a mixture ratio, r_{unif}, of 1.27 to satisfy the uniform mixing criterion. Since N_2O_4 + 50/50 fuel achieves a peak theoretical c^* at a mixture ratio of approximately 1.6, operation at 1.27 is off-optimum chemically. Hence, prior to the commencement of the resonant combustion experiments, it was necessary to evaluate the steady-state combustion performance of the injector when operated with the latter propellants over this range of mixture ratios.

The results of the performance evaluation are summarized in Fig. 2, which presents the experimental values of c^*, I_s, C_F and combustion roughness (P_cRMS) as determined at the Edwards Test Station. For comparison
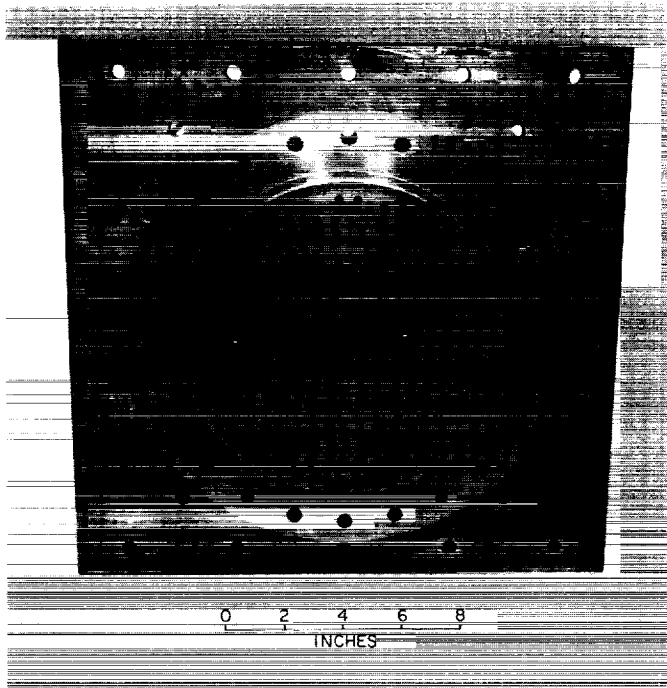
Fig. 1. Face of RMIR Injector 7 as used for Resonant Combustion Program

the pertinent theoretical performance curves are also shown. The values of experimental c^* and C_F are based on a one-dimensional isentropic stagnation pressure, which is computed from a static pressure measurement at the nozzle inlet. Examination of these results shows an essentially flat relative performance (c^*) of approximately 97% throughout the mixture ratio range of 1.27 to 2.0.

For the resonant combustion experiments a transition from the steady to the oscillatory mode was induced by a 13.5 grain high-explosive bomb pulse initiated approximately 400 msec after engine ignition. This firing procedure allowed the engine to attain essentially steady-state operation before the onset of the oscillatory mode and thus made possible the measurement of the operating conditions existing at the time of the bomb-induced combustion perturbation. The firings were terminated after approximately 100 msec of resonant combustion. The resulting total firing duration of 500 msec was within the thermal limitations of the uncooled Kistler pressure transducers used for high-response pressure measurements. Further details of the experimental procedures and instrumentation system are discussed in Ref. 2.

Portions of typical pressure-versus-time records obtained during fully developed resonant operation are

NOMINAL OPERATIONS CONDITIONS (RUNS B961–971)

p_c = 300 psia (BASED ON $p_{c_{eff}}$ FOR NOZZLE ENTRANCE)

p_a = 13.5 psia

THRUST = 20,000 lbf

A_c/A_t = 2.0

ϵ = 4.58

THEORETICAL DATA BASED ON:

$(\Delta H_f)_{N_2O_4}(\ell) = -4.676$ kcal/mole

$(\Delta H_f)_{N_2H_4}(\ell) = 12.050$ kcal/mole

$(\Delta H_f)_{UDMH}(\ell) = 12.724$ kcal/mole

— · — THEORETICAL EQUILIBRIUM
— — — THEORETICAL FROZEN
——— EXPERIMENTAL
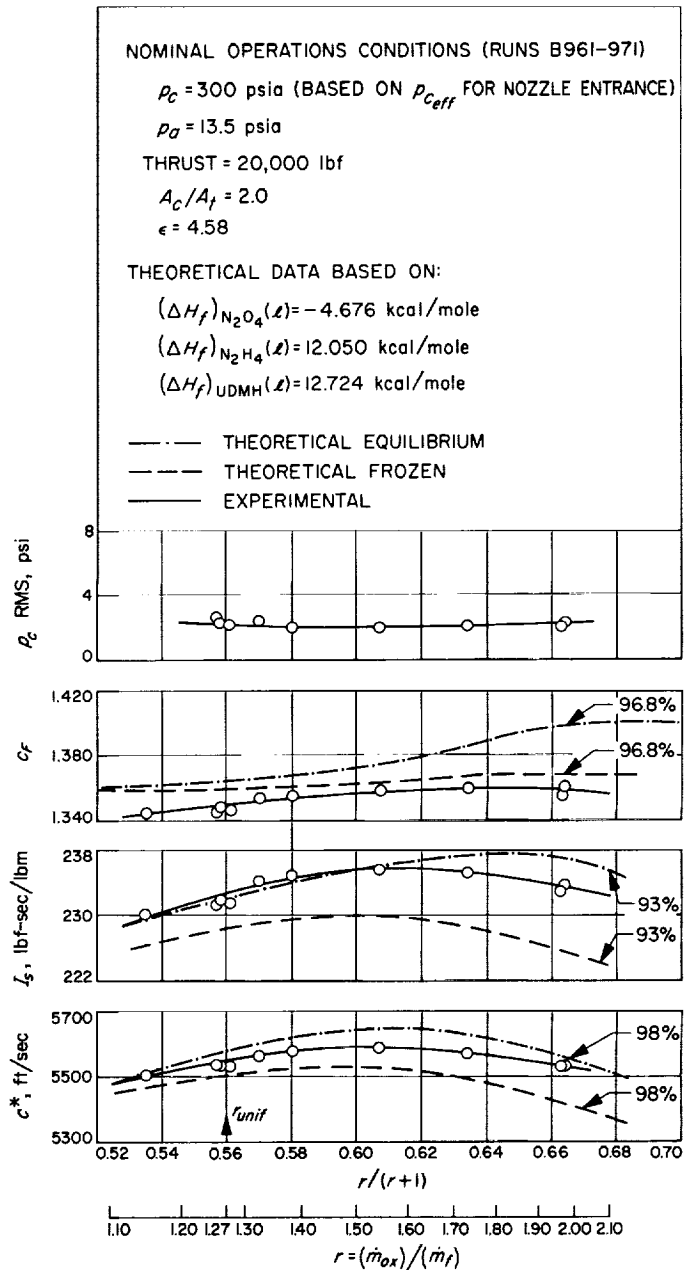
Fig. 2. Experimental performance, RMIR Injector 7 with N_2O_4 + 50/50 fuel (UDMH/N_2H_4)

illustrated on the upper right-hand side of Figs. 3 and 4, which show, respectively, a set of simultaneous measurements down the length of the chamber and across the injector radius. For comparison, analogous measurements obtained for Injector 5 are shown on the upper left-hand sides of the same figures. The geometric orientation of the several measurements is shown schematically in the lower portion of each figure. All traces are DC analog records of Kistler transducer outputs.
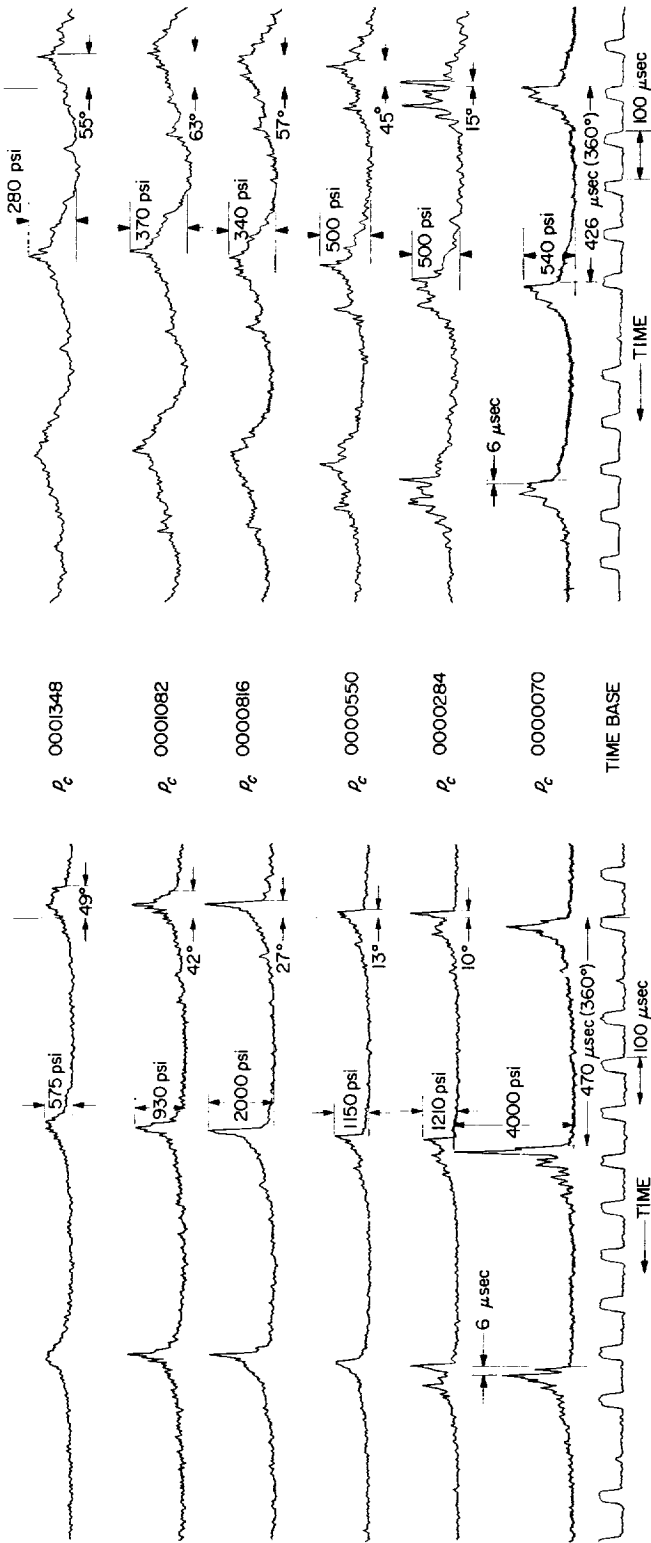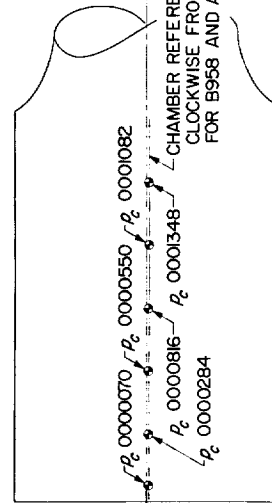
Fig. 3. Typical pressure distribution along the chamber wall versus time during resonant combustion, RMIR Injectors 5 and 7
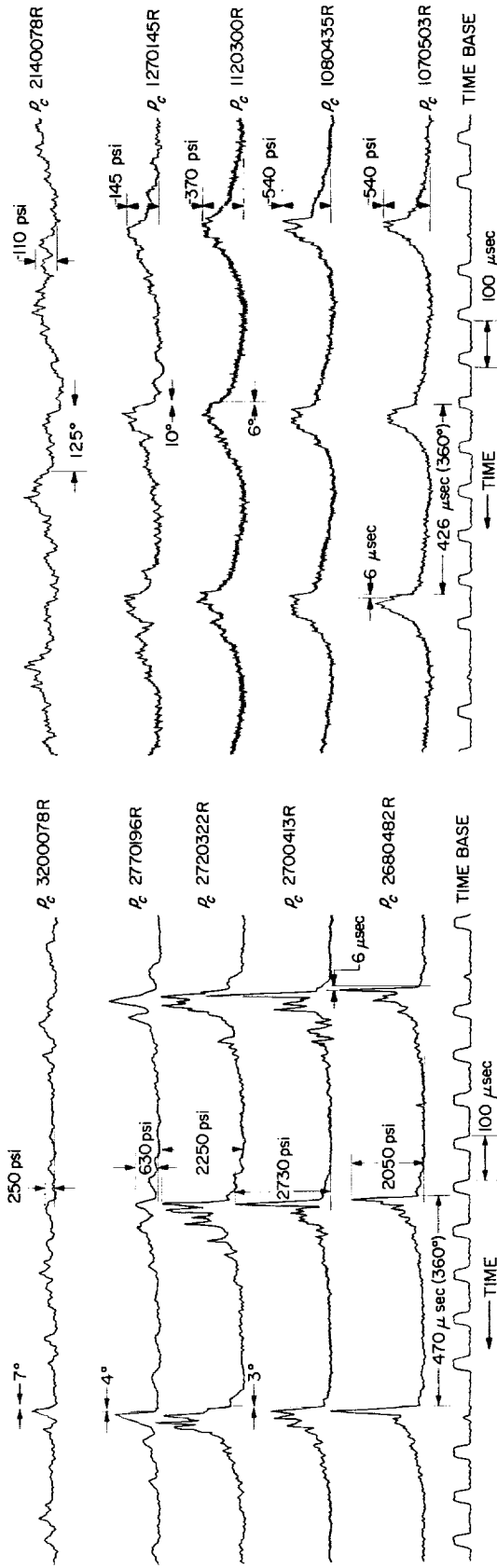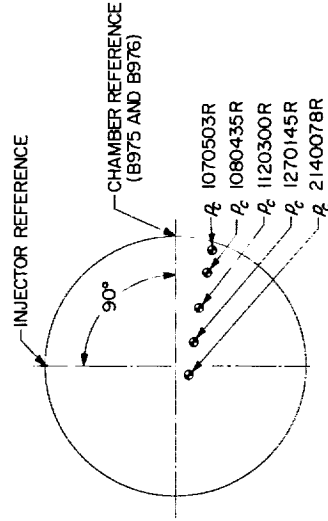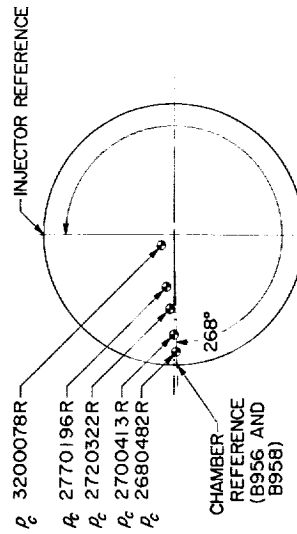
Fig. 4. Typical pressure distribution across the injector face versus time during resonant combustion, RMIR Injectors 5 and 7

The transient characteristics of the oscillatory combustion mode exhibited by both engine configurations are very similar. The description of these characteristics for Injector 7 may be summarized by essentially the same words which were used in Ref. 2 to describe the transient behavior of the disturbance for Injector 5. Briefly, this description is:

(1) A single disturbance rotates with supersonic velocity about the combustion chamber axis.

(2) The period of the combustion pressure oscillation, as observed at the chamber boundaries, is essentially identical at all locations on the face and wall for given engine operating conditions (i.e., during any particular engine firing).

(3) The peak pressure amplitude of the disturbance varies axially and radially, the greatest amplitudes being exhibited near the injector end of the chamber wall and in the outer half radius of the injector face.

(4) The waveform of the disturbance, particularly at locations near the injector end of the chamber, exhibit a steep-frontedness which is characteristic of shock waves.

As a result of this description and certain quantitative measurements of the shape of the wave-to-boundary intersection obtained for these engines, an artist's conception of this mode of combustion is presented in Fig. 5. The drawing is a pictorial representation of the "rotating
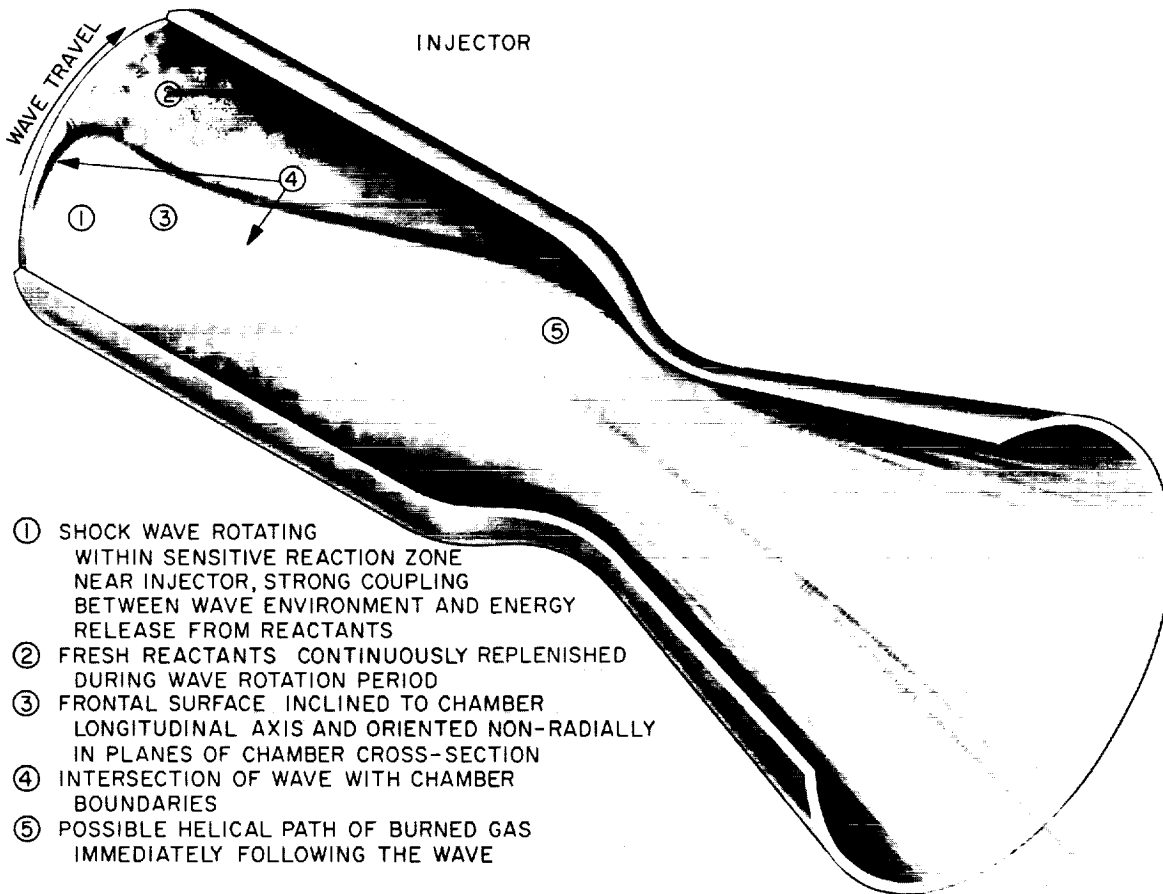
① SHOCK WAVE ROTATING
    WITHIN SENSITIVE REACTION ZONE
    NEAR INJECTOR, STRONG COUPLING
    BETWEEN WAVE ENVIRONMENT AND ENERGY
    RELEASE FROM REACTANTS
② FRESH REACTANTS CONTINUOUSLY REPLENISHED
    DURING WAVE ROTATION PERIOD
③ FRONTAL SURFACE INCLINED TO CHAMBER
    LONGITUDINAL AXIS AND ORIENTED NON-RADIALLY
    IN PLANES OF CHAMBER CROSS-SECTION
④ INTERSECTION OF WAVE WITH CHAMBER
    BOUNDARIES
⑤ POSSIBLE HELICAL PATH OF BURNED GAS
    IMMEDIATELY FOLLOWING THE WAVE

Fig. 5. Artist's conception of rotating detonation-like wave front

detonation-like wave" concept which is discussed in Refs. 1 and 2.

Figs. 3 and 4 exemplify disturbances which traveled in a clockwise direction (as viewed from the nozzle end of the chamber). Qualitatively identical pressure waves have been observed traveling in a counterclockwise direction in more recent experiments with Injector 7. The direction of travel appears to be related to the circumferential orientation of the radially positioned bomb with respect to the injected propellant distribution near the chamber walls. Although it is presently not well substantiated, there is some indication that the rotational period is somewhat longer ($\sim$1% at r_{unif}) for counterclockwise travel. Other differences have not been quantitized at this time. There is also evidence that (at least for the clockwise direction) the rotational period varies with gross mixture ratio and reaches a minimum value near the uniformity mixture ratio.

The instantaneous disturbance-to-chamber boundary intersection for Injector 7 as deduced from records similar to those depicted in Figs. 3 and 4 is shown in Fig. 6 where it is compared to the intersection observed for Injector 5. The data is presented for engine operation near the uniformity mixture ratio of 1.27. The intersection geometries shown are for clockwise wave rotation; however, the relative shape of the intersection does not appear to be substantially altered for rotation in the opposite direction. To be noted here is the general similarity in the shape of the intersections for both injectors. In particular it is noted that: (1) the nozzle end of the intersection leads the injector end, (2) the wave-to-injector face intersection is nonradial and (3) the intersection curves "steepen" considerably in the proximity of the face and wall boundary junction (i.e., the "corner" of the chamber). While no attempt will be made here to discuss the quantitative dissimilarities between the disturbance-to-boundary intersections for the two injectors, it is believed that these differences are the result of the particular distribution of energy and mass sources available to the pressure wave in the respective combustion systems.

The pressure amplitude distribution on the boundary of the chamber is shown in Fig. 7. Again the results are compared to those observed for Injector 5. The trend of the pressure distribution for Injector 7 is similar to that for No. 5; that is, the maximum average values of peak amplitude ($\tilde{p}_{pk}$) and pressure ratio $[(\tilde{p}_{pk} + 13.5)/(\tilde{p}_{min} + 13.5)]$ occur at locations in the proximity of the "corner" of the combustion chamber. However, there is a substantial decrease in the magnitude of these parameters for Injector 7.

Ultimately, these amplitude differences (as well as the intersection shape differences previously noted) must be correlated with the properties of the actual chamber environment and the response of this environment to the passage of the pressure wave. At this time, however, there is an almost total lack of quantitative information on the details of the combustion processes. It is hoped that the analytical program (discussed below) will eventually provide an insight to those environmental properties which serve to control the response of the processes to pressure waves. In the interim, we can point out only the most obvious gross difference in these two particular combustion systems and postulate the resulting influence on the observed boundary pressure transient characteristics.

Based on the rotating detonation-like wave concept which was depicted in Fig. 4 and which states that the disturbance consists of a combustion-supported shock wave which rotates about the chamber circumference, we presume that the strength (i.e., pressure ratio) of the wave is governed by the various forms of energy imposed upon it. If these energies are categorized as: (1) driving energies and (2) dissipative or loading energies, then it seems clear that a balance between these two energy categories for a particular combustion system will produce a wave of a particular strength.

While the dissipative energies are probably coupled to the driving energies through the motion and character of the wave itself and hence may not remain constant for different levels of driving energy, we feel intuitively that reducing the driving energy will reduce the wave strength. The source of driving energy is the chemical energy of the propellants, released by augmented combustion immediately behind the wave front. Certainly then, one of the gross parameters controlling the available driving energy must be the quantity of propellant mass present (assuming that it is present at a combustible mixture ratio). The mass throughput for Injector 7 is approximately 13% smaller than that for Injector 5, owing to the higher level of c^* performance of the N_2O_4 + 50/50 fuel. Additionally, the wave rotation period for this engine is some 9% shorter (426 μsec) than the corresponding period for Injector 5. Thus, it is felt that a portion of the amplitude reduction is the result of a decrease in energy available to the wave by virtue of the decreased mass present in its path.

That the wave strength and velocity of wave propagation are dependent upon the local conditions in the environment through which the wave passes is believed to be illustrated by the results from the Injector 7 experiments shown in Fig. 8. The uppermost plot on this figure
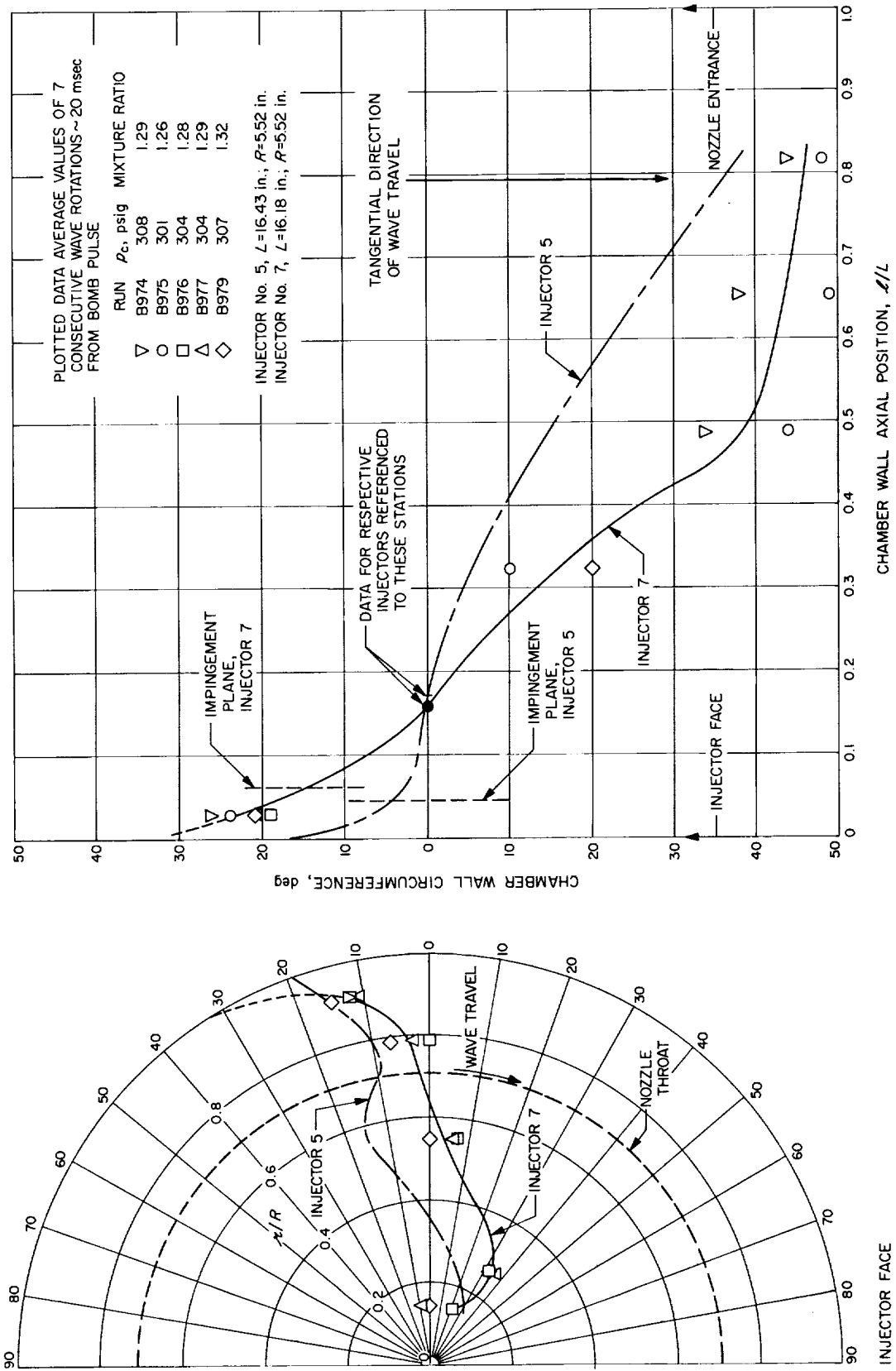
Fig. 6. Intersection of resonant combustion wave with chamber boundary, RMIR Injector 7 with N_2O_4 + 50/50 fuel (UDMH/N_2H_4)

1. PLOTTED DATA AVERAGE VALUES OF 7
   CONSECUTIVE WAVE ROTATIONS ~20 msec
   FROM BOMB PULSE

2.
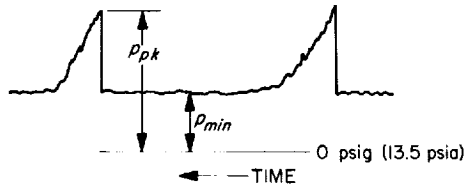
	RUN	p_c, psig	MIXTURE RATIO
▽	B974	308	1.29
○	B975	301	1.26
□	B976	304	1.28
△	B977	304	1.29
◇	B979	307	1.32

3. INJECTOR No. 5 ——————, L=16.43 in.; R=5.52 in.
   INJECTOR No. 7 ——————, L=16.18 in.; R=5.52 in.
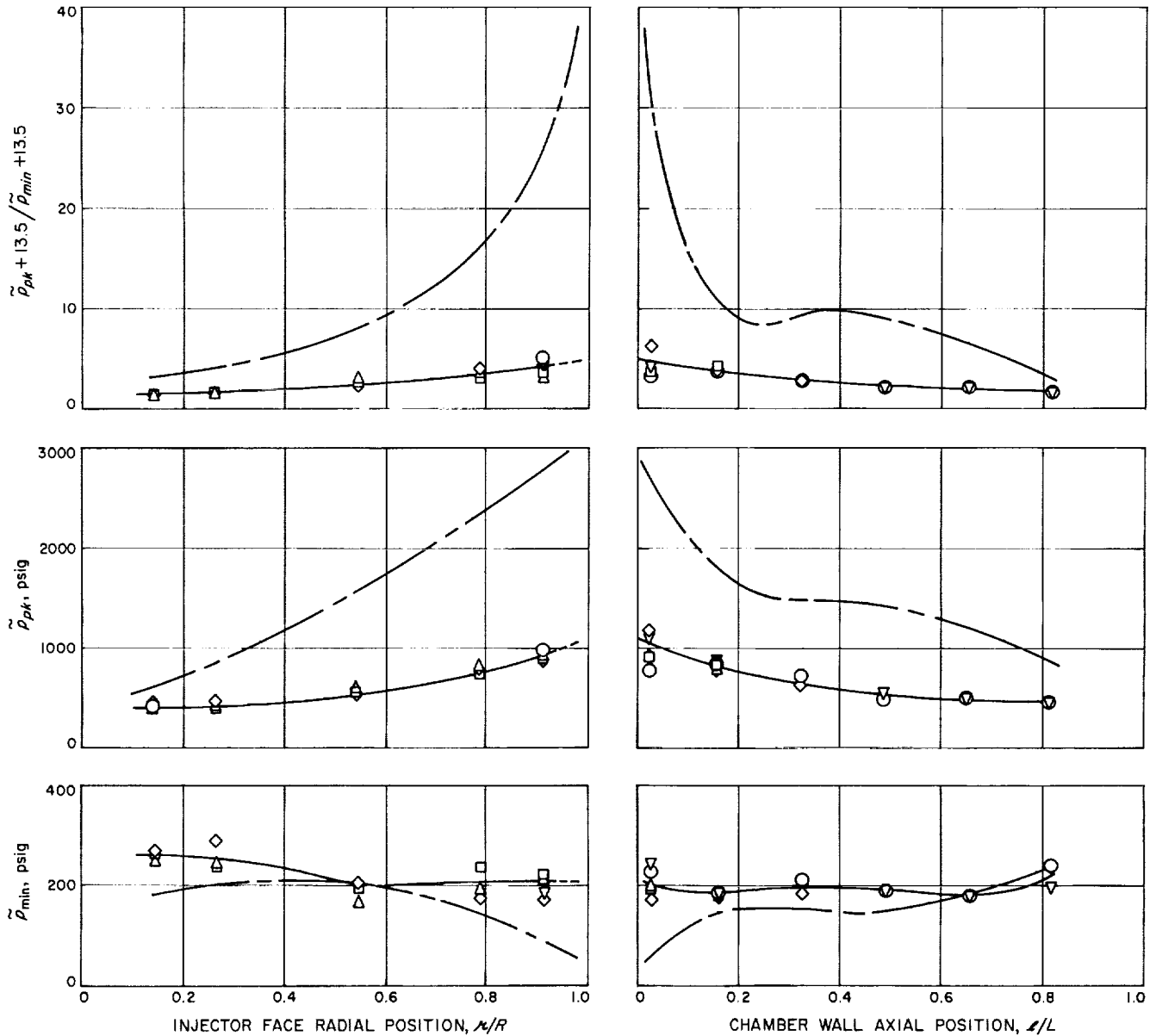
Fig. 7. Resonant combustion pressure amplitude versus chamber boundary position,
RMIR 7 with N_2O_4 + 50/50 fuel (UDMH/N_2H_4)

AXIAL–MASS–FLUX DISTRIBUTION FOR RMIR INJECTOR 7 AT A STATION 3.11 in.
FROM FACE ($\ell/L = 0.192$) FOR r_{unit} CONDITIONS—VIEW LOOKING UPSTREAM
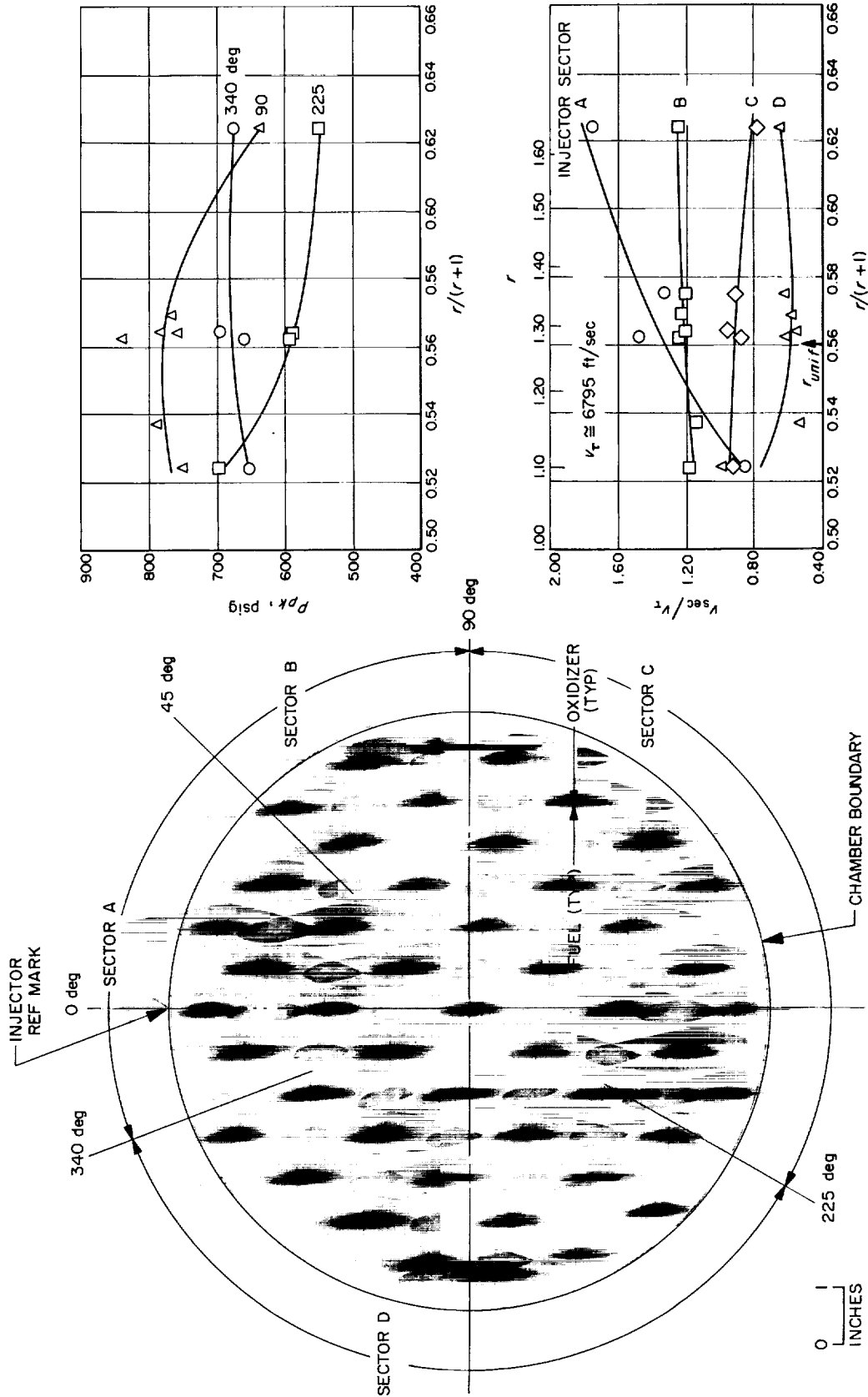
Fig. 8. Circumferential variations of amplitude and tangential velocity versus gross mixture ratio at a station 2.57 in. from face ($\ell/L = 0.159$), RMIR Injector 7 with N_2O_4 + 50/50 fuel

shows the variation in peak amplitude with gross mixture ratio at three circumferential locations on the chamber wall at a common station 2.57 in. from the injector face. The data are for clockwise wave travel. Shown on the left-hand side of the figure is the circumferential orientation of the measurements relative to the injection scheme. This representation of the injection scheme is fully described in Ref. 3 and is based on the mass distribution produced by the individual impinging elements when flowed with nonreacting fluids at the uniformity mixture ratio. If it is assumed that as a first approximation combustion effects on these nonreactively determined spray properties can be neglected, then it is apparent that considerable circumferential nonuniformity in the mass distribution is present in the vicinity of the chamber walls. Evidently then, the desirable correlation is the relation of the local wave strength to local mass-and-mixture ratio rather than to gross mixture ratio. Unfortunately, this correlation is not possible at the present time, since these local conditions have not been quantitized.

It can be shown, however, that the essentially elliptical spray pattern (i.e., mass distribution) produced by these particular elements at their uniformity mixture ratio distort from near-symmetry about the major axis of the pattern when operated at other mixture ratios. For example, at fuel-rich ratios the increased relative momentum of the fuel jet diverts the resultant spray momentum for each element to the right, as viewed in Fig. 8. Converse distortion to the mass distribution is produced at oxidizer-rich ratios. Thus, as the injector is operated at various gross mixture ratios, nonlinear changes in local mass and mixture ratio occur, a situation which is probably most significant in the regions of the combustion volume in the proximity of the chamber walls not too distant from the injector.

The lower plot of Fig. 8 shows the ratio of sector tangential velocity of the wave-to-wall intersection (v_{sec}) to the average velocity for a complete wave rotation (v_T) plotted against gross mixture ratio. The sectors of interest are shown relative to the injection scheme on the mass distribution to the left. Again, it is felt that the significant correlation would be with local environment conditions rather than gross mixture ratio. To be noted on this plot is not only the wide circumferential variation in velocity at a fixed mixture ratio (r_{unif}, for instance), but also the nonlinear variation in sector velocity with mixture ratio.

While the results presented in Fig. 8 fall far short of describing a truly definitive relationship between the high-amplitude, steep-fronted combustion disturbance

observed in these engines and the energy coupling mechanisms which sustain and control them, it is believed that the information demonstrates the localized nature of the coupling. Indeed, a reduction of peak amplitudes by approximately 30% is apparently possible, based on examination of the limited data shown in Fig. 8, if the local conditions leading to the minimum amplitude observed at the 225-deg location were understood and generated uniformly about the chamber boundary.

It is noted that circumferential variations to the extent illustrated in Fig. 8 were not observed for the Injector 5 engine. This is not to say that variations may not exist, since they may not have been exposed by the range of measurements and operating conditions that were covered.

2. Analytical Program

An attempt is being made to describe analytically the sustaining mechanism for the rotating detonation-like wave. This work is being done for JPL by V. D. Agosta and S. Z. Burstein, consultants. A brief description of the method will be given here.

It is assumed that the combustion process, originally operating under steady-state conditions, is disturbed by a wave (e.g., from a bomb). A short time later the combustion process will react to the disturbance, and the object is to discover under what conditions the wave can be sustained by the resulting time-dependent release of mass and energy into the gas phase.

In order to specify the conditions at any point in the chamber at any instant, we need the local values of at least five variables: pressure, temperature, velocity, density, and combustion rate. The five simultaneous equations required can be derived using the following laws:

(1) Conservation of mass.

(2) Conservation of momentum.

(3) Conservation of energy.

(4) Ideality of the gases and temperature-dependency of the liquid densities.

(5) Time and space dependency of combustion rate (this relation is extremely complex, and the generation of a satisfactory approximation to it is a major part of the project).

Using these concepts, the work is divided into the following six tasks:

a. Generation of the "wave equations" in three dimensions. This task has been accomplished, and essentially consisted of mathematically expressing the three conservation laws, listed above, for the two-phase medium.

b. Determination of the axial droplet distribution from a one-dimensional nonoscillatory aerothermochemical analysis. This gives the conditions in the combustion chamber prior to introduction of the wave. Ignoring variations in the radial and tangential directions and assuming an initial drop-size distribution, one can calculate (using such empirical laws as are available) gas and droplet axial velocities, pressure, temperature, the amount of unburned propellant in the gas phase, etc. A computer program for this purpose has been written by the consultants, and is now being used in the Resonant Combustion Program. Effects of compressibility are included, and droplet combustion rates are calculated with the aid of the evaporation-rate relations of Ranz and Marshall (Ref. 5). Unfortunately, the size of the propellant droplets formed in the atomization process is not known as a function of injector design. This is a major weakness which plagues all liquid-propellant combustion theories at present. The difficulty is circumvented by choosing a statistical distribution of drop sizes and a *mean* drop size which is chosen so that the calculated combustion rates match with the known performance of the real engine. With the basic assumption that this procedure gives satisfactory results, one can then use the program to compare different combustion chambers which use the same injector. (Conceivably, such calculations could be useful in an engine development program by helping to optimize the length and shape of the combustion chamber.)

c. Determination of the radial and tangential droplet distribution from the injector geometry. This will involve the use of cold-flow studies of the injector elements used in the experimental program. Data for the mass flux and mixture ratio distribution produced by Injector RC1 (Ref. 1) have already been obtained, utilizing the nonreactive spray properties of the individual elements.

d. The determination of the relaxation time from a nonsteady droplet evaporation analysis and a diffusion analysis of the evaporated fuel vapor through the film surrounding the droplet. The evaporation model was formulated by Grossman (Ref. 6), and can be used to calculate changes in evaporation rate for heat inputs of the form expected from the passage of a shock wave through the combustion zone. The heat entering the film

around a drop (plus any heat released by vapor decomposition inside the film) goes into heating the drop, evaporating the liquid, and further heating the vapors. The heat balance achieved determines evaporation rate. Spherical symmetry is assumed in order to reduce the complexity of the analysis.

e. From (b), (c), and (d), calculation of the excess mass release caused by the wave. For mathematical convenience, this mass release, instead of being continuous, will be considered to occur at some discrete instant after passage of the wave; the "time lag" will be computed from the nonsteady analysis mentioned in *d*.

f. Applying (e) to the wave dynamic equations of (a) and solving numerically. This presents a formidable mathematical problem whose solution would be a major advance in combustion theory.

B. Advanced Liquid Propulsion Systems

R. N. Porter and H. B. Stanford

1. Introduction, R. N. Porter

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

Periodic reports in the SPS's (starting with SPS 37-8) describe the progress of work on the various parts of the ALPS system. Recent accomplishments are outlined below. These include (1) completion of the investigation to determine the degree to which combustion processes interfere with liquid-liquid propellant mixing, (2) performance of some laboratory tests to verify the theory that increased permeation through expulsion bladder materials will result from downstream freezing-out of the permeant, (3) the fabrication of a new composite bladder material, and (4) the construction of some recycleable, all-metal reversing expulsion diaphragms.

2. Injector Development, H. B. Stanford

The original ALPS 2000-lbf thrust injector was designed to use two single doublet elements. With this design, 0.236-in. diameter streams of nitrogen tetroxide and hydrazine (N_2O_4–N_2H_4) were impinged at an angle of 60 deg. Design chamber pressure was 150 psia. Initial tests with this injector gave an average characteristic exhaust velocity c^* of only 67% of the theoretical value, indicating a very inefficient combustion process. To explore the reasons for this poor performance, an experiment was designed to investigate a hypothesis made by Elverum (Ref. 7), in which poor mixing of N_2O_4–N_2H_4 in some injectors was believed to be caused by the effect of rapid liquid-phase reactions upon injected propellant streams. A description of this experiment at the 2000-lb thrust level, together with the test results, is presented in Ref. 8, and the results of similar experiments at the 100-lbf thrust level were reported in SPS 37-31, Vol. IV.

Briefly, the experiment involved firing a single doublet injector in a chamber which was divided into two longitudinal channels by a baffle plate. The baffle was oriented so as to be in a plane perpendicular to the plane in which the injector orifices were located. The top of the baffle was situated a short distance from the injector plate so as to avoid interference with the impingement process, and the bottom of the baffle was in the plane of the nozzle throat. Two full-cone spray nozzles were located in the chamber wall, one on each side of the baffle. Provision was also made to mount turbulence rings in the chamber, downstream of the spray nozzles. The experiment consisted of determining the difference in measured performance when the auxiliary propellants sprayed from the side nozzles, one spraying fuel and the other oxidizer, were reversed. The concept involved is that if the propellants flowing from the main injector are repelled from each other to form a fuel-rich zone on the fuel orifice side and an oxidizer-rich zone on the oxidizer orifice side of the chamber, then the baffle should prevent secondary mixing due to turbulence and diffusion. Thus, one channel should contain an oxidizer-rich flow and the other a fuel-rich flow of gases. Spraying oxidizer into fuel-rich gases and fuel into the oxidizer-rich gases (labelled "unlike" propellants henceforth), should increase performance, while spraying fuel into fuel-rich gases and oxidizer into oxidizer-rich gases (labelled "like" propellants henceforth) should reduce performance. If the streams from the main injector do not blow apart, and a nearly uniform mixture ratio distribution exists in the chamber, performance should remain relatively unchanged when the propellant sprays are reversed. If the streams penetrate through each other, the fuel-rich and oxidizer-rich channels will be reversed, and the performance changes should indicate this condition also.

In summary, the results of the tests indicate that the 0.236-in.-D streams did react and repel each other at the impingement points, so that considerable quantities of fuel and oxidizer did not undergo mixing and combustion. However, firing tests of a 100-lbf-thrust injector employing ten doublets indicated that a minimal amount of these effects might be found with such low-thrust elements. Because of this evidence it was decided to conduct similar experiments at the 100- and 10-lbf thrust levels, in order to obtain data on the effect of injector element scale on the phenomenon. The experimental program at the 100-lbf thrust level, utilizing both 0.064-in.-D streams and impinging sheet elements, revealed that considerable reaction effects were still being encountered with the combustion from the mixing of the round streams, although the sheet element combusion appeared unaffected (see SPS 37-31, Vol. IV, for further detail).

To investigate this phenomenon with even smaller diameter streams, an uncooled engine of nominal 10-lbf thrust was designed and built. This engine was linearly scaled down from the 100-lbf engine, referred to previously, and is shown in Fig. 9. The length-to-diameter and contraction ratios were held constant, but the characteristic length L^* was changed from 50 to 20 in. A divider plate, or baffle, was positioned in the center of the chamber, oriented 90 deg to the plane of the impingement angle, and held in place by a turbulence ring with a total open area of 2½ times the throat area. The baffle extended from slightly below the stream impingement zone in the chamber to the 2-to-1 contraction area ratio plane.

Two flat-plate injectors were used with the 10-lbf chamber section. One impinged 0.020-in. round streams at an included angle of 60 deg. These streams were formed by equal-area straight tubes of 0.020-in. ID with length-to-diameter ratios of 100. In the other injector these tubes were positioned to flow the streams onto curved deflectors, which formed flat sheets, again impinging at a 60-deg angle. Side injection was accomplished through full-cone spray nozzles located in the combustion chamber on either side of the baffle plate and well above the turbulence ring.

Initial firings with this chamber indicated two problem areas. First, the thin baffle plates and turbulence rings, required in scaling down to this small chamber, lacked sufficient mass to withstand the heat generated within the

0 1 2 3 4 5 6 7
INCHES
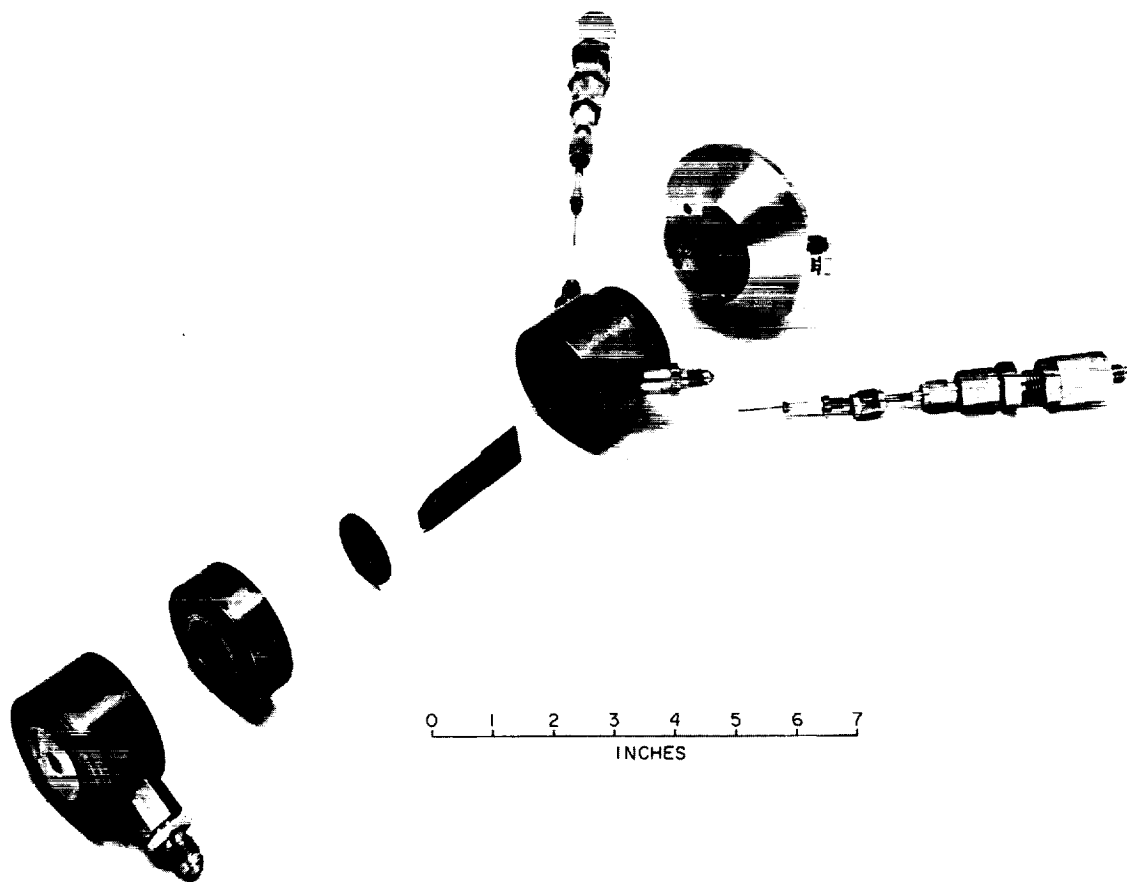
Fig. 9. Engine (10-lb thrust) used for combustion effects experiment

chamber in a 3-sec firing. Baffles and turbulence rings made of copper burned away, and those made of steel melted. Several materials were considered, but it was decided to try pyrolytic graphite, which was available from discarded free-standing pyrolytic graphite rocket engine chambers. This material proved to be entirely satisfactory. It was easily machined to the shapes required and was able to withstand five or more 3-sec firings without serious erosion. The other problem involved the metering of the low propellant flows required for this engine, particularly at the low side-flow conditions. To meter these low flows (0.0006 to 0.004 lbm/sec) restrictors were used which were made from capillary tubing with inside diameters of 0.008- to 0.015-in., and with lengths of from 3 to 5¾ in. Total flow was measured by means of calibrated orifice plates and differential pressure transducers.

In all of the tests where side flows were used, the main injector flow was held constant at 0.030 lbm/sec, and in all cases an overall mixture ratio of 1.2 (W_{ox}/W_f) was

maintained. Performance was calculated on the basis of the total flowrate of propellant injected into the chamber, the measured throat diameter, and the chamber pressure. This latter parameter was obtained from pressure taps located on either side of the baffle but below the turbulence ring. To measure the datum performance level, the engine was fired at design flow rate with each injector type, both with and without the baffle, but with no side injection. Firings were then conducted with each injector, additional propellant being injected through the spray nozzles at the side of the chamber. In some tests like propellant was injected at the sides, while in others unlike propellant was injected. In each of these series, tests were made in which the side flows were adjusted to values ranging from 6 to 25% of the total flow.

The results of this test series are shown graphically in Fig. 10. These curves show little difference in performance with impinging sheets or impinging round streams. They also show that in both cases performance is optimum without side spray and tends to fall off with either
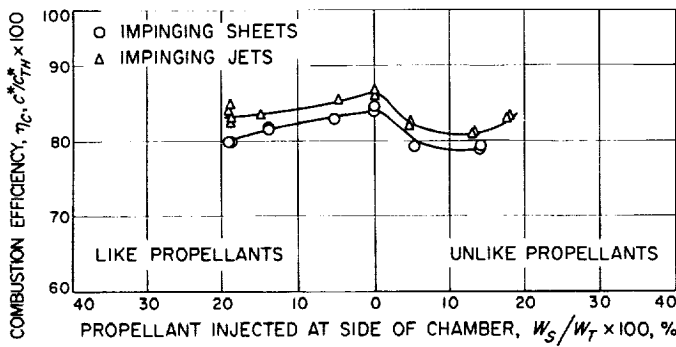
Fig. 10. Experimental results of combustion effects testing at 10-lb thrust level

like or unlike propellants injected into the chamber. The decrease, however, is less with the injection of like propellants. From these observations the following conclusions can be drawn with regard to the 10-lbf engine tests:

(1) Propellant mixing is equally as effective with round streams of 0.020-in. D as with flat sheets of thin cross section.

(2) The effect of rapid liquid-phase reaction is not significant with impinging streams of 0.020-in. D.

(3) A small amount of stream penetration occurs with streams of small cross section, resulting in the lower rate of performance decrease when like propellant side spray is added.

Fig. 11 shows the data for all three sizes of impinging doublet elements that have been tested to date. The absolute levels of combustion efficiency for each curve should not be compared relative to each other, since the data are derived from combustion chamber pressure mea-
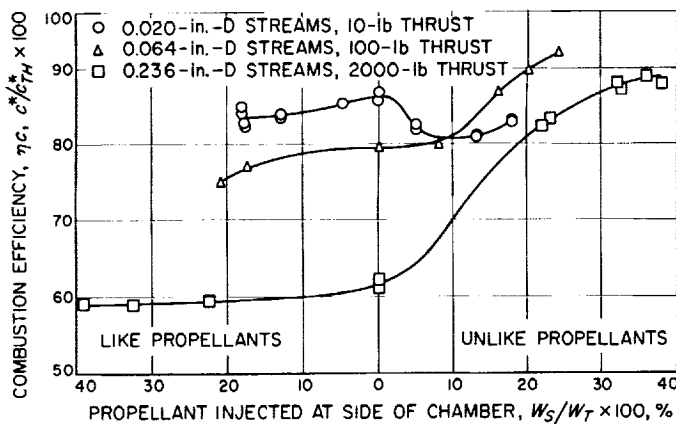
Fig. 11. Combustion effects test results for three sizes of doublet elements

surements, and in each case the installation varied slightly. The relative shape of each curve is meaningful, however, and it is evident that the larger, 0.236-in.-D streams are severely disrupted by the rapid reactions at the impingement interface, whereas the 0.020-in.-D streams penetrated and mixed well, as evidenced by the occurrence of an optimum performance value for the case without side spray flow. The exact size at which the disruption process starts to dominate the propellant mixing is at some size between the 0.020- and the 0.064-in. diameters.

These experiments conclude the investigations relative to the reaction effects phenomenon with hypergolic propellants. A formal report describing this program is now being prepared.

3. ALPS Expulsion Device Development, H. B. Stanford

a. Cryogenic pumping. As part of the study of a hydrazine–nitrogen tetroxide bipropellant propulsion system in a space environment, the effect of temperature extremes on exposed parts of system plumbing was questioned. Of particular concern was the effect known as "cryogenic pumping" on the rate transfer of propellant through a permeable bladder wall if a part of the ullage upstream of the bladder, such as a length of metal tubing, was exposed to low temperature. It was theorized that such an area of low temperature would condense the propellant vapor to liquid as it vaporized from the bladder surface, and would then freeze the condensate solid if the temperature was low enough. This condensation procedure would cause a reduction in partial pressure, prohibit the saturation of the ullage area by propellant vapor, and would effect a continuous pumping action.

To investigate this theory, a simple piece of chemistry hardware was designed and built (Fig. 12). This equipment consisted of two flanged stainless steel cups approximately 1½ in. deep by 1½ in. ID which could be bolted together with a sheet of Teflon film clamped between them. One cup had a fitting which could be tightly sealed after the injection of N_2O_4 on one side of the Teflon sample. To the other cup was attached a 2-in. length of ¼-in. Kovar steel tubing, and a 4-in. length of ¼-in. glass tubing with a ½-in.-D by 6-in.-long glass bulb at the end. The assemblage could be mounted on a ring stand with both sides of the Teflon sample at ambient temperature, or the glass bulb could be inserted into a Dewar bottle filled with a liquid coolant to simulate the condition in question. With this test setup, gravity was a constant source of pressure differential. However, its effect was originally considered to be negligible in this test.
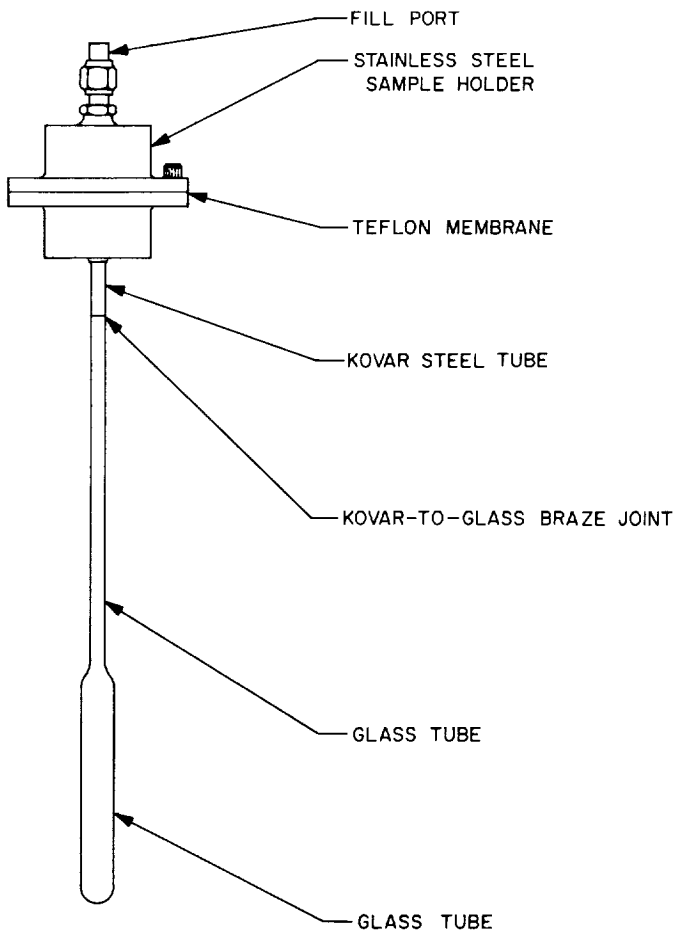
Fig. 12. Cryogenic pumping test fixture

increase of 15 mg/in.^{2}/hr over that normally expected. The permeation rate during the 2-wk period following removal of the bulb from the LN$_2$ was 0.61 mg/in.^{2}/hr. This figure, although small, is more than would be expected if a true concentration equilibrium had been reached. The discrepancy is believed to be partially caused by gravity, perhaps augmented by the fact that the apparatus was stored in a ventilated hood where a draft may have lowered the temperature in the bulb to slightly less than the temperature of the N$_2$O$_4$ over the Teflon membrane.

An attempt to reverse the process and return the N$_2$O$_4$ to its original position above the Teflon film, by heating the bulb, ended in failure; the small length of ¼-in. Kovar steel tubing attached to the lower cup eroded through, allowing the N$_2$O$_4$ to dissipate.

It was noted, as had been theorized, that the N$_2$O$_4$ vapor condensed and froze solid at the first point cold enough to do so; in this case the ¼-in. tube above the bulb. These frozen slugs of N$_2$O$_4$ were solid and certainly sufficient to clog a pressure line. It should be noted, however, that the temperatures imposed during this experiment were extreme compared to the temperatures normally expected to be encountered in a well-designed spacecraft system.

b. Composite material. Much of the ALPS bladder development effort has been directed toward the search for materials that are compatible, impermeable, and durable enough to be used in the fabrication of bladders for use with the propellants N$_2$O$_4$ and N$_2$H$_4$.

A sheet of 0.010-in.-thick skived Teflon film was clamped in the apparatus, and 20 cm^3 of N$_2$O$_4$ was injected into the cup over it. After two days at ambient temperatures the glass bulb became saturated with N$_2$O$_4$ vapor (determined by visual reference only), but no liquid N$_2$O$_4$ had formed. At this time the glass bulb was inserted into a Dewar bottle of liquid nitrogen (LN$_2$). After 24 hr in LN$_2$, ½ cm of liquid (a figure later determined to represent 0.33 cm^3) had accumulated in the glass bulb (Fig. 13). After seven days 3.63 cm^3 of N$_2$O$_4$ had collected in the glass bulb. The Dewar bottle of LN$_2$ was then removed, and the apparatus was allowed to stand at ambient temperatures for a 2-wk period. Little change was noted for several days; however, at the end of the 14-day period the volume of N$_2$O$_4$ in the bulb had increased to 3.88 cm^3.

The normal permeation rate for 0.010-in.-thick skived Teflon film with N$_2$O$_4$ at ambient temperature is 2 to 3 mg/in.^{2}/hr (Ref. 9). The permeation rate during the test for cryogenic pumping averaged 17.50 mg/in.^{2}/hr, an

As part of this effort the idea of a composite fabric was conceived. This fabric would be made up of several compatible materials, each of which would contribute a particular capability to the composite. Individually these materials would not necessarily make satisfactory bladders. The composite fabric, as it was originated, was composed of gold foil, FEP Teflon film and Teflon felt joined in some quilted pattern. The gold foil was to provide a barrier to permeation of the propellants; the FEP, in several discrete layers, was to contain the other elements and strengthen the composite structure; and the Teflon felt was to be saturated with, and serve as a vehicle for, a reactant chemical (manganese hydroxide) which would capture any permeating propellant that penetrated the gold foil barrier through pinholes. (This reactant, in theory, was to be inert to N$_2$H$_4$). As an additional benefit, it appeared that the Teflon felt would serve to cushion the metal foil and protect it from sharp bends and folds.
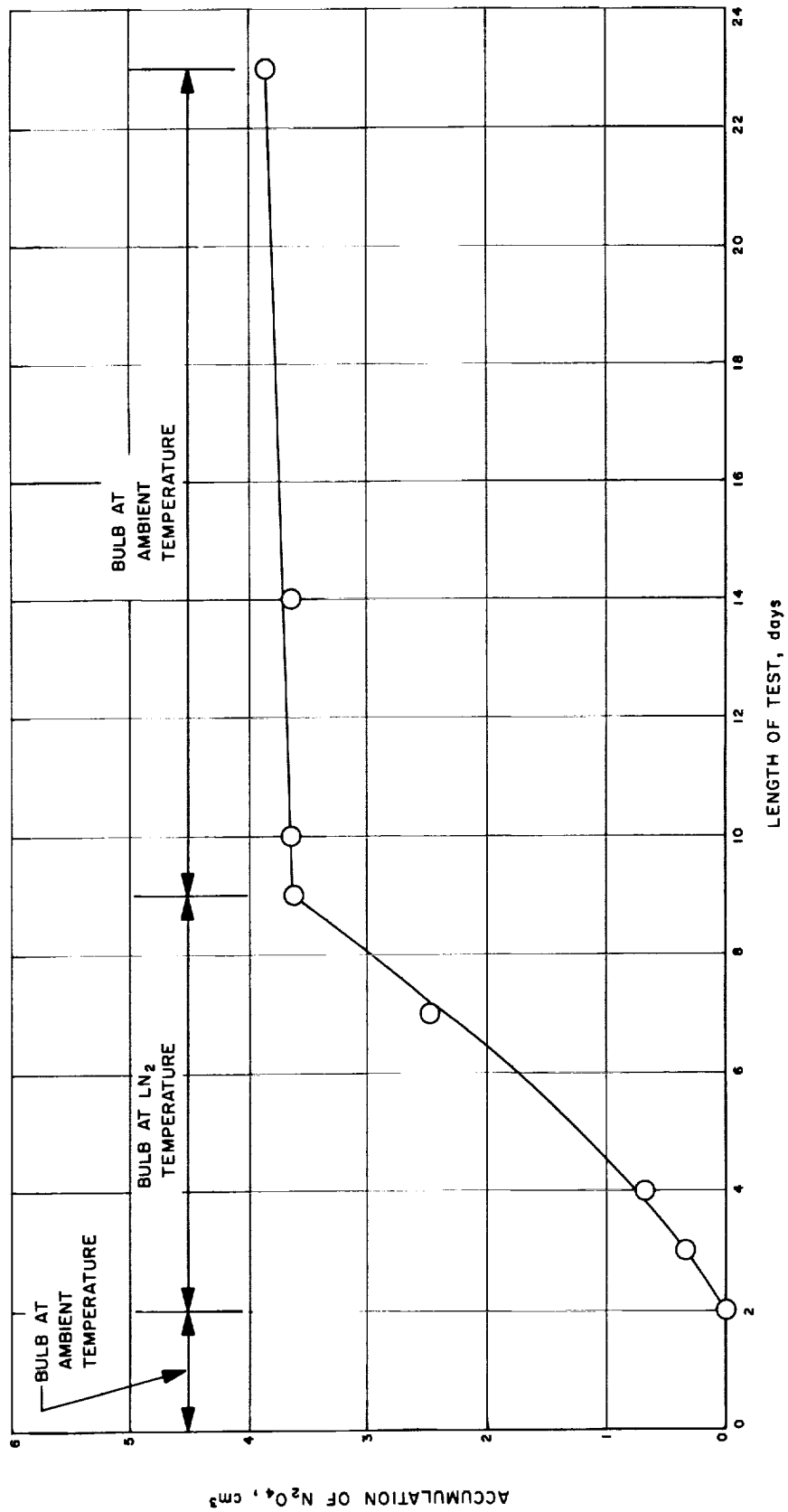
Fig. 13. N_2O_4 cryogenic pumping test results on 0.010-in. TFE Teflon skived film

To investigate this composite fabric and the problems of bladder fabrication inherent to it, a contract was placed with Dilectrix Corp. of Farmingdale, Long Island. Their proposal retained the basic concepts of the original composite but altered the means of attaining it. By using the Dilectrix process of spraying and sintering Teflon, it is possible to form the material over spherical or other curved surfaces. The metal foil (in this case, aluminum) is formed and applied in overlapping gores, and the Teflon felt can be readily formed to the shape desired. The various elements are bonded together, as they are assembled, by sprayed and sintered layers of Teflon. Because the contract was designed to study initial fabrication problems only, it was decided not to include the reactant chemical in the first structures.

Several combinations of the composite material have been fabricated into test "pipes," which are cylinders 5 in. in diameter and 12 in. long. Two typical composite fabrications are as follows:

Composite with ¼-mil aluminum foil	Composite with ½-mil aluminum foil
TFE (3-mils) with FEP (1-mil), laminated	FEP, 4 mils
Aluminum foil, ¼-mil	Aluminum foil, ½-mil
FEP, 2-mils	FEP, 2 mils
Aluminum foil, ¼-mil	TFE felt, 0.090 in.
FEP, 2 mils	FEP (3-mils) with TFE (1-mil), laminated
TFE felt, 0.090-in.	
FEP (3-mils) with TFE (1-mil), laminated	

Other combinations use various arrangements of these and other elements. All have the 0.090-in. layer of Teflon felt.

A standard JPL N_2O_4 permeation test (Ref. 9), but modified to adequately seal the edges of a flat disc-shaped test sample containing Teflon felt, was run by Dilectrix on a composite sample containing ½-mil foil. In 24 hr no N_2O_4 permeation was detected. Helium leak tests on the pipes were run by Dilectrix in a cylindrical fixture designed to seal the pipes at both ends. The helium leak test apparatus was calibrated against a standard leak device rated at 6.1 × 10⁻⁷ cm³/sec at atmospheric pressure (atm cm³/sec) and its sensitivity recorded for each test reading. The sensitivity readings are the basis of comparison against the actual leak rate readings, their difference being the actual leak rate of the samples. All of the

helium leak tests were continued for 1 hr, at which time it appeared that the leak rates were constant. Some selected samples were run for 2 hr to assure that the leak rates had stabilized within the first hour. For some typical results of these tests, see Fig. 14. Helium leak rates through the total surface of the pipe varied from 3410 × 10⁻⁹ atm cm³/sec for the worst measurable sample, down to 320.0 × 10⁻⁹ atm cm³/sec[1]. The average leak rate for the 12 pipes tested was 1075 × 10⁻⁹ atm cm³/sec. The test results were then compiled on the basis of aluminum foil content. Seven pipes contained two layers of ¼-mil aluminum foil and five pipes contained one layer of ½-mil aluminum foil. The ½-mil samples averaged 1405 × 10⁻⁹ atm cm³/sec, and those containing two layers of ¼-mil aluminum foil averaged 839 × 10⁻⁹ atm cm³/sec. The lowest individual permeation rate was a sample containing a single ½-mil aluminum foil layer.

One pipe of each of the eight composite structures was delivered to JPL, where crease endurance tests were run with the JPL crease test device (SPS 37-28, Vol. IV).

---

[1] One test pipe was damaged in the etching tank, and three others leaked enough to drive the sensing appartus off scale.
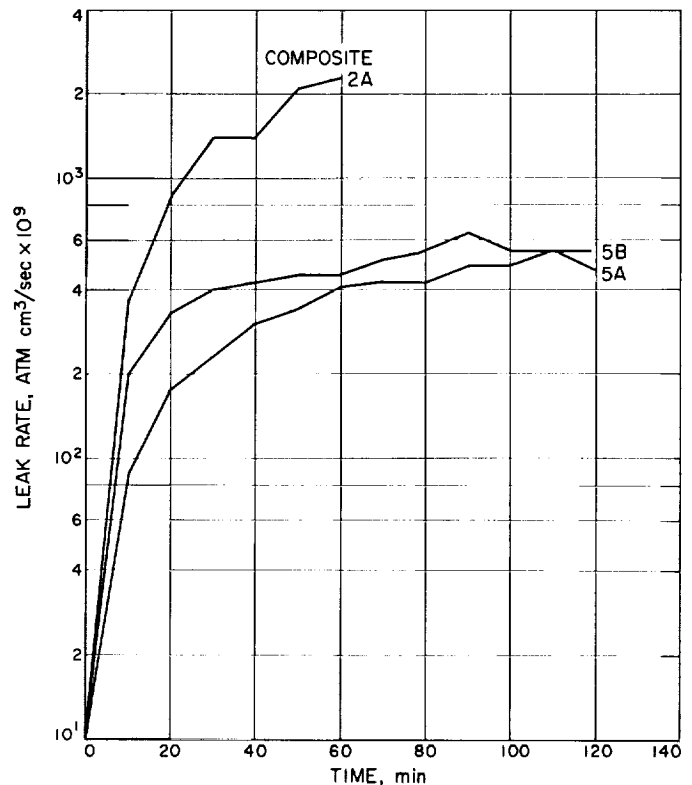
Fig. 14. Three typical helium leak rates through composite bladder materials

Because of the thickness of the composite materials, round blades 0.190 in. in diameter were used. These blades were placed at a vertical angle of 30 deg and a horizontal angle of 2 deg with a 0.104-in. gap left between the ends. Cable tension was set at 5 lb. Final failure was determined by the use of dye penetrant, which was applied at frequent intervals until 100 cycles had been completed, and at every 100 cycles thereafter. The test samples were closely observed for the first occurrence of metal failure, delamination, and rupture of surface layers of Teflon.

Cycle life of the various materials varied from 400 to 1300 cycles, as determined by dye penetration. Metal failure as indicated by visible cracks occurred between 5 and 15 cycles in all cases, and delamination started between 15 and 20 cycles (Fig. 15). As cycling continued, the metal fractures grew, and eventually the outer layer of Teflon film was ruptured as well. Delamination, in this case loosening of the Teflon felt rather than separation of discrete layers, tended to limit itself to parallel horizontal folds. In no case did a complete separation occur. These tests, while not conclusive from all standpoints, did indicate that composites with a TFE Teflon base layer and two layers of $\frac{1}{4}$-mil aluminum foil in the structure were the most durable.

Compatibility tests were run with both N_2O_4 and N_2H_4 by placing 1-in. squares of composites in vials of the propellants. After 24 hr both samples had started to delaminate at the Teflon film–aluminum foil interfaces; they were, therefore, removed from the propellants. After

outgassing, the sample which had been in contact with N_2O_4 separated by itself. The sample which had been in contact with N_2H_4 did not delaminate by itself, but could be pulled apart. There was no tendency for the material to delaminate at the Teflon felt–Teflon film interfaces.

Development of these materials is continuing, and complete bladders will be made of those composite materials determined by permeation and crease tests to be best suited to the purpose.

c. Reinforced metal hemispheres. Metallic expulsion devices have been considered to be secondary (or backup components) in the ALPS bladder development program, because compatible and impermeable polymeric bladders offered the more ideal solution to the expulsion device problem. However, completely satisfactory polymeric bladders have not been forthcoming; consequently, work has been continued on convoluted metal diaphragms (SPS 37-18, 37-22, and 37-25, Vol. IV). This device, although it meets the basic ALPS requirement for expelling two different propellants from the same tank, is limited, since two diaphragms are required to completely expel the contents of a spherical tank. This is the case because it has not been possible to make a convoluted diaphragm reverse itself. All liquid propulsion systems require some ullage in the propellant tanks to allow for liquid expansion due to temperature changes. Systems to be heat-sterilized must have up to 35% ullage. This means that the expulsion device will be more or less distended as requirements vary, and that convoluted diaphragms
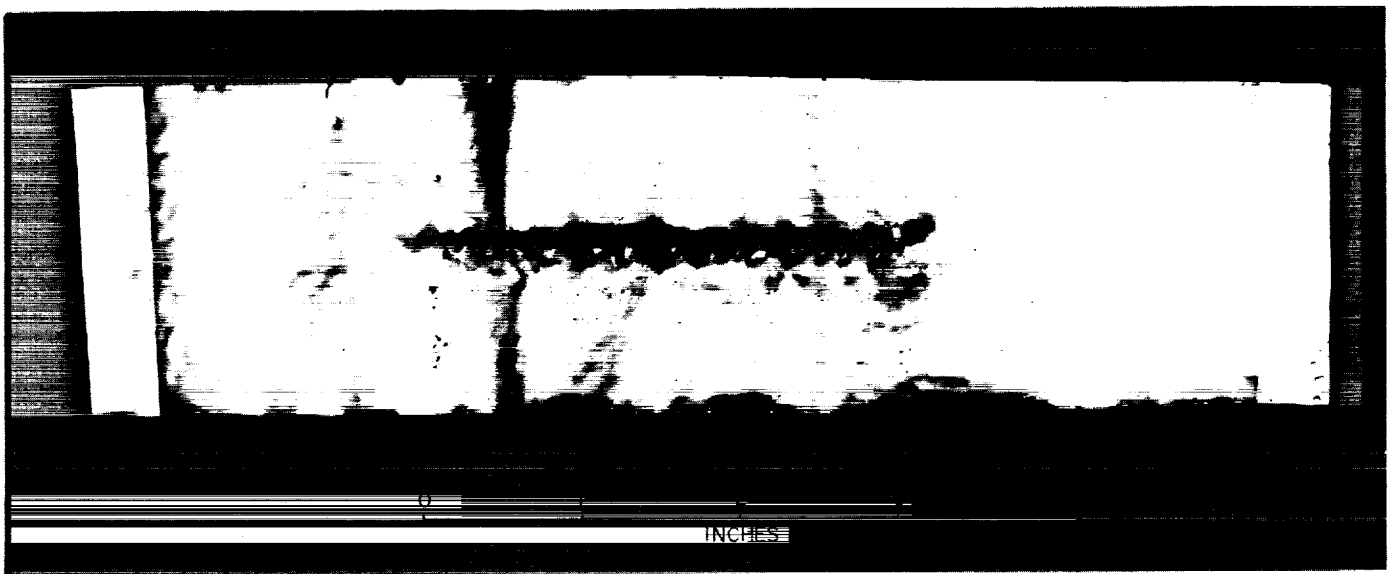
Fig. 15. Dilectrix–JPL composite bladder material crease test sample after 600 cycles

(which must be positioned near the center plane of the tank) will be separated by a considerable space and supported only by the pressurizing gas. Whether they are oriented perpendicular or parallel to the axis of thrust, one or both of them will be subject to distortion by hydrostatic pressure exerted by the propellants during the high g period of booster operation.

A recent development in metallic expulsion devices is the thin walled, wire-reinforced stainless steel hemisphere proposed and demonstrated in 6-in.-D size by Arde Inc., of Paramus, N. J. This idea conceptually satisfies the requirement for a single metallic device to completely expel the liquid from a spherical tank. Because of its mode of operation it need not be placed in a position where it can be distorted by high gravity loading during booster acceleration, even though it is partially distended because of ullage requirements. In 1964 a contract was placed with Arde, Inc., to demonstrate the feasibility of the concept in the ALPS 18-in.-D test size. Hemispherical shells of this diameter with 2-in. flanges and wall thickness of 0.008 to 0.010 in. made from 321 stainless steel were obtained from Jaycraft, Inc., of El Cajon, Calif. Concentric wires, approximately 0.090-in. D spaced at ½-in. intervals, were furnace-brazed with copper to the shells by Arde, Inc. This configuration is shown in Fig. 16.

A cycle test with one shell of this construction gave three complete reversals before a pinhole leak developed. This leak was repaired with silver solder, and another reversal was obtained. During the fourth reversal some of the reinforcing wires came loose, and no further attempts to cycle the hemisphere were made. Apparently
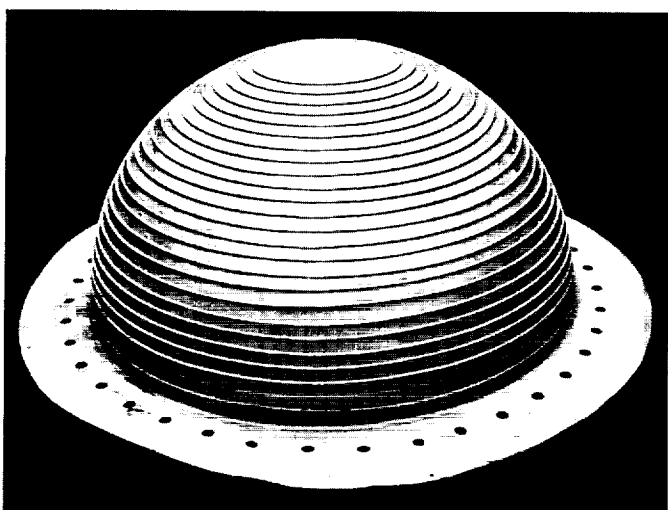
Fig. 16. Hemispherical expulsion diaphragm, 18-in.-D wire-reinforced stainless steel

the wires came loose because of poor brazing, caused by surface contamination of the hemispherical shells. This contamination probably originated in early processing, either from insufficient cleaning or a faulty annealing oven. A method of cleaning the shells has been evolved, and subsequent attachment of the wire to the shells should be improved.

Investigation of this concept is continuing, both at JPL and at other agencies. It is understood that Arde, Inc., has recently fabricated and successfully tested reinforced stainless steel hemispheres of this type in 23-in.-D size for another customer. They have also made and tested a welded spherical tank assembly 13 in. in diameter, with the two tank halves and the diaphragm joined by a single weld bead.

C. Combustion Effects in Sprays

J. W. Woodward

Studies (Refs. 10–12) have shown that in a liquid propellant rocket engine, the injection scheme has considerable influence on the overall combustion efficiency of the engine, on heat transfer to the combustion chamber walls, and in the case of ablative chamber materials, on the erosion of the chamber wall. An extensive study of the properties of sprays of nonreactive fluids has been conducted (Refs. 13–15) and these results applied to injector designs, based on the assumption that the sprays formed by reactive liquids are similar to the sprays formed by nonreactive liquids. This assumption is probably quite good for many propellant combinations, but there is considerable experimental evidence (Refs. 12, 16, 17) which indicates that sprays of certain highly reactive propellant combinations, such as nitrogen tetroxide–hydrazine, deviate considerably from the sprays of nonreactive liquids. In particular, for the case of a single, unlike-impinging-doublet injector element, the spray data for nonreactive liquids indicates that the liquids penetrate each other; that is, the spray is fuel-rich on the side opposite the fuel orifice and is oxidizer-rich on the side opposite the oxidizer orifice. However, experiments with engines using the N_2O_4–N_2H_4 propellant combination indicate that, under certain circumstances, the liquid streams separate; that is, the resulting spray is fuel-rich on the fuel orifice side and is oxidizer-rich on the oxidizer

side. The experiments indicate that the degree of separation is affected by the size of the orifice with large orifices, producing a greater separation effect than small orifices.

A major cause of the "stream separation" is thought to be gas evolution, as a result of rapid-liquid phase reactions at the interface of the two propellant streams. If true, a study of the effect of chamber pressure on the degree of separation could provide useful design information, and it would seem possible that some range of operating conditions could be found where the rapid liquid phase reactions would enhance mixing, that is, some intermediate between the observed stream penetration and the stream separation observed under certain conditions for very reactive propellant combinations.

A series of experiments has been planned to measure the local fuel mixture ratio distribution produced by a single pair of unlike-impinging streams in a combustion chamber using a double sonic orifice sampling probe similar to the ones used by Lewis (Refs. 18, 19) and a high-speed mass spectrometer. The experiment is depicted schematically in Fig. 17. An uncooled combustion chamber with an inside diameter of 3 in., using a single, unlike impinging-doublet injector element, will be fired at pressures ranging from atmospheric to 300 psi. The sampling probe can be mounted at different axial stations and will be traversed across the chamber during a single firing, with gas samples being withdrawn continuously. The sample will be fed directly to a Jarrell-Ash Quadrupole Mass spectrometer which can operate at scanning rates up to one scan every 50 msec. Thus in
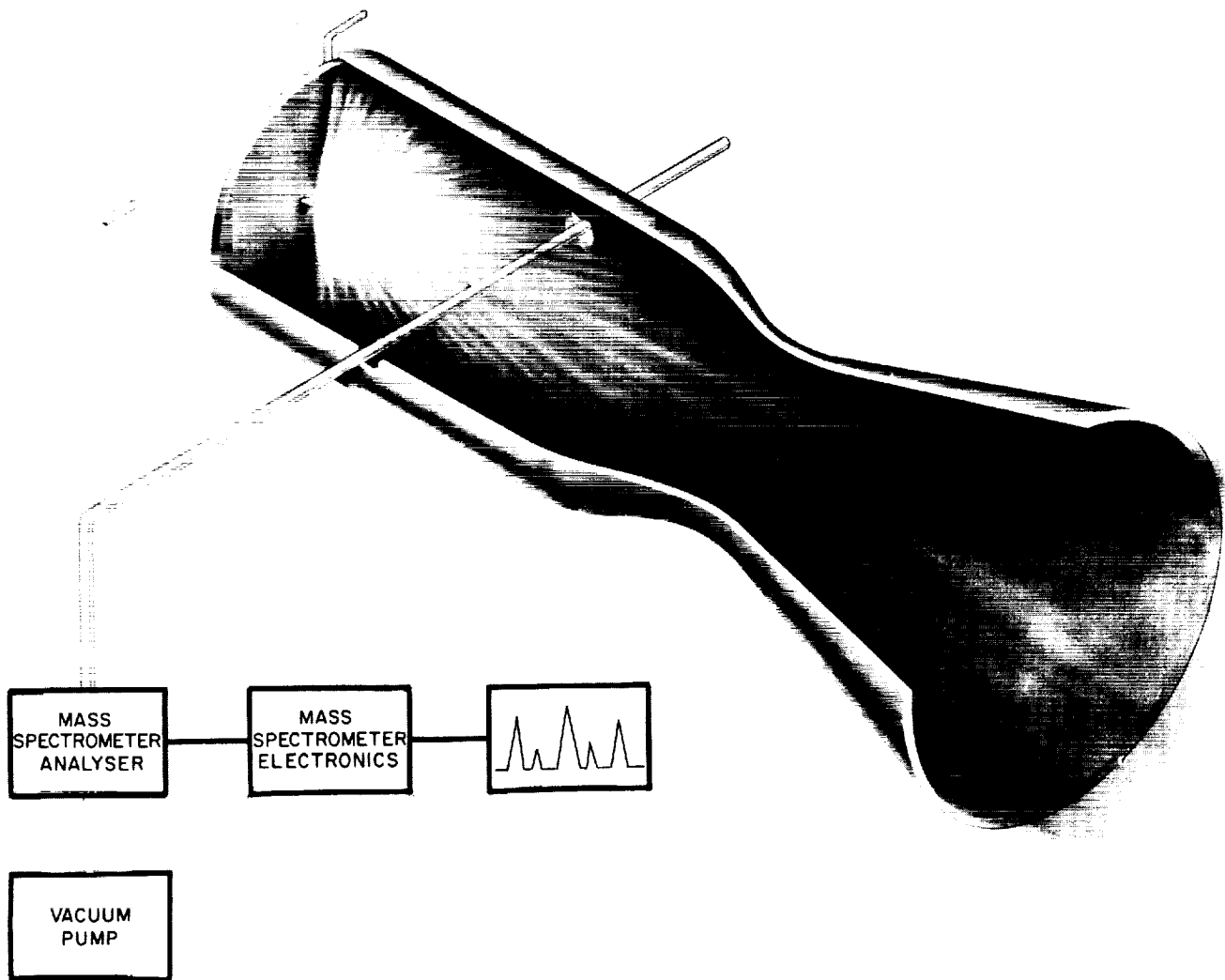
Fig. 17. Experiment to determine combustion effects in sprays

a single 5-sec traverse across the chamber, 100 analyses can be completed. An error analysis of the calculations necessary to reduce the mass spectrometer output to the value of the local mixture ratio indicates that the accuracy should be 10%, or better.

The mixture ratio distribution obtained under actual combustion conditions will be compared to the distribution predicted from nonreactive spray data, and attempts will be made to correlate any differences with chamber pressure, orifice size, impingement angle, and propellant reactivity.

D. Ignition of Ultra-Fine Powdered Boron in Air, Nitrogen, Carbon Dioxide, and Mixtures of Nitrogen and Carbon Dioxide

R. A. Rhein

Boron is considered to have potential as a fuel in the atmosphere of Mars (Ref. 20). Consequently, a series of experiments was conducted to determine the ignition temperature of powdered boron in air, nitrogen, carbon dioxide, and mixtures of nitrogen and carbon dioxide corresponding to the presumed atmospheres of Venus and Mars.

The boron (Ultra-Fine boron powder, Callery Chemical Company, Callery, Pennsylvania, Lot 2782-23-2, assay 99% pure) was heated in a Vycor tube. The gas was passed through the tube at 100 ml/min. This procedure was previously described (Ref. 21).

1. Results

a. Ignition of boron in nitrogen. In nitrogen (Linde Extra-Dry High-Purity nitrogen), there was evidence for ignition at 1059°C in one experiment, but in three other experiments, there was no evidence of ignition when heated to 1370, 1440, and 1220°C.

b. Ignition of boron in carbon dioxide. In carbon dioxide (The Matheson Co., Coleman Grade) boron ignited and burned vigorously. The measured ignition temperatures were 721, 871, 922, and 1017°C.

c. Ignition of boron in a simulated Venus atmosphere. In a gas mixture of 9.17% N_2, 96.73% CO_2, and 4.10% Ar, representing the Venusian atmosphere, ignition occurred at 858, 1000, and 1256°C. The combustion was reasonably vigorous.

d. Ignition of boron in a simulated Mars atmosphere. In a gas mixture containing 11.21% CO_2, 87.63% N_2, and 2.16% Ar, the boron ignited at 1172, 1203, and 1344°C. The combustion was observed to be reasonably vigorous.

e. Ignition of boron in air. In the open air, boron ignited at 203 and 215°C in two separate experiments, and burned reasonably vigorously.

2. Conclusions

The ultra-fine powdered boron ignited and burned in the simulated Mars and Venus atmospheres, in air, and in carbon dioxide. There was evidence for ignition in nitrogen, also. This information indicates the feasibility of using boron as a fuel in the Mars or Venus atmosphere, but a considerable amount of development work remains to be done.

References

1. Clayton, R. M., and Rupe, J. H., "Resonant Combustion," *SPS 37-30*, Vol. IV, pp. 123–130, December 31, 1964.

2. Clayton, R. M., and Rogero, R. S., "Experimental Measurements on a Rotating Detonation-Like Wave Observed During Liquid Rocket Resonant Combustion," Technical Report No. 32-788, Jet Propulsion Laboratory, Pasadena, California, August 15, 1965.

References (Cont'd)

3. Rupe, J. H., "A Correlation Between the Dynamic Properties of a Pair of Impinging Streams and the Uniformity of Mixture Ratio Distribution in the Resulting Spray," Progress Report No. 20-209, Jet Propulsion Laboratory, Pasadena, California, March 28, 1956.

4. Rupe, J. H., "An Experimental Correlation of the Nonreactive 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 No. 32-255, Jet Propulsion Laboratory, Pasadena, California, July 15, 1965.

5. Ranz, W. E., and Marshall, W. R., Jr., "Evaporation from Drops," *Engineering Progress*, Vol. 48, No. 3, March 1952, and No. 4, April 1952.

6. Grossman, B., "Droplet Vaporization: Heat Transfer Limited Model," Polytechnic Institute of Brooklyn, 1964.

7. Elverum, G. W., Jr., and Morey, T. F., "Criteria for Optimum Mixture Ratio Distribution Using Several Types of Impinging Stream Elements," Memorandum No. 30-5, Jet Propulsion Laboratory, Pasadena, California, February 25, 1959.

8. Johnson, B. H., "An Experimental Investigation of the Effects of Combustion on the Mixing of Highly Reactive Propellants," Technical Report 32-689, Jet Propulsion Laboratory, Pasadena, California, July 15, 1965.

9. Vango, S. P., "Determination of Permeability of Cast Teflon Sheet to Nitrogen Tetroxide and Hydrazine," Technical Memorandum No. 35-55, Jet Propulsion Laboratory, Pasadena, California, 1961.

10. Rowley, R. W., and Tyler, W. H., AIAA Paper No. 65-586, presented at the AIAA Propulsion Joint Specialist Conference, Colorado Springs, Colo., June 1965.

11. Rupe, J. H., and Jaivin, G. I., Technical Report 32-648, Jet Propulsion Laboratory, Pasadena, California, October 1, 1964.

12. Stanford, H. B., and Tyler, W. H., Space *Programs Summary* No. 37-31, Vol. IV, Jet Propulsion Laboratory, Pasadena, California, February 28, 1965, p. 192.

13. Rupe, J. H., Progress Report 20-195, Jet Propulsion Laboratory, Pasadena, California, August 6, 1953.

14. Rupe, J. H., Progress Report 20-209, Jet Propulsion Laboratory, Pasadena, California, March 28, 1956.

15. Rupe, J. H., Technical Report 32-255, Jet Propulsion Laboratory, Pasadena, California, July 15, 1965.

16. Elverum, G. W., and Staudhammer, P., Progress Report 30-4, Jet Propulsion Laboratory, Pasadena, California, August 25, 1959.

17. Johnson, B. H., Technical Report 32-689, Jet Propulsion Laboratory, Pasadena, California, July 15, 1965.

18. Lewis, J. D., and Merrington, A. C., Proceedings of the 7th Symposium (International) on Combustion, p. 124, Butterworths Scientific Publications, London, 1958.

References (Cont'd)

19. Lewis, J. D., and Harrison, D., Proceedings of the 8th Symposium (International) on Combustion, p. 366, 1960.

20. Mueller, K. H., and Weber, J. O., "Theoretical Performance of Metals Burning in Nitrogen for Propulsion in the Martian Atmosphere," Chemical Engineering Techniques in Aerospace, Chemical Engineering Progress, Vol. 60, Symposium Series, No. 52, New York, American Institute of Chemical Engineers, 1964, pp. 17–22.

21. Rhein, R. A., "The Ignition of Powdered Metals in Nitrogen and in Carbon Dioxide," Technical Report 32-679, Jet Propulsion Laboratory, Pasadena, California, December 1964.

XII. Solid Propellant Engineering

A. Sterilization or Biodecontamination by Diffusion of Freox Gas Through Propellant

L. C. Montgomery

In initial propellant sterilization studies two approaches to chemical sterilization or biodecontamination were taken: (1) to sterilize the propellant by mixing with ethylene oxide sterilant and, (2) to diffuse ethylene oxide gas through cured propellant. The following information is that determined from investigations of the second approach. (SPS 37-17, Vol. IV.)

In the diffusion approach the fire-safe, highly effective sterilant, Freox (a gaseous mixture of 12% ethylene oxide and 88% Freon 12) was diffused through cured propellant. As indicated in the referenced SPS, inoculated propellant samples were subjected to Freox under a pressure of ½ to 1 in. of water. Penetration occurred, and sterilization of inoculated samples resulted. The penetration rate into polyurethane binder was approximately ¼₄ in./hr, and penetration into propellant was shown to be slower. The test results are shown in Table 1.

In these tests two types of samples were used, one to establish the penetration rate, the second to determine if sterilization had been accomplished. The test specimen to establish penetration rate was an ETO sensor sealed beneath binder or propellant of measured thicknesses (from ¼₆ to ½ in.). The samples to test for sterilization by penetration were small aluminum blocks with holes drilled ½ in. in diameter to depth of ¼₆, ¼ and ½ in. In the first set of blocks tested, the blocks were first inoculated with 1.1×10^6 spores of Bacillus subtilis var. niger (BG). Then polyurethane propellant or binder was cast into the holes and cured. In the second set inert polyurethane propellant was inoculated with 10^7 spores/cm^3 and then cast into the holes in the aluminum blocks and cured.

The results of these experiments show that all samples ¼₆-in. thick had been sterilized by the 11-hr Freox exposure whether inoculated internally or between the propellant and the bottom wall. Binder specimens up to ½-in. in thickness were sterilized. The growth results obtained with the thicker propellant samples indicated that viable BG inoculum recovery could be accomplished with the 10^6 and 10^7 inoculations by the coarse breakup of hand drill cuttings. These experiments indicate that

Table 1. Initial ethylene oxide penetration rate test results

Sample	Thickness, in.	Exposure time, hr	Sensor color[a]
X500-A303B[b]	0.060	4½	Green
		4¾	Light blue
		5	Blue
X500-A303B	0.106	9¾	Light green
		11	Blue
X500-A303B	0.239	11	Showed no change during test. Turned blue after the test
X500-A303B	0.517	11	No change
X500-A203[c]	0.084	5¼	Very light green. Never turned blue
X500-A203	0.125	11	No change
X500-A203	0.234	11	No change
X500-A203	0.500	11	No change
X500-A303B	0.0960	7	Began to change
		8	Light green
		8½	Light blue
		9½	Medium blue
		11	Dark blue
X500-A303B	0.1080		Did not change during test[d]. Dark blue 48 hr after test
X500-A203	0.1133	11	No change
X500-A203	0.1800	11	No change
Thiokol HC	0.0820	11	Did not change during test[d]. Light blue 48 hr. after test
Thiokol HC	0.1170	11	No change

[a] ETO sensor is yellow and requires exposure of 250 mg/liter for 2 hr to turn blue.
[b] Polyurethane binder.
[c] Polyurethane propellant.
[d] Next observation was 48 hr after finished test.

Freox can perform biodecontamination and possibly sterilization of binder and propellants through penetration.

To investigate the feasibility of using this technique on thicker sections, such as those contained in solid rocket motors, the following experiments were conducted.

Propellant samples of Thiokol PBAA-HC-*Surveyor* propellant and JPL *Syncom* propellant were case-bonded into a JPL 5 X 6-in. motor case. A special transparent pressure cap was made for one end, under which was placed an ETO sensor tab. The other end of the chamber was pressurized with Freox gas at about 34-psi pressure.

To prevent transfer of the gas along the case wall bonding area, an extra large buildup of propellant and bonding material was used along the wall; thus the only path by which the gas could get to the sensor was by diffusion through a given thickness of propellant.

To determine the length of time required for diffusion through the propellant, time-lapse photography was set up to view the sensor each hour. From the photographs thus taken, a time could be established when the sensor color began to change, indicating that ETO had reached the sensor. The experimental technique used served its purpose but the results are not exact because of the slow response to the color changes of the sensor. Also, the Freox was not sampled at the pressure chamber to determine ETO content, although the mixture in the supply bottle was agitated periodically in an effort to maintain a uniform mixture of 12% ETO and 88% Freon 12.

The pressurization tests of thick samples of the propellant indicated that this Freox diffusion approach was not feasible. In the first of this series of tests, diffusion through the propellant occurred in 12 to 18 days for ½-in. propellant thickness. However, thick propellant sections required a much longer time for penetration, and Freox degradation of both JPL and Thiokol propellants resulted as indicated in Figs. 1 and 2. Since the propellants used

Fig. 1. Freox degradation of JPL *Syncom* propellant
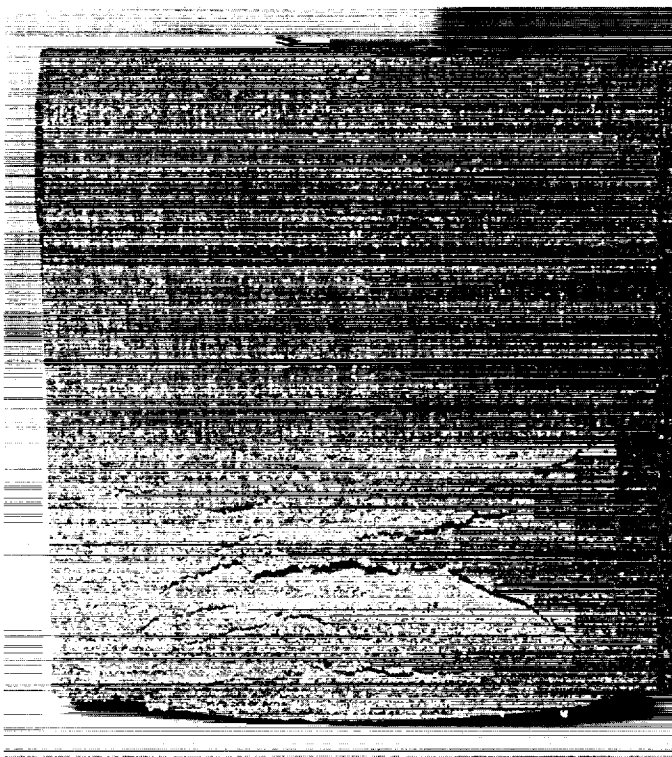
Fig. 2. Freox degradation of Thiokol HC Surveyor propellant

did not show compatibility with the Freox mixture, this approach was abandoned.

Other items noted, but not necessarily related, to the failure are as follows:

(1) Diffusion occurred more rapidly through the Thiokol propellant than through the JPL propellant.

(2) A brown liquid substance condenses out on pressure chamber walls and propellant surface during the propellant exposure (this was also noted in the propellant mixture when Freox was used for sterilization in the mixture).

(3) Ammonium perchlorate was heat-sterilized before mixing into the propellant to sterilize any occlusions in the crystals (145°C for 36 hr). (This was also done in many other batches of propellant with no evidence of adverse effects.)

The preceding effort is admittedly a very preliminary study but the results indicate that long periods of exposure of these two propellants to Freox under pressure will cause polymer breakdown. It is possible that other polymeric materials may also be attacked by this mixture

which suggests an area that should be investigated when ETO exposure is contemplated.

B. Applications Technology Satellite (ATS) Motor Development

R. G. Anderson

1. Introduction

In January 1963 the Jet Propulsion Laboratory initiated a development program to provide a solid propellant apogee rocket motor for a second-generation *Syncom* satellite. This program, under the management of the Goddard Space Flight Center, was designated *Advanced Syncom*. It was to result in a spin-stabilized, active repeater communications satellite weighing about 750 lb, operating at synchronous altitude (22,300 mi) which would handle voice communications, teletype, and monochrome and color television signals.

In January 1964 the *Advanced Syncom* communication program was redirected to include a number of experimental instruments in addition to the original communication instruments. This expanded program is the Applications Technology Satellite (ATS) program and will result in a general-purpose satellite capable of operation at synchronous altitude with experimental instruments in the areas of meteorology, communications, radiation, navigation, gravity gradient stabilization, and various engineering experiments. For those satellites to be placed in synchronous orbit, JPL will provide a solid propellant rocket motor to provide the final required velocity increment at the apogee of the elliptical transfer orbit. This rocket motor is designated the JPL SR-28-1 (steel chamber) or JPL SR-28-3 (titanium chamber) rocket motor. It is presently intended that only the JPL SR-28-3 unit will be delivered for flight use.

Previous reports of progress on the development of this motor have been published in *SPS 37-20* to *37-33*, Vol. V and *SPS 37-34* to *37-35*, Vol. IV.

2. Program Status Summary

The motor development program calls for static firing of four heavywall motors and 25 flightweight motors,

including two with flight-design titanium chambers, prior to conducting a nine motor qualification program. To date, the four heavywall motors plus 18 flightweight motors have been static-fired, four of which were under simulated high-altitude conditions at Arnold Engineering Development Center (AEDC) Tullahoma, Tennessee. All of the flightweight motors tested to date have been with Type 410 chromium steel chambers, with the exception of Dev. G-8T, which used the first titanium chamber in a static test.

During the period October to November 1965, one apogee motor (G-8T) was static-tested, and the dynamic-model (D-2T), using a titanium chamber was put through the required vibrational environmental test at Edwards Test Station (ETS).

3. Vibrational Test of Dev. D-2T

The inert loaded motor T-3 (Dev. D-2T) was put through the required environmental vibration test for the ATS JPL-SR-28-3 solid propellant rocket motor at the ETS facility on November 15 to 17. The purpose of the test included the following:

(1) To demonstrate the use of an automatic equalizer system on the random noise portion of the test.

(2) To obtain the motor response to the required input levels and compare these with past tests using the Type 410 chromium steel chamber.

(3) To establish detail procedures for conducting the vibrational test at ETS for the nine qualification motors.

The motor was attached to a special head-end fixture which was mounted to the shaker as shown in Fig. 3 (lateral test mode). For the axial test the shaker is rotated 90 deg, and the motor is mounted on top of the shaker.

For the sinusoidal vibrational test, the test article was vibrated in the lateral and axial directions through a 2.0 octave/min frequency sweep rate to the following input levels:

Frequency, cps	Level, g rms
15–250	2.12
250–400	3.54
400–2000	5.30
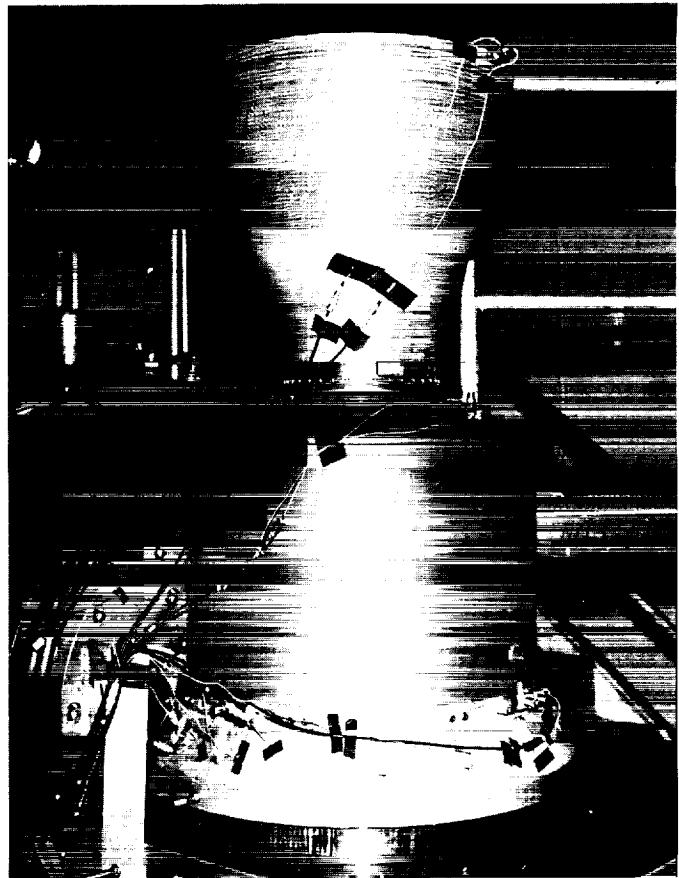
Fig. 3. Lateral axis installation of Dev. D-2T for a vibration test

For the random noise portion of the test the following inputs were used:

Lateral direction		
Band-limited frequency, cps	PSD[1]	g rms
20–80	0.04 g^2/cps	
80–1280	Linear increase from 0.04 g^2/cps to 0.07 g^2/cps at 0.61 db/octave	11.2
1280–2000	0.07 g^2/cps	

Axial direction		
Band-limited frequency, cps	PSD[1]	g rms
20–1000	0.1 g^2/cps	9.9

[1] Power spectral density

For the random tests the acceleration cutoff rate at both frequency extremes must be 24 db/octave, or greater.

Preliminary analysis of the data indicates that all of the input requirements of the test have been met, with the exception of the levels in a frequency band between 1700 and 2000 cps for the random lateral-axis test. In this region there is some noise that the servo-control system is not able to control, resulting in a 2- to 3-db gain above the required tolerances. An investigation of this situation is in progress, but at this time no absolute cause has been determined.

4. Static Test of Dev. G-8T

On November 16, 1965 the first titanium motor chamber was employed during the static test of Dev. G-8T. The motor was instrumented with 25 thermocouples on

Table 2. Dev. G-8T Static test data summary

Test conditions	
Type of test	Atmospheric
Test location	ETS, E-60
Test date	November 9, 1965
Run no.	E-485
Physical hardware	
Chamber S/N	T-4
Nozzle S/N	F-36
Igniter S/N	SYC-264
Squib S/N	S/N-61
Pressure data	
Characteristic velocity, $\overline{W}^*$ (ft/sec)	4967.2
Chamber pressure integral, psia-sec	8898.8
Igniter basket peak pressure psia (msec)	2863 at 12.0
Chamber ignition peak pressure, psia (msec)	296 at 22.0
Chamber starting pressure, psia (sec)	99.3 at 0.21
Chamber run peak pressure, psia (sec)	262.7 at 30.8
Time data	
Ignition delay, msec	3.0
Action time, sec	41.76
Run time, sec	41.92
Nozzle dimensions	
Nozzle throat diameter, in.	4.084 (initial)
	4.114 (final)
	4.099 (average)
Nozzle exit diameter, in.	24.138
Nozzle throat erosion, %	1.47

the case, 2 thermocouples on the nozzle attachment ring and the normal pressure transducers and load cell associated with the performance portion of the test. A summary of the important test data is given in Table 2. A summary of temperature-time histories of the thermocouples located on the case have been plotted in Figs. 4, 5, 6, 7 with each thermocouple location given in Fig. 8.

The thermocouples used on this test were employed to evaluate the present insulation configuration and titanium chamber combination. The criterion for evaluating the insulation is based on the margin of safety between the actual working stress in the chamber and the titanium yield strength at the operating temperature. Fig. 9 depicts typical chamber temperature–chamber stress titanium-yield-strength curves for the test of Dev. G-8T. Thermocouple B-5, which gave the minimum margin of safety, was used for the above analysis. Fig. 9 shows that the insulation configuration employed in this test provides adequate chamber temperature control during the motor burning phase. Actual stress levels were kept well below the minimum yield strength for titanium.

5. Storage Motors

Three ATS motors have been processed and placed in long-term storage at ambient conditions. The units represent the current ATS apogee motor, except that the motor chambers are Type 410 chromium steel rather than the flight-type 6Al-4V titanium. The motors were subjected to the same pre- and post-loading requirements as presently planned for the flight units. The empty motor assemblies were balanced prior to propellant casting. After propellant loading, the units received radiographic inspection, hardware alignment, center-of-gravity and moment-of-inertia determinations.

At 3-mo intervals each unit will be disassembled and inspected for any anomalies which may have occurred during storage. After 1 yr of storage, or prior to casting the first flight unit, one motor will be removed from storage and subjected to the environments of temperature cycle, booster acceleration, and booster vibration prior to static-firing. The remaining units will be removed from storage after 2 yr, subjected to the test environments and static-fired.

6. Static Test of Dev. G-9T

Dev. G-9T represents the final motor of the environmental series to be tested prior to casting the nine qualification charges. The motor was cast on October 27, 1965

MOTOR

POSITION
POTENTIOMETER

WORM GEAR
AND CAM

GRATING SHAFT
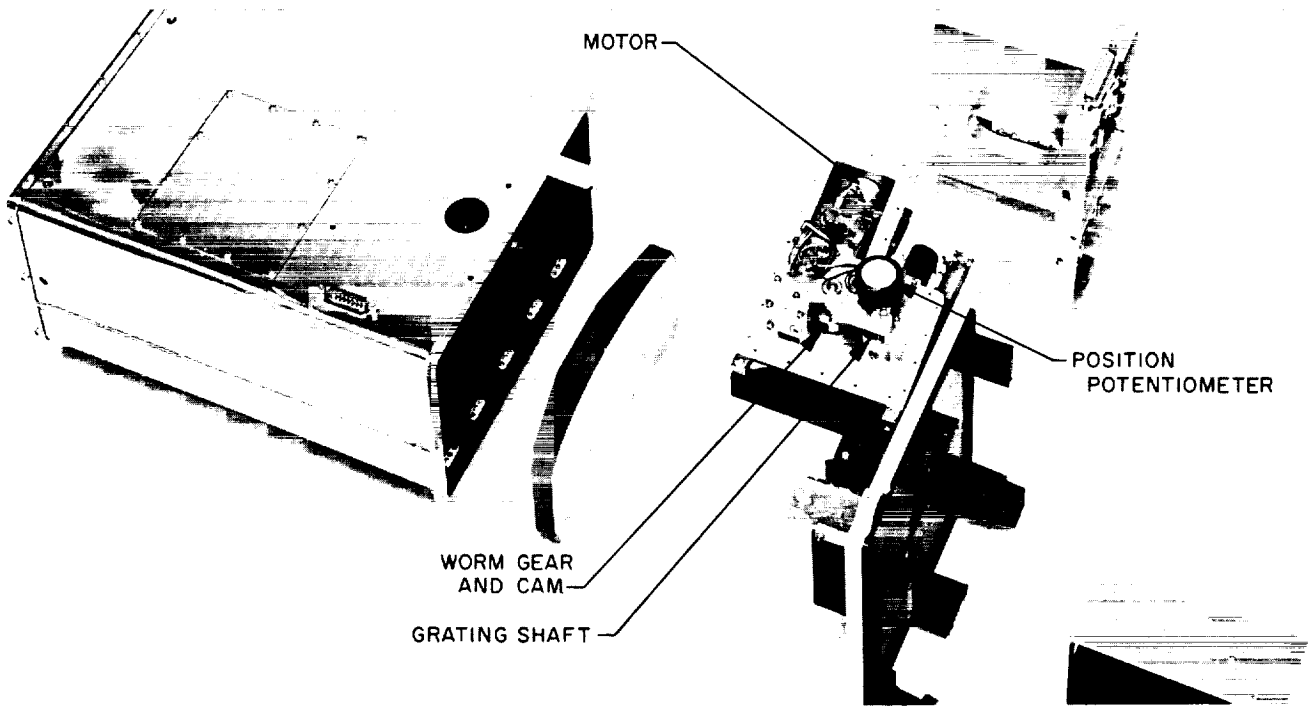
Fig. 4. External chamber temperature measurement, Dev. G-8T, Thermocouples A-1 to A-4
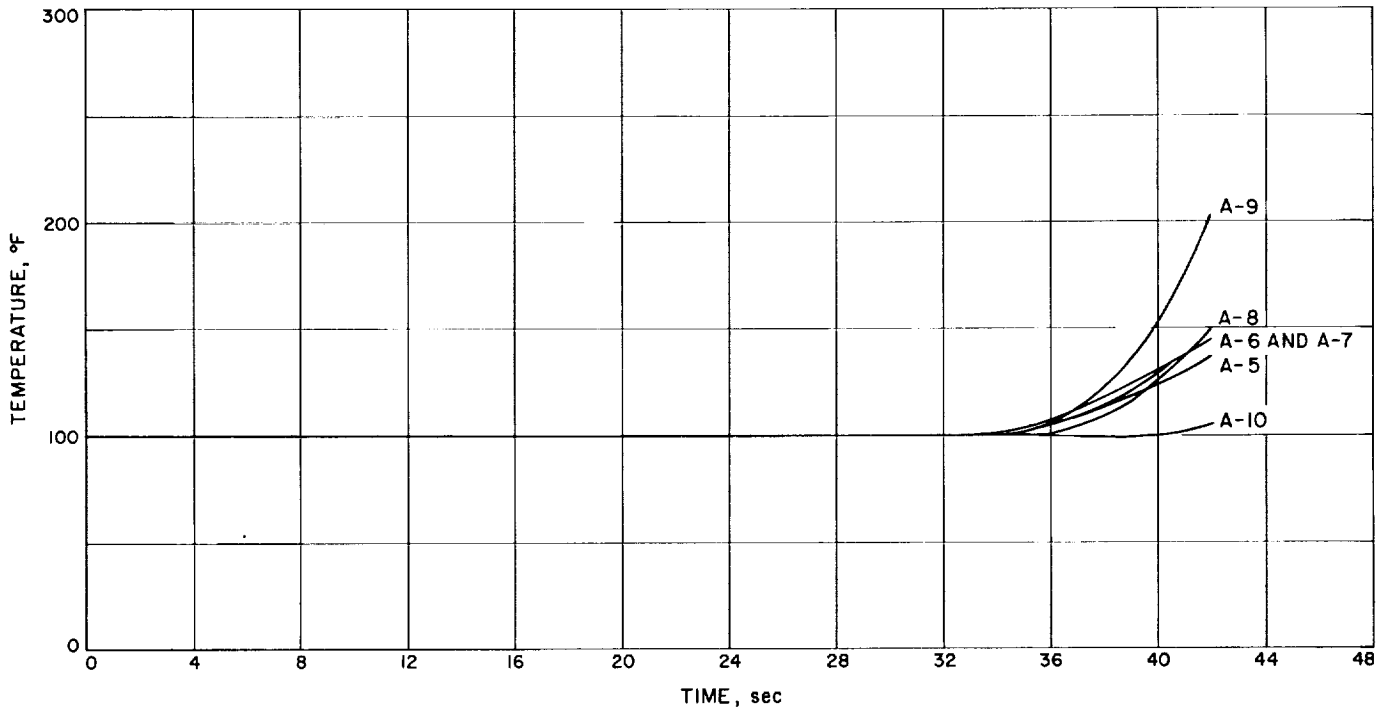
Fig. 5. External chamber temperature measurement, Dev. G-8T, Thermocouples A-5 to A-10
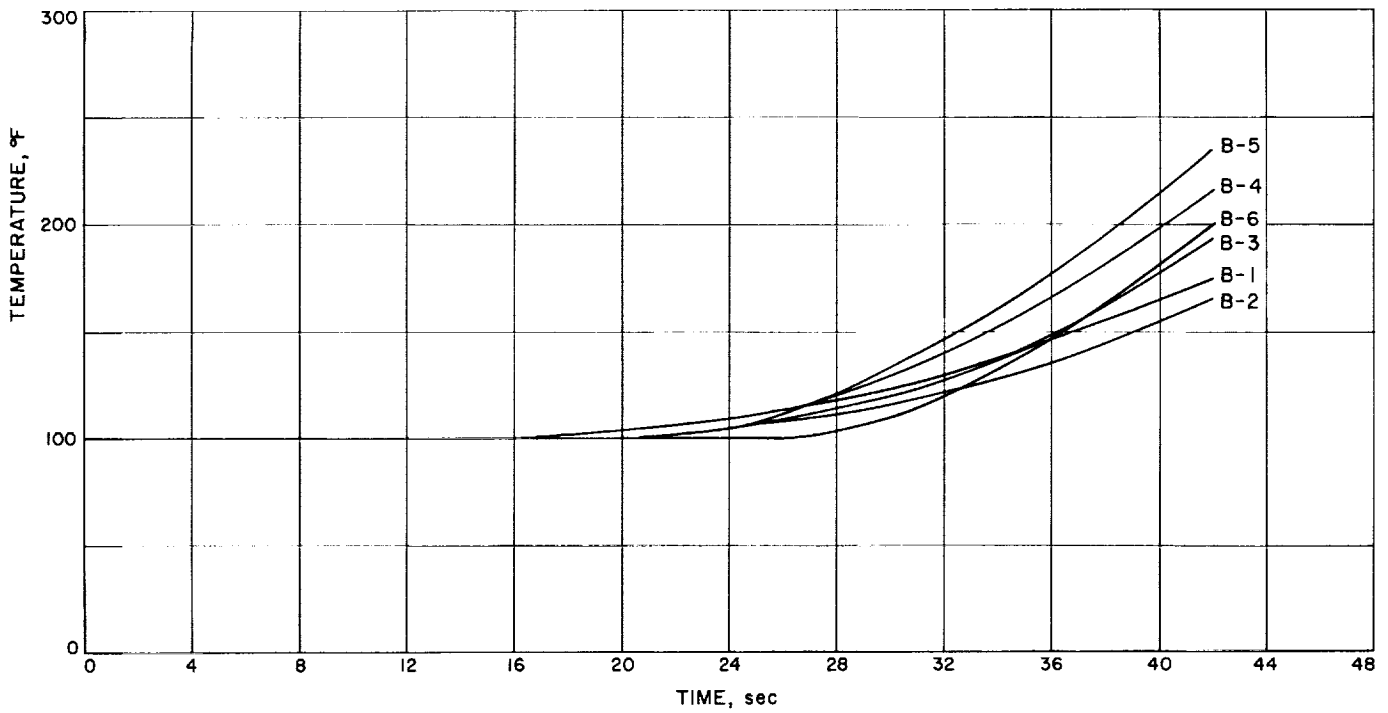
Fig. 6. External chamber temperature measurement, Dev. G-8T, Thermocouples B-1 to B-6
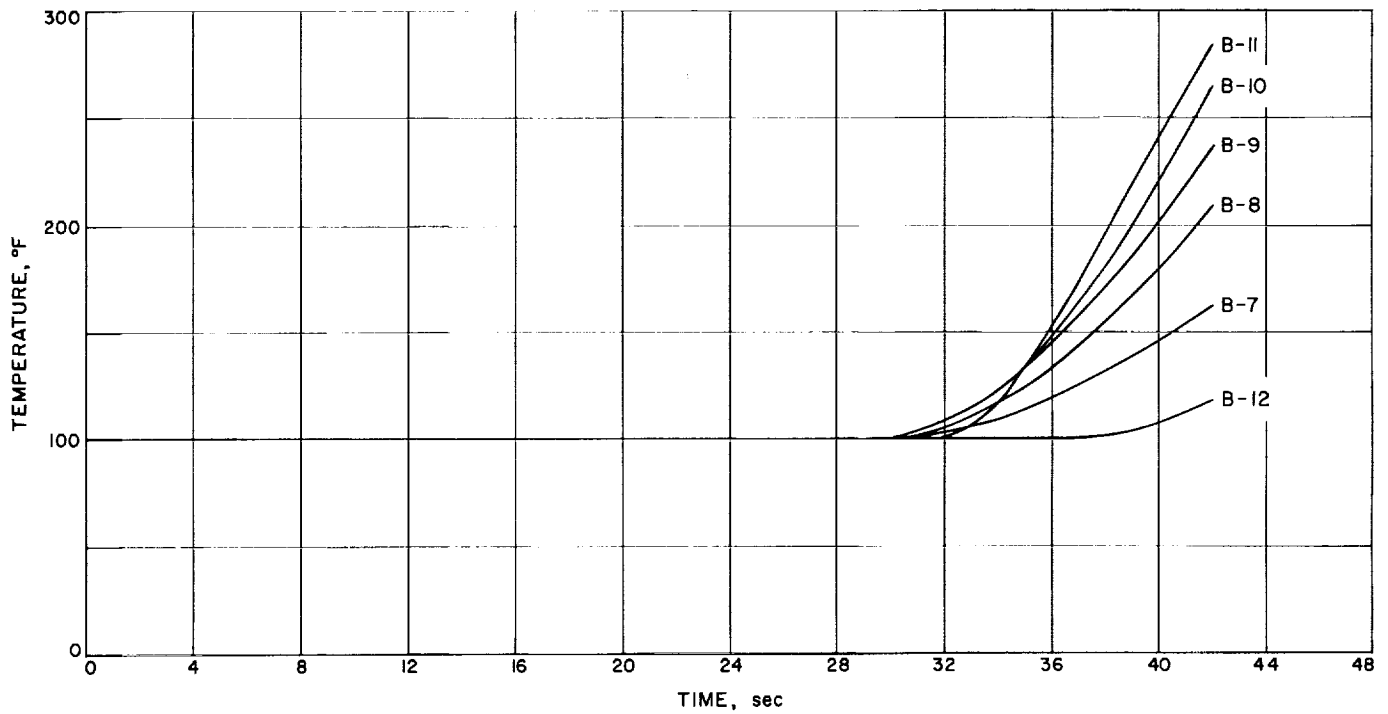
Fig. 7. External chamber temperature measurement, Dev. G-8T, Thermocouples B-7 to B-12
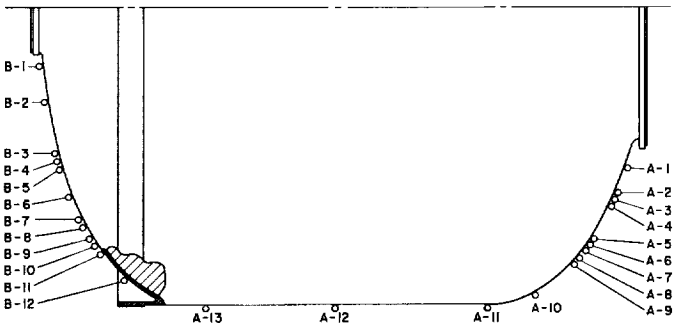
Fig. 8. Chamber thermocouple placement development code Dev. G-8T

Fig. 9. Chamber temperature and strength data at thermocouple location B-5, Dev. G-8T

and is presently going through the temperature cycling environmental test. It will start a simulated booster acceleration test on December 7, 1965 and a simulated booster vibration test on December 14, 1965, after which its static-firing will be conducted on the spin test stand at ETS on January 4, 1966. For this test, the motor will

be temperature-conditioned to a temperature of 10°F and static-fired at a spin rate of 150 rpm.

SPACE SCIENCES DIVISION

XIII. Space Instruments

A. Infrared Spectrometer for High-Altitude Balloon Experiments

P. W. Schaper and D. D. LaPorte

1. Introduction

A spectrometer has been developed which is suitable for making observations of the infrared region of the spectrum from aircraft and high-altitude balloons. The instrument scans the spectral region from 3.6 to 4.8 μ in the first order of the diffraction grating with a spectral resolution of about 100 ($\lambda/\Delta\lambda$). Second-order information can also be obtained on a separate detector system.

The capability of the instrument was demonstrated on a balloon flight in March 1965. This flight was the culmination of a design effort which started as a result of a proposal to search for features in the infrared spectrum of the planet Mars between 5 and 6.6 μ in the first order and 2.5 and 3.3 μ in the second order. The instrument was designed to be part of the scientific payload of the *Mariner IV* spacecraft; however, the final payload did not include an infrared experiment.

Subsequently, scientific interest shifted to the properties of the Martian atmosphere. A determination of the atmospheric temperature profile from the intensity variations in the 4.3-μ CO_2 band was proposed for the next Mars fly-by. The old *Mariner IV* spectrometer was to be used as a breadboard to demonstrate the feasibility of the temperature sounding by this technique. The instrument was modified as necessary to cover the shorter wavelength region. This article describes the instrument as it existed for the balloon flight in March 1965.

2. Instrument Description

a. Entrance optics. The spectrometer optics are arranged in the manner suggested by Ebert, employing a single collimating mirror and straight entrance and exit slits. The field of view of the instrument has been measured to be 12 × 12 deg square. It has an aperture of f/3. An exploded view of the spectrometer is shown in Fig. 1.

Radiation emitted by the unknown source is directed onto the entrance aperture of the instrument by a plane mirror. This reflection is required to keep the instrument
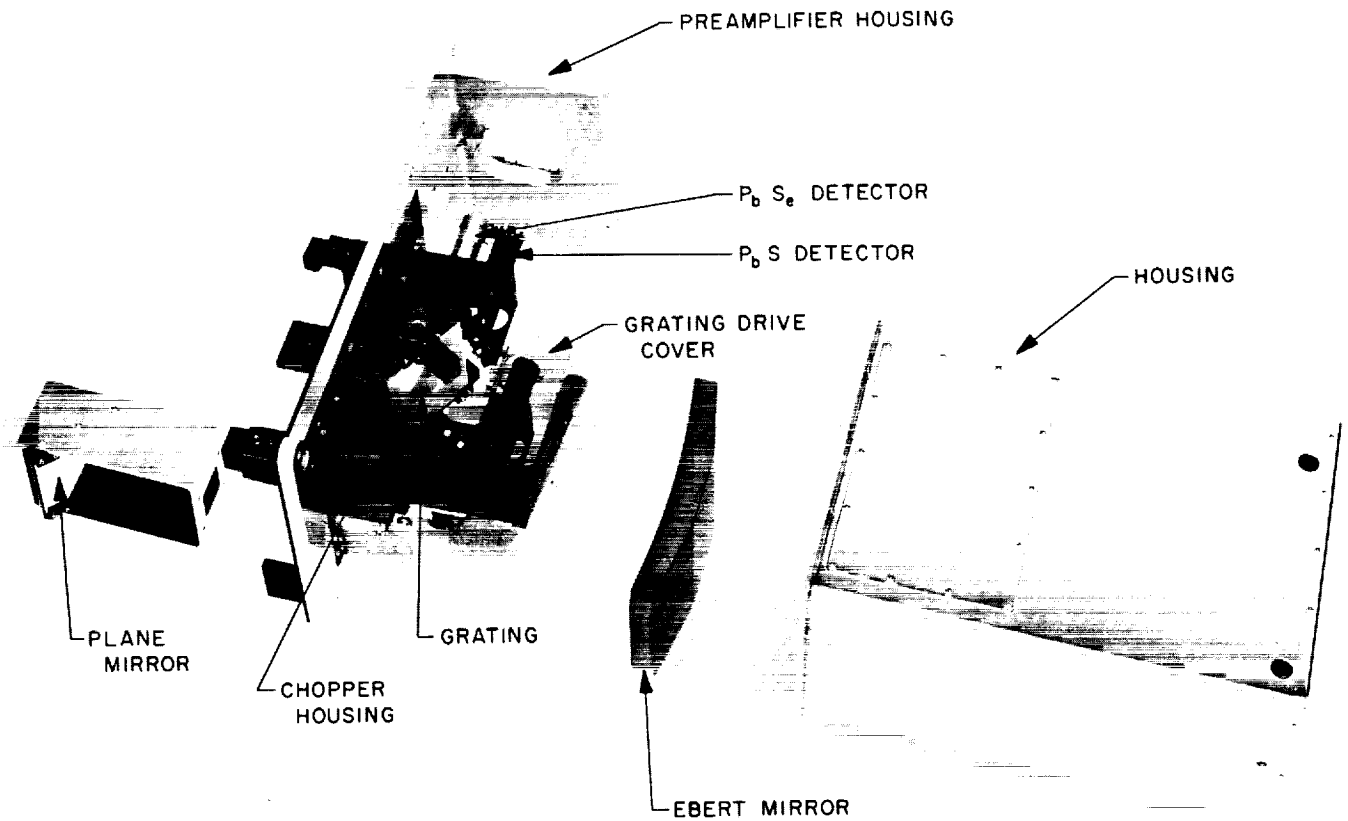
Fig. 1. Infrared spectrometer exploded view

in a horizontal orientation while looking downward. The horizontal orientation of the spectrometer is necessary to keep liquid nitrogen in the detector flask, mounted vertically in the instrument. A highly reflective gold coating on the mirror keeps its radiant energy contribution at a low level.

The radiation entering the instrument is periodically interrupted by a tuning fork chopper. A view of this device as seen from the radiation source is shown in Fig. 2. The chopper blades are reflecting toward the source and painted black on the side facing the detectors. The fork is tuned to a frequency of 330 cps. Chopped radiation entering the instrument is limited by the entrance slit, which is 10 mm high and 3 mm wide.

Periodically, during the course of an experiment, a shutter is placed in front of the chopper entrance slit assembly. The side of the shutter facing the chopper acts as Lambert diffuser for controlled radiation from an

in-flight calibration lamp. This lamp consists of a tungsten filament encased in a silicon envelope. The shutter assembly can be seen in Fig. 3, positioned over the entrance aperture, with chopper and entrance slit removed.

Radiation passing through the entrance slit falls on a 10-in. focal length mirror and is collimated and directed toward the diffraction grating. The mirror is gold-plated on a quartz base of such size that the grating is filled completely by the collimated beam.

The grating is a replica with 240 lines/mm, blazed at 3.75 μ. It is driven by a phase-locked, speed-controlled, dc motor through a cam arrangement which permits a rapid 1-min scan to be made through the spectrum for automatic gain setting purposes and subsequent data scan of 9-min duration. Grating position is sensed by a potentiometer coupled to the grating shaft. Thus the potentiometer measures grating angle. The grating drive mechanism is shown in Fig. 4.
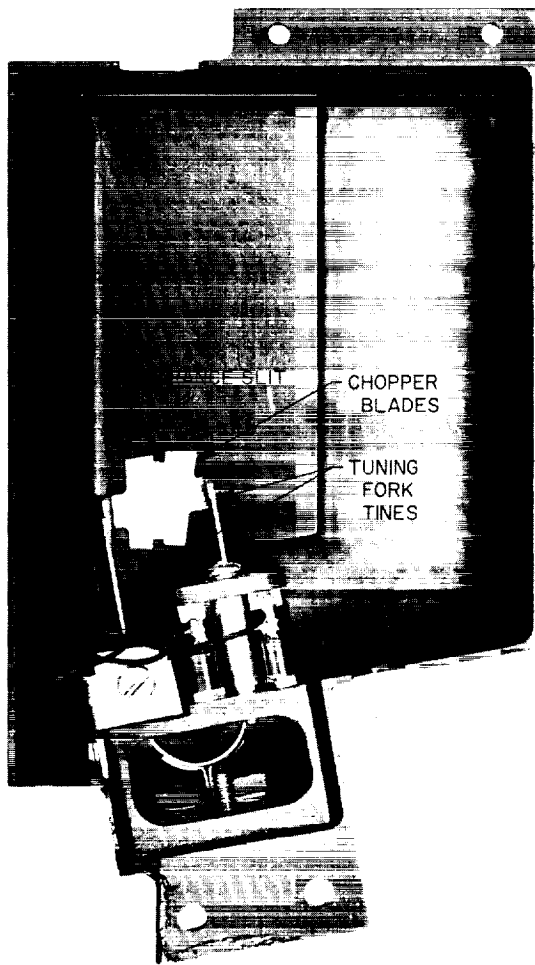
Fig. 2. Infrared spectrometer tuning fork chopper

Fig. 3. Infrared spectrometer details

b. Exit optics and data conditioning. The Ebert mirror focuses monochromatic radiation from the grating on the detectors. First-order radiation passes through a dichroic beam splitter which deflects second- and higher-order radiation at 90 deg to the incident beam. It is then deflected by a plane mirror onto an exit slit which is evaporated on the PbSe detector assembly. The detector assembly consists of the elements shown in Fig. 5. The field lens and hemispherical immersion lens form an image of the grating on the detector flake. The transmission characteristics of the dichroic beam splitter are shown in Fig. 6. Fig. 7 shows the detectivity of the PbSe detector at liquid nitrogen temperature.

Second-order radiation is deflected by the dichroic beam splitter, as indicated by the curve in Fig. 8, and directed through a slit onto a PbS detector with an optical system similar to that of the PbSe detector. The PbS detector performance is shown in Fig. 9. It is operated at ambient temperature.

As the block diagram in Fig. 10 indicates, both detectors employ basically the same data conditioning system. The signal from the detector is fed through a variable gain preamplifier and amplifier, demodulated in synchronism with the chopper and filtered. The dc voltage from the post detection filter is sequenced into an analog-to-pulse-width converter along with information from other data channels and then recorded or telemetered, as the experiment requires. All data channels are sampled once at the beginning of each run; the rest are sampled in sequence 4, 3, 2, 1, 4, 3, 2, 1, etc. The sampling frequency is determined by the time constant of the data system.

The gain of the preamplifiers is set during the rapid scan through the spectrum. At the beginning of this scan the gain is set at the maximum value. As a larger output of the detector is sensed by the peak detector, the gain setting is reduced. Before the end of the rapid scan is reached, a level sensor on the grating position potentiometer disconnects the memory from the peak detector and the gain indicated at that time remains constant for the entire run.
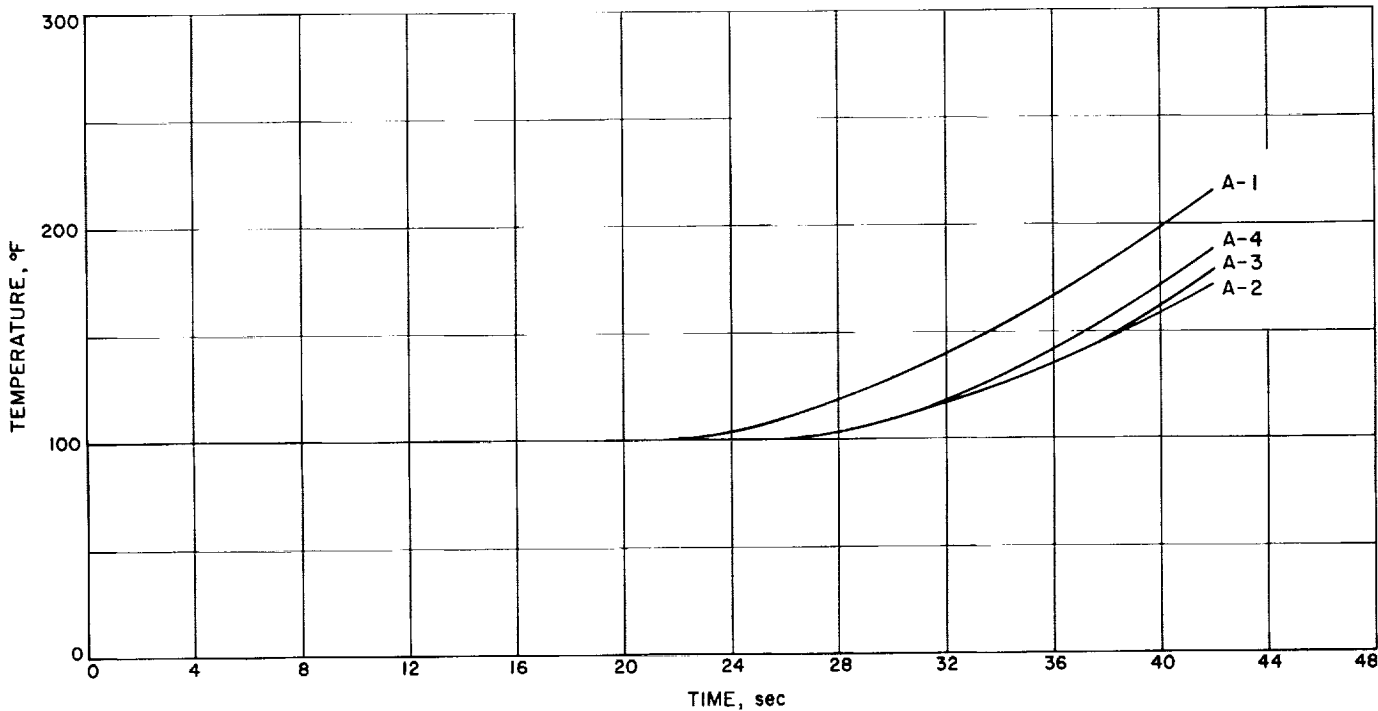
Fig. 4. Infrared spectrometer grating drive mechanism

Fig. 5. PbSe detector
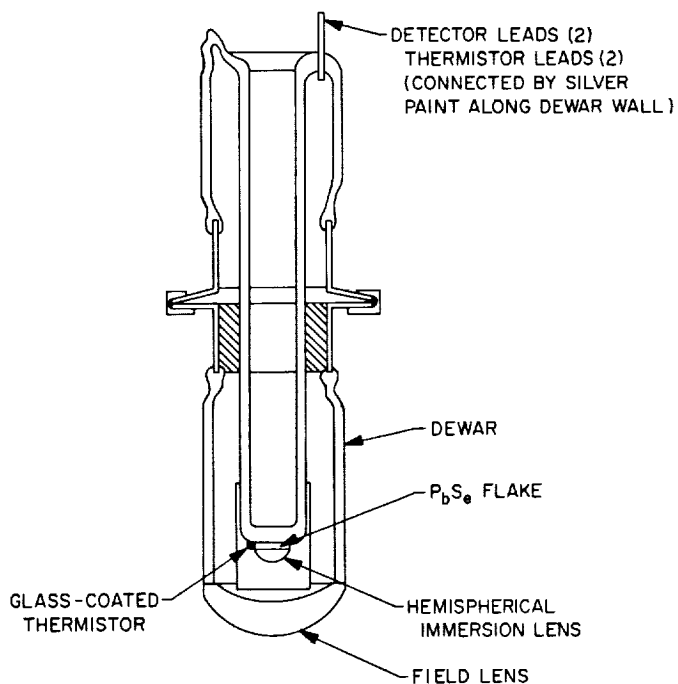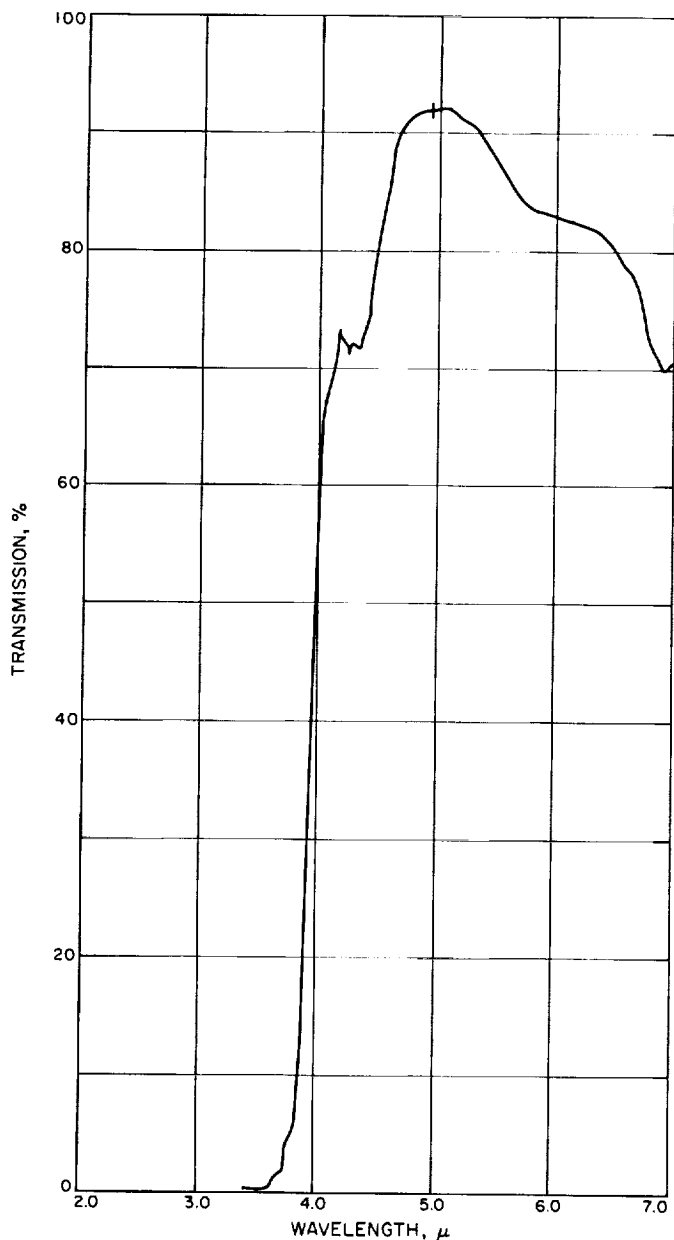
Fig. 6. Transmission characteristics of the dichroic beam splitter

Fig. 7. Detectivity of the PbSe detector

Simultaneously a "high calibrate" signal is inserted at the detectors with the detector bias voltage turned off. A fixed amount of time later a "low calibrate" signal is inserted. This is subsequently removed and the detector bias voltages are turned on again as the grating has started on the slow scan of the spectrum. In this manner the gain of the system and its linearity have been determined and pulse-width data can be reduced to detector output voltage.
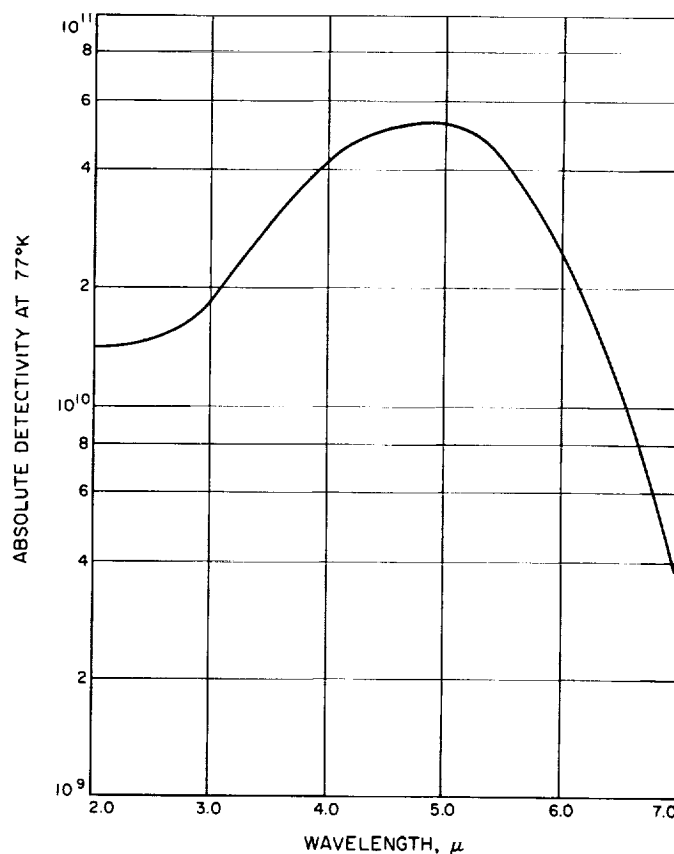
3. Calibration and Data Reduction

The pre-flight and post-flight calibrations consisted of establishing the instrument profile and transmission characteristics under simulated conditions of temperature and ambient pressure. A "black body" at temperatures between 0 and $-20°C$ was placed in front of the instrument and the instrument output was recorded. Instrument temperatures were controlled by varying the temperature inside the calibration chamber.

Early tests showed that the gain of the preamplifiers was set by noise pulses which did not show up on the data during the slow scan period. A noise-clipping network was added, unfortunately without being previously tested due to lack of time. This network caused considerable non-linearities in the data system. Another problem occurred in the calibrate-level system. The signal out of the analog-to-pulse-width converter did not remain constant long enough for a true level to be determined by the computer. Furthermore, the phase relation between the reference and calibrate signal was uncertain when the instrument was operated under high-altitude
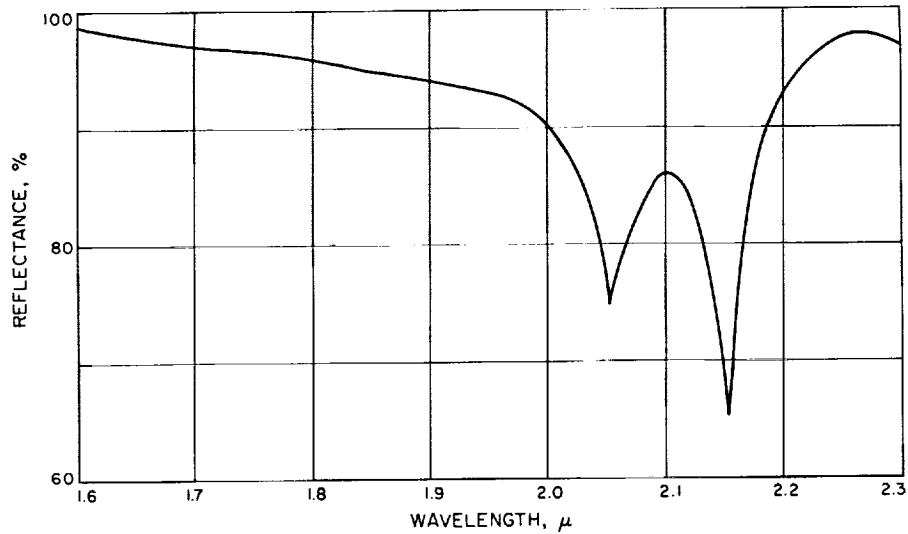
Fig. 8. Reflectance of the dichroic beam splitter

environment and thus the calibrate level could not be used as a true voltage calibration.

For these reasons the calibration lamp was used as a secondary energy standard in the data reduction. In post-flight calibrations lamp energies were determined from experimental comparisons with laboratory black body data. The laboratory run was made with the same gain setting as the flight "calibrate" run, as determined from the calibrate signal level. Subsequently, the "data" runs were reduced by taking ratios of calibrate signal levels and applying these ratios to the pulse-width data to convert it to energies.

Wavelength data were obtained by recording the grating position potentiometer output when monochromatic radiation was registered by the detector. The monochromatic radiation was intense enough compared to the black body background the instrument saw, so that direct peak reading was possible.

4. Flight Results

The first flight tests of the instrument occurred on a U-2 high-altitude aircraft in October 1964. The purpose of the flight was to get an engineering evaluation of the instrument under high-altitude conditions. Several engineering problems were discovered, such as the setting of the gain on noise pulses, an unreliable sequencing mechanism, non-constant motor speeds, etc. These were corrected during the testing period. The most significant positive result obtained was the fact that the instrument proved to be more sensitive than expected.
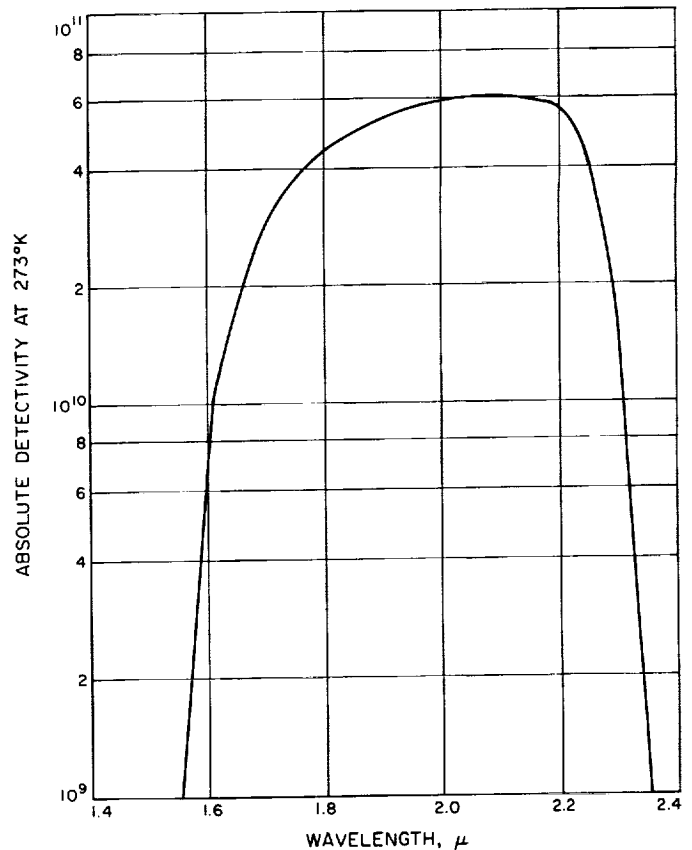
Fig. 9. Detectivity of the PbSe detector

All the pulse-width data had to be tape-recorded on a new airborne machine purchased for flight duty. A failure in the play-back mechanism of the machine made reduction of the pulse-width data impractical. Analog
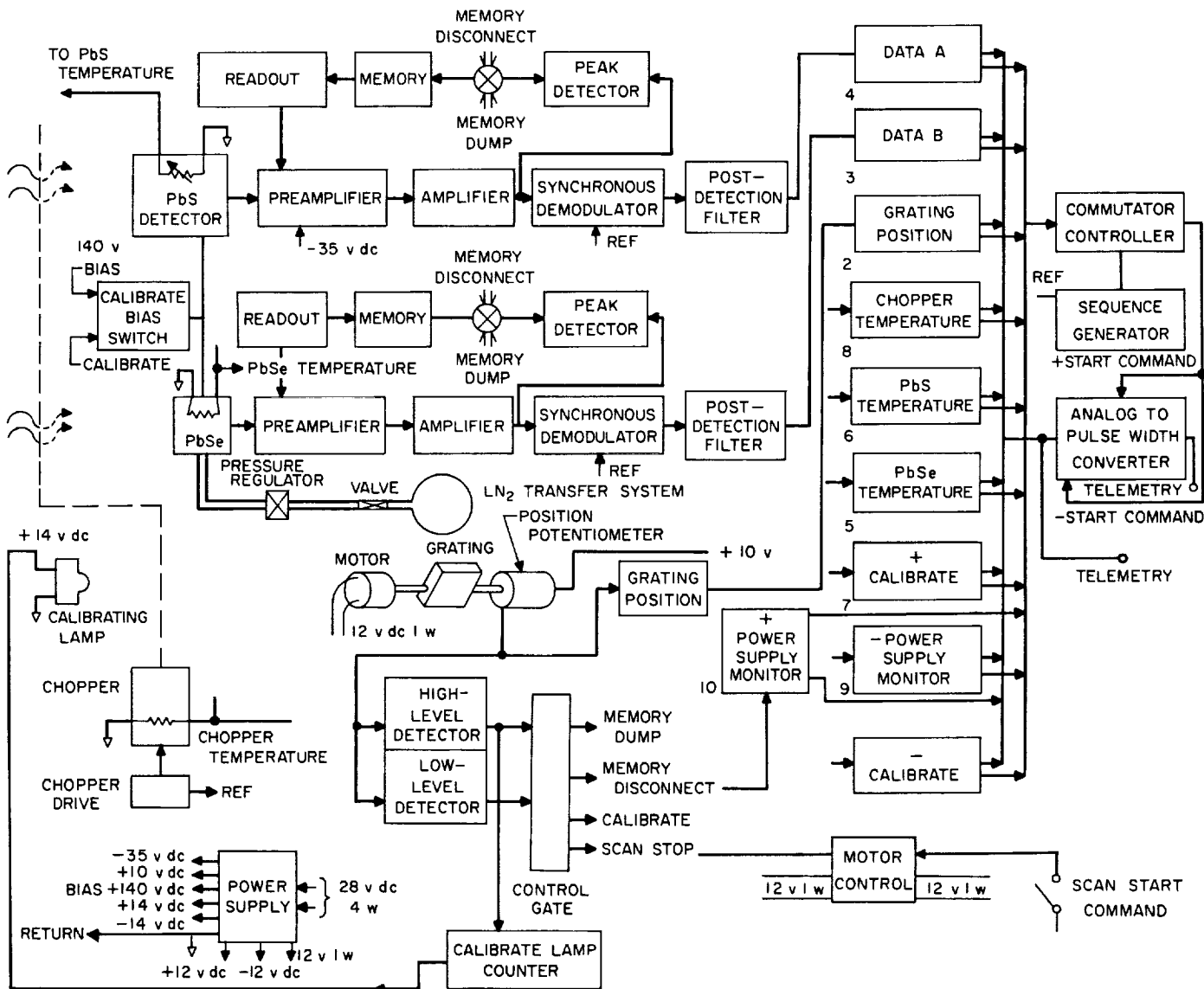
Fig. 10. Infrared spectrometer block diagram

data were recorded during flight on an oscillograph recorder and were analyzed for reproducibility and quality. Lack of accuracy of these recorded data did not permit their reduction to intensities.

On March 10, 1965 the spectrometer was flown on a high-altitude balloon from Sioux Falls, South Dakota. The balloon reached an altitude of 100,000 ft. Final results of this flight will be published at a later date, when a complete analysis of the data has been made. Preliminary results show that the instrument performed well under the high-altitude environment and meaningful results were obtained. A typical intensity plot is shown in Fig. 11.
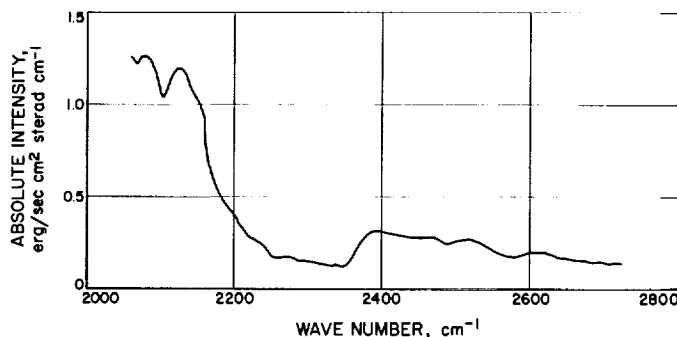
Fig. 11. Typical spectral data from the March 1965 balloon flight

In a radiometric sense the data appear to be quite uncertain when compared to intensities calculated from atmospheric temperatures. In retrospect it is evident that more precise temperature measurements are required to give better quality radiometric data. As a result of the flight and subsequent data analysis, critical temperature control points have been determined.

Wavelength calibration of the spectrometer has been shown to be quite satisfactory, as evidenced by the proper position of some of the known atmospheric absorption bands. No use has been made of the raw data obtained from the PbS detector from either flight, since this information is of little interest to the temperature sounding experiment.

B. Precision Testing of Pulse Height Analyzers

W. J. Schneider

The objective of this supporting research and advanced development program is the development of a pulse height analyzer capable of supporting nuclear experiments in orbiting, impacting, or soft landing vehicles. The following material relates to the determination of some of the pertinent properties of such an analyzer but omits any discussion of the analyzer mechanism.

A pulse height analyzer converts pulse height into numerical values for digital processing. Pulse height is a continuous variable while only a discrete set of numerical values are allowed.

For the purpose of example, consider an analyzer which accepts pulses from 0 to +10 v in amplitude and provides a 9-bit binary output. The question of data interpretation is: If a pulse height is indicated as having magnitude n (the nth quantization number or channel), what was the height in volts? In the ideal case the answer is

$$n \cdot 10 \cdot 2^{-9} < V_p < (n + 1) \cdot 10 \cdot 2^{-9} .$$

The fact is that the channel boundaries are not linearly related as the inequality indicates, neither are they as stable with time and temperature nor as distinct as indicated. Linearity and stability as used here have the usual connotation of analog circuit design since analog portions of the analyzer are required to have linearity and stability at least an order of magnitude better than the channel width or about 0.01%. The discrete character of the channel boundary is destroyed by noise within the analyzer. In the analyzer of the example an rms noise level of 1 mv results in the boundaries being normally distributed with a standard deviation equal to 0.2 channel.

Gamma-ray spectrometer performance may be evaluated by making a calibration spectrum with a known radiation source. The results of such a test do not provide definitive information on the pulse height analyzer performance. The most easily implemented test of the analyzer alone is the sliding pulser differential linearity measurement. The sliding pulser consists of a linear ramp generator, a sampling switch, and a pulse-forming network. The ramp voltage spans the input range of the analyzer slowly while the sampling and pulse generation is performed rapidly. The ideal result is a rather dense, uniformly distributed, pulse height spectrum. The number of pulses occurring in each quantization level or channel, as determined by the analyzer, is then plotted. The number of pulses in each channel is linearly related to the channel width and hence to the slope of the quantization function.

Difficulties here are twofold: (1) this method is extremely sensitive to systematic errors; i.e., the ramp generator and sampling mechanism must be an order of magnitude more linear than the analyzer; and (2) the method is insensitive to non-systematic errors; i.e., the presence of noise, in either the pulser or analyzer, does not destroy the uniformity of the pulse height spectrum and hence goes undetected.

Analytical testing seems to require that the results of each pulse analysis be determined on an individual basis. Automatic equipment capable of performing such tests is described below.

1. Equipment Description

The calibration equipment consists of three elements—pulse generator, data handling equipment, and a controller (Fig. 12). As its name indicates, the controller
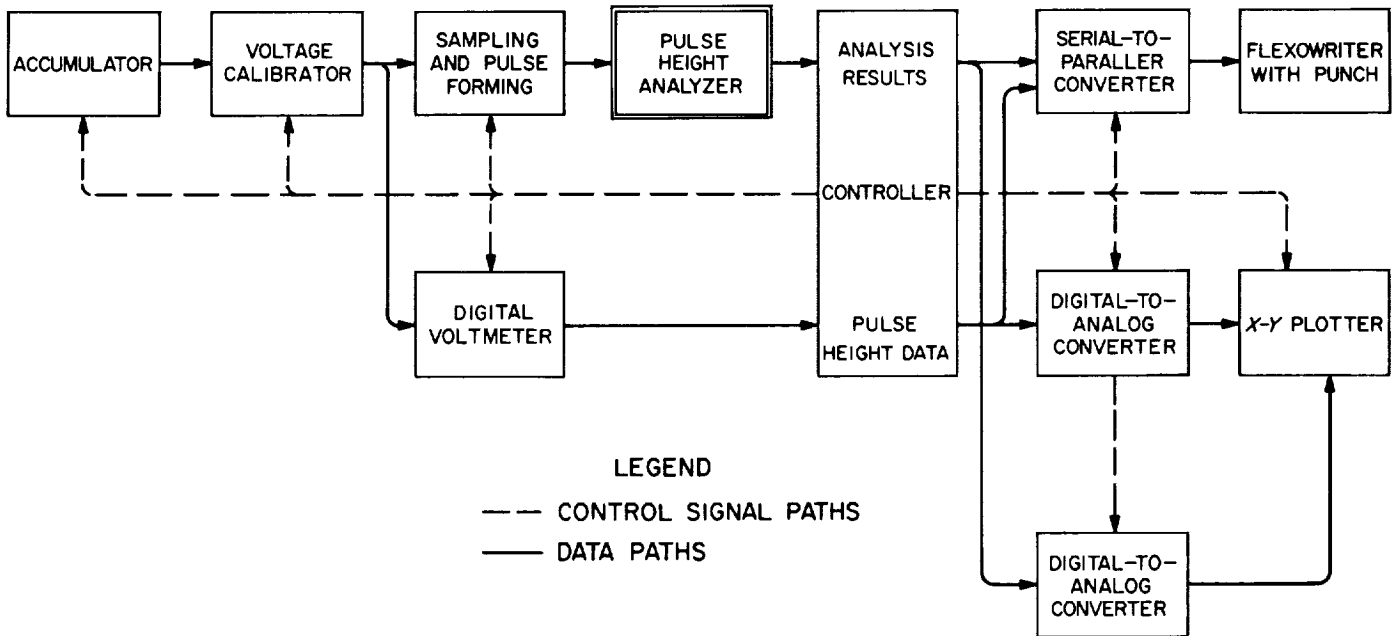
Fig. 12. Pulse height analyzer precision calibrator block diagram

controls the pulse height analyzer under test and the other elements of the calibration equipment as follows:

(1) Generates a pulse.

(2) Increases pulse level.

(3) Determines results of pulse height analysis.

(4) Records input pulse height and result of pulse height analysis.

(5) If pulse count is not complete, returns to (1).

(6) Stop. (Quick look data resulting from the sequence is shown in Fig. 13.)

a. Pulse generator. The pulse generator consists of three commercially available equipments: (1) a 6-digit decimal preset accumulator; (2) a programmable voltage calibrator supplying 0 to 50 v with 10-μv resolution and 2-amp capacity; and (3) a pulse generator consisting of a mercury relay and linear, passive pulse-forming network.

$5 \cdot 10^5$ distinct pulse levels are available, nearly 1,000 uniformly spaced amplitudes within each quantization level of the pulse height analyzer of the example. The accumulator provides the programming to the voltage calibrator for increasing or decreasing the pulse level between preset limits with decade adjustable steps.
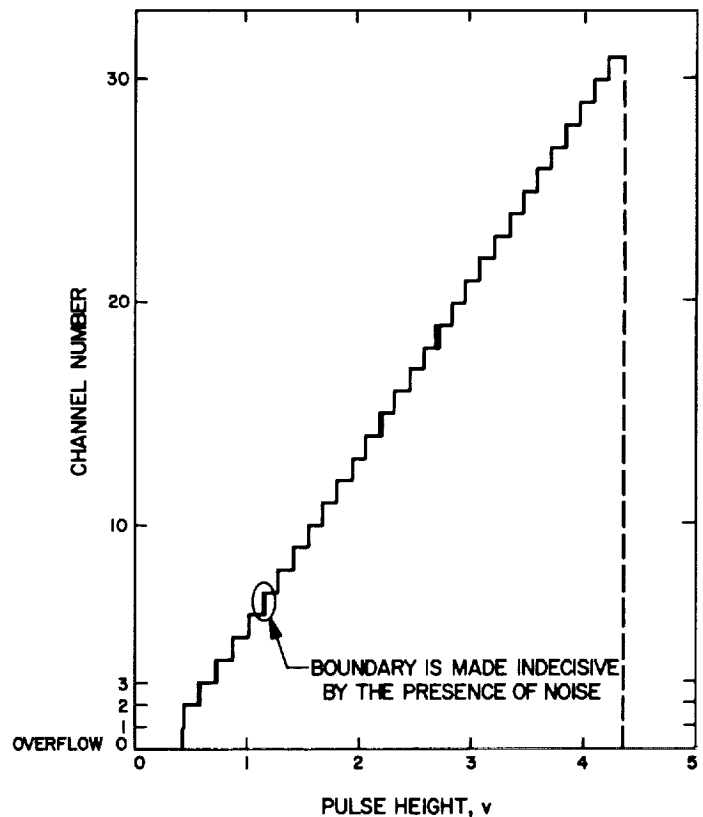
Fig. 13. Quick look data resulting from precision calibration of the pulse height analyzer

An alternate and simpler mechanization was achieved utilizing a charge pump-type stair-step generator for the voltage source. In this case increments of a few millivolts are quite easily obtained, affording 50 or more pulse amplitude levels within each quantization level of the pulse height analyzer of the example. In either case the voltage source is instrumented by a digital voltmeter whose output is used as a measure of the peak pulse voltage.

For the most precise work a mercury relay is used to switch a sampling capacitor from the charging voltage to the pulse-forming network. An optional pulser incorporates a semiconductor chopper and is capable of pulse rates of 10^5 per second, at reduced precision.

b. Controller. The form in which the results of the pulse height analyses are available is determined by the equipment under test. If a quantizer alone is under test, then either a gated clock signal or a binary register will contain the required information. If a complete analyzer is under test, then its memory must be searched for the address corresponding to the analysis quantization number. The required waveform interface equipment is tailored to the equipment under test and to the controller.

The controller itself is sufficiently versatile to accommodate input data in a number of forms and to provide a uniform output format. A change of modes allows the controller to perform the standard sliding pulser test for differential linearity.

c. Data output. Raw data is available in tabulated form, in graphic plots for quick look, and as punched tape for computer analysis. The paper tape format is compatible with tape-controlled card punch and with Fortran II.

2. Equipment Usage

The calibration equipment will be used throughout the JPL pulse height analyzer development program. The ability to obtain the required information from the quantizer or from the completed analyzer has been provided early in the program. The objective of this effort is to provide uniform measure of performance capability and a continuity of data from development through spacecraft testing. The definitive history of analyzer performance is intended to provide a sound basis for the interpretation of experimental data.

C. Phase Stable Frequency Doubler

J. S. Bunn

1. Introduction

This article describes a method of generating a second harmonic phase reference from a given signal at the fundamental frequency, f_0.

Certain instruments, such as the fluxgate magnetometer, are characterized by sensor excitation at f_0 and sensor output signal at the second harmonic, $2f_0$. In order to achieve a high signal-to-noise ratio and convert the sensor signal $(2f_0)$ to dc, it is often desirable to utilize a synchronous demodulation technique. Therefore, a signal at the frequency $2f_0$ is required to provide a frequency and phase reference for the demodulator.

2. The Method Proposed

The method proposed here for generating a reference second harmonic signal is based on the polarity relationships in the fundamental signal and its quadrature. Fig. 14a shows these two signals. Fig. 14b is a plot of the relative polarity of the two signals in Fig. 14a. Note that the function shown in Fig. 14b is a second harmonic square wave. If, therefore, we are given a sinewave of frequency f_0, we can generate the second harmonic reference in two steps: (1) phase shift the original signal
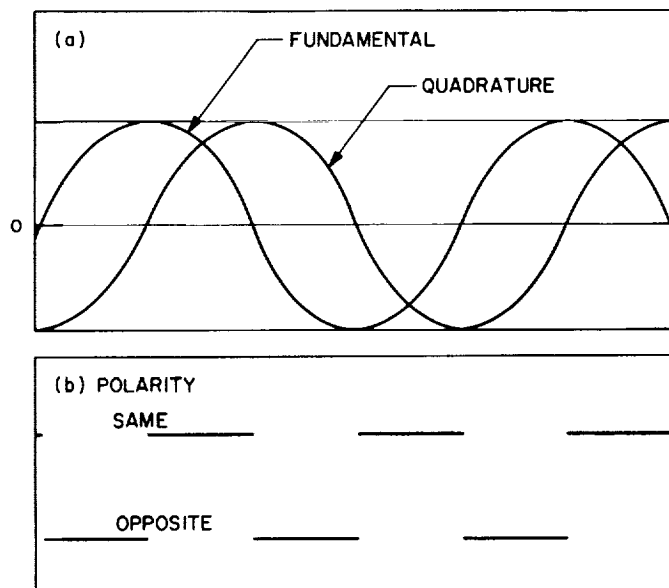
Fig. 14. Timing relationships

$\pi/2$ radians to obtain a quadrature signal, and (2) pass the signal and its quadrature through an *exclusive or* gate to obtain the relative polarity. Fig. 15 shows a block diagram which performs these necessary steps. The waveforms at various points in the block diagram are shown in Fig. 16. The waveform at E can be used to drive a switching type demodulator.

3. Other Methods

Second harmonic generators in general use in space instruments are of two basic types, as shown in Figs. 17a and 17b. Fig. 17c shows a third type which is not generally used due to mechanization problems, i.e., it is difficult to perform the operation in a distortion-free manner over wide temperature extremes.

The scheme shown in Fig. 17a accumulates phase drift in the filter following the binary ($\div$ 2) circuit. The phase

Fig. 15. Method for generating a reference second harmonic signal
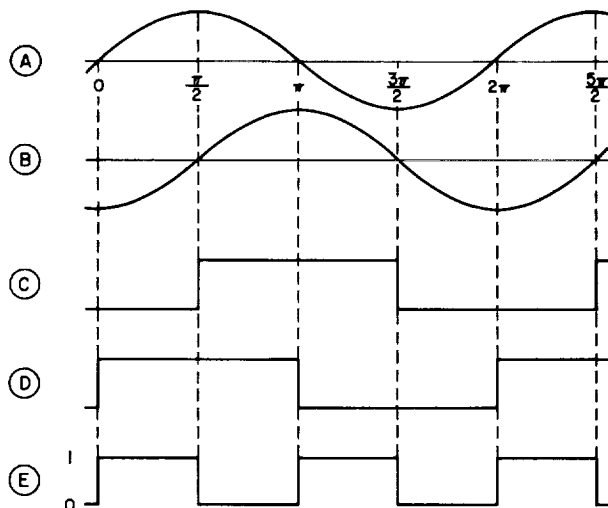
Fig. 16. Waveforms for Fig. 15

Fig. 17. Second harmonic demodulator reference schemes

drift in this filter is doubled when referenced to the second harmonic. The method of Fig. 17b requires a fairly high Q-tuned filter following the nonlinear element in order to extract the second harmonic. Thus, the control of phase drift over a wide temperature range is a very real problem.

4. Advantages of the Method Proposed

The method proposed (Fig. 15) avoids these phase stability problems by using elements which can be easily designed to be phase stable over large temperature excursions and even for large ($\pm$ 50%) frequency changes.

Consider the circuits required. The phase shifter can be built using simple operational amplifier techniques such that the phase shift is 90 deg (within a fraction of a degree) over a wide frequency range. The squaring circuit is an ac-coupled zero crossing detector. This device and the *exclusive or* circuit are designed to have sufficient speed to satisfy the phase accuracy requirements.

The technique offers advantages even where phase stability is not important because it is easy to implement and no tuning procedures are required.

5. Typical Application

A fluxgate magnetometer can be used to illustrate the use of the second harmonic generator described. Fig. 18 is a block diagram of this instrument using the technique described. A typical value for f_0 might be about 20 kHz.

The circuitry required to perform the functions inside the dashed line of Fig. 18 is shown in Fig. 19. The 2N2432 and 2N943 transistors in Fig. 19 constitute a composite shunt demodulator and may be inverted if lower offset voltage is desired.

6. Summary

A method has been described which generates a second harmonic phase reference from a fundamental frequency. The method can be implemented without the use of tuned circuits.
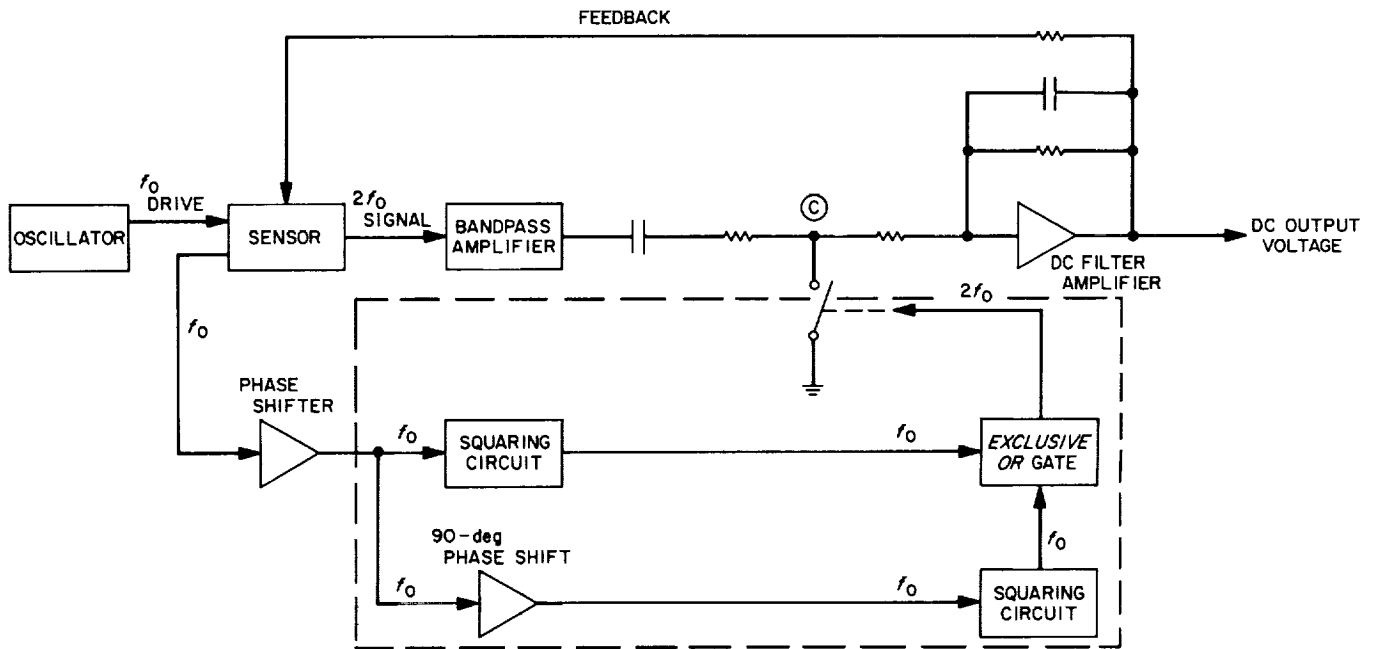
Fig. 18. Fluxgate magnetometer application
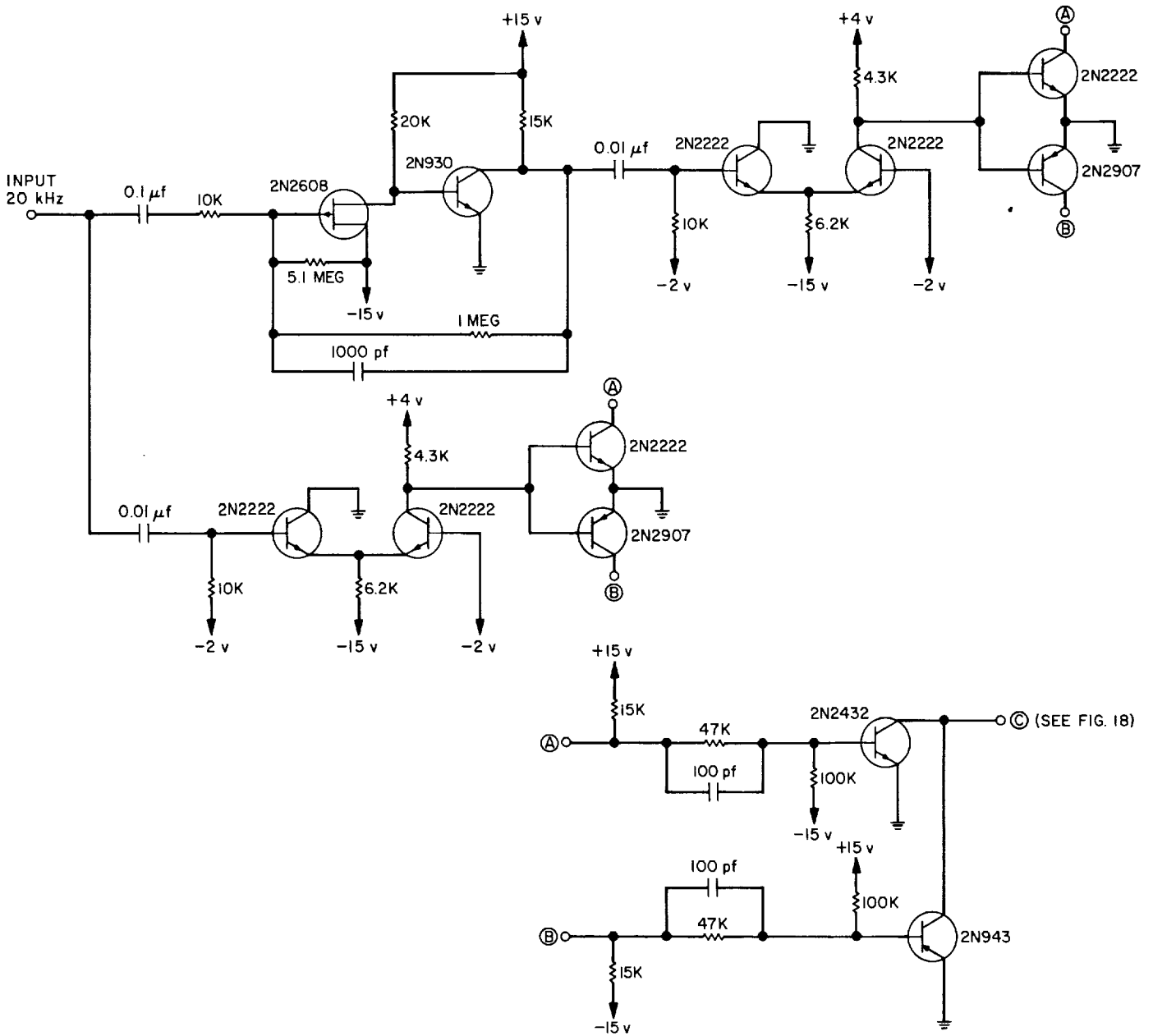
Fig. 19. Circuitry for 20-kHz doubler

XIV. Lunar and Planetary Sciences

A. Observed Venus Microwave Brightness Temperature in the Wavelength Internal 1.25–1.45 cm

D. E. Jones

From July 3 through July 19, 1964, Venus was observed at eleven frequencies between 20.6 and 24.0 Gc/sec (20.60, 20.63, 21.0, 21.5, 22.0, 22.20, 22.23, 22.50, 23.0, 23.5, and 24.0). The measurement program was carried

out using the 30-ft millimeter wave radio telescope that is located at the JPL Goldstone Facility.

Average values of Venus brightness temperatures, T_{BV}, are plotted in Fig. 1. The error flags noted are those due to statistical fluctuations only. For purposes of comparison of the results at neighboring frequencies, the relative uncertainties associated with the antenna efficiency determination amount to about $\pm 7^\circ$K. The absolute temperature accuracy due to uncertainties in the antenna
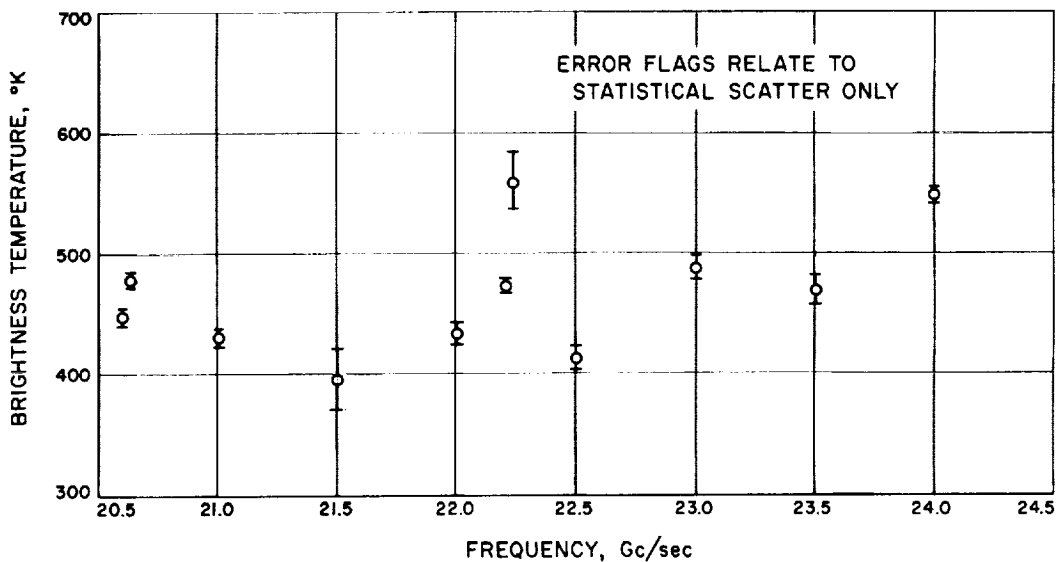
Fig. 1. Venus brightness temperature vs frequency

efficiency is estimated to be about $\pm\,25°K$. A broad minimum centered at between 21.5 and 22.0 Gc/sec is suggested as well as the possibility of fine structure.

Table 1 lists the reduced brightness temperatures and atmospheric zenith opacity as a function of date, as well as the number of observations and relative probable

Table 1. Venus brightness temperature vs date and frequency

Date (1964)	Frequency, Gc/sec	No. of readings	Atmospheric extinction coefficient τ, db/air mass	Venus brightness temperature T_{BV}, °K	Probable error PE, °K		Date (1964)	Frequency, Gc/sec	No. of readings	Atmospheric extinction coefficient τ, db/air mass	Venus brightness temperature T_{BV}, °K	Probable error PE, °K
7/3	21.0	24	0.0286	437.0	10.8			22.50	41	0.0610	377.3	7.8
	22.0	28	0.0420	496.3	15.0			24.00	40	0.0406	547.0	15.9
	22.5	18	0.0450	421.8	7.7		7/11	21.00	39	0.0505	436.7	9.3
	23.0	38	0.0405	463.6	10.1			22.23	40	0.0850	492.7	14.8
	24.0	14	0.0290	504.5	18.8		7/12		(Poor observing condition)			
7/4	20.63[a]	20	0.0255	445.4	15.08		7/13	21.00	18	0.0675	516.1	22.4
	20.63[b]	16	0.0255	505.8	23.8			24.00	20	0.0682	555.4	27.5
	22.00	20	0.0440	436.28	19.9		7/14	20.60	20	0.0277	437.6	16.5
	23.50	41	0.0377	403.60	11.4			21.50	20	0.0440	317.4	8.0
7/5	21.50	40	0.0400	385.0	9.8			22.00	20	0.0523	413.5	21.3
	22.50	40	0.0483	423.1	9.7			22.20	20	0.0550	477.0	13.9
	24.00	32	0.0320	579.0	25.0			22.50	20	0.0543	359.6	11.3
7/6	20.60	38	0.0255	467.1	12.9			23.50	18	0.0420	423.9	24.0
	21.50	16	0.0400	388.8	13.4			24.00	19	0.0353	525.4	19.4
	22.00	33	0.0470	435	17.6		7/15	21.00	23	0.0368	457.3	6.4
	23.50[c]	19	0.0390	440.2	16.3			22.00	24	0.0555	409.7	10.9
	23.50[d]	20	0.0390	489.4	12.0			22.50	23	0.0576	423.6	10.6
7/7	21.50	40	0.0382	356	7.7			23.00	24	0.0528	469.6	13.5
	22.00	38	0.0450	450.8	12.6			24.00[e]	23	0.0380	520.8	18.2
	22.23	32	0.0460	619.8	17.2			24.00[f]	20	0.0380	570.1	39.1
	23.00	40	0.0430	541.7	15.6		7/16	20.60	38	0.0300	448.9	14.3
7/8	20.6	39	0.0288	421.6	10.1			21.50	47	0.0486	529.7	9.8
	22.50	37	0.0562	430.12	13.3			22.20	18	0.0600	471.4	17.5
	24.0	42	0.0377	578.2	17.7			23.50	39	0.0470	518.4	21.6
7/9	21.0	36	0.0580	403.0	18.1		7/17	21.00	39	0.0370	433.2	9.0
	22.0	18	0.0844	449.1	15.7			22.00	38	0.0560	442.6	10.0
	22.23	40	0.0890	567.1	18.8			23.00	39	0.0530	492.8	14.6
	23.50	41	0.0877	511.18	18.7			23.50	38	0.0452	508.0	25.5
7/10	20.60	40	0.0320	459.8	17.4		7/18	24.00	38	0.0340	548.2	18.2
	22.00	28	0.0585	271.8	11.9		7/19	22.50	33	0.0502	458.3	12.4

[a] 11:46–12:32 PDT
[b] 14:18–14:54 PDT
[c] 8:24– 9:20 PDT
[d] 11:26–12:46 PDT
[e] 9:10–10:08 PDT
[f] 15:09–16:05 PDT

errors of the daily means due to scatter alone. The data at 22.23 Gc/sec, while not entirely reliable because of a radiometer baseline sampling incompatibility, are nevertheless of interest because of the implications one obtains when the nature of the sampling incompatibility is studied in greater detail.

The data in Table 1 suggests a time variation which is particularly pronounced at several frequencies—the most noteworthy being 21.5 and 23.5 Gc/sec. It is seen, for example, that between July 14 and 16 the change in the Venus 21.5 Gc/sec temperature amounted to over 20 probable errors. A comparison of the data obtained at other frequencies on these two days, on the other hand, shows little change. Because of this fact and a careful consideration of the measurement procedure, it is not clear how such a variation can have been caused by any other than a change in the brightness temperature of the planet.

When both the 2800-Mc/sec solar flux and the Venus data are plotted versus date there is the suggestion of a correlation between the solar flux and the 21.5 Gc/sec data corresponding to a lag of between one and two days (Fig. 2). This is also true at 23.5 Gc/sec, but to a lesser

degree. Following this suggestion the cross-correlation coefficients were computed for all of the frequencies observed except 22.20 and 22.23 Gc/sec, the latter being unreliable for such a study. The results for 20.6, 21.5, 22.0, and 23.5 Gc/sec are shown in Fig. 3. Again a lag of between one and two days for the 21.5 and 23.5 Gc/sec data is suggested.

In interpreting the results of such a cross-correlation, one must bear in mind the relatively few number of Venus data points. As a result, the 99% confidence intervals are quite large and indeed the suggested correlation may be purely a result of poor statistics. If borne up by a subsequent longer measurement program during the coming conjunction, however, very exciting possibilities exist in regards to suggested constituents in the Venus atmosphere.

Because the magnetometer aboard the *Mariner II* spacecraft failed to detect the presence of any magnetic field attributed to Venus (Refs. 1 and 2), one is tempted to interpret the brightness temperature results obtained here in terms of charged particle excitation of microwave rotation lines of trace constituents in the Venus atmosphere. The upper limit values for the magnetic moment

Fig. 2. Venus brightness temperature and 2800-Mc/sec solar flux vs date

Fig. 3. Cross correlation between Venus brightness temperature and 2800-Mc/sec solar flux

Assuming that solar protons do impinge on the Venus atmosphere, they will obviously be stopped in the regions of the atmosphere where the "solar-wind" pressure is balanced by the ion pressure, presumably where the atmospheric pressure is quite low. One would then expect that if molecules are stimulated which have microwave rotational lines, the observed planetary temperatures at these lines would show a tendency to be correlated with solar activity, whereas (with little pressure broadening at these high altitudes) at nearby frequencies, the effect would be less. This appears to be the case as one readily notes from various aspects of the data. One would also infer that the degree of correlation should increase as more of the illuminated disc of the planet is observed. Unfortunately, the observing program could not extend beyond the several weeks after conjunction because of other antenna commitments, and to a lesser extent because of the sensitivity of the radiometer.

Negative voltages for the output baseline level of the radiometer at 22.23 Gc/sec were observed from time to time and consequently data at this frequency are in general considered to be unreliable because the data sampling equipment used did not sense the voltage polarity. The temperatures at 22.23 should, in fact, be considered as lower limit values. This was not the case at any of the other frequencies. A study of the 22.23 Gc/sec daily histograms suggests that the temperature values listed in Table 1 for July 7 (620 ± 17°K) and for July 11 (493 ± 15°K) are probably all right. At 22.20 Gc/sec it is noted that the temperatures of 477 ± 14°K on July 14 and of 471 ± 18°K on July 16 are obtained. Comparing these results and those obtained at other nearby frequencies suggests that the water line may have been observed in emission and that either the observed temperature differences with time at 22.23 Gc/sec are due to errors in setting the local oscillator frequency, or that the brightness temperature of the planet at 22.23 Gc/sec varies with time. The latter seems to be the most likely possibility.

of Venus indicated by the *Mariner II* results are 1/10 or 1/20 that of the Earth (Refs. 1 and 2). However, a smaller magnetic moment is compatible with the data and indeed if the magnetic moment of Venus were less than about 0.6% that of the Earth, a solar proton flux with $N \sim 20/cm^3$ (Ref. 3) and $v \sim 800$ km/sec (average velocity for a travel time to the planet of $\sim 1\frac{1}{2}$ days) could be impinging directly upon the atmosphere of the planet. One could, in fact, turn the problem around and say that, if the suggested correlation of Venus temperatures with solar activity is confirmed, it will imply an even smaller value for the upper limit of the planet's magnetic moment.

Since the radiometer IF frequency and bandwidth were 60 Mc/sec and 10 Mc/sec, respectively, radiation from the planet was being received in two sidebands spaced 60 Mc/sec either side of the indicated observing frequency. At 21.550 and 23.445 Gc/sec are two of the microwave rotational lines of methyl alcohol, which are quite close to the upper and lower sidebands of 21.50- and 23.50-Gc/sec observing frequencies, respectively. Microwave lines due to organic molecules have been a possibility suggested by Kaplan (Ref. 4) and was one interpretation of the results of the *Mariner II* microwave

radiometer experiment[1] (Refs. 5 and 6). Kuzmin (Ref. 7) has also postulated the possible presence of lines of organic molecules to explain the indicated abruptness in the microwave spectrum obtained from terrestrial-based observations. A mechanism for the production of organic-type molecules as suggested by Robbins (Ref. 8) is also of importance. It is to be noted that other workers observing at wavelengths shorter than 1.5 cm have noted anomalous behavior with time in their data[2,3] (Ref. 9) and some have suggested a possible correlation with solar activity[2] (Ref. 9). Clearly, more measurements, particularly at greater phase angles, are required.

---

[1]Jones, D. E., Ph.D. dissertation, Brigham Young University, Provo, Utah, 1964 (unpublished).

[2]Copeland, J., private communication, 1962.

[3]Barrett, A. H., private communication, 1965.

If the interpretation suggested by the data presented here is substantiated by additional observations that will be conducted next conjunction, it is clear that:

(1) A means for identifying molecules in the Venus atmosphere for microwave rotational lines would be available and subsequent narrowband, closely spaced frequency measurements should allow estimates of the upper atmospheric structure (i.e., down to where the flux of particles from the Sun are stopped).

(2) Solar particle flux on Venus could be indirectly monitored by observing the brightness temperature at several closely spaced frequencies. Such information would contribute greatly to the study of the degree of homogeneity, etc., of solar particle-magnetic field phenomena in interplanetary space

References

1. Coleman, P. J., Sonett, C. P., Smith, E. J., and Davis, L., Jr., *Science*, Vol. 139, No. 3559, pp. 905–910, March 8, 1963.

2. Smith, E. J., Davis, L., Jr., Coleman, P. J., and Sonett, C. P., "Magnetic Measurements near Venus," *Journal of Geophysical Research*, Vol. 70, No. 7, April 1, 1965.

3. Neugebauer, M., and Snyder, C., "Solar Wind Measurements near Venus," *Journal of Geophysical Research*, Vol. 70, No. 7, April 1, 1965.

4. Kaplan, L. D., *Proceedings of the International Symposium on Molecular Structure and Spectroscopy*, Tokyo, Science Council of Japan, B407, 1–4, 1962.

5. Jones, D. E., *Transactions, American Geophysical Union*, Vol. 44, No. 4, pp. 886–887, December 1963.

6. Jones, D. E., *A Measurement and Analysis of the Brightness Temperature Distribution over the Planet Venus at Wavelengths of 13.5 and 19 Millimeters Utilizing Mariner II Spacecraft*, Technical Report No. 32-722, Jet Propulsion Laboratory, Pasadena, California, October 1, 1965.

7. Kuzmin, A. D., "On Venus Model with Cold Absorbing Atmosphere," *Izvestiya Vysshikh Uchehnykh Zavedenii, Radiofizika*, Tom 7, No. 6, pp. 1021–1031, 1964; translated in Document No. ST-LPS-10314, NASA, Goddard Space Flight Center, April 6, 1965.

8. Robbins, R. C., "The Reactive Products of Solar Hydrogen and Components of the High Atmosphere of Venus — A Possible Source of the Venusian Clouds," *Planetary and Space Science*, No. 12, pp. 1143–1146, 1964.

9. Tolbert, C. W., and Straiton, A. W., "35 GC/S, 70 GC/S and 94 GC/S Cytherean Radiation," *Nature*, Vol. 204, pp. 1242–1245, 1964.

XV. Bioscience

A. An Electrostatic Theory of Gaseous Anesthesia

J. King, Jr. and S. W. Benson

There has been much speculation concerning the role of gases in the mechanism of anesthesia (Refs. 1–4). Any attempt to explain gaseous anesthesia must account for the anesthetizing property of the rare gases, particularly xenon. The anesthetic potency of xenon is best known from the investigations of Cullen and Gross (Ref. 5), who demonstrated conclusively that the chemically inert gas[1] is an anesthetic agent for man at atmospheric pressure.

Of the theories which attempt to explain the mechanism of anesthesia, the hydrate theories of Pauling (Ref. 3) and Miller (Ref. 4) are the best known and most widely accepted. Pauling proposes the formation, mainly in the synaptic region of the nerve network, of hydrate microcrystals which are formed by crystallization of the encephalonic fluid. He suggests that these hydrate microcrystals could trap some of the electrically charged side-chain groups of proteins and some of the ions of the encephalonic fluid, interfering with their freedom of motion and

with their contribution to the electrical oscillation. This interference could increase the impedance offered by the network to the electrical wave and cause the level of electrical activity of the brain to be restricted to that characteristic of anesthesia and unconsciousness.

Pauling supports his hypothesis by correlating both the partial pressure of anesthetic agents in equilibrium with their hydrate crystals and ordinary ice and water at 0°C and the anesthetizing partial pressure for mice with the molar refraction of the anesthetic agents. However, he stated that the mechanism of anesthesia could not be the single formation of hydrate crystals in the brain, since none of the hydrates of anesthetic gases thus far studied is stable at physiological temperature and pressure. In order to account for the formation of the hydrate microcrystals, Pauling suggested that some substance, or substances, in addition to the anesthetic molecules, are able to take part in hydrate formation and add stability to the hydrate structure. Charged side-chains of protein which are present in the brain fluid could possibly stabilize the hydrate microcrystals at temperatures not much lower than normal mammalian body temperature (37°C).

Miller also considered the possible role of hydrates in anesthesia. He found a correlation between the dissociation pressure of gaseous hydrates at 0°C and anesthetic

---

[1]Although xenon has been found to form chemical compounds with the halides, it is extremely improbable that xenon's anesthetic properties are based on any type of chemical reaction. We, therefore, consider it as being chemically inert.

potency, but he estimated that 34 times the pressure needed at 0°C is necessary for the existence of the hydrates of the smaller gases at body temperature and 1,000 times the pressure for the larger gases. Because of this and the fact that several of the gases would liquefy at 37°C, Miller concluded that it is doubtful that crystalline hydrates could form during anesthesia even if they are stabilized by protein side-chains as suggested by Pauling. As an alternative, Miller proposed the formation of "icebergs" or "ice covers," which he described as being formed when gas molecules orient water around themselves, much as proteins do, so that the molecules are surrounded by a shell of water which is in a more highly ordered state than is the remainder of the water. The support for this type of "iceberg" comes from the fact that liquid water is known to have some type of ice-like structure, which may or may not be like the structures of the hydrates, but with "free" water molecules occupying the center of the cages (Ref. 6). Miller also showed a correlation between the anesthetic partial pressure and that portion of a hypothetical surface covered with structured water (ice-cover) in the different stages of anesthesia. Miller contends that his "icebergs" differ from Pauling's microcrystals in being much smaller and not crystals in the usual sense.

However, the two theories are quite similar and suffer from the same deficiencies. One of these is the stability of the "hydrates" at 37°C. It is rather doubtful that the "icebergs" are any more stable at body temperature than Pauling's microcrystals.

A second and somewhat more serious drawback to the theories is the fact that many gases which have anesthetic properties do not form hydrates. The most notable of these is diethyl ether but the list also includes hydrocarbons higher than propane, aromatic hydrocarbons, various ethers, and halogenated derivatives of these compounds. Also, there are a number of gases, such as H_2S, SO_2, Cl_2, O_2, $CHCl_2F$, C_2H_5F, etc., which form hydrates, but are not known to have anesthetic properties. In all cases, these gases exhibit acute toxic reactions at pressures lower than those necessary for narcosis.

As an alternative to the hydrate theories of Pauling and Miller, we suggest that direct electrostatic interaction between the anesthetizing gases and the charged sites on the non-aqueous phase of the encephalonic fluid could lead to the symptoms of anesthesia and unconsciousness. This suggestion is based on an extension of the electrostatic theory of physical adsorption (SPS 37-33, Vol. IV, pp. 208–212). Although the theory was originally devised to explain the adsorption of gases on solid surfaces, it can also be applied to the problem of gaseous anesthesia.

According to the electrostatic theory, the gases, as they approach a solid surface possessing either a strong electric field or charged sites, are attracted to the surface; the energy of attraction being

$$\Phi_{att} = -\frac{\alpha}{2} E_z^2 \tag{1}$$

or

$$\Phi_{att} = -\frac{\alpha}{2} \frac{C_{eff}^2}{z^4} \tag{2}$$

where α is the polarizability of the adsorbed molecule, C_{eff} is an effective surface charge, z is the distance the molecule is from the surface, and E_z, which is usually a strong function of z, is the electric field intensity normal to the surface.

In gaseous anesthesis if, instead of hydrate formation, the gases interact directly with some charged sites, the interaction would be a function of the polarizabilities of the gases in accordance with Eq. (2). While Pauling used the very striking correlation between the partical pressure necessary to anesthetize and the molar refraction[2] to support his suggestions of "hydrate" formation, this correlation provides equally strong support for an "adsorptive" model of anesthetic behavior in which the adsorptive forces are primarily electrostatic. Pauling's correlation is reproduced in Fig. 1. We have added diethyl ether, which was excluded from Pauling's original curve

[2] Molar refraction $= \frac{4}{3}\pi N\alpha$ where N is Avogadro's number.
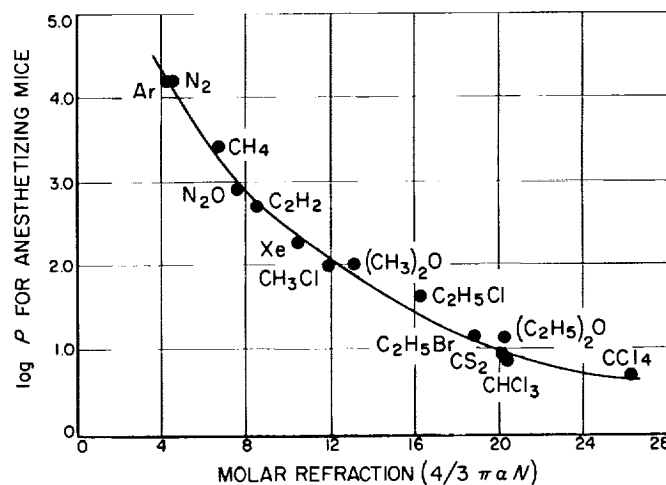
Fig. 1. **Anesthetizing pressure vs molar refraction**

since it does not form a hydrate. The fact that both dimethyl and diethyl ether fall along Pauling's curve supports the inclusion of the non-hydrate-forming gases. Thus, the electrostatic theory does not suffer from the deficiencies of the "hydrate" theories in that it does not exclude molecules which do not form hydrates but are anesthetic agents. Also, it is not surprising that helium, hydrogen, and probably neon have no anesthetic effect (Ref. 4) since these gases have the lowest polarizability values.

Whatever the detailed mechanism of nerve conduction may ultimately be, any model must result from considerations of the surface potentials at the interface between an essentially organic, non-conducting, lipid cell wall and an aqueous, electrolytic solution. The electrostatic model described here suggests that polarizable molecules such as argon or cyclopropane which are dissolved in the aqueous phase will tend to be preferentially adsorbed at charged sites on the organic phase and that this sorption will follow the polarizabilities in much the same way that it does on solid surfaces. Such reversible sorption would be expected to alter the surface potentials and impede the conduction of charges across the interfaces, thereby providing a pathway for anesthetic action.

While both Miller's and Pauling's theories attempt to relate the anesthetic reaction to an immobilization of fluid H_2O and both invoke (although for different reasons) charged sites on surfaces, neither attempts to critically examine the problem of why such charged sites would be particularly involved with the anesthetic molecule or how nerve conduction would be involved. The minor changes in bulk electrical conductivity in the cerebral fluid which the extremely small mole fractions of dissolved agents would produce seem hardly likely to be significantly related to nerve conduction. It appears to us that the specific site adsorption suggested by the electrostatic viewpoint is a much more reasonable mechanism for the blocking of nerve conduction.

It should be pointed out that while it is not too difficult to make a detailed model and calculation of gas adsorption on a solid substrate (SPS 37-33, Vol. IV, pp. 208–212), it is much more difficult to do so for partition between an aqueous solution and a wetted surface. The reason for this is that one has to consider the free energy charge in the process of agent A going from aqueous phase to wetted surface site S with displacement of water:

$$A(aq) + S(aq) \rightleftharpoons S \cdot A(aq) + \text{free water} \qquad (3)$$
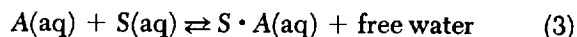
For these reasons we have not attempted a general calculation of the equilibrium constant for such a process and the direct application of Eq. (2) can be misleading.

That sorption of anesthetics at aqueous interfaces is large and important at the partial pressures effective in anesthesia has been demonstrated by the measurements of Clements and Wilson (Ref. 7).

Since all of the accepted theories of gaseous anesthetics involve in some way molecular polarizabilities and molecular sizes, the electrostatic "adsorption" theory simplifies the situation by eliminating the necessity of "hydrate" or other aggregate formations.

B. The NMR and Microwave Spectrum of Some Deutero Derivatives of 2,4-Dicarbaclovoheptaborane(7)

R. L. Poynter, T. Onak,[3] G. B. Dunks,[3] and R. A. Beaudet[4]

1. Introduction

The structure of the dicarbaclovoheptaborane(7), prepared from pyrolysis of 2,3-dicarbahexaborane(8) (Refs. 8 and 9), has been determined to be that of a pentagonal bipyramid with the two carbon atoms in the 2,4 positions (Ref. 10) (Fig. 2). Although the structure is known, an unambiguous assignment of certain ^{11}B and ^1H nuclear magnetic resonance (NMR) lines has not, heretofore, been possible.

The ^{11}B NMR spectrum consists of three sets of doublets with an area ratio of 1:2:2 (Fig. 3). Either one of the two doublets of equal area ($\delta = 2.0$ or $+ 23.5$) may arise from proton coupling to the two apical boron atoms and the remaining doublet may arise from proton coupling to the two equivalent basal borons. There is no a priori reason to select one assignment over the other. For although apical boron resonances in the higher boron hydrides (e.g., B_5H_9, B_6H_{10}) are generally found at high

[3] Department of Chemistry, California State College at Los Angeles, Los Angeles, California.

[4] Department of Chemistry, University of Southern California, University Park, Los Angeles, California.
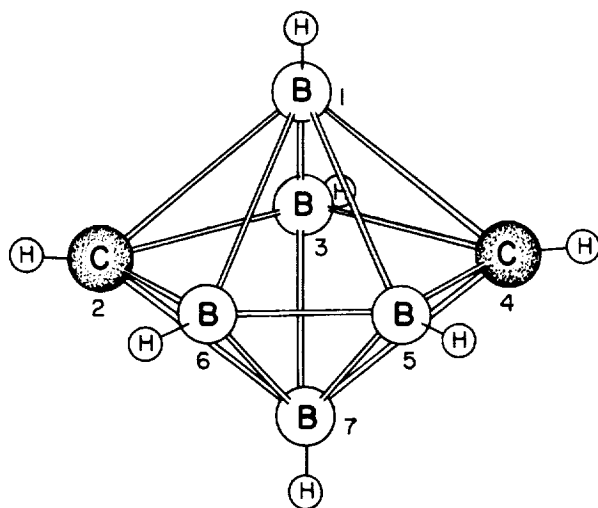
Fig. 2. Molecular structure of 2,4-C₂B₅H₇

Fig. 3. ¹¹B NMR spectra of 2,4-C₂B₅H₇: (a) undeuterated; (b) 3,5,6-deuterated; (c) 1,3,5,6,7-deuterated

fields (Ref. 11), the apical boron resonances of $B_{10}H_{10}^{2-}$, a compound isoelectronic with the hypothetical $C_2B_8H_{10}$ carborane, are found at lower fields than the frequencies of the equatorial borons (Ref. 12).

The ¹H NMR spectrum poses a similar assignment problem. Proton resonances at $\tau = 6.00$ and 9.85 are assigned to the apical hydrogens and to the two equivalent boron-attached equatorial hydrogens, but not necessarily in that order (Ref. 9).[5]

Early studies on the deuteration of other small carboranes, 1,5-$C_2B_3H_5$ and 1,6-$C_2B_4H_6$ (Refs. 13 and 14), suggested that some or all of the boron-attached hydrogens of 2,4-$C_2B_5H_7$ might exchange with deuterodiborane. With appropriate deuterium substitution it was expected that a solution to the NMR assignments would result from an unambiguous structural assignment provided by microwave analysis coupled with changes in the ¹¹B and ¹H NMR spectrum.

2. Experimental

Deuterodiborane was prepared by adding boron trifluoride ethyl etherate dropwise to a diglyme solution of lithium aluminum deuteride (98% D)[6] using a standard workup and purification scheme (Ref. 15). *2,4-Dicarbaclovoheptaborane(7)* was prepared by the pyrolysis of 2,3-dicarbahexaborane(8) (Refs. 8 and 9).

---

[5] All previously reported ¹H chemical shifts for $C_2B_5H_7$ have been converted to the τ reference system in order to avoid negative values where trimethyl silane is selected as the zero reference.

[6] Metal Hydrides, Inc., Beverly, Massachusetts.

a. Hydrogen-deuterium exchange between deuterodiborane and 2,4-dicarbaclovoheptaborane(7).

Ambient temperature; preparation of B,B,B-trideutero-2,4-dicarbaclovoheptaborane(7). An initial H,D exchange between a 1:1 mixture of B_2D_6 and 2,4-$C_2B_5H_7$ at ambient temperature was followed by infrared spectrometry.

Minor changes in the B–H and B–D stretch vibrational frequency regions, as well as other regions of the spectrum, were evident within 2 hr. After 24 hr the infrared spectrum of the mixture had changed quite significantly. Very little additional exchange occurred over longer periods of time. For other physical measurements a 0.20-m mole quantity of the deuterated carborane was prepared by three 24-hr exchange reactions with equimolar quantities of B_2D_6. After each exchange the deuterium-enriched carborane was separated from the partially protonated deuterodiborane by vacuum fractionation from traps at -110 and $-190°C$. The contents of the $-110°C$ trap were mixed with fresh deuterodiborane for the next exchange reaction. The infrared spectrum of the carborane fraction after the final exchange exhibited absorption frequencies at 2630 cm^{-1} (s) (B–H stretch), 1990 cm^{-1} (vs) (B–D stretch), 1190 cm^{-1} (m), 1162 cm^{-1} (m), 1060 cm^{-1} (s), and additional weak peaks. The mass spectrum exhibited a sharp cutoff at $m/e = 89$ which corresponds to $^{12}C_2\,^{11}B_5\,^1H_4\,^2D_3^+$. An infrared analysis (Ref. 15b) of the deuterodiborane from the final exchange indicated ca 90% D enrichment of all equilibrated positions in the carborane.

Exchange at 100°C; preparation of B,B,B,B,B-penta-deutero-2,4-dicarbaclovoheptaborane(7). A mixture of 1.0 mmole 2,4-$C_2B_5H_7$ and 1.0 mmole B_2D_6 was sealed in a 500-ml flask and heated at 100°C for a 5-hr period. Using the procedure given above for ambient temperature, the partially deuterated carborane was separated from the mixture and reacted with fresh deuterodiborane. After a fourth exchange an infrared analysis (Ref. 15b) of the deuterodiborane fraction indicated ca 90% D enrichment of all equilibrated positions in the carborane. The carborane fraction contained small amounts of deuterated higher boron hydride impurities (probably formed from the pyrolysis of the deuterodiborane) which were conveniently removed by gas-liquid chromatography. The infrared spectrum of the purified deuterated carborane exhibited absorptions at 2630 cm^{-1} (m), 1990 cm^{-1} (vs), 1180 cm^{-1} (m), 1050 cm^{-1} (s), and other weak peaks. The mass spectrum exhibited a cutoff at $m/e = 91$ which corresponds to the $^{12}C_2\,^{11}B_5\,^1H_2\,^2D_5^+$ ion.

b. Microwave spectrum of B,B,B-trideuterated-2,4-dicarbaclovoheptaborane(7). The microwave spectrum of the trideuterated species was studied in the region from 18 to 36 Gc with a conventional Stark modulated microwave spectrometer. Frequencies were measured by the interpolation technique, using a Collins 51S-1 receiver. The standard frequencies were generated by a

Hewlett-Packard Model 104AR quartz oscillator which was calibrated against WWV.

The carborane sample was studied in a 12-ft absorption cell cooled to dry ice temperatures. The sample pressure was kept between 20–40 μ. No decomposition was observed under these conditions.

Boron, as it occurs in nature, consists of two isotopes, ^{11}B and ^{10}B, with relative abundances of 81 and 19%, respectively. Hence there are a variety of isotopic $C_2B_5H_7$ molecules, distinguished from each other by the boron isotopic concentration ratio. The most abundant species (35%) contains all ^{11}B atoms (normal species). Referring to Fig. 2, the isotopic species with a ^{10}B atom substituted at the (1) position and the species with a ^{10}B atom substituted at the (5) position each make up 16% of the molecules. (In the future, we shall refer to these two isotopic species as the B(1) and B(5) species, respectively.) The isotopic species with a ^{10}B atom substituted at the (3) position (designated as the B(3) species) accounts for only 8% of the molecules. All of the doubly and higher substituted species account for the remaining 10% of the molecules. In this analysis, the spectra of only the normal, the B(1), and the B(5) molecular species were assigned. The lines of the other isotopic species were much weaker and were not assigned in this study. The assignments were made by comparing the Stark effects of the lines with those of the undeuterated species (Ref. 16).

Since the molecular structure of the undeuterated 2,4-dicarbaclovoheptaborane(7) has already been accurately determined (Refs. 10 and 16), the rotational constants could be calculated for the two possible trideuterated species which are consistent with the NMR data. This substitution may occur at either the 1,3,7 or the 3,5,6 positions. The rotational constants were obtained by a two-step process. Using the molecular structure (Ref. 10) the moments of inertia were calculated for the undeuterated and the deuterated species. The difference between the moments of inertia of the deuterated and undeuterated species were then added to the observed moments of inertia of the undeuterated species. This procedure minimized errors due to the zero-point vibrations.

The calculated rotational constants are compared with the experimentally determined constants in Table 1. From an examination of this table it may be seen that deuteration has occurred in the 3,5,6 positions. The equatorial hydrogens appear to exchange more readily with deuterium.

Table 1. Rotational constants of $B_5C_2H_4D_3$

Species		Estimated rotational constants, Mc/sec		Observed rotational constants, Mc/sec
		3,5,6-deuterated	1,3,7-deuterated	
All-B^{11}	A	4601	4350	4590.93
	B	4031	3949	4047.09
	C	3296	3618	3304.50
5-B^{10}	A	4634	4381	4622.46
	B	4068	3987	4085.35
	C	3337	3670	3346.30
1-B^{10}	A	4659	4402	4648.87
	B	4074	3990	4091.81
	C	3296	3618	3304.32

3. Results and Discussion

From the above evidence all five of the boron-attached hydrogen atoms in 2,4-dicarbaclovoheptaborane(7) undergo an exchange reaction with deuterodiborane at 100°C. To summarize:

(1) The mass spectrum of the deuterated carborane exhibits a sharp cutoff at $m/e = 91$ which corresponds to the $^{12}C_2$ $^{11}B_5$ $^1H_2$ $^2D_5^+$ ion.

(2) A strong infrared absorption at 1990 cm^{-1} is present in the exchanged material and is assigned to the B–D stretching region. The absence of a C–D stretch is not significant because the C–H stretching frequency is infrared inactive.

(3) The ^1H NMR spectrum (Fig. 4c) exhibits only a strong carbon-attached hydrogen resonance.

(4) The collapse of all B–H doublets in the ^{11}B NMR spectrum (Fig. 3c) indicates that all five of the boron-attached hydrogen atoms have been exchanged.[7]

Under ambient temperature conditions, however, the exchange reaction does not proceed to the same extent as at the higher temperatures. This is evident from the mass spectrum which shows a sharp cutoff corresponding to the $^{12}C_2$ $^{11}B_5$ $^1H_4$ $^2D_3^+$ ion. Further, only the low field B–H doublets ($\delta = -2.0$, -5.0) (Ref. 8) collapse in the ^{11}B NMR spectrum; the high field doublet ($\delta = + 23.5$) is unchanged. In the ^1H NMR, the high field quartet

centered at $\tau = 9.85$ (Fig. 4b) is the only H–^{11}B resonance present.

The low field doublet ($\delta = -5.0$) in the ^{11}B spectrum and the low field quarter ($\tau = 5.25$)[8] in the ^1H spectrum of $C_2B_5H_7$ may be assigned to the lone equatorial B–H group on the basis of area considerations alone. The two remaining doublets in the ^{11}B NMR spectrum and the quartets in the ^1H NMR spectrum are of equal area and can be equivocally assigned to positions 1,7 or 5,6.

Analysis of the microwave spectrum conclusively shows that in $B_5C_2H_4D_3$ the three deuterium atoms are attached to the equatorial borons (3,5,6). Hence the remaining doublet in the ^{11}B NMR ($\delta = + 23.5$) can be assigned to the apical borons at positions 1 and 7. The resonance at ($\delta = - 2.0$) belongs to the equatorial borons.

Similarly in the ^1H NMR spectrum, the quartet centered at $\tau = 9.85$ can be assigned to the protons on the apical borons. The remaining unassigned quartet at $\tau = 6.00$ belongs to the equatorial borons at positions 5 and 6.

Hence the apical boron resonances are located at higher field than the equatorial boron resonances in agreement with the corresponding resonances found in related systems, e.g., B_5H_9, B_5H_{11}, B_6H_{11}, B_6H_{10}, and $C_2B_4H_8$.

---

[7] The ^{11}B-D spin-spin coupling is not well resolved in the ^{11}B NMR spectrum of the deuterated $C_2B_5H_7$.

[8] Naturally-occurring boron consists of 81% ^{11}B ($I = 3/2$; magnetogyric ratio = 1366 cycles/gauss-sec) and 19% ^{10}B ($I = 3$; magnetogyric ratio = 457.5 cycles/gauss-sec). The quartets assigned to the ^{11}B-attached hydrogens in the ^{1}H spectrum are usually well defined whereas the smaller septets of the ^{10}B-attached hydrogens are buried under the labyrinth of all other resonances.
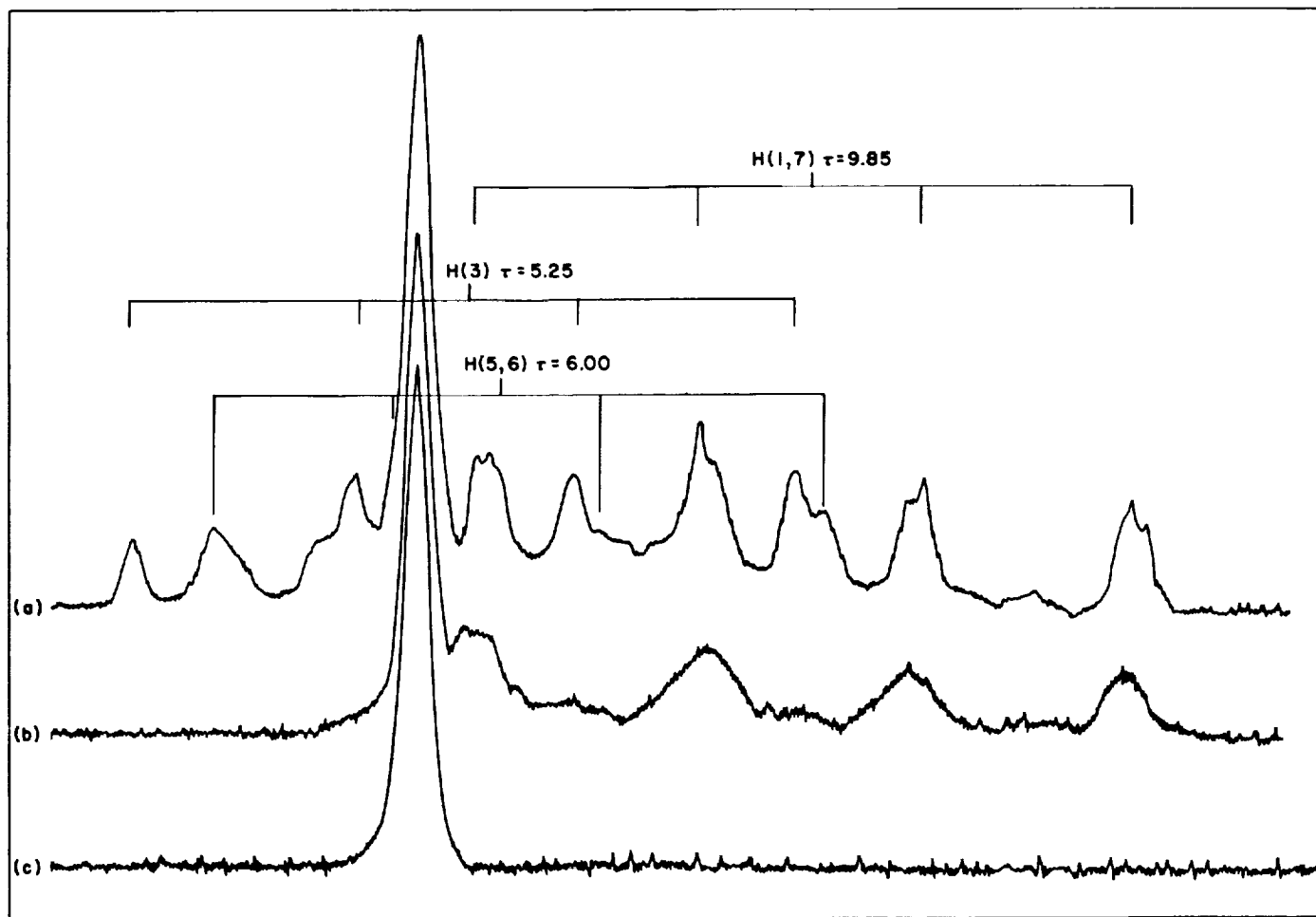
Fig. 4. ¹H NMR spectra: (a) undeuterated; (b) 3,5,6-deuterated; (c) 1,3,5,6,7-deuterated

References

1. Butler, T., *Pharmacological Reviews*, Vol. 2, p. 121, 1950.

2. Featherstone, R. M., and Muehlbacher, C. A., *Pharmacological Reviews*, Vol. 15, p. 97, 1963.

3. Pauling, L., *Science*, Vol. 134, p. 15, 1961.

4. Miller, S. L., *Proceedings of the National Academy of Sciences*, Vol. 47, p. 1515, 1961.

5. Cullen, S., and Gross, E., *Science*, Vol. 113, p. 580, 1951.

6. Pauling, L., *The Nature of the Chemical Bond*, 3rd Edition, Cornell University Press, Ithaca, N.Y., p. 469, 1960.

7. Clements, J. A., and Wilson, K. M., *Proceedings of the National Academy of Science*, Vol. 48, p. 1008, 1962.

References (Cont'd)

8. Onak, T., Gerhart, F. J., and Williams, R. E., *Journal of the American Chemical Society,* Vol. 85, p. 3378, 1963.

9. Onak, T., Drake, R. P., and Dunks, G. B., *Inorganic Chemistry,* Vol. 3, p. 1686, 1964.

10. Beaudet, R. A., and Poynter, R. L., *Journal of the American Chemical Society,* Vol. 86, p. 1258, 1964.

11. (a) Onak, T., Landesman, H., Williams, R. E., and Shapiro, I., *Journal of Physical Chemistry,* Vol. 63, p. 1533, 1959; (b) Williams, R. E., Gibbins, S. G., and Shapiro, I., *The Journal of Chemical Physics,* Vol. 30, p. 320, 1959; (c) Williams, R. E., Gibbins, S. G., and Shapiro, I., *The Journal of Chemical Physics,* Vol. 30, p. 333, 1959.

12. Muetterties, E. L., Balthis, J. H., Chia, Y. T., Knoth, W. H., and Miller, H. C., *Inorganic Chemistry,* Vol. 3, p. 444, 1964.

13. Shapiro, I., Good, C. D., and Williams, R. E., *Journal of the American Chemical Society,* Vol. 84, p. 3837, 1962.

14. Shapiro I., Keilin, B., Williams, R. E., and Good, C. D., *Journal of the American Chemical Society,* Vol. 85, p. 3167, 1963.

15. (a) Shapiro, I., Weiss, H. G., Schmich, M., Skolmik, S., and Smith, G. B. L., *Journal of the American Chemical Society,* Vol. 74, p. 901, 1952; (b) Lehmann, W. J., Ditter, J. F., and Shapiro, I., *The Journal of Chemical Physics,* Vol. 29, p. 1248, 1958.

16. Beaudet, R. A., and Poynter, R. L., *The Journal of Chemical Physics,* Vol. 43, p. 2166, 1965.

XVI. Fluid Physics

A. The Inviscid Stability of the Laminar Compressible Boundary Layer for Three-Dimensional Disturbances

L. M. Mack

All of the work reported previously in the *Space Programs Summary*, Vol. IV, on the stability of the laminar boundary layer has been based on the use of two-dimensional disturbances. This restriction was made because at finite Reynolds numbers three-dimensional disturbances require the solution of an eighth-order system of differential equations rather than a sixth-order system, with a consequent large increase in the running time of the computer program. For incompressible viscous flow, the transformation of Squire (Ref. 1) transforms the system of equations for three-dimensional disturbances into an equivalent two-dimensional system. A generalization of this same transformation is valid for the simplified compressible-flow equations of Dunn and Lin (Ref. 2). The neutral-stability curves obtained by Dunn and Lin indicate that the boundary layer at low Mach numbers is more unstable to three-dimensional disturbances than to two-dimensional disturbances. This same result is obtained at higher Mach numbers by Brown (Ref. 3), who calculated neutral-stability curves directly from the complete eighth-order system of linear stability equations.

Since the generalized Squire transformation applies to the Dunn-Lin simplified equations, it also applies to the

inviscid equations. Consequently, the inviscid equations for a three-dimensional disturbance can be reduced to two-dimensional equations, and eigenvalues and eigenfunctions of the three-dimensional disturbances can be found from the same computer program that was used previously for the two-dimensional equations. Since it has been determined that the instability of the boundary layer at high Mach numbers is essentially an inviscid instability, the inviscid three-dimensional results will provide information concerning amplification rates and unstable frequencies for three-dimensional disturbances at finite Reynolds numbers.

The inviscid equations of motion for the complex amplitude functions of a sinusoidal disturbance are

$$\rho \left[i \left(u - c \right) f + u' \phi \right] = - \frac{i\pi}{\gamma M_1^2}$$

$$\rho i \alpha^2 \left(u - c \right) \phi = - \frac{\pi'}{\gamma M_1^2}$$

$$\rho \alpha \left(u - c \right) h = - \beta \frac{\pi}{\gamma M_1^2}$$

$$i\alpha \left(u - c \right) r + \rho' \alpha \phi + \rho \left(i\alpha f + i\beta h + \alpha \phi' \right) = 0$$

$$\rho \left[i\alpha \left(u - c \right) \theta + T' \alpha \phi \right] = - \left(\gamma - 1 \right) \left(i\alpha f + i\beta h + \alpha \phi' \right)$$

$$\pi = \frac{r}{\rho} + \frac{\theta}{T}$$

The quantities appearing in these equations are the same as in the two-dimensional equations except for the addition of h, the amplitude function for the transverse

velocity, and β, the wave number in the transverse, or z, direction. The form of the three-dimensional disturbance is

$$Q\,(x, \eta, z, t) = q\,(\eta)\,\exp\left[i\,(\alpha x + \beta z - \alpha c t)\right]$$

and represents an oblique wave with the wave normal at an angle σ to the free-stream direction (x-axis), where

$$\cos \sigma = \frac{\alpha}{(\alpha^2 + \beta^2)^{\frac{1}{2}}}$$

The real part of the wave velocity is the phase velocity c_r in the x direction.

The Dunn-Lin transformation is

$$\alpha f + \beta h = \tilde{\alpha}\tilde{f}$$
$$\alpha \phi = \tilde{\alpha}\tilde{\phi}$$
$$\alpha \pi = \tilde{\alpha}\tilde{\pi}$$
$$\alpha r = \tilde{\alpha}\tilde{r}$$
$$\alpha \theta = \tilde{\alpha}\tilde{\theta}$$
$$c = \tilde{c}$$
$$(\alpha^2 + \beta^2)^{\frac{1}{2}} = \tilde{\alpha}$$
$$\alpha M_1 = \tilde{\alpha}\tilde{M}_1$$

The resulting equations are identical to two-dimensional equations for the tilde quantities. The eigenvalues are $\tilde{\alpha}$ and $\tilde{c}$, where $\tilde{\alpha}$ is the wave number in the direction of the wave normal, and $\tilde{c}$ is the wave velocity in the same direction made dimensionless with respect to $U \cos \sigma$ The transformed free-stream Mach number $\tilde{M}_1$ is the component of the true free-stream Mach number in the direction of the wave normal. The dimensionless mean-flow boundary-layer quantities u and T are unchanged. Consequently, for a given free-stream Mach number, M_1, eigenvalues and eigenfunctions for all values of σ can be obtained from the existing computer program for the solution of the two-dimensional stability equation. It is only necessary to change the input free-stream Mach number from M_1 to $M_1 \cos \sigma$ while retaining the boundary-layer profiles for the Mach number M_1.

The quantity of most immediate interest is the amplification rate of the disturbance and its variation with wave angle. The time rate of amplification, $d(\log |Q|)/dt$, is given by αc_i. The amplification rates of the first and second modes have been computed at $M_1 = 4.5, 5.8, 8.0$, and 10.0 for several wave angles for the insulated-wall boundary layer. As an example, Fig. 1 gives the results obtained at $M_1 = 4.5$. The amplification rate is plotted against the frequency αc_r. The first mode is most unstable
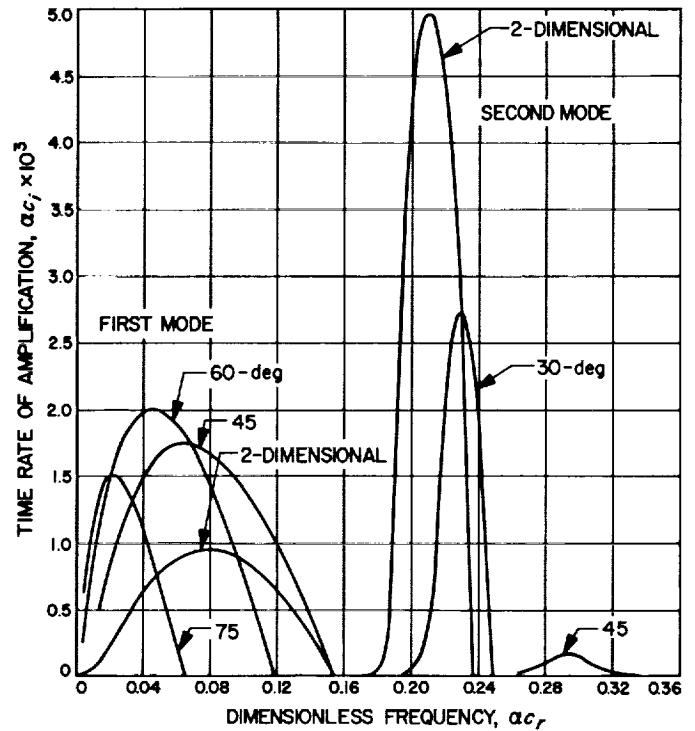
Fig. 1. Time rate of amplification of first and second modes vs frequency for several wave angles at $M_1 = 4.5$

for a wave angle near 60 deg, where the maximum amplification rate is approximately double that of the two-dimensional disturbance. The most unstable frequency at $\alpha = 60$ deg is only about 40% of the value at $\sigma = 0$ deg. As σ increases beyond 60 deg, both the amplification rates and the frequencies tend to zero. The second mode has a very different behavior. As σ increases from zero, the amplification rates drop sharply and the most unstable frequency increases approximately 50% from $\sigma = 0$ deg to $\sigma = 45$ deg.

From Fig. 1 it is possible to prepare a plot of the maximum amplification rate as a function of wave angle. This procedure has been carried out at $M_1 = 5.8, 8.0$, and 10.0 as well as 4.5, and the results are shown in Fig. 2. For the first mode the variation of $(\alpha c_i)_{max}$ with wave angle is similar at all Mach numbers. The wave angle at which the amplification rate is a maximum is between 50 and 60 deg, and the value of the maximum amplification rate at this angle is approximately double its value for the two-dimensional wave. For the second mode the distribution of amplification rate with wave angle becomes flatter at the higher Mach numbers, at least over the range of σ for which computations were
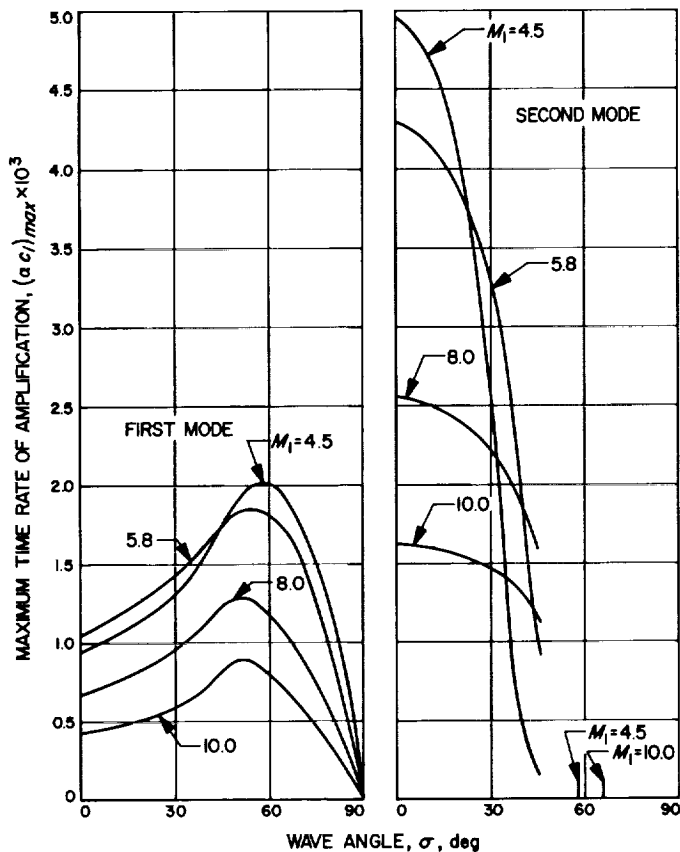
Fig. 2. Effect of wave angle on maximum time rate of amplification of first and second modes at $M_1 = 4.5, 5.8, 8.0,$ and 10.0

made. It is difficult to compute the maximum second-mode amplification rate at wave angles above 45 deg for the higher Mach numbers, and these computations have not yet been carried out.

The different behavior of the first- and second-mode amplification rates as functions of wave angle reflects the different physical nature of the two modes. For the first mode, the entire boundary layer is important, and, particularly, the relative positions of the generalized inflection point, η_s, where $(u'/T)' = 0$, and the point η_0, where $u = 1 - 1/M_1$. With increasing σ, $\tilde{M}_1$ decreases and η_0 moves toward $\eta = 0$, whereas η_s remains fixed because the boundary-layer profiles are fixed. Consequently, the amplified solutions have increasingly large values of c_i. Initially this increase leads to an increased amplification rate. However, α must decrease to zero in the limit $\sigma \to 90$ deg and so must the amplification rate. On the other hand, the second mode depends for its existence on the presence within the boundary layer of a region of supersonic mean flow relative to the phase

velocity c_r; i.e., when $T > \tilde{M}_1^2 (u - c)^2$ throughout the boundary layer, there is no second mode. For the four values of M_1 of Fig. 2 this condition holds beyond some σ which lies between 55 and 65 deg, depending on the Mach number. It is less clear why the maximum amplification rate decreases monotonically to zero at all M_1, because for the two-dimensional disturbances a decrease in M_1 for $M_1 > 5$ results in an increase in the maximum amplification rate. It will be necessary to know more about the nature of second-mode amplified solutions before this question can be answered.

B. Shock-Wave Strengthening by Area Convergence

D. A. Russell

A decrease in the cross-sectional area of a duct will result in an increase in the strength of the transmitted shock wave. The theoretical strengthening for an ideal gas is shown in Fig. 3 as function of the area ratio, the ratio of specific heats of the test gas, γ, and the initial shock Mach number, M_0. The lower curves are obtained
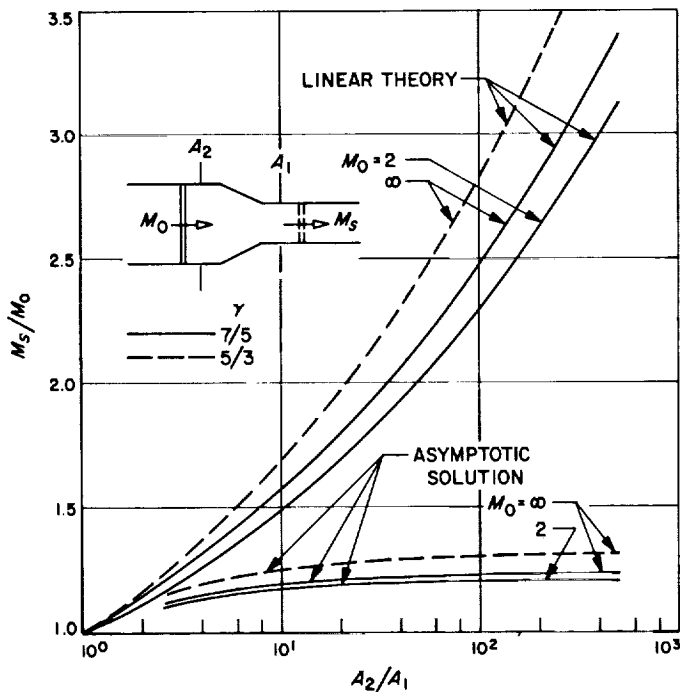
Fig. 3. Theoretical strengthening (ideal gas)
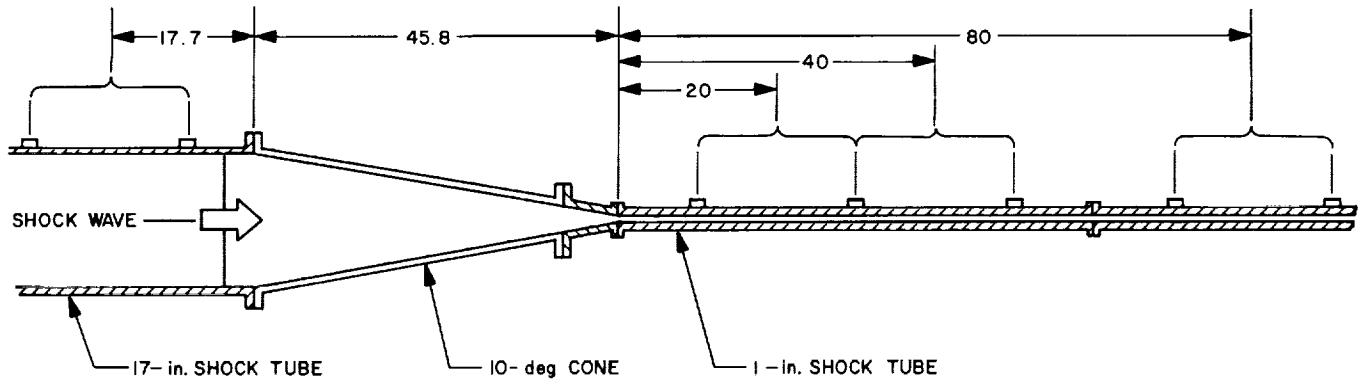
DIMENSIONS IN INCHES

Fig. 4. Scale drawing of the experiment (the mid-point of each pair of gages is indicated)

with a model which treats the convergence as an instantaneous area change. It is the solution reported in several monographs on shock tube theory and operation (Ref. 4), and it is expected to be asymptotically valid far downstream of a region of varying area. The upper set of curves is the linearized one-dimensional theory due independently to Chester, Chisnell, and Whitham (Ref. 5). This theory should be valid within a region of monotonically decreasing area. Thus, a shock wave is expected to emerge from a monotonic area convergence at a speed given by the linear theory, and then attenuate due to second-order disturbances until it reaches the asymptotic solution.

An experiment was set up in order to test these predictions and to study the possibilities of using area convergence as a means of extending a shock tube's ability to produce high-temperature plasma. The GALCIT 17-in. shock tube was coupled by a 10-deg half-angle cone to a 1-in. diameter tube, providing an area convergence of nearly 300 (Fig. 4). Pairs of thin-film gages were used to measure the speed of the shock wave. The individual gages were spaced 20 in. apart, and each pair was centered as shown on Fig. 4. M_s/M_0 was obtained by dividing each downstream velocity by the upstream value for the same run. The results are plotted versus M_0 in Fig. 5, where the initial pressure for each run and the downstream position of each data point are indicated. It is seen that the shock speed is substantially increased by the convergence, and that it decays with distance downstream.

The experimental data of Fig. 5 depart farther from the linear theory as M_0 is raised. This is due to real-gas effects (which reduce the value of the linear-theory solution), and also due to the viscous attenuation encountered
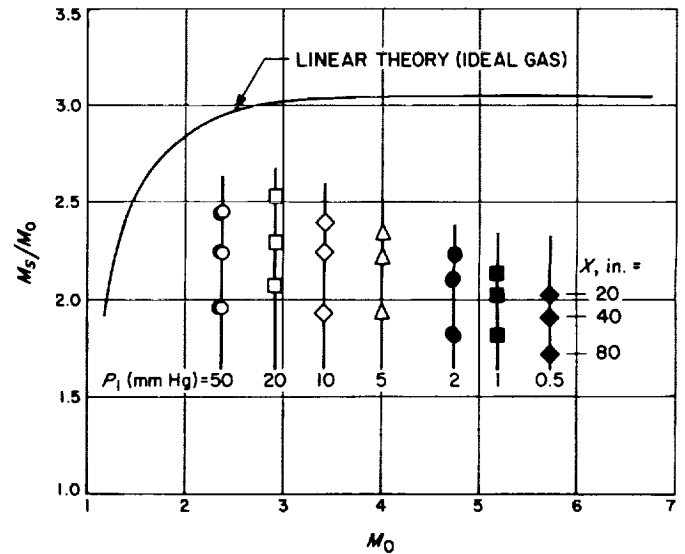
Fig. 5. Uncorrected data for air (initial pressure and downstream distance are indicated for each point)

at the low pressures used to obtain the higher values of M_0. Calculation procedures have been developed for predicting both of these effects. They will be presented in the completed version of this work.

After correction for viscous attenuation, the data indicate that each shock wave attained more than 95% of the real-gas linear-theory prediction at the nozzle exit. The parametric dependence of the subsequent decay is being explored further. Initial results are that the decay requires a distance equivalent to 5–15 nozzle lengths to complete, and that its rate is approximately constant over this distance. This rate of deceleration is at least an order of magnitude less than the acceleration rate near the

nozzle exit, which suggests that a much higher degree of flow uniformity may be obtained where the flow is dominated by the decay process. Fine-wire heat transfer measurements indicate that this is the case, and that a convergence can be designed for a desired degree of flow uniformity downstream. It is concluded that area convergence can provide a useful means of extending the performance of a shock tube.

References

1. Squire, H. B., "On the Stability for Three-Dimensional Disturbances of Viscous Fluid Flow between Parallel Walls," *Proceedings of the Royal Society London* (A), Vol. 142, pp. 621–628, 1933.

2. Dunn, D. W., and Lin, C. C., "On the Stability of the Laminar Boundary Layer in a Compressible Fluid," *Journal of the Aeronautical Sciences,* Vol. 22, pp. 455–477, 1955.

3. Brown, W. B., "Stability of Compressible Boundary Layers, Including the Effects of Two-Dimensional Mean Flows and Three-Dimensional Disturbances," *Bulletin of the American Physical Society,* Vol. 10, p. 682, 1965.

4. Glass, I. I., and Hall, J. G., "Shock Tubes," *Handbook of Supersonic Aerodynamics,* NAVORD Report 1488, Vol. 6, Sec. 18, 1959.

5. Whitham, G. B., "On the Propagation of Shock Waves through Regions of Non-Uniform Area or Flow," *Journal of Fluid Mechanics,* Vol. 4, pp. 337–360, 1958.

XVII. Physics

A. Nanosecond Rectangular-Wave High-Voltage Pulser for Spark Chamber

L. L. Lewyn

An omni-directional spark chamber is currently under development for the purpose of investigating the energy and angular distribution of solar and galactic particles. This program is described and details of the device are sketched in Ref. 1.

The chamber operates by applying a brief high-voltage pulse to a series of parallel-plane wire-grid systems with alternate grids perpendicular to each other. The high-voltage pulse is applied when a coincidence pulse from an associated scintillation telescope indicates that an energetic charged particle has passed through the chamber. A spark appears at the intersection of grid wires nearest the ionization wake left by the particle. The spark locations therefore trace the particle's trajectory.

The parameters of the device which supplies the high-voltage pulse are crucial for determining the sparking characteristics of the system. Specifically, the design aim of the present pulser is to achieve a very narrow rectangular pulse with an amplitude of several kilovolts. This article reports the successful development of a laboratory model pulser which serves as a prototype for balloon-flight experiments presently in preparation.

A fast high-voltage pulser was constructed to modulate a small, laboratory, parallel-plate spark chamber. The pulser is shown in Fig. 1. It is capable of generating a 3- to 4-kv pulse within a delay of 60 nsec. The no-load rise and fall times are on the order of 10 nsec and the pulse-width ranges from 140 nsec at 3 kv to 60 nsec at 4 kv.

The pulser is composed of two main parts, a solid-state amplifier and a gap driver. The function of the solid-state amplifier is to generate a 4-kv trigger pulse which will fire the triggered spark gap G_1. The gap G_1 applies
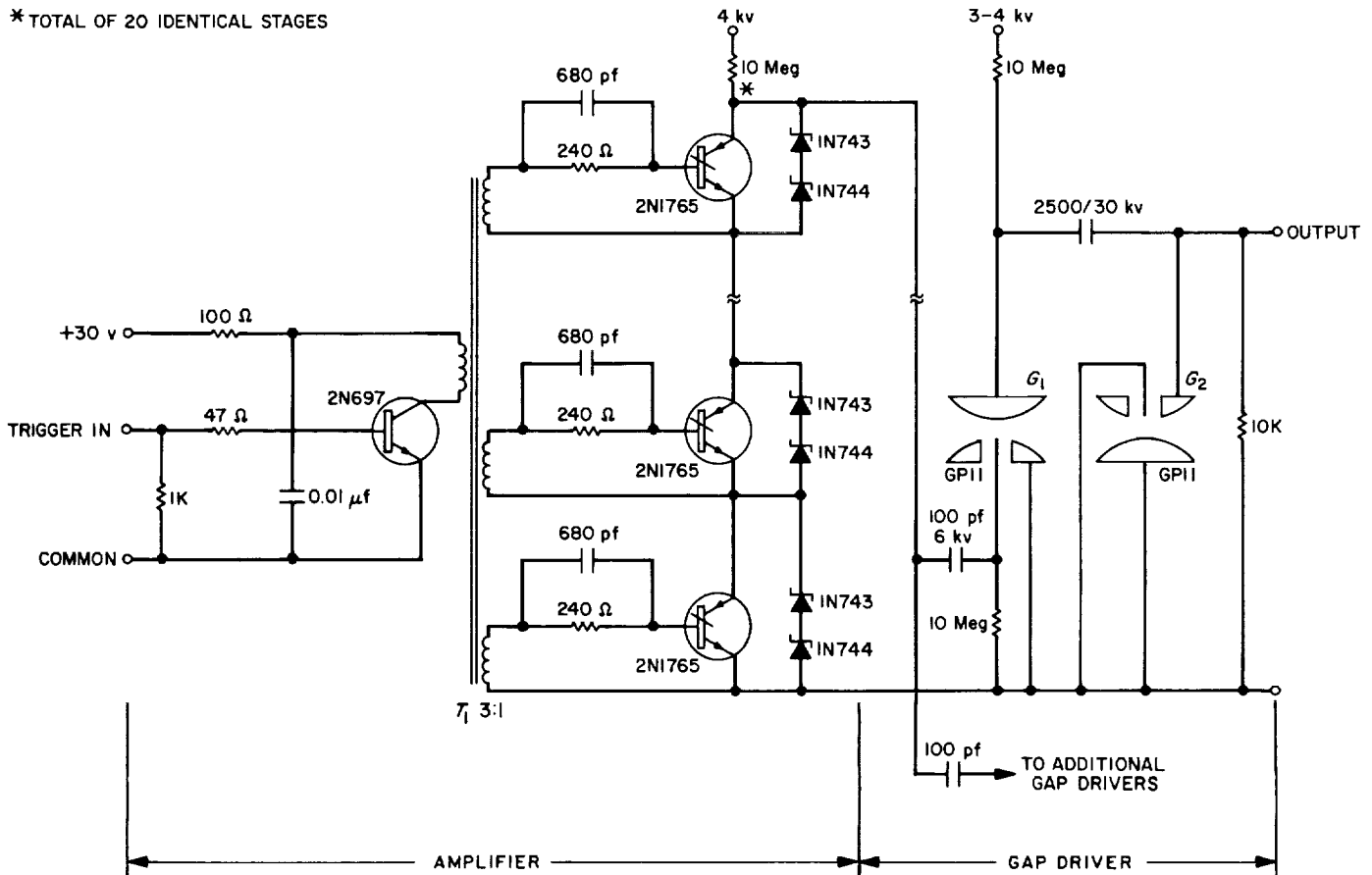
Fig. 1. High-voltage pulser

the high voltage to the pulser output. The voltage is removed by gap G_2. The use of the second gap to actively and quickly remove the high voltage from the chamber is a unique feature of this pulser which was suggested by Mr. Max Gumpel of JPL. Ordinarily, spark chamber voltages are allowed to tail off exponentially. This prolonged presence of the high voltage tends to increase the probability of spurious sparking, that is, sparking at a position other than along the trajectory of the particle which triggered the chamber.

The output section of the amplifier is composed of silicon-controlled rectifiers (SCRs) driven from a common transformer. The leakage currents of the SCRs vary widely. Ballast resistors consuming several times the leakage current of the SCRs are ordinarily connected across each SCR to assure quiescent voltage-drop uniformity. Zener diodes which add little current to the SCR leakage are used instead of resistors to keep the SCR voltages below the avalanche region.

The triggered gap-driver utilizes two ceramic-metal spark gaps capable of switching currents in excess of 5000 amp. After a delay of 30 nsec following the amplifier trigger pulse output, the gap G_1 conducts, applying high voltage to the pulser output.

Gap G_2 is connected in a configuration which has an inherently longer delay than that of gap G_1. The high-voltage pulse remains at the output during the delay time of gap G_2 and is clamped to ground as G_2 fires. While the output capacitor is being discharged by gap G_2, gap G_1 is also conducting, thus giving rise to the ringing observed on the trailing edge of the pulser output waveform which is shown in Fig. 2.

The pulser is capable of driving several grid systems in parallel. However, a separate gap-driver must be added for each additional grid system. The laboratory pulser was successfully operated for several months driving two grid systems.
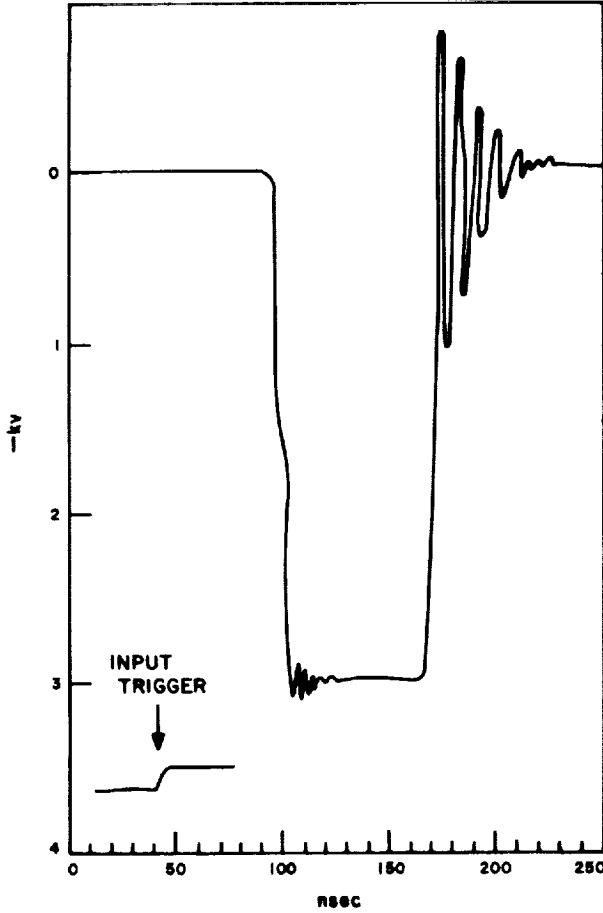
Fig. 2. Spark chamber modulator waveform

B. Transverse Electromagnetic Instabilities in a Plasma

C.-S. Wu

Recently, several authors (Refs. 2 and 3) have independently reported theoretical predictions of unstable modes of transverse electromagnetic waves in a plasma. A common conclusion reached by all these authors is that instability will occur if electron temperature in the direction transverse to the wave vector, hereafter called $T_{e\perp}$, is higher than that in the direction parallel to the wave vector, hereafter called $T_{e\parallel}$. The purpose of this article is to report some new conditions of instability, from which we shall see that, under certain conditions, instability may still occur even if $T_{e\perp} = T_{e\parallel}$ or $T_{e\perp} < T_{e\parallel}$ and, moreover, under different conditions, instability may not exist even if $T_{e\perp} > T_{e\parallel}$.

To proceed with our discussion, we consider a multicomponent plasma for which the dispersion equation concerning the transverse waves can be readily derived from the Vlasov equations. Here we shall omit the derivation and merely state the result; i.e.,

$$k^2 (c^2 - w^2) + \sum_s \omega_s^2$$
$$\times \left[1 + \int_{-\infty}^{+\infty} dv_1 v_1^2 \int_{-\infty}^{+\infty} dv_3 \frac{\partial \tilde{F}_s}{\partial v_3} \frac{1}{(v_3 - w)} \right] = 0 \quad (1)$$

In Eq. (1), the subscript s designates the particle species, and v_1 and v_3 are the velocity components parallel to the electric field and the wave vector, respectively. We have also used the notations that $w = \omega/k$ (ω is the complex frequency and k the wave number), $\omega_s^2 = 4\pi n_s e_s^2/m_s$ (n stands for the number density, e the charge, and m the particle mass), and

$$\tilde{F}_s (v_1, v_3) = \int_{-\infty}^{+\infty} dv_2 F_s (v_1, v_2, v_3)$$

where $v_2 = \vec{v} \cdot \hat{a}$ and $\hat{a} = (\vec{k} \times \vec{E})/kE$ (thus v_2 is the velocity component perpendicular to both $\vec{E}$, the transverse electric field, and the wave vector $\vec{k}$). Furthermore, we remark that Eq. (1) is defined for $Imw > 0$. For $Imw < 0$, Eq. (1) may be defined by analytic continuation across the real axis in the complex w-plane.

Making use of the Nyquist hodograph method (Ref. 4), we can discuss the stability criteria of the present problem. It is found that the necessary and sufficient conditions for an instability are that, for some value of real w,

$$\frac{1}{w^2 - c^2} \sum_s \omega_s^2 \left[\int_{-\infty}^{+\infty} dv_1 v_1^2 \frac{\partial \tilde{F}_s}{\partial v_3} \right]_{v_3 = w} = 0 \quad (2)$$

$$\frac{\partial}{\partial w} \left\{ \frac{1}{w^2 - c^2} \sum_s \omega_s^2 \left[\int_{-\infty}^{+\infty} dv_1 v_1^2 \frac{\partial \tilde{F}_s}{\partial v_3} \right]_{v_3 = w} \right\} > 0 \quad (3)$$

$$\frac{1}{w^2 - c^2} \sum_s \omega_s^2 \left[1 + \int_{-\infty}^{+\infty} dv_1 v_1^2 P \int_{-\infty}^{+\infty} dv_3 \frac{\partial \tilde{F}_s}{\partial v_3} \frac{1}{v_3 - w} \right] > 0 \quad (4)$$

In Eq. (4), P denotes the principal value of the integral. Now, let us apply these conditions and examine two fundamental cases:

(a) *Anisotropic plasma without current.* In this case, we postulate that the distribution functions $\tilde{F}_s(v_1, v_3)$ take

the following form:

$$\tilde{F}_s(v_1, v_3) = \left(\frac{m_s}{2\pi k T_{s\parallel}}\right)^{1/2} \left(\frac{m_s}{2\pi k T_{s\perp}}\right)^{1/2}$$

$$\times \exp\left[-\frac{m_s v_3^2}{2kT_{s\parallel}} - \frac{m_s v_1^2}{2kT_{s\perp}}\right] \quad (5)$$

From Eqs. (2), (3), and (4), we conclude readily that the condition of instability is

$$\sum_s \omega_s^2 \left[1 - \frac{T_{s\perp}}{T_{s\parallel}}\right] < 0 \quad (6)$$

Condition (6) states clearly that for a multispecies plasma, even when electrons possess isotropic distribution, unstable modes may still exist due to the anisotropic ion distribution if $T_{i\perp}/T_{i\parallel} < 1$.

(b) Anisotropic plasma with current flowing parallel to the wave vector. Let us consider the third case, in which

$$\tilde{F}_s(v_1, v_3) = \left(\frac{m_s}{2\pi k T_{s\parallel}}\right)^{1/2} \left(\frac{m_s}{2\pi k T_{s\perp}}\right)^{1/2}$$

$$\times \exp\left[-\frac{m_s(v_3 - v_s)^2}{2kT_{s\parallel}} - \frac{m_s v_1^2}{2kT_{s\perp}}\right] \quad (7)$$

It can be visualized that in this case the condition of instability depends not only upon the ratios of $T_{s\perp}/T_{s\parallel}$, but also upon the drift velocities v_s. Thus, the situation is slightly more complicated than in the previous two cases.

Let us make use of Expression (7) and rewrite Eq. (4) as follows:

$$\frac{1}{w^2 - c^2}\sum \omega_s^2 \left[\left(1 - \frac{T_{s\perp}}{T_{s\parallel}}\right) - \frac{T_{s\perp}}{T_{s\parallel}}(w - v_s)P\right.$$

$$\left.\times \int_{-\infty}^{+\infty} dv_3 \frac{\bar{F}_s(v_3)}{v_3 - (w - v_s)}\right] > 0 \quad (8)$$

where

$$\bar{F}_s(v_3) = \left(\frac{m_s}{2\pi k T_{s\parallel}}\right)^{1/2} \exp\left[-\frac{m_s v_3^2}{2kT_{s\parallel}}\right]$$

However, in order to satisfy Eq. (2), w, in this case, can be estimated in the range $0 < w < v_{s\,max}$. Here $v_{s\,max}$ denotes the maximum drift velocity among the many

species. Hence it must be true that $w < c$ and Condition (8) is equivalent to

$$\sum \omega_s^2 \left[\left(1 - \frac{T_{s\perp}}{T_{s\parallel}}\right) - \frac{T_{s\perp}}{T_{s\parallel}}(w - v_s)P\right.$$

$$\left.\times \int_{-\infty}^{+\infty} dv_3 \frac{\bar{F}_s(v_3)}{v_3 - (w - v_s)}\right] < 0 \quad (9)$$

Since (Ref. 5)

$$\beta P \int_{-\infty}^{+\infty} dv_3 \frac{\bar{F}_s(v_3)}{v_3 - \beta} < 0$$

for all values of real β, we conclude that: (1) if $T_{s\perp} = T_{s\parallel}$ for all species, there is no instability, and (2) even if $T_{s\perp} > T_{s\parallel}$, there may still be no instability since the presence of current tends to stabilize the plasma, as we see from Condition (9).

The fact that ions can play important roles is not surprising, since the phase velocity of the unstable mode may be very small. Thus, in this sense, the neglect of ions in the theory may not be justifiable.

C. Interaction of an Electron with the Photon Field

M. M. Saffren

In the usual textbook discussions of the electromagnetic field interacting with matter, the electromagnetic field is usually assumed to be represented by only one photon. This photon is either absorbed, emitted, or scattered by the matter. However, the advent of lasers as electromagnetic sources has made it possible to generate electromagnetic fields that contain more than one photon, and it is of interest therefore to consider processes in which more than one single photon takes part.

For a photon beam that is intense, or that is coherent, semi-classical approximations in which the photon beam is treated as a classical beam appear to be adequate. The interaction of an electron with a correlated many-photon state does not, however, in general appear to have any semi-classical analog, as we attempt to show below. This

result is not too surprising. The applicability of the semi-classical approximation may be understood as a simple consequence of the correspondence principle. Since an intense beam contains many nearly identical quanta, it is a state with a large quantum number, and the correspondence principle assures us that the beam can be described classically. For a correlated state, however, the many quanta can be distributed over a large number of "single photon states" so that no individual "state" is occupied by a large number of quanta and we are no longer assured of a classical correspondence for the beam. Nevertheless, there probably are limiting cases where the interaction of a correlated state with an electron does have a classical correspondence. We do not analyze such cases in this article, however.

In order to study the interaction of a correlated photon field with matter, we wish to study the scattering of an electron by such a field. Here we present the formal expression for the amplitude appropriate for this scattering.

We begin by reviewing the description of correlated many-photon states of the electromagnetic field. A many-photon state has the form

$$|\text{photon}> = \sum_{n=0}^{\infty} A_n \, |x_n>$$ (1)

Here $|x_n>$ denotes an n-photon state. This state in Fock space is equivalent to the n-photon function $x(\underline{k}_1, \cdots \underline{k}_n)$. This function is symmetric in the $\underline{k}$, since photons are bosons. The symbol $\underline{k}$ not only denotes a photon of wave-vector $\underline{k}$ but is to be understood as standing for the wave vector of the photon and its polarization as well ($+$ or $-$ for left- or right-handed polarization, respectively); in expressions where the symbol of integration, $[d\underline{k}]$, appears the symbol is meant to indicate summation over polarization as well. Using this convention, the state can now be written as

$$|x_n> = \frac{1}{\sqrt{n'}} \int [d\underline{k}_1] \cdots [d\underline{k}_n] \, x(\underline{k}_1, \cdots \underline{k}_n) \, C_{k_1}^+ \cdots C_{k_n}^+ |0>$$ (2)

where C_k^+ denotes the creation operator for a photon of momentum and polarization k, and $|0>$ is the vacuum. The C_k satisfy the commutation relations $[C_{k'}, C_k^+] = \delta(\underline{k} - \underline{k}')$, $[C_{k'}^+, C_{k'}^+] = 0$ (the δ function is to be understood as having absorbed the polarization δ function). The function $x(\underline{k}_1 \cdots \underline{k}_n)$ is normalized as

$$\int [d\underline{k}]^n \, | \, x(\underline{k}_1, \cdots \underline{k}_n)|^2 = 1$$ (3)

Here $[d\underline{k}]^n \equiv [d\underline{k}_1] \cdots [d\underline{k}_n]$. With this normalization $|x_n>$ is normalized as $<x_n|x_n> = 1$. If now

$$\sum_{n=0}^{\infty} |A_n|^2 = 1$$ (4)

then $<\text{photon}|\text{photon}> = 1$. The vector potential $\underline{A}(\underline{x})$ can be expressed in terms of C's and is in fact,

$$\underline{A}(\underline{x}) = \gamma \int \frac{[d\underline{k}]}{\sqrt{|\underline{k}|}} \left[C_k^+ \, \underline{A}_k \, e^{i\underline{k}\cdot\underline{x}} + C_k \, \underline{A}_k \, e^{-i\underline{k}\cdot\underline{x}} \right]$$ (5)

(the underline denotes a 4-vector). The number operator $dn(\underline{k})$ appropriate for the photons in the interval $[d\underline{k}]$ can also be expressed in terms of the C's and is

$$dn_k = C_k^+ \, C_k \, [d\underline{k}]$$ (6)

Notice that because

$$<\text{photon}|C_k^+ C_k|\text{photon}> =$$

$$\sum_{n=0} |A_n|^2 \, n \int [d\underline{k}]^{n-1} \, |x \, (\underline{k}_1, \cdots \underline{k}_{n-1}, \underline{k})|^2$$ (7)

the number of photons in a given frequency interval is compatible with many different states of the photon field. Thus, even with $n(\underline{k})$ specified, the photon field is not uniquely determined. However, if $x(\underline{k}_1, \cdots \underline{k}_n) = \phi(\underline{k}_1) \cdots \phi(\underline{k}_n)$, then $n(\underline{k}) = \Sigma |A_n|^2 n |\phi(k)|^2$, and the dependence of $n(\underline{k})$ on the wave function is quite simple. This wave function is that of an intense but uncorrelated beam.

With these preliminaries done, we now go to a discussion of an electron in interaction with the photon field. The state of the system comprised by the electron and the photon field can be written as

$$|\text{electron photon}> =$$

$$\sum_n \int [d\underline{k}]^n \, [d\underline{p}] \, x(\underline{p}; \underline{k}_1, \cdots \underline{k}_n) \, A_p^+ C_{k_1}^+ \cdots C_{k_n}^+ |0>$$ (8)

Here $\underline{p}$ refers to the electron momentum, and A_p^+ is the appropriate creation operator. The function $x(\underline{p}; \underline{k}_1, \cdots \underline{k}_n)$ can be expanded in a complete set of electron wave functions, say $\phi_i(\underline{p})$, so that

$$x(\underline{p}; \underline{k}_1, \cdots \underline{k}_n) = \sum B_i \, \phi_i(\underline{p}) \, x_i(\underline{k}_1, \cdots \underline{k}_n)$$ (9)

where B_i has been inserted to keep x_i normalized. We see, therefore, that there is no loss in generality if we choose to write the states of the photon-electron system as $\phi(p)\, x\ (\text{photon})$; if the S matrix element we wish to consider is

$$< x'\ (p';\ \text{photon})\ |S|\ x(p;\ \text{photon})> \qquad (10)$$

then evidently this can be written

$$\sum_{i',\,i} B_{i'}'^+\, B_i\, < \psi_{i'}\ (p')\ x_{i'}'\ (\text{photon})\ |S|\ x_i\ (\text{photon})\ \psi_i\ (p) > \qquad (11)$$

To study the electron-photon scattering, then we need only evaluate

$$< \psi'\ (p')\ x'\ |S|\ x\psi\ (p) > \equiv$$

$$\int [dp']\ [dp]\ \psi'^+\ (p')\ \psi\ (p) < x'\ |S\ (p,p')|\ x > \qquad (12)$$

The S matrix, $S(p',p)$, with radiative corrections ignored can be written as

$$S(p',p) = \sum_{m,\,m'} \int [d\kappa']^{m'}\ [d\kappa]^m$$

$$\times\ \sigma(p', \kappa', \cdots \kappa'_{m'}; \kappa_1, \cdots \kappa_m; p)\ C_{\kappa_1}^+ \cdots C_{\kappa_{m'}}^+\, C_{\kappa_1} \cdots C_{\kappa_m} \qquad (13)$$

From standard S matrix theory we know that the function consists of sums of time-ordered Feynman graphs. In each graph there are the two external electron lines— p entering and p' leaving. And also in each graph are the m external photon lines $\kappa_1 \cdots \kappa_m$ that terminate at absorption vertices, and the m' external photon lines $\kappa'_1 \cdots \kappa'_m$ that begin at emission vertices. The sum implied in σ is a sum over all time orderings of these $m + m'$ vertices.

By expanding x' and x we can now write our matrix element as

$$\sum_{n',\,n} A_{n'}'^+ A_n \int [dk']^{n'}\ [dk']^n\ [d\kappa]^{m'}\ [d\kappa]^m$$

$$\sigma\ (p'; \kappa', \cdots \kappa'_m; \kappa, \cdots \kappa_m; p)$$

$$\frac{x'^+\ (k'_1 \cdots k'_{n'})\ x(k_1 \cdots k_n)}{\sqrt{n'}\qquad\qquad \sqrt{n'}}$$

$$< C_{k_1'} \cdots C_{k'_{n'}}\, C_{\kappa_1'}^+ \cdots C_{\kappa'_{m'}}^+\, C_{\kappa_1} \cdots C_{\kappa_m}\, C_{k_1}^+ \cdots C_{k_n}^+ > \qquad (14)$$

Evidently this vanishes unless $n \geq m$ and $n' \geq m'$ and unless $n - m = n' - m' \equiv q$. We have, in fact, for each n and n'

$$\int [dk']^{m'}\ [dk]^m\ [d\lambda]^q \binom{n}{q}\binom{n'}{q} \frac{m'!}{\sqrt{n'!}}\ \frac{m!\, q!}{\sqrt{n!}}$$

$$x^+(k'_1, \cdots k'_{m'}; \lambda_1, \cdots \lambda_q)\, x(k_1, \cdots k_m; \lambda_1, \cdots \lambda_q)$$

$$\sigma(p'; k'_1, \cdots k'_{m'}; k_1, \cdots k_n; p) \qquad (15)$$

The combinatorial factors in Eq. (15) are a consequence of the equation that now follows, itself a consequence of the commutation relations:

$$C_{\kappa_i} \cdots C_{\kappa_m} C_{k_1}^+ \cdots C_{k_n}^+ |0> =$$

$$\sum_{j_{m+1},\, j_{m+2},\, \cdots j_n}^{n'} C_{k_{j_{m+1}} \ldots\ k\, j_n} C_{k_{j_{m+1}}}^+ \cdots C_{k_{j_n}}^+ |0> \qquad (F1)$$

Here $C_{k_{j_{m+1}} \ldots\ k\, j_n}$ is a sum over products of δ functions

$$\sum_{j_1,\, j_2,\, \cdots j_m}^{n} \delta_{\kappa_1 k_{j_1}}\, \delta_{\kappa_2 k_{j_2}} \cdots \delta_{\kappa_m k_{j_m}} \qquad (F2)$$

The prime on a sum denotes that the summation is to be taken with no index repeated. Clearly there are m' terms in each

$$C_{k_{j_{m+1}} \ldots\ k\, j_n}$$

and there are $\binom{n}{q}$ states in the sum [Eq. (F2)]. Since the σ are symmetric in the κ, the m' terms in

$$C_{k_{j_{m+1}} \ldots\ k\, j_n}$$

yield the same result on integration and give the factor m'. Finally, each of the $\binom{n}{q}$ states leads to one non-zero inner product with one of $\binom{n'}{q}$ states in the corresponding sum in the primed variables, and each such inner product yields the factor q'.

We write the matrix finally as

$$\sum_{m,\,m'} \int [dk']^{m'}\ [dk]^m\ \sigma(p'; k'_1, \cdots k'_m; k_1, \cdots k_m; p)$$

$$\times X\ (k'_1, \cdots k'_m; k_1, \cdots k_m) \qquad (16a)$$

$$X = \sum_q A_{m'+q}^+ A_{m+q} \sqrt{\frac{(m'+q)!}{q!}} \sqrt{\frac{(m+q)!}{q!}}$$

$$\times \int [d\lambda]^q\, x^{+}\ (k'_1, \cdots k'_{m'}; \lambda_1, \cdots \lambda_q)$$

$$\times\ x\ (k_1, \cdots k_m; \lambda \cdots \lambda_q) \qquad (16b)$$

This is our main formal result.

As a simple application we now examine Eq. (16) for processes in which the electron scatters an intense beam. The process we consider is one in which the incoming beam scatters as a whole into an outgoing beam without breakup of the beam. This process is a "semi-classical" one in the sense that the quantum nature of the photon beam is not apparent — the individual quanta do not disclose themselves in the scattering.

The n-photon wave function of the intense beam before scattering we take to be

$$\chi'(\underline{k}'_1 \cdots \underline{k}'_{n'}) = \psi(\underline{k}_1) \cdots \psi(\underline{k}_n) \tag{17}$$

and after the scattering

$$\chi'(\underline{k}'_1 \cdots \underline{k}'_n) = \psi'(\underline{k}'_1) \cdots \psi(\underline{k}'_{n'}) \tag{18}$$

Then

$$X = \left\{ \sum_q <\psi'|\psi>^q A^{1\dagger}_{m'+q} A_{m+q} \sqrt{\frac{(m'+q)!}{p!}} \sqrt{\frac{(m+q)!}{p!}} \right\}$$
$$\times \psi'^+(\underline{k}'_1) \cdots \psi'^+ \underline{k}'_{m'} \psi(\underline{k}_1) \cdots \psi(\underline{k}_m) \tag{19}$$

If we suppose that there are exactly $\bar{n}$ photons in the incident beam and $\bar{n}'$ photons in the scattered beam, then $A'_{m'+q} = \delta_{m'+q,\bar{n}'}$, and $A_{m+q} = \delta_{m+q,\bar{n}}$. Then the term in braces in Eq. (19) becomes

$$<\psi'|\psi>^{\bar{n}-m} \sqrt{\frac{\bar{n}'!}{(\bar{n}'-m')!}} \sqrt{\frac{\bar{n}!}{(\bar{n}-m)!}} \tag{20}$$

with $\bar{n}'-m' = \bar{n}-m$. If the photons of the incident beam and scattered beam are orthogonal—$<\psi'|\psi> = 0$—then there is only one term and that is for $\bar{n}'=m'$ and $\bar{n}=m$. If we use Stirling's approximation for the factorial, this becomes

$$2\pi \sqrt[4]{\bar{n}'\bar{n}} \left(\frac{\bar{n}'}{e}\right)^{\frac{\bar{n}'}{2}} \left(\frac{\bar{n}}{e}\right)^{\frac{\bar{n}}{2}} \tag{21}$$

This does not agree with what one would get by using the semi-classical approximation. In that approximation this term would be proportional to $(\bar{n}')^{\bar{n}'/2}(\bar{n})^{\bar{n}/2}$. Suppose now that the photons of incident and scattered beam are *not* orthogonal. Then the term in braces in Eq. (19) yields

$$<\psi'|\psi> \exp\left(\bar{n} + \frac{\bar{n}'}{2}\right) \left(\frac{\bar{n}'}{<\psi'|\psi>}\right)^{\frac{m'}{2}} \left(\frac{\bar{n}}{<\psi'|\psi>}\right)^{\frac{m}{2}} \tag{22}$$

If we now write

$$\psi' = \frac{\psi + \lambda\phi}{\sqrt{1 + |\lambda|^2}} \tag{23}$$

where

$$<\psi|\phi> = 0 \tag{24}$$

then

$$<\psi'|\psi> = \frac{1}{\sqrt{1 + |\lambda|^2}} \tag{25}$$

If we retain only first-order terms in λ in the expansion of the S matrix, we retrieve the semi-classical approximation, since in that order $<\psi'|\psi> = 1$. If both the incoming and scattered beams are coherent, then we can write

$$A_{m+p} = \exp\left(\frac{-\bar{n}^{-2}}{2}\right) \frac{\bar{n}^{(m+q/2)}}{\sqrt{(m+q)!}} \tag{26}$$

As a result the term in braces in Eq. (19) becomes

$$(\bar{n}')^{\frac{m}{2}} (\bar{n})^{\frac{m}{2}} \exp\left(-\frac{1}{2} \left\{\bar{n}^2 + \bar{n}'^2 - 2 <\bar{n}'\psi'|\bar{n}\psi>\right\}\right) \tag{27}$$

Thus, except for the exponential, we have the semi-classical expression. We can write the exponent as

$$\exp i\phi \; e^{-\frac{1}{2} <\bar{n}'\psi'-\bar{n}\psi|\bar{n}'\psi'-\bar{n}\psi>} \tag{28}$$

where $e^{i\phi}$ has to do with the relative phase of $\bar{n}' \psi'$ and $\bar{n}\psi$. If, however, $\bar{n}' \sim \bar{n}$ and $\psi' \sim \psi$, the term in braces is essentially a phase factor and the semi-classical expression holds exactly.

Thus, even in these simple examples, where the state of the photon field is uncorrelated, the semi-classical approximation is valid only in special circumstances. It would seem then that when the photon field is in a truly correlated state, the semi-classical approximation breaks down completely.

References

1. McDonald, W. S., "The JPL Spark Chamber Program," *Space Programs Summary No. 37-35*, Vol. IV, Jet Propulsion Laboratory, Pasadena, California, pp. 248–251, October 31, 1965.

2. Weibel, E., *Physical Review Letters*, Vol. 2, p. 83, 1959.

3. Sudan, R. N., *The Physics of Fluids*, Vol. 6, p. 57, 1963.

4. Penrose, O., *The Physics of Fluids*, Vol. 3, p. 258, 1960.

5. Fried, B. D., and Conte, S. D., *The Plasma Dispersion Function*, Academic Press, New York, N. Y., 1961.

TELECOMMUNICATIONS DIVISION

XVIII. Spacecraft Telemetry and Command

A. The Effect of Sensor Noise on Quantization Accuracy During Analog-to-Digital Conversion

G. L. Fultz

In many spacecraft engineering data handling and television subsystems, there is a requirement to digitize analog information. If a signal is applied to an ideal analog-to-digital encoder, as shown in Fig. 1, the resultant binary code or data number (DN) will identify a particular voltage level for each sample. Unfortunately, some noise is always associated with the input signal.

Thus, the probability that the signal will be encoded into a DN which does not correspond to the noise-free encoded DN will be derived.

Consider the input signal uniformly distributed over the smallest quantization interval (voltage level $2a$) of the analog-to-digital converter and independent of the noise voltage. Let the signal have value z at the sampling time, and let the noise have power σ^2. Referring to Fig. 2, the probability (PW) that the signal is encoded improperly, given z, is the shaded area under the curve.
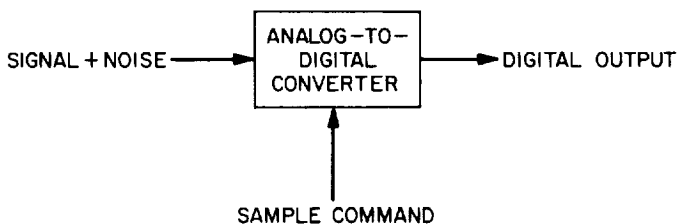
Fig. 1. Signal digitization
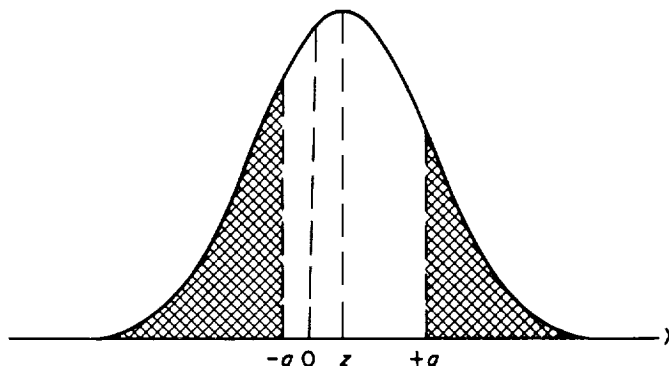
Fig. 2. Mathematical model

Let $PC = 1 - PW$ be the probability of correct encoding. First, the probability distribution of signal plus noise is found by convolving the probability distribution of signal $P(z)$ and noise $P(x)$. Let

$$\lambda = x + z$$

$$P(z) = \frac{1}{2a} \qquad -a \leq z \leq +a$$

$$P(x) = \frac{1}{(2\pi)^{1/2} \sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \qquad -\infty \leq x \leq +\infty$$

Thus,

$$P(\lambda) = \frac{1}{2a} \frac{1}{(2\pi)^{1/2} \sigma} \int_{-a}^{a} \exp\left[\frac{-(\lambda - x)^2}{2\sigma^2}\right] dx$$

$$= -\frac{1}{2a} \frac{1}{(\pi)^{1/2}} \int_{(\lambda+a)/(2)^{1/2}\sigma}^{(\lambda-a)/(2)^{1/2}\sigma} \exp(-y^2)\, dy$$

Let

$$\text{erf}(A) = \frac{2}{\pi^{1/2}} \int_{0}^{A} \exp(-y^2)\, dy$$

Thus,

$$P(\lambda) = \frac{1}{4a}\left\{\text{erf}\left[\frac{\lambda + a}{(2)^{1/2}\sigma}\right] - \text{erf}\left[\frac{\lambda - a}{(2)^{1/2}\sigma}\right]\right\}$$

and

$$Pc = \frac{1}{4a} \int_{-a}^{a}\left\{\text{erf}\left[\frac{\lambda + a}{(2)^{1/2}\sigma}\right] - \text{erf}\left[\frac{\lambda - a}{(2)^{1/2}\sigma}\right]\right\}$$

Let $z = (\lambda + a)/(2)^{1/2}\sigma$ for the first integration, and $z = (\lambda - a)/(2)^{1/2}\sigma$ for the second integration.

Therefore,

$$Pc = \frac{(2)^{1/2}\sigma}{4a} \int_{0}^{2a/(2)^{1/2}\sigma} \text{erf}(z)\, dz - \frac{(2)^{1/2}\sigma}{4a} \int_{-2a/(2)^{1/2}\sigma}^{0} \text{erf}(z)\, dz$$

$$= \frac{(2)^{1/2}\sigma}{2a} \int_{0}^{2a/(2)^{1/2}\sigma} \text{erf}(z)\, dz$$

$$= \frac{(2)^{1/2}\sigma}{2a}\left\{\frac{2a}{(2)^{1/2}\sigma}\text{erf}\left(\frac{2a}{(2)^{1/2}\sigma}\right)\right.$$

$$\left. - \frac{1}{\pi^{1/2}}\left[1 - \exp\left(-\left[\left(\frac{2a}{(2)^{1/2}\sigma}\right)^2\right]\right)\right]\right\}$$

Thus,

$$Pc = \text{erf}\left[\frac{2a}{(2)^{1/2}\sigma}\right] - \left(\frac{2}{\pi}\right)^{1/2}\frac{\sigma}{2a}\left[1 - \exp\left(-\frac{2a^2}{\sigma^2}\right)\right]$$

and $PW = 1 - Pc$. The signal-to-noise ratio (SNR) of the input is

$$SNR = \left(\frac{2^n 2a}{\sigma}\right)^2 = \frac{\text{peak-to-peak signal power}}{\text{rms noise power}}$$

where 2^n is the number of encoding levels.

Before PW is evaluated, a more general relation will be derived. Let $PW(L)$ be the probability that the noisy signal is encoded more than $\pm L$ DN's from the no-noise encoded DN. Using the same method of analysis as before,

$$Pc(L) = \frac{1}{4a} \int_{-(2L+1)a}^{(2L+1)a}\left\{\text{erf}\left[\frac{\lambda + a}{(2)^{1/2}\sigma}\right] - \text{erf}\left[\frac{\lambda - a}{(2)^{1/2}\sigma}\right]\right\} d\lambda$$

Thus,

$$Pc(L) = (L+1)\,\text{erf}\left[\frac{2a(L+1)}{(2)^{1/2}\sigma}\right] - (L)\,\text{erf}\left[\frac{2aL}{(2)^{1/2}\sigma}\right]$$

$$- \frac{(2)^{1/2}\sigma}{2a(\pi)^{1/2}}\left\{\exp\left[-\frac{2a^2L^2}{\sigma^2}\right] - \exp\left[-\frac{2a^2(L+1)^2}{\sigma^2}\right]\right\}$$

In order to calculate $PW(L)$, let

$$SNR' = \left(\frac{2a}{\sigma}\right)^2$$

and

$$a = \frac{1}{2}$$

Then

$$SNR = (2^n)^2 \left(\frac{1}{\sigma}\right)^2 = (2^n)^2\, SNR'$$

The results are shown in Fig. 3.

The principal application for these results is in estimating the photometric accuracy of a digital television system. The *Mariner IV* TV system will be used as an example.

Consider picture 21. This picture was taken when the television system was viewing a portion of Mars beyond the terminator or dark side of the planet. Since the error rate caused by the noisy communication channel was very small, the DN contained in picture 21 should be random. This randomness is caused by the sensor and vidicon electronics noise. The distribution of noise in picture 21 can be related to the other pictures.
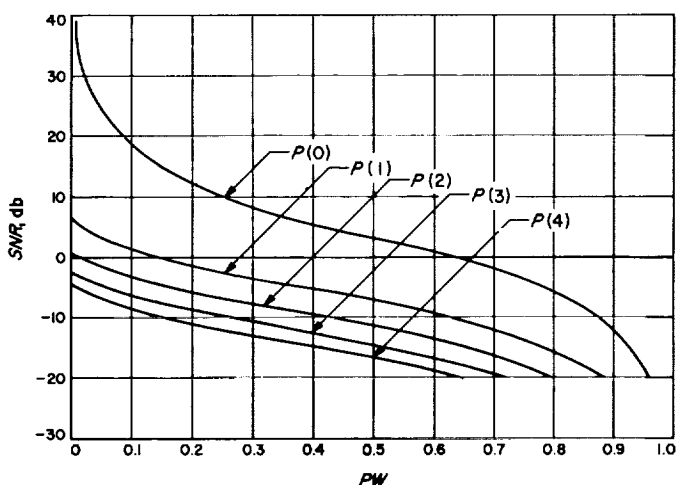
Fig. 3. Probability of incorrect in coding

DEVIATION MEASURED, DN
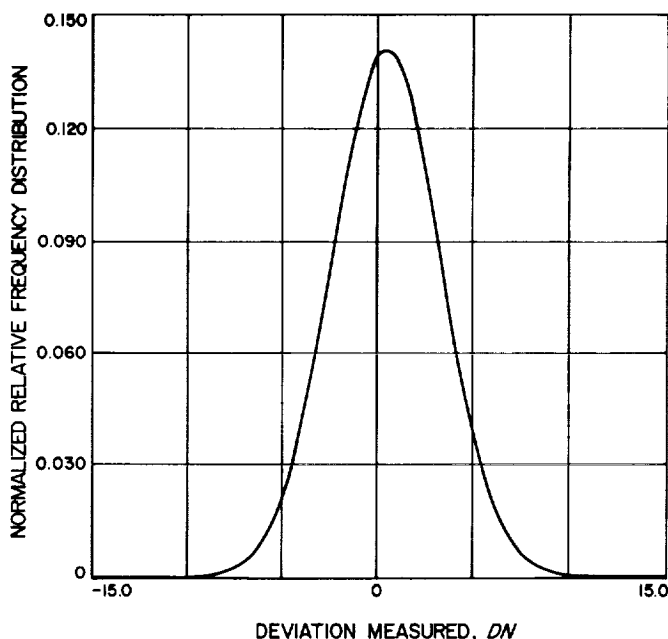
Fig. 4. Noise amplitude distribution

Pictures 1–18 were taken with a constant gain setting of the video amplifier. The gain of the video amplifier for picture 21 was $(2.8)^3$ above that for the first eighteen pictures.

To determine the noise distribution for picture 21, the mean and variance of the DN cannot be measured directly due to shading in the vidicon and the fiducial marks. To determine the noise distribution, use the following procedure:

(1) Eliminate the data samples corresponding to fiducial information.

(2) Assume the shading is a linear function of position along a television line.

(3) Fit a curve $Y_p = a + bX$ in the mean square sense to each television line where Y_p is the predicted intensity DN, and X is the position along the television line.

(4) Compute the deviation of the actual television data from Y_p.

(5) Form a relative frequency distribution as each line is processed.

(6) Compute the mean and variance of the resultant relative frequency distribution.

(7) Plot the experimental data and overlay a Gaussian curve with the same mean and variance.

The resulting noise distribution is shown in Fig. 4 with $\sigma = 2.82$ DN's. Scaling this result to the first 18 pictures, $\sigma = 0.13$ DN. Thus, $SNR' = (1/\sigma)^2 = 17.7$ db. Referring to Fig. 3, $PW\,(L = 0) \cong 0.11$ and $PW\,(L \neq 0) \cong 0$.

Thus, when examining brightness information in the first 18 *Mariner IV* pictures, 89% of the DN's will be within one DN of the true scene brightness, and approximately none of the DN's will be more than one DN from the true scene brightness.

B. Advanced Data Processing Systems

R. F. Trost

1. Introduction

a. Background. Early in 1965 a decision was made to develop a breadboard engineering data handling system that uses data compression. The model chosen for this development is a typical *Voyager* orbiter, a choice influenced by the need to do an *a priori* design process for a data compression system. Previous work had modeled only past missions.

b. Objectives. The primary objective of this development is to demonstrate the feasibility of an advanced engineering data handling system (EDHS) that uses data compression and its attendant functions. Data compression is defined as the removal of redundancy (as

defined by the data user) from sensor signals. Functions attendant to data compression are confidence sampling (automatic and commanded), controlled redundancy introduction for operational measurements, data buffering, and priority selection during buffer overloading in addition to the standard functions of multiplexing, signal conditioning, and quantizing. Feasibility will be demonstrated for:

(1) Hardware:

    (a) Design feasibility.

    (b) Reliability, weight, and power costs.

(2) Efficiency of information transmission.

A secondary objective is to provide meaningful information for later flight system development, design, and fabrication.

c. Methods. The hardware design feasibility objective will be pursued by the design and fabrication of a breadboard data compression system using modules that are as representative as possible of flight hardware, and by making flight reliability, weight, and power cost appraisals.

The hardware reliability, weight, and power cost objectives will be approached in three basic steps:

(1) Software implementation of the system transfer function in parallel with the system hardware design.

(2) Showing agreement of outputs for identical input data to establish confidence in the software implementation.

(3) Further, using software simulation to derive other quantitative results from data that cannot easily be inputted to a hardware system (for example, telemetry data stored on magnetic tape is more suitable for input to a software simulation, while real time data generation on an individual source basis is more suitable for input to a hardware multiplexer).

The secondary objective acknowledges the importance of a good information turnover in a development process. Data compression systems are more user oriented. Consequently, data requirements cannot be specified as in past systems, and new design methods must be established. This objective also recognizes that a data handling system input is probably the most variable interface on the spacecraft. Even though this interface will change, the design philosophy and certain desirable functional characteristics of the system must not be compromised during later development and flight design.

In summary, information generated for the turnover process must allow for knowledgeable deviation from the specific system while ensuring that functional integrity is preserved. Accordingly, the secondary objective will be realized by:

(1) The compilation of a specific set of system input measurements in consultation with engineering data users.

(2) The compilation and classification of data user processing requirements and the methods used to derive these requirements.

(3) The outlining of a system design philosophy, design criteria, and functional specifications based on the compiled data processing requirements.

d. Results. By June 1965, the mission assumptions had been written, a measurement list had been compiled, and the measurement processing requirements were in the process of being defined in detail. These results were then used to describe the system functional requirements. By August 1965, the system design criteria and the subsystem functional specifications for the EDHS had been completed. As of November 1965, the computer simulation of the EDHS has been initiated while the hardware implementation of it continues on schedule. A block diagram and brief description of the basic EDHS will now be presented.

2. The EDHS

As shown in Fig. 5, the EDHS is composed of three basic subsystems: (1) the absolute rate commutator, (2)
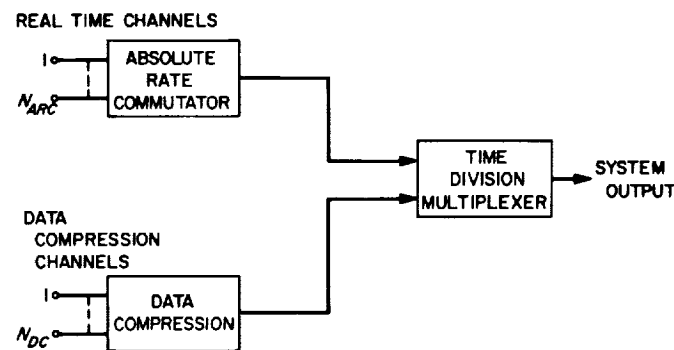
Fig. 5. Engineering data handling system block diagram

the data compression subsystem, and (3) the time division multiplexer.

a. Absolute rate commutator subsystem. It is the function of the absolute rate commutator to sample the real time measurements at an absolute rate independent of the communication link bit rate. For example, if a particular measurement requires a sampling rate of 1 sample/min, then the absolute rate commutator will sample this measurement once every minute regardless of the telemetry bit rate. The assumption, of course, is that the lowest telemetry bit rate is at least adequate to support the highest sampling rate available on the absolute rate commutator. This point will be expanded upon later.

It should be mentioned that the output from the absolute rate commutator subsystem contains both an address and sample value for each input channel. Thus, an output word from it is of the same length as an output word from the data compression subsystem.

b. Data compression subsystem. The data compression subsystem basically contains a processor, a processing memory, a buffer control routine, and a buffer memory. For each input measurement there is a set of parameters stored in the processing memory. These parameters essentially describe various characteristics of a particular measurement so that a maximum amount of information may be derived from it by proper processing. The processor applies the compression algorithm to the incoming data, and then significant data, together with its address, is temporarily stored in the buffer memory. At the same time, the buffer control routine monitors the fullness of the buffer memory and, if necessary, alters the processing rate to control the number of samples entering the buffer memory.

c. Time division multiplexer. The time division multiplexer shown in Fig. 5 combines the outputs from the other two subsystems. However, this is not an ordinary multiplexer because it must be capable of giving priority to the absolute rate commutator whenever it produces a sample for transmission. During this time, the data compression subsystem can wait, since it has an inherent buffering capability.

The output of the multiplexer operates periodically with a fixed, absolute period T_{EDHS}. As shown in Fig. 6, a fraction of this period, T_{ARC}, is assigned to the absolute rate commutator and the remainder, T_{DC}, is assigned to

T_{ARC} TIME ALLOCATED TO THE ABSOLUTE RATE COMMUTATOR,
T_{DC} TIME ALLOCATED TO THE DATA COMPRESSION SYSTEM
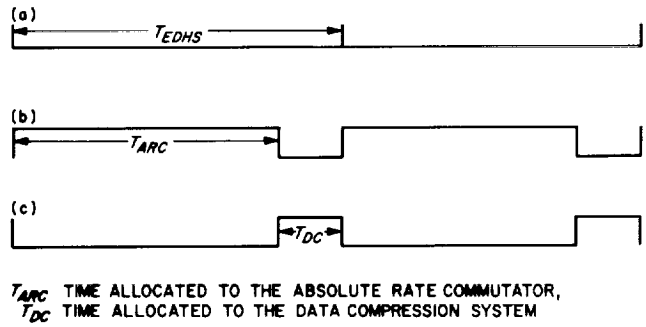
Fig. 6. Time sharing scheme by the time division multiplexer

the data compression subsystem; clearly,

$$T_{EDHS} = T_{ARC} + T_{DC} \qquad (1)$$

The times T_{ARC} and T_{DC}, the number of inputs on the absolute rate commutator, N_{ARC}, and the maximum sampling rate, R_{MAX}, must all be chosen to be compatible at the lowest telemetry word rate, W_{MIN}. Each real time measurement is sampled, at most, only once during each T_{ARC} cycle. This implies that

$$R_{MAX} \leq 1/T_{EDHS}. \qquad (2)$$

Furthermore, at the minimum word rate it must be possible for all N_{ARC} channels to be sampled at $R_{MAX} = 1/T_{EDHS}$. This implies that

$$N_{ARC} \leq W_{MIN} \cdot T_{ARC}, \qquad (3)$$

where:

R_{MAX} is the maximum sampling rate available on the absolute rate commutator in samples per second.

N_{ARC} is the total number of input channels to the absolute rate commutator.

T_{EDHS} is the effective cycle time of the time division multiplexer.

T_{ARC} is the absolute time in seconds allocated to the absolute rate commutator.

T_{DC} is the absolute time in seconds allocated to the data compression subsystem.

W_{MIN} is the minimum telemetry word rate in words per second.

Eq. (2) is the condition for the EDHS to support the maximum sampling rate on the absolute rate commutator and is independent of the total number of inputs to the subsystem. Once R_{MAX} has been chosen, Eq. (3) gives the trade-offs between the total number of input channels
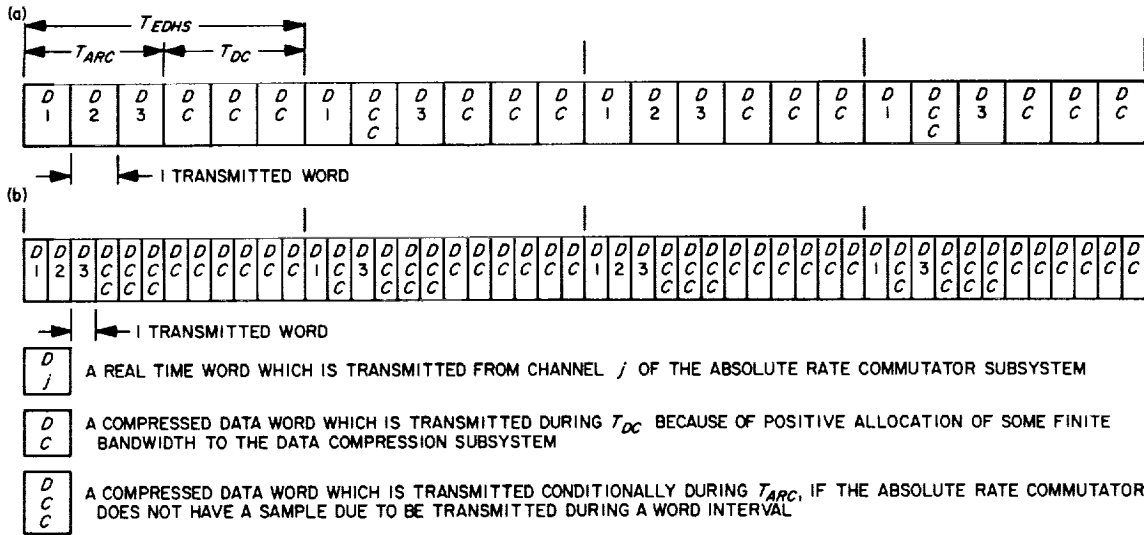
Fig. 7. Typical example of the EDHS system output at (a) the minimum bit rate B, and (b) at bit rate 2B

and the time available to sample them for a given minimum telemetry word rate.

At this point a numerical example might further illustrate these relationships. Assume the following typical values: Total word length (address + sample value) = 20 bits; minimum telemetry bit rate = 50 bits/sec; desired $R_{MAX} = 1/30$ (one sample per 30 sec) and, $T_{ARC} = T_{DC} = (1/2)T_{EDHS}$.

By Eq. (2), $T_{EDHS} \leq 30$ sec, therefore $T_{ARC} \leq 15$ sec.

By Eq. (3), $N_{ARC} \leq (50/20) \, 15 = 37.5$

Therefore, with the assumed constraints, the absolute rate commutator would be able to support 37 input channels at the maximum rate.

d. System output. In general, the absolute rate commutator will have some input channels being sampled at rates slower than R_{MAX}. When this condition occurs, there will be some word intervals during T_{ARC} which contain no output from the absolute rate commutator. This absence is an indication to the time division multiplexer to fill in this gap with a word from the data compression subsystem. This type of fill-in word is termed a "conditional data compression" word, and it is necessary to maintain continuity in the system output; it also tends to increase the effective bandwidth allocated to the data compression subsystem.

Fig. 7 illustrates the data stream at the system output for a minimum bit rate B, and also 2B. For simplicity, $N_{ARC} = 3$; $T_{ARC} = (1/2)T_{EDHS}$; $R_{MAX} = 1/T_{EDHS}$; $R_{MIN} = (1/2)R_{MAX}$; and $W_{MIN} = 6/T_{EDHS}$. Channels 1 and 3 on

the absolute rate commutator have an assigned rate of R_{MAX}, while the sampling rate for channel 2 is R_{MIN}.

Note in Fig. 7(b) that the system output contains additional data compression samples when the bit rate increases while the absolute rate commutator produces a constant output. This is the main advantage of the absolute rate commutator over the normalized rate commutators that were used in the *Ranger* and *Mariner* spacecraft. Fig. 8 compares the bandwidth available to a data compression subsystem when it is used with these two types of real time commutators.

Finally, it should be pointed out that although this discussion assumed that the EDHS system output was continuous, it is not a necessary assumption. Indeed, if the system output of the EDHS must time share the communication channel with another subsystem (i.e., scientific
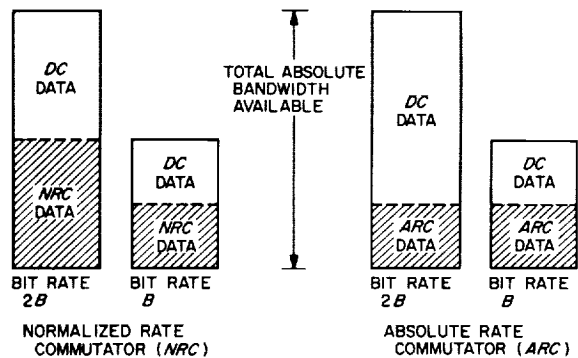
Fig. 8. Comparison of percentage bandwidth available to a data compression system for two different types of commutators as the bit rate increases

measurements) for T_0 seconds, then the discussion presented herein is still applicable by defining $T_{DC} = T_{DC}' + T_0$ in Eqs. (1)–(3). In this case T_{DC}' is still defined as the absolute time in seconds allocated to the data compression subsystem.

3. Conclusion

The purpose of this article is to introduce the basic concepts involved in the EDHS, currently being developed. Additional details concerning the individual subsystems will appear in subsequent issues of the SPS. Present and future effort on the data compression transfer function simulation will also be reported in the SPS.

C. A Note on Signal-to-Noise and Signal-to-Noise Spectral Density Ratios at the Output of a Filter-Limiter Combination

J. C. Springett

1. Introduction

The bandpass limiter has been the subject of a great deal of analysis in the past and has resulted in a number of widely quoted publications, the most notable of which is probably Davenport's (Ref. 1). The initial objective of this article is to provide the reader with the signal-to-noise spectral density ratio $(S/N_0)_0$ at the output of the limiter. Motivation for the analysis stems from the fact that in many cases, the limiter is followed by a narrow filter (e.g., a phase-locked loop, matched filter, etc.) rather than a zonal filter. However, during the course of analysis several additional items were investigated which proved or showed the following:

(1) The signal plus total noise power in any zone at the output of the limiter is a constant; i.e., it does not change as one varies the input signal-to-noise ratio (SNR). Further, this constant is independent of the input noise covariance function.

(Note: In this presentation the input process of sine wave plus noise is always considered to be bandpass and narrowband, i.e., $2\Delta\omega << \omega_0; \omega_0 =$ center frequency; $2\Delta\omega =$ input bandwidth, such that there is no significant overlap of the spectral zones.)

(2) The suppression factors for the signal components in the various zones are different, and, therefore, the SNR's in each zone for a given input SNR are different.

(3) Based upon (1) above, the first zone SNR was calculated and disagrees with Davenport's result over a rather wide range of input SNR.

The following summary will present the highlights of the analysis and the calculated results.

2. Analysis Summary (Refs. 2, 3, 4, 5)

For the reader who is not intimately familiar with the analysis that leads to an expression for the output second moment function, it is presented here in summary form.

We consider an ideal limiter with transfer characteristics shown in Fig. 9, where

$$z = g(y) = \begin{cases} +1 \; y(t) > 0 \\ -1 \; y(t) < 0 \end{cases}, \text{ and } y(t) = S(t) + N(t).$$
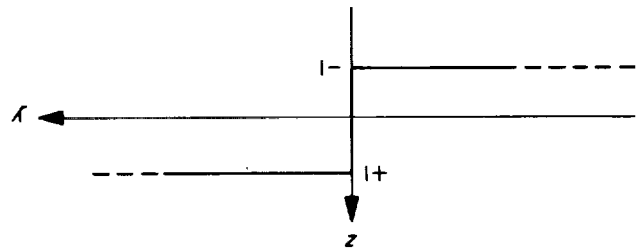
Fig. 9. Limiter transfer characteristic

The input signal $S(t) = \sqrt{2A} \cos(\omega_0 t + \phi)$, with ϕ being uniformly distributed over one cycle of ω_0. The noise $N(t)$ will be gaussian and narrowband, i.e., $2\Delta\omega << \omega_0'$, $\omega_0 =$ the noise spectrum center frequency about which the spectrum is symmetrical; and $2\Delta\omega =$ the noise bandwidth. The covariance function of the noise is given by $K_N(\tau) = K_0(\tau) \cos \omega_0 \tau$, where $K_0(\tau)$ is a lowpass covariance function.

The second-moment function of a process that has undergone a nonlinear transformation is given by:

$$M_z(\tau) = \frac{1}{(2\pi)^2} \int_c \int f(j\xi_1) f(j\xi_2) F_2(j\xi_1, j\xi_2; \tau) d\xi_1 d\xi_2 \qquad (1)$$

where $f(j\xi)$ is the complex Fourier transform of the nonlinear transfer characteristic, and F_2 is the second-order characteristic function of the input process. In the case at hand because of odd symmetry

$$F_2(j\xi_1, j\xi_2; \tau) = F_2(j\xi_1, j\xi_2; \tau) - F_2(j\xi_1, -j\xi_2; \tau)$$

and $F_{2(S+N)} = F_{2(S)} F_{2(N)}$, since it will be assumed that the input signal and noise processes are independent.

The second order characteristic functions for signal and noise are, respectively,

$$F_{2(S)} = \sum_{m=0}^{\infty} (-1)^m \varepsilon_m J_m (\sqrt{2}A\xi_1) J_m (\sqrt{2}A\xi_2) \cos m\omega_0\tau \quad (2)$$

and

$$F_{2(N)} = \exp\left[-\frac{\sigma^2}{2} (\xi_1^2 + \xi_2^2) \right] \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \xi_1^n \xi_2^n K_N^n (\tau) \quad (3)$$

where ε_m = Neumann factor, $\varepsilon_0 = 1$, $\varepsilon_m = 2 (m \neq 0)$, $J_m ( )$ is a Bessel function of the first kind, and $\sigma^2 = N_0\Delta\omega/\pi$, the input noise power with N_0 = single-sided noise spectral density prior to filtering.

Substituting into Eq. (1) we find integrals of the type

$$h_{m,n} = \int_c \exp\left(-\frac{\sigma^2\xi^2}{2}\right) \xi^{n-1} J_m (\sqrt{2}A\xi) \, d\xi \quad (4)$$

and also that the function exists only for $m + n$ odd. As a result we obtain

$$M_z (\tau, X) = \frac{1}{\pi^2} \sum_{m+n=1,3,5,\cdots} \frac{\varepsilon_m}{n!} h_{m,n}^2 (X) K_N^n (\tau) \cos m\omega_0\tau \quad (5)$$

where X = the input signal-to-noise ratio A^2/σ^2.

Now

$$K_N (\tau) = \frac{N_0\Delta\omega}{\pi} \rho_N (\tau) = \frac{N_0\Delta\omega}{\pi} \rho_0 (\tau) \cos \omega_0\tau$$

where $\rho_0 (\tau)$ is the normalized lowpass noise autocorrelation function, $\rho_0 (0) \equiv 1$, and we define the amplitude factors $C_{m,n}$ as

$$C_{m,n} = \frac{\varepsilon_m}{\pi^2 n!} \left(\frac{N_0\Delta\omega}{\pi}\right)^n h_{m,n}^2 = \frac{\varepsilon_m}{\pi^2 n!} \sigma^{2n} h_{m,n}^2$$

$$= \frac{2^n \varepsilon_m}{n!} X^m \frac{{}_1F_1^2 \left[\dfrac{m+n}{2}; m+1; -X\right]}{\Gamma^2 (m+1) \Gamma^2 \left(1 - \dfrac{m+n}{2}\right)} \quad (6)$$

where ${}_1F_1 [\ ]$ is the confluent hypergeometric function, and $\Gamma ( )$ is the gamma function. Thus, we have

$$M_z (\tau, X) = \sum_{m+n=1,3,5,\cdots}^{\infty} C_{m,n} \rho_N^n (\tau) \cos m\omega_0\tau \quad (7)$$

Partially expanding, we obtain

$$M_z (\tau, X) = \sum_{m=1,3,5,\cdots}^{\infty} C_{m,0} \cos m\omega_0\tau$$

$$+ \sum_{n=1,3,5,\cdots}^{\infty} C_{0,n} \rho_N^n (\tau)$$

$$+ \sum_{m+n=3,5,\cdots}^{\infty} C_{m,n} \rho_N^n (\tau) \cos m\omega_0\tau \quad (8)$$

Eq. (8) breaks up the second moment function into sets of terms representing the interaction of the input signal with itself ($S \times S$), the interaction of the input noise with itself ($N \times N$), and the interaction of the input signal and noise ($S \times N$). Taking the Fourier transform to obtain the intensity spectrum we have

$$\mathcal{W}_z (f, X) = \sum_{m=1,3,5,\cdots}^{\infty} \frac{C_{m,0}}{2} [\delta (f - mf_0) + \delta (f + mf_0)]$$

$$+ \sum_{n=1,3,5,\cdots}^{\infty} \frac{C_{0,n}}{2} [\mathcal{W}_N (f - f_0)$$

$$+ \mathcal{W}_N (f + f_0)]^{*(n-1)}$$

$$+ \sum_{m+n=3,5,\cdots}^{\infty} \frac{C_{m,n}}{4} [\mathcal{W}_N (f - f_0)$$

$$+ \mathcal{W}_N (f + f_0)]^{*(n-1)}$$

$$* [\delta (f - mf_0) + \delta (f + mf_0)] \quad (9)$$

where $\mathcal{W}_N (f - f_0)$ is the single-sided bandpass spectrum of the noise, $*$ denotes convolution, and $[\ ]^{*(n-1)}$ denotes the $(n - 1)$ fold convolution of the quantity within the brackets.

Examining the terms in the first zone, we obtain:

$$\mathcal{W}_z (f, X)|_{f=f_0} = \frac{C_{1,0}}{2} [\delta (f - f_0) + \delta (f + f_0)]$$

$$+ \sum_{n=1,3,5,\cdots}^{\infty} \frac{C_{0,n}}{2} [\mathcal{W}_N (f - f_0)$$

$$+ \mathcal{W}_N (f + f_0)]^{*(n-1)}_{f=\pm f_0}$$

$$+ \sum_{n=2,4,6,\cdots}^{\infty} \sum_{m=0,2,4,\cdots}^{n} \frac{C_{m+1,n}}{4} [\mathcal{W}_N (f - f_0)$$

$$+ \mathcal{W}_N (f + f_0)]^{*(n-1)}_{\{f=\pm [mf_0]\}+\{f=\pm [(m+2)f_0 \lessgtr nf_0]\}}$$

$$+ \sum_{n=1,3,5,\cdots}^{\infty} \sum_{m=1,3,5,\cdots}^{n} \frac{C_{m+1,n}}{4} [\mathcal{W}_N (f - f_0)$$

$$+ \mathcal{W}_N (f + f_0)]^{*(n-1)}_{\{f=\pm [mf_0]\}+\{f=\pm [(m+2)f_0 \lessgtr nf_0]\}}$$

$$(10)$$
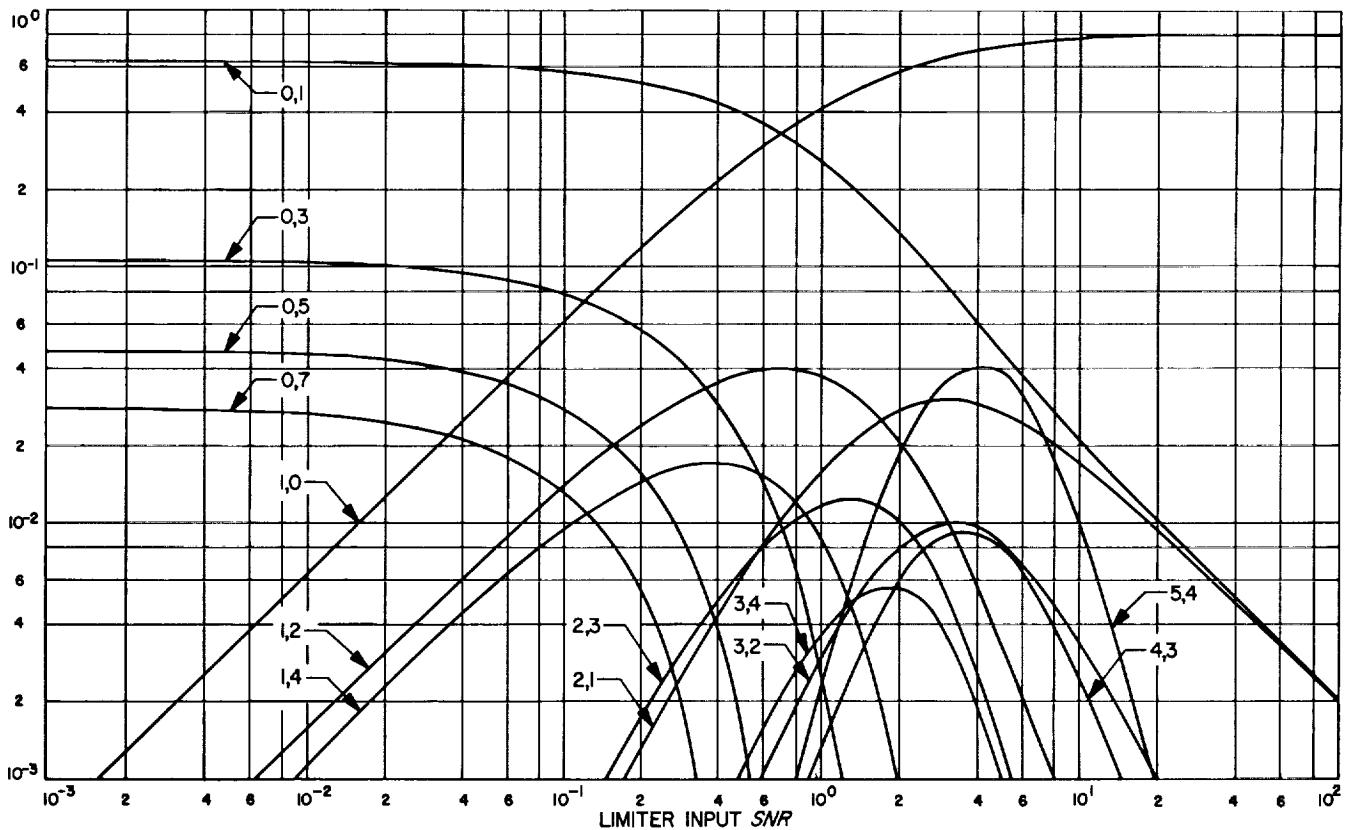
Fig. 10. Amplitude factors

where $\{f \pm [mf_0]\} + \{f \pm [m + 2)f_0 \leq nf_0]\}$ means evaluation and sum of the convolution at the $\pm mf_0$ and $\pm (m + 2)f_0$ zones, with the latter omitted if $(m + 2)$ $f_0 > nf_0$.

3. Results

To obtain the signal-to-noise spectral density $(S/N_0)_0$ in the first zone, the convolutions are evaluated only at the appropriate $\pm mf_0$ points. This was done here assuming a rectangular noise spectrum of unit height and width $2\Delta f$, for $(n - 1)$ up to 6. The corresponding amplitude factors were calculated using a computer and are plotted in Fig. 10. Fig. 11 shows the result plotted in the form of $(S/N_0)_0/(S/N_0)_i$ against X.

It can be seen from the amplitude factor plots of Fig. 10 that at very high input SNR the only significant noise terms are $(0,1)$ and $(2,1)$, which correspond to nothing more than weighting the input noise spectrum. As a result, there is no appreciable amount of spectral spreading at high input SNR. It is, therefore, not surprising to see that the ratio $(S/N_0)_0/(S/N_0)_i$ approaches $+3.0$ db, just as Davenport's zonal result does.
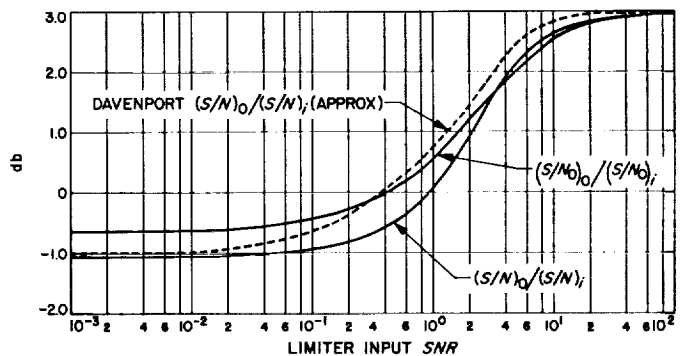
Fig. 11. Limiter 1st zone signal-to-noise ratios

At low input SNR, the significant factors are the $(0,1)$, $(0,3)$, etc., and we observe a definite spectral spreading of the noise. As can be seen from Fig. 11, $(S/N_0)_0/(S/N_0)_i$ is on the order of -0.65 db.

For the input SNR range between 0.01 and 10, the more significant $S \times N$ terms must be taken into account. Since the input SNR to the limiter is generally between 0.1 and 1 in most practical applications, it can be seen from Fig. 11 that the gain or loss resulting from following the

limiter with a narrow band filter (i.e., its bandwidth is much less than $2\Delta\omega$) is small, and probably would not be observed in most situations because of measurement inaccuracy.

We now turn our attention to some of the properties of the zonal SNR. It has been conjectured for a long time that the signal plus noise power in each zone totals to a constant independent of X. Proof of this is given by Tausworthe SPS 37-35, Vol. IV, p. 307. An alternate proof is to differentiate the second moment function representing the zone with respect to X, set $\tau = 0$, and show that the result is zero. With this result then, and since for an input $SNR = \infty$ the output of the limiter is a square-wave of ± 1 amplitude, we have $S_1 + N_1 = 8/\pi^2$, $S_3 + N_3 = 8/9\pi^2$, etc., where the subscripts refer to the zones.

The first zone output SNR is then calculated using

$$\left(\frac{S}{N}\right)_0 = \frac{S_1}{N_1} = \frac{S_1}{\dfrac{8}{\pi^2} - S_1} = \frac{C_{1,0}}{\dfrac{8}{\pi^2} - C_{1,0}}$$

The plot appears in Fig. 11. Davenport's result is also plotted, as best as can be interpreted from the rather small graph in Ref. 1, for comparison. As can be seen, there is a rather large discrepancy between the two results.

Finally, $(S/N)_0/(S/N)_i$ was calculated for the third zone and appears in Fig. 12. Note the rather large output SNR loss (6.6 db) for the relatively high input SNR of 100, and the fact that for low input SNR, the ratio falls off at 20 db/octave. In Fig. 13, the signal power suppression factors for the first and third zones are plotted.
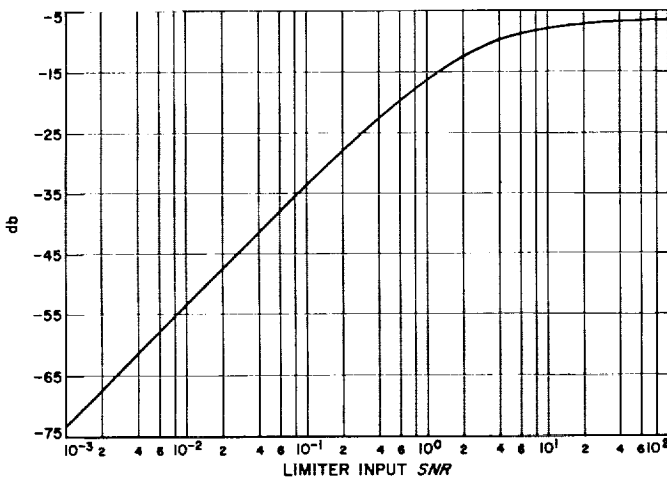
Fig. 12. $(S/N)_0/(S/N)_i$ 3rd zone
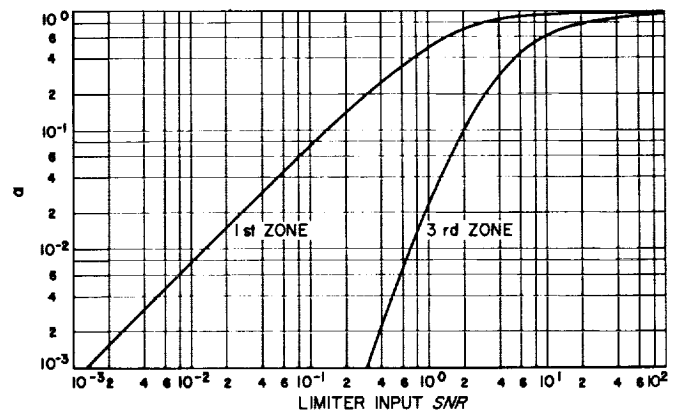
Fig. 13. Signal power suppression factors

D. Error Probability for Transmission of M-ary Orthogonal Signals Over a Partially Coherent Channel

N. P. Shein

An extensive treatment of a partially coherent binary communication system has been given in a recent paper by Viterbi (Ref. 6). In this article a method of Nuttall (Ref. 7) is used to obtain an expression for the error probability of a partially coherent channel for the case when M orthogonal, equally probable, equal energy signals are transmitted.

The system we will consider is described in Fig. 14. (Underbars are used to indicate quantities in complex
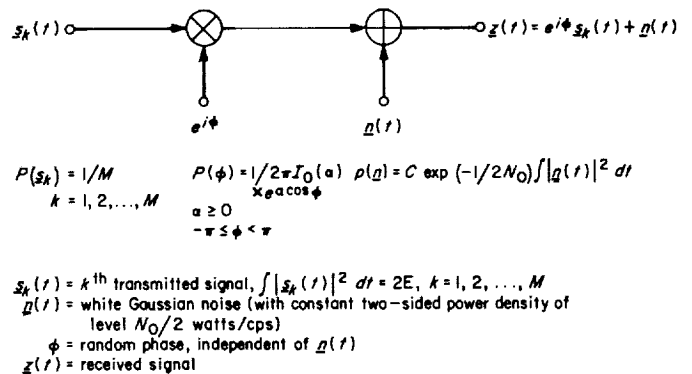
Fig. 14. Model of a partially coherent communication system

analytic form, e.g., $\underline{z}(t) = z(t) + i\hat{z}(t)$, "^" denotes Hilbert transform.) We assume the signals to be given by

$$\underline{s}_k(t) = s_k(t) + i\hat{s}_k(t) \qquad k = 1, 2, \cdots, M \qquad (1)$$

where

$$\int s_j(t)\, s_k(t)\, dt = \int \hat{s}_j(t)\, \hat{s}_k(t)\, dt = \begin{cases} E, & j = k \\ 0, & j \neq k \end{cases} \qquad (2)$$

$$\int s_j(t)\, \hat{s}_k(t)\, dt = -\int \hat{s}_j(t)\, s_k(t)\, dt = 0, \qquad \text{all } j \neq k \qquad (3)$$

It can be easily shown that the Bayes decision rule for this case becomes

$$\text{decide } \underline{s}_k(t) \text{ sent if } \left| \frac{\alpha_k}{N_0} + \alpha \right| > \left| \frac{\alpha_j}{N_0} + \alpha \right|, \qquad \text{all } j \neq k \qquad (4)$$

where

$$\alpha_k = \int \underline{z}^*(t)\, \underline{s}_k(t)\, dt \qquad k = 1, 2, \cdots, M \qquad (5)$$

One possible detector implementation of this decision rule is given in Fig. 15.

In order to obtain the probability of correct detection for the system of Fig. 14, we consider first the statistics of the quantities

$$q_j = \int \underline{n}^*(t)\, \underline{s}_j(t)\, dt = Re\, q_j + i\, Im\, q_j$$

$$j = 1, 2, \cdots, M \qquad (6)$$
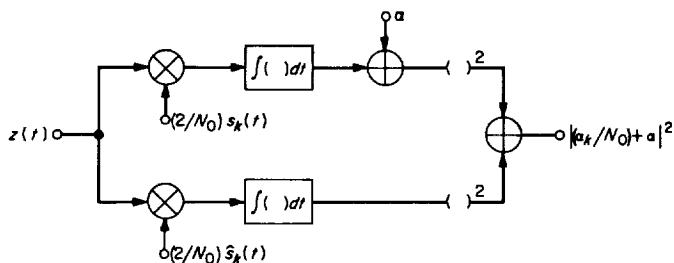
Fig. 15. Implementation of detector for partially coherent reception of M orthogonal signals

It is readily established that the q_j's have the joint "probability density"

$$p(q_1, q_2, \cdots, q_M)$$

$$\underset{\underline{=}}{\triangleq} p(Re\, q_1, Im\, q_1; Re\, q_2, Im\, q_2; \cdots; Re\, q_M, Im\, q_M)$$

$$= \frac{1}{\pi^M (4N_0 E)^M} \exp \left\{ -\frac{1}{4N_0 E} \sum_{j=1}^{M} |q_j|^2 \right\} \qquad (7)$$

Next we determine the joint probability density of the α_k's, subject to the condition (without loss of generality) that signal $s_1(t)$ has been transmitted. In this case,

$$\alpha_1 = \int \underline{z}^*(t)\, \underline{s}_1(t)\, dt = 2E \cdot e^{i\phi} + q_1 \qquad (8)$$

$$\alpha_k = \int \underline{z}^*(t)\, \underline{s}_k(t)\, dt = q_k, \qquad k = 2, 3, \cdots, M \qquad (9)$$

Then noting that

$$p(\alpha_1, \alpha_2, \cdots, \alpha_M | s_1) = \int p(\phi)\, p(\alpha_1, \alpha_2, \cdots, \alpha_M | s_1, \phi)\, d\phi \qquad (10)$$

we have for the term $p(\alpha_1, \alpha_2, \cdots, \alpha_M | \underline{s}_1, \phi)$

$$p(\alpha_1, \alpha_2, \cdots, \alpha_M | \underline{s}_1, \phi)$$

$$= p(q_1 = \alpha_1 - 2E\, e^{i\phi}, q_2 = \alpha_2, \cdots, q_M = \alpha_M)$$

$$= \frac{1}{\pi^M (4N_0 E)^M} \exp\left\{ -\frac{E}{N_0} \right\} \exp\left\{ \frac{1}{N_0} Re\, [\alpha_1^* \, e^{i\phi}] \right\}$$

$$\times \exp\left\{ -\frac{1}{4N_0 E} \sum_{k=1}^{M} |\alpha_k|^2 \right\} \qquad (11)$$

This gives for the conditional distribution of the α_k's

$$p(\alpha_1, \alpha_2, \cdots, \alpha_M | \underline{s}_1)$$

$$= \frac{1}{\pi^M (4N_0 E)^M} \exp\left\{ -\frac{E}{N_0} \right\} \exp\left\{ -\frac{1}{4N_0 E} \sum_{k=1}^{M} |\alpha_k|^2 \right\}$$

$$\times \int_{-\pi}^{\pi} \frac{1}{2\pi I_0(\alpha)} \exp\left\{ Re\left[\left(\frac{\alpha_1}{N_0} + \alpha \right)^* e^{i\phi} \right] \right\} d\phi$$

$$= \frac{1}{\pi^M (4N_0 E)^M} \exp\left\{ -\frac{E}{N_0} \right\}$$

$$\times \exp\left\{ -\frac{1}{4N_0 E} \sum_{k=1}^{M} |\alpha_k|^2 \right\} \frac{I_0\left(\left| \frac{\alpha_1}{N_0} + \alpha \right| \right)}{I_0(\alpha)} \qquad (12)$$

where we have made use of the formula

$$I_0(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\{x \cos\theta\}\, d\theta \qquad (13)$$

Introducing the transformation of variables

$$\beta_k = \alpha_k + N_0\alpha \qquad k = 1, 2, \cdots, M \qquad (14)$$

we obtain for the conditional probability density of the β_k's

$$p(\beta_1, \beta_2, \cdots, \beta_M | s_1) = \frac{1}{\pi^M (4N_0 E)^M} \exp\left\{-\frac{E}{N_0}\right\}$$

$$\times \exp\left\{-\frac{1}{4N_0 E} \sum_{k=1}^{M} |\beta_k|^2\right\} \exp\left\{-\frac{M N_0^2 \alpha^2}{4N_0 E}\right\}$$

$$\times \exp\left\{\frac{2N_0\alpha}{4N_0 E} \sum_{k=1}^{M} Re\,\beta_k\right\} \frac{I_0\left(\frac{|\beta_1|}{N_0}\right)}{I_0(\alpha)} \qquad (15)$$

Let us now return to the decision rule given by Eq. (4). If we transform the β_k's to polar coordinates, i.e.,

$$\beta_k = r_k e^{i\theta_k} \qquad k = 1, 2, \cdots, M \qquad (16)$$

we obtain for the probability of correctly deciding that $s_1(t)$ has been transmitted

$$P(r_1 > r_k, \text{all } k \neq 1 | s_1(t) \text{ sent})$$

$$= \int_{-\pi}^{\pi} d\theta_1 \int_0^{\infty} r_1\, dr_1 \int_{-\pi}^{\pi} d\theta_2 \int_0^{r_1} r_2\, dr_2 \cdots \int_{-\pi}^{\pi} d\theta_M \int_0^{r_1} r_M\, dr_M$$

$$\times \frac{1}{\pi^M} \exp\left\{-\frac{E}{N_0}\right\} \exp\left\{-\frac{\alpha^2 N_0^2 M}{4N_0 E}\right\}$$

$$\times \exp\left\{-\frac{1}{4N_0 E} \sum_{k=1}^{M} r_k^2\right\}$$

$$\times \exp\left\{\frac{2N_0\alpha}{4N_0 E} \sum_{k=1}^{M} r_k \cos\theta_k\right\} \frac{I_0\left(\frac{r_1}{N_0}\right)}{I_0(\alpha)} \qquad (17)$$

Eq. (17) integrates to

$$P(r_1 > r_k, \text{all } k \neq 1 | s_1(t) \text{ sent})$$

$$= \int_0^{\infty} \frac{1}{2N_0 E} r_1 \exp\left\{-\frac{E}{N_0}\right\} \exp\left\{-\frac{\alpha^2 N_0^2}{4N_0 E}\right\}$$

$$\times \exp\left\{-\frac{1}{4N_0 E} r_1^2\right\} \times I_0\left(\frac{\alpha}{2E} r_1\right) \frac{I_0\left(\frac{r_1}{N_0}\right)}{I_0(\alpha)}$$

$$\times \left[1 - Q\left(\frac{\alpha N_0}{(2N_0 E)^{1/2}}, \frac{r_1}{(2N_0 E)^{1/2}}\right)\right]^{M-1} dr_1 \qquad (18)$$

where

$$Q(x, y) = \int_y^{\infty} t \exp\left[-\tfrac{1}{2}(t^2 + x^2)\right] I_0(xt)\, dt \qquad (19)$$

Since we have assumed that the $s_k(t)$ are equally probable, we obtain for the probability of correct detection

$$P_c = \sum_{k=1}^{M} \frac{1}{M} P(\text{decide } s_k | s_k \text{ sent})$$

$$= \exp\{-\tfrac{1}{2}\rho\} \cdot \exp\left\{-\tfrac{1}{2}\frac{\alpha^2}{\rho}\right\}$$

$$\times \int_0^{\infty} x \exp\{-\tfrac{1}{2}x^2\} I_0\left(\frac{\alpha x}{\rho^{1/2}}\right) \frac{I_0[(\rho)^{1/2} x]}{I_0(\alpha)}$$

$$\times \left[1 - Q\left(\frac{\alpha}{\rho^{1/2}}, x\right)\right]^{M-1} dx \qquad (20)$$

where $\rho = 2E/N_0$.

Some checks on the preceding result will now be given. In particular, we will consider the limiting cases $\alpha = 0$ and $\alpha \to \infty$. When $\alpha = 0$, Eq. (20) becomes

$$P_c = \exp\{-\tfrac{1}{2}\rho\} \int_0^{\infty} x \exp\{-\tfrac{1}{2}x^2\} I_0[(\rho)^{1/2} x]$$

$$\times [1 - \exp(-\tfrac{1}{2}x^2)]^{M-1} dx$$

$$= \exp\{-\tfrac{1}{2}\rho\} \sum_{n=0}^{M} (-1)^{n-1} \binom{M}{n} \exp\left\{\frac{\rho}{2n}\right\} \qquad (21)$$

which checks with the result for the totally noncoherent reception of m orthogonal signals (Ref. 7, p. 31, Eq. 3.29). To study the case $\alpha \to \infty$, we make use of the asymptotic form of the function $I_0(x)$

$$I_0(x) = \frac{\exp x}{(2\pi x)^{1/2}}, \qquad x \gg 1 \qquad (22)$$

This gives for the Q function

$$1 - Q\left(\frac{\alpha}{\rho^{1/2}}, x\right) \simeq \int_0^x t \exp\left\{-\frac{1}{2}\left(t^2 + \frac{\alpha^2}{\rho}\right)\right\} \frac{\exp\{t\alpha/\rho^{1/2}\}}{(2\pi)^{1/2}(t\alpha/\rho^{1/2})^{1/2}} dt$$

$$\simeq \int_0^x \frac{t}{(2\pi)^{1/2}} \frac{\exp\left\{-\frac{1}{2}\left(t - \frac{\alpha}{\rho^{1/2}}\right)^2\right\}}{[t\alpha/(\rho)^{1/2}]^{1/2}} dt$$

$$\simeq \int_0^x \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}\left(t - \frac{\alpha}{\rho^{1/2}}\right)^2\right\} dt \tag{23}$$

since most of the contribution occurs in the vicinity of the point $t = \alpha/(\rho)^{1/2}$.

For Eq. (20) we have

$$P_c = \exp\{-\tfrac{1}{2}\rho\} \int_0^\infty x \exp\{-\tfrac{1}{2}x^2\} \frac{\exp\{\alpha x/(\rho)^{1/2}\}}{(2\pi)^{1/2}[\alpha x/(\rho)^{1/2}]^{1/2}}$$

$$\times \frac{I_0[(\rho)^{1/2}x]}{\exp\alpha/(2\pi\alpha)^{1/2}} \exp\left\{-\frac{1}{2}\frac{\alpha^2}{\rho}\right\}$$

$$\times \left[\int_0^x \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}\left(t - \frac{\alpha}{(\rho)^{1/2}}\right)^2\right\} dt\right]^{M-1} dx$$

$$\cong \int_0^\infty x \exp\left\{-\tfrac{1}{2}\left(x - \frac{\alpha}{(\rho)^{1/2}}\right)^2\right\}$$

$$\times \frac{\exp\left\{-\frac{1}{2}\rho\right\}}{(2\pi)^{1/2}[\alpha x/(\rho)^{1/2}]^{1/2}} \frac{I_0[(\rho)^{1/2}x]}{e^\alpha/(2\pi\alpha)^{1/2}}$$

$$\times \left[\int_0^x \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}\left(t - \frac{\alpha}{(\rho)^{1/2}}\right)^2\right\} dt\right]^{M-1} dx \tag{24}$$

Similarly, we recognize here that the major contribution to the above integral occurs in the vicinity of $x = \alpha/(\rho)^{1/2}$. By changing the variable of integration in Eq. (24) to

$$y = x - \frac{\alpha}{(\rho)^{1/2}} \tag{25}$$

and writing the $I_0[(\rho)^{1/2}x]$ term in its asymptotic form, we obtain

$$P_c \cong \int_{-\frac{\alpha}{(\rho)^{1/2}}}^\infty \frac{1}{(2\pi)^{1/2}} \exp\{-\tfrac{1}{2}y^2\} \exp\{-\tfrac{1}{2}\rho\} \exp\{(\rho)^{1/2}y\}$$

$$\times \left[\int_{-\frac{\alpha}{(\rho)^{1/2}}}^y \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}z^2\right\} dz\right]^{M-1} dy \tag{26}$$

In the limit as $\alpha \to \infty$ this expression becomes

$$P_c = \int_{-\infty}^\infty \frac{1}{(2\pi)^{1/2}} \exp\{-\tfrac{1}{2}u^2\}$$

$$\times \left[\int_\infty^{u+(\rho)^{1/2}} \frac{1}{(2\pi)^{1/2}} \exp\{-\tfrac{1}{2}z^2\} dz\right]^{M-1} du \tag{27}$$

which agrees with the well known result for the purely coherent reception of M orthogonal signals.

References

1. Davenport, Jr., W. B., "Signal-to-Noise Ratios in Band-Pass Limiters," *Journal of Applied Physics*, Vol. 24, No. 6, June 1953.

2. Middleton, David, *An Introduction to Statistic Communication Theory*, McGraw-Hill Book Company, Inc., New York, 1960.

References (Cont'd)

3. Davenport, Jr., W. B., and Root, W. L., *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill Book Company, Inc., New York, 1958.

4. Rice, S. O., "Mathematical Analysis of Random Noise," *Selected Papers on Noise and Stochastic Processes*, Dover Publications, Inc., New York, 1954.

5. *Handbook of Mathematical Functions*, Edited by M. Abramowitz and I. A. Stegun, U.S. Dept. of Commerce, National Bureau of Standards, Applied Mathematics Series-55, June 1964.

6. Viterbi, A., "Optimum Detection and Signal Selection for Partially Coherent Binary Communication," *IEEE Transactions on Information Theory*, Vol. IT-II pp. 239–246, April 1965.

7. Nuttall, A., *Error Probabilities for Non-orthogonal M-ary Signals Under Phase-Coherent and Phase-Incoherent Reception*, Technical Report 61-1-BF, Litton Systems, Inc., Waltham, Mass, June 1961.

XIX. Communications Elements Research

A. Microwave Electronics

W. H. Higa

1. Hydrogen Maser Frequency Standard

Two hydrogen masers were recently purchased by the Telecommunications Division of JPL and placed into operation. One will be operated by the Communications Systems Research Section while the other will be operated by the Communications Elements Research Section. The primary objective shall be to evaluate the hydrogen maser for possible future use in the DSIF as a primary frequency standard.

The two masers were fabricated by the Quantum Electronic Devices division of Varian Associates in Beverly, Mass. They were then shipped to the National Bureau of Standards at Boulder, Colorado for extensive testing with the Bureau's cesium standards. Subsequently, they were shipped via truck to JPL and immediately placed in operation. Observe that a good beginning has been made in the direction of a transportable device.

The vacuum chamber which contains the quartz bulb and resonant cavity (*SPS 37-35*, Vol. IV, p. 259) was sealed at the factory, and Vacion pumps have reliably maintained the vacuum ever since. Of the two machines received at JPL, one was placed in immediate operation,

while the other required minor maintenance work on the auxiliary electronic equipment. Frequency comparison tests subsequently indicated that both machines were properly tuned; that is, the beat frequency indicated both machines to be tuned to approximately a part in 10^{12} of each other. A precise tuning procedure has been developed by Varian Associates to ensure correct tuning of the resonant cavity. This method requires two masers for frequency comparison.

These are the kinds of problems that need to be solved before the hydrogen maser can be used operationally as a primary standard. Absolute tuning of the cavity, short term stability characteristics and spectral purity are important problems that will be studied.

B. Antennas for Space Communications

A. C. Ludwig

1. Multiple Frequency Dual Mode Horn

a. Introduction. A design of a feed horn which provided dual mode horn performance at two discrete frequencies has been previously described (*SPS 37-34*, Vol. IV,

pp. 235-237). This design required a phasing section which is roughly 45 in. long at S-band frequencies. For application as a feed in the JPL advanced antenna system (AAS), this created severe packaging problems in the feed cone. Therefore, the question of bandwidth of a normal dual mode horn was studied quantitatively, and a compromise design was evolved that provided reasonable performance at AAS operating frequencies without a long phasing section.

b. Dual mode horn performance. As described in the above reference, dual mode horn performance is dependent on maintaining a certain phase relationship between the TE_{11} and TM_{11} cylindrical waveguide modes in the horn aperture. For a dual mode horn designed to have correct phasing at 2388 Mc (R&D frequency), the phase of the TM_{11} mode lags the correct value by 49 deg at 2295 Mc (DSIF receive frequency). By shortening the phasing section of the normal dual mode horn, which decreases the phase length of the horn more for the TE_{11} mode than for the TM_{11} mode, the performance at these two frequencies may be more nearly balanced. At 2116 Mc (DSIF transmit frequency), the TM_{11} mode is below cutoff in the phasing section; by shortening the phasing section, more of this mode is allowed to propagate. Since the phase of the TM_{11} mode at this frequency cannot be adjusted without degrading the performance at the other two frequencies, this mode will in general degrade the performance at 2116 Mc. The previously described dual frequency design, with a very long phasing section, suppressed this mode entirely. The relative phasing of the TE_{11} and TM_{11} modes affect the phase patterns of the horn as well as the amplitude patterns. This effect turned out to be fairly serious.

Amplitude patterns of the dual mode compromise design are shown in Fig. 1. The phase patterns, which are not shown, are somewhat distorted at 2388 Mc and particularly at 2116 Mc. The resulting antenna efficiency of a paraboloidal antenna illuminated by these feed patterns, assuming optical reflection from a hyperboloid, is shown in Table 1. Performance levels for the dual-frequency design (see above reference for radiation patterns) are shown for reference. The efficiencies shown in Table 1 include loss due to spillover, nonuniform illumination, cross polarization, and phase errors. Not included are antenna surface tolerance and aperture blockage. A hyperboloid edge-angle of 14 deg is selected.

It is seen from Table 1 that very reasonable performance is obtained from the compromise design at 2388 and 2295 Mc; fairly serious degradation occurs at
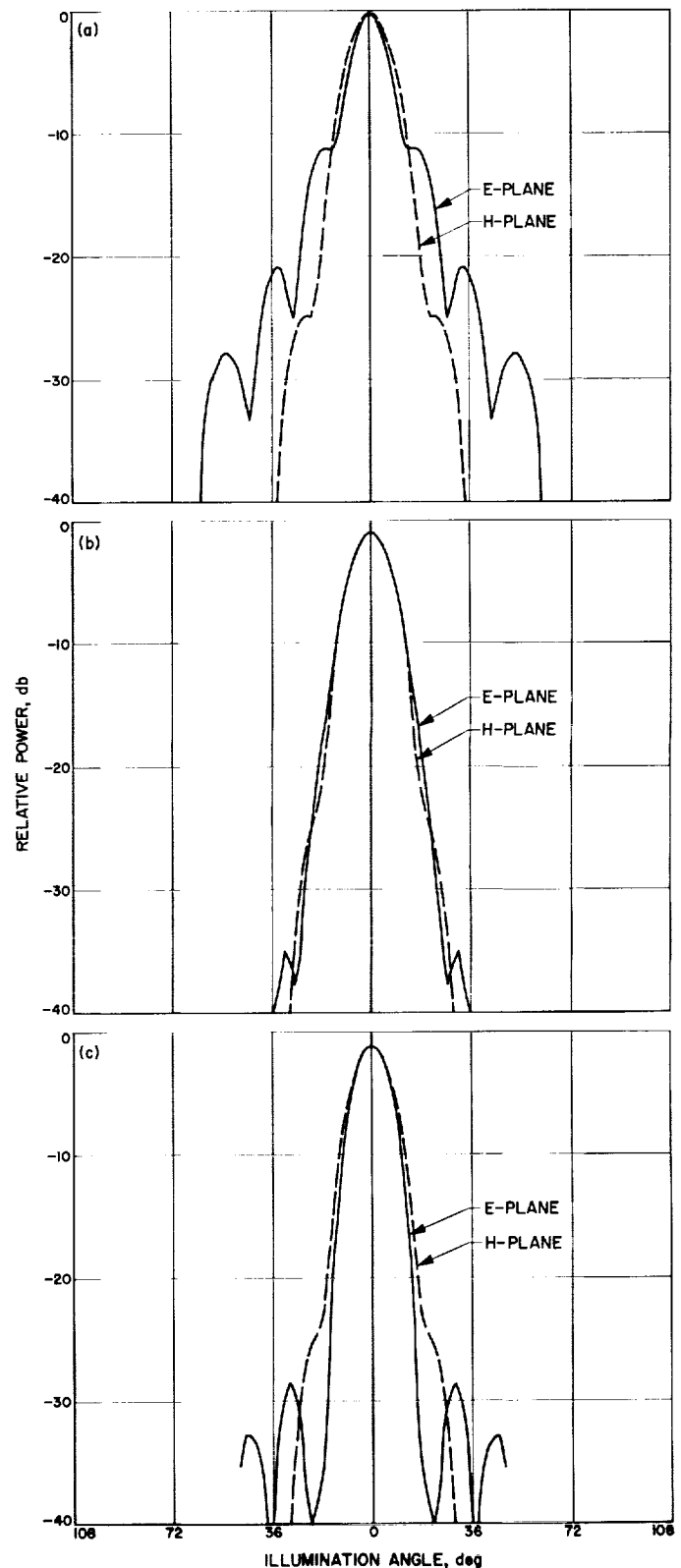
Fig. 1. Dual mode horn, compromise design, radiation patterns: (a) 2116 Mc; (b) 2295 Mc; (c) 2388 Mc

Table 1. Feed horn performance

Frequency, Mc	Dual frequency design efficiency, %	Compromise design efficiency, %	Performance of compromise design relative to dual frequency design, db
2116	71.0	58.0	−0.88
2295	79.6	80.2	+0.03
2388	80.6	74.9	−0.32

2116 Mc. These results are considered adequate for an initial AAS installation; if future implementation allows the inclusion of a longer phasing section, the dual-frequency design will be used.

2. Mode Generation in Cylindrical Waveguide

a. Introduction. The improvement in antenna efficiency obtainable with use of higher-order cylindrical waveguide modes in the antenna feed has been the subject of a previous article (*SPS 37-33*, Vol. IV, pp. 261–266). A major difficulty in realizing the potential gains predicted by theory is the generation and control of higher order waveguide modes. In this article, an antenna pattern of a dual mode horn (Ref. 1) is expanded as a sum of radiation patterns from individual waveguide modes. The power in each mode is then computed by pattern integration, and the relative mode strengths are compared to theoretical predictions of mode generation by the step discontinuity used in the dual mode horn.

b. Analysis of dual mode horn pattern. The physical configuration of a dual mode horn is shown in Fig. 2. The TE_{11} cylindrical waveguide mode is incident from the left; the step discontinuity generates higher order modes. The design is such that of the modes generated, only the TE_{11} and TM_{11} modes will propagate. An experimental radiation pattern of the dual mode horn, with a

specific phasing of the two modes, is shown in Fig. 3. Following the analysis given in a previous article (*SPS 37-19*, Vol. IV, pp. 190–196) where it is shown that the field at certain specific points is due to a single mode only, the pattern shown in Fig. 3 may be resolved into component patterns of individual modes. It is found that three modes are present[1]; the theoretical component patterns are shown in Fig. 4, and the sum of the component patterns is shown in Fig. 5. It is seen that the amplitude and phase patterns or the sum of the theoretical component patterns are an excellent match to the experimental patterns. The patterns shown in Figs. 4 and 5 were integrated using a computer program previously described (*SPS 37-26*, Vol. IV, pp. 200–208). The results are summarized in Table 2.

Table 2. Mode power in theoretical patterns

Mode	100 · Mode power/total power in sum pattern, %
TE_{11}	85.6
TM_{11}	10.7
TE_{12}	2.3
Total	98.6ᵃ

ᵃPowers do not add to 100% due to errors in pattern integration.

c. Theoretical mode generation. The theoretical investigation of mode generation by a step discontinuity has been made by Yeh (Ref. 2) and Nagelberg and Shefer (Ref. 3). In Ref. 3, a mode conversion coefficient C is defined as the ratio of the transverse E-fields of the TM_{11} and TE_{11} modes, measured in the E-plane at the wall of the waveguide. The mode coefficients defined in Ref. 2 are related to this quantity by $C = 1.2744 \cdot (Q_1/P_1)$, where Q_1 and P_1 are mode coefficients of the TM_{11} and TE_{11}

---

[1]The third mode, the TE_{12}, most likely arises in the aperture due to a nonplanar phase front in the aperture.

$2a$ $2b$
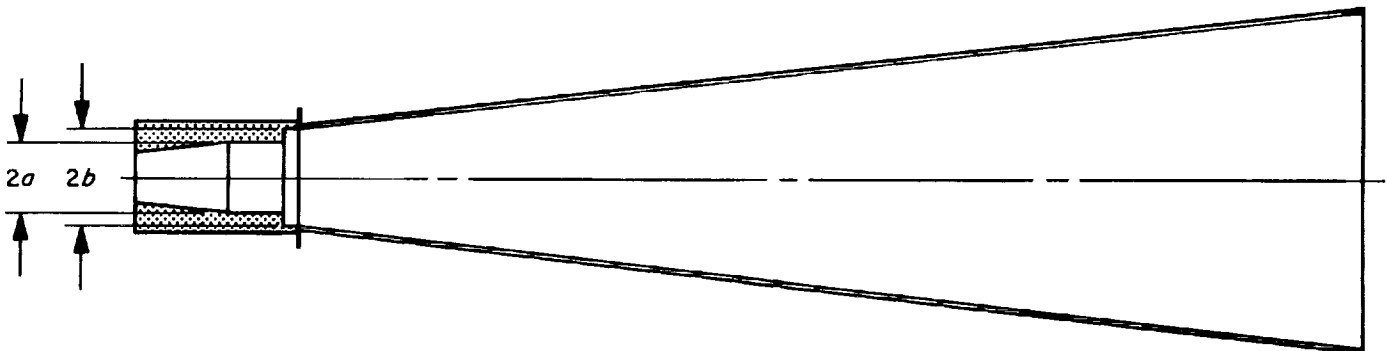
Fig. 2. Dual mode horn
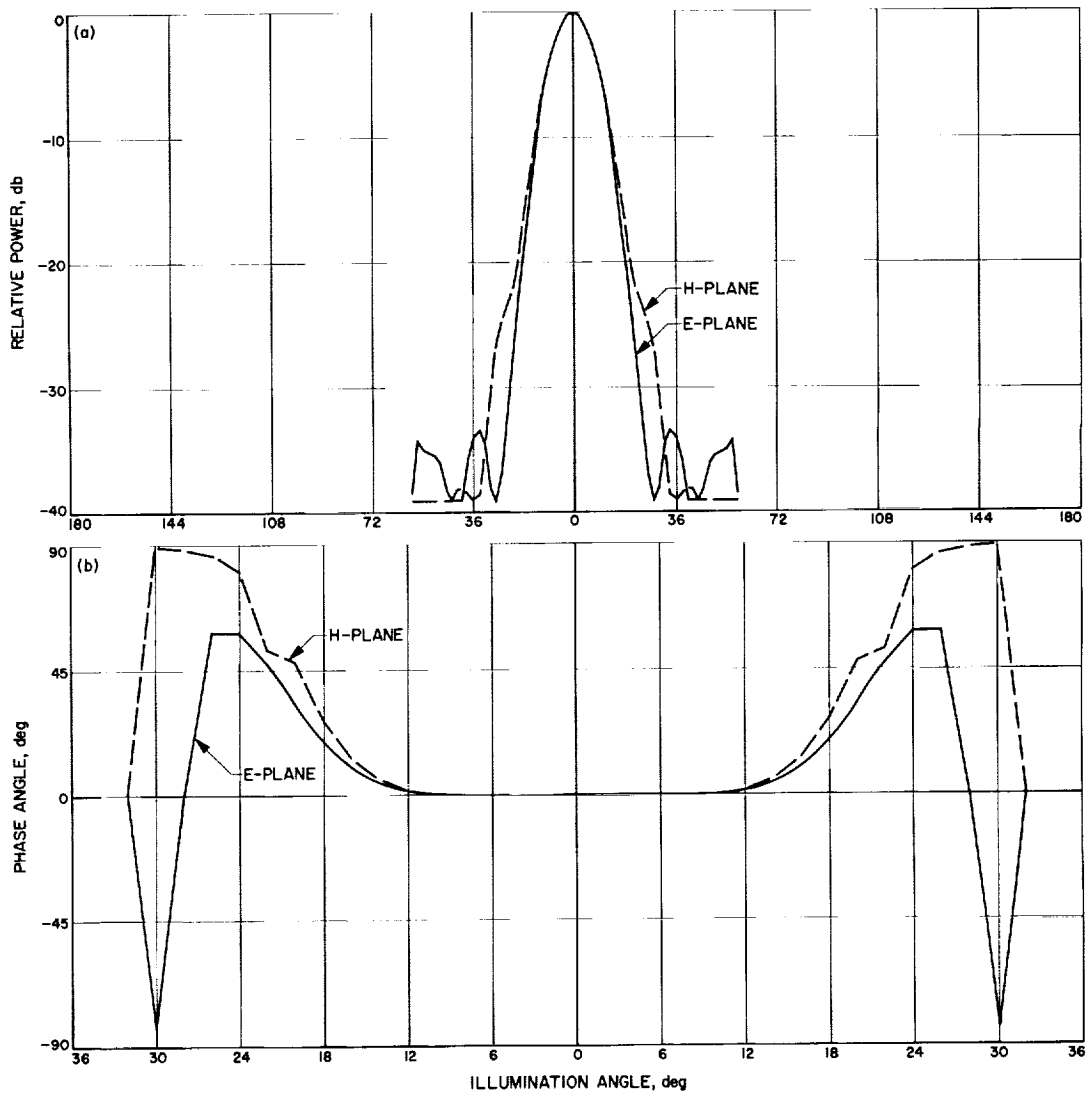
Fig. 3. Dual mode horn radiation pattern: (a) amplitude pattern; (b) phase pattern
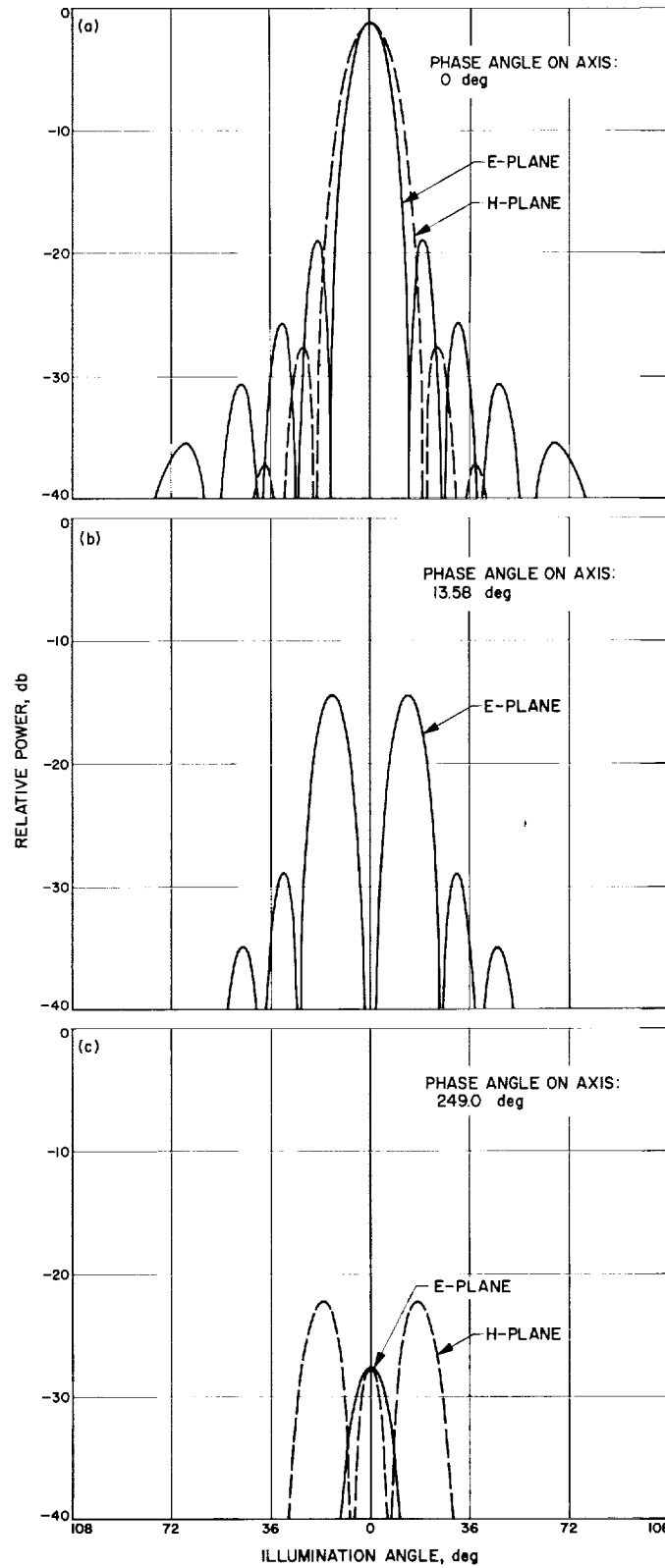
Fig. 4. Theoretical component patterns of dual mode horn: (a) TE$_{11}$ mode; (b) TM$_{11}$ mode; (c) TE$_{12}$ mode
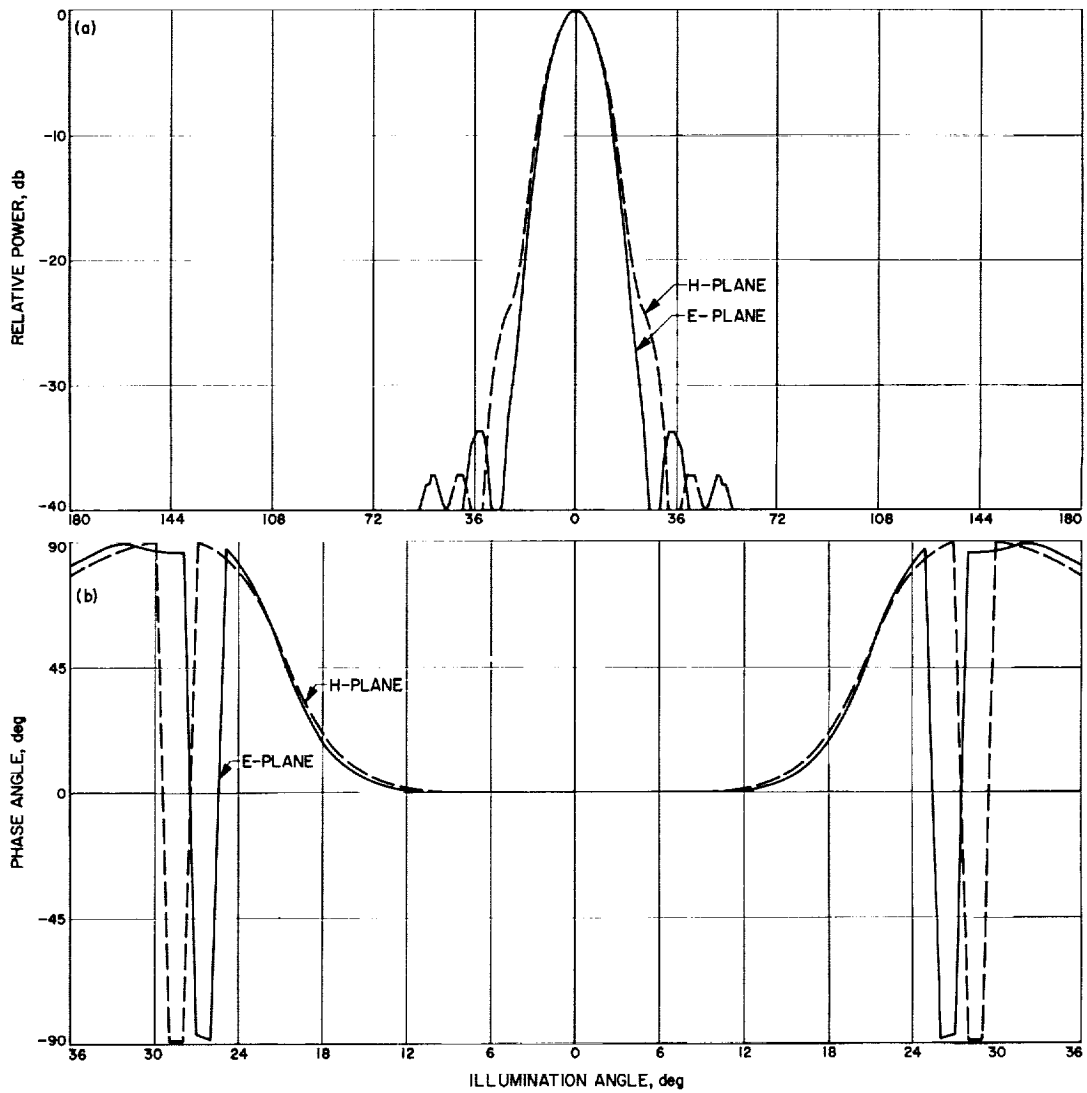
Fig. 5. Sum of theoretical component patterns: (a) amplitude pattern; (b) phase pattern

modes respectively, as defined in that article. The quantity C, as determined in the above references, is plotted in Fig. 6 vs the normalized step height. Although both articles use basically the same method, some approximations made in Ref. 3 lead to somewhat different results.

C as determined in Ref. 3 is independent of frequency; the data taken from Ref. 2 assumed $(2\pi/\lambda) \cdot b = 5.251$ and $(2\pi/\lambda) \cdot b = 4.051$, which gave virtually identical results for C when plotted against normalized step height. A conversion between mode power and the quantity C may be easily determined, and is found to be

$$\frac{\text{Power in TM}_{11} \text{ mode}}{\text{Power in TE}_{11} \text{ mode}} = 2.390\, C^2 \cdot \left[1 - \left(3.832\,\frac{\lambda}{2\pi b}\right)^2\right]^{\frac{1}{2}}$$

$$\times \left[1 - \left(1.841\,\frac{\lambda}{2\pi b}\right)^2\right]^{\frac{1}{2}} \qquad (1)$$

for the dual mode horn pattern in Fig. 3, $2\pi a/\lambda = 3.195$, $2\pi b/\lambda = 4.090$. For this case $(b - a)/a = 0.218$ and the power ratio is $0.73C^2$. Using the data in Table 2, the experimental value for $C = -7.67$ db. This is plotted as a circle in Fig. 6. There is reasonably good agreement with the theory.

Interpolating and scaling experimental data given in Ref. 3, which were obtained by directly probing the fields with a slotted line, rather than by pattern expansion, a value for $C = -7.7$ db was obtained for the same dual mode horn configuration. So the experimental agreement
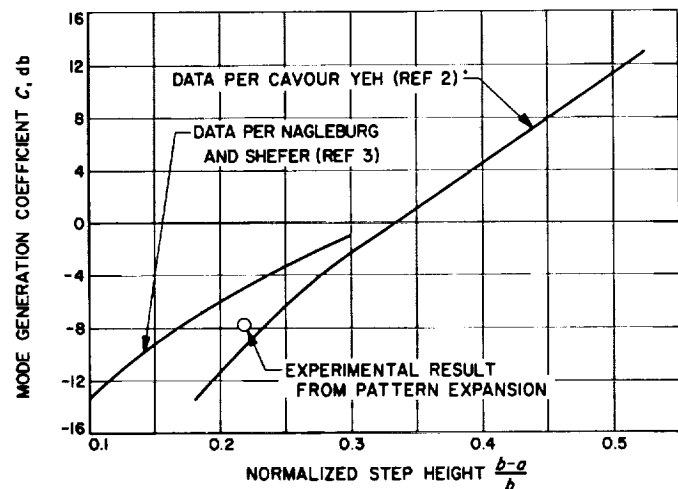
Fig. 6. Mode conversion coefficient vs normalized step height

between the two methods is outstanding. Although the theoretical values are not good enough to adequately predict dual mode horn performance, they are valuable for predicting the dependence on step height, and as a starting point in an experimental development.

C. RF Techniques

C. T. Stelzried, W. V. T. Rusch, B. Seidel, G. S. Levy,
T. V. Otoshi and M. S. Reid

1. 90-Gc mm Wave Work, C. T. Stelzried and W. V. T. Rusch

a. Introduction. The objective of the millimeter wave work is to investigate millimeter wave components and techniques to ascertain the future applicability of this frequency range to space and communications and tracking. This involves the development of instrumentation for accurate determination of insertion loss, VSWR, power and equivalent noise temperatures of passive elements. Millimeter wave circuit elements are being evaluated in a radio telescope system consisting of a 60-in. antenna and a superheterodyne radiometer (*SPS 37-33*, Vol. IV, p. 245). The radio telescope was used to observe the 90-Gc temperature of the Moon during the December 30, 1963 eclipse (*SPS 37-26*, Vol. IV, p. 181) and the most recent eclipse, December 18, 1964. These experiments were joint efforts made by personnel from JPL and the Electrical Engineering Department of the University of Southern California.

b. Observations and Data Reduction. Radiometric observations of the sub-Earth point on the Moon were carried out daily during the period August 3 to September 12 (*SPS 37-35*, Vol. IV, p. 270). Due to atmospheric attenuation, the measured radiometric temperature is a function of elevation angle. An analysis, using a least squares best fit technique, was made (*SPS 37-28*, Vol. IV, p. 148) which involved taking logarithms of the equations relating the measured temperatures to the actual Moon temperature as a function of elevation angles. This was necessary in order to linearize the equations. However, an unknown weighting factor was introduced (Ref. 4, p. 247) due to the logarithms. A new analysis has been made using the least squares best fit technique in which linearization is achieved by using the first order correction term obtained from a Taylor expansion (Ref. 4, p. 256). All measurements are equally weighted with
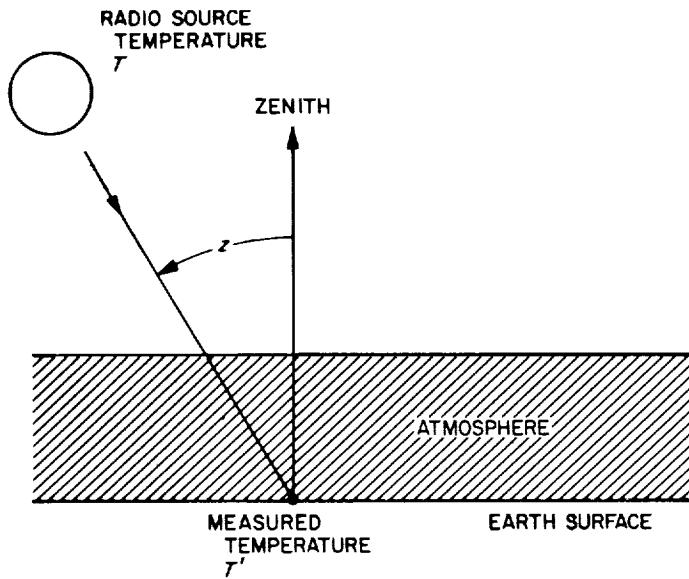
Fig. 7. Geometry used with flat Earth for computation of atmospheric attenuation and actual radio source temperature

this technique. Assuming a flat Earth and an atmosphere height l_0, the path length at zenith angle z is (Fig. 7)

$$l = l_0 \sec z \tag{1}$$

With an atmospheric attenuation constant α, the loss through the atmosphere at zenith is

$$L_0 = e^{\alpha l_0} \tag{2}$$

and that at an angle z

$$L = e^{\alpha l} \tag{3}$$

so that

$$L = (L_0)^{\sec z} \tag{4}$$

The equivalent noise temperature T of a radio source is related to the measured temperature T' by

$$T' = T (L_0)^{-\sec z} \tag{5}$$

The problem is to determine the best value for T and L_0 from a series of measurements of T' and z at various zenith angles.

Rewriting Eq. (5) in terms of constants a and b and data points x and y,

$$y = ab^{-\sec x} \tag{6}$$

Linearizing by taking the logarithms,

$$\ln y = \ln a - \sec x \ln b \tag{7}$$

The variance is given by (all summations in the following are from 1 to N, the number of observations)

$$(\sigma')^2 = \frac{1}{N-2} \sum (\ln y - \ln a + \sec x \ln b)^2 \tag{8}$$

Differentiating with respect to $\ln a$ and $\ln b$ and equating to zero,

$$\left. \begin{array}{l} \sum (\ln y - \ln a + \sec x \ln b) = 0 \\ \sum (\ln y - \ln a + \sec x \ln b) \sec x = 0 \end{array} \right\} \tag{9}$$

Solving for the constants and identifying the first estimate with primes,

$$\left. \begin{array}{l} \ln a' = \dfrac{\begin{vmatrix} \sum \ln y & -\sum \sec x \\ \sum \sec x \ln y & -\sum \sec^2 x \end{vmatrix}}{\Delta'} \\[20pt] \ln b' = \dfrac{\begin{vmatrix} N & \sum \ln y \\ \sum \sec x & \sum \sec x \ln y \end{vmatrix}}{\Delta'} \end{array} \right\} \tag{10}$$

where

$$\Delta' = \begin{vmatrix} N & \sum \sec x \\ \sum \sec x & \sum \sec^2 x \end{vmatrix}$$

Expanding Eq. (6) in a Taylor series and retaining only the 1st order terms,

$$y = y' + \left. \frac{\partial y}{\partial a} \right|_{a=a'} (a - a') + \left. \frac{\partial y}{\partial b} \right|_{b=b'} (b - b') \tag{11}$$

where

$$y' = a' (b')^{-\sec x}$$

the variance is now

$$\sigma^2 = \frac{1}{N-2} \Sigma \left[y - y' - y' \left(\frac{a - a'}{a'} \right) + y' \sec x \left(\frac{b - b'}{b'} \right) \right]^2 \tag{12}$$

Differentiating with respect to a and b, and equating to zero,

$$\left. \begin{array}{l} \Sigma \left[y - y' - y' \left(\dfrac{a - a'}{a'} \right) + y' \sec x \left(\dfrac{b - b'}{b'} \right) \right] y' = 0 \\[3mm] \Sigma \left[y - y' - y' \left(\dfrac{a - a'}{a'} \right) + y' \sec x \left(\dfrac{b - b'}{b'} \right) \right] y' \sec = 0 \end{array} \right\} \tag{13}$$

Solving for the constants

$$\left. \begin{array}{l} a = a' + \dfrac{\begin{vmatrix} \Sigma (y - y') y' & -\dfrac{1}{b'} \Sigma (y')^2 \sec x \\[3mm] \Sigma (y - y') y' \sec x & -\dfrac{1}{b'} \Sigma (y')^2 \sec^2 x \end{vmatrix}}{\Delta} \\[12mm] b = b' + \dfrac{\begin{vmatrix} \dfrac{1}{a'} \Sigma (y')^2 & \Sigma (y - y') y' \\[3mm] \dfrac{1}{a'} \Sigma (y')^2 \sec x & \Sigma (y - y') y' \sec x \end{vmatrix}}{\Delta} \end{array} \right\} \tag{14}$$

where

$$\Delta = \begin{vmatrix} \dfrac{1}{a'} \Sigma (y')^2 & -\dfrac{1}{b'} \Sigma (y')^2 \sec x \\[3mm] \dfrac{1}{a'} \Sigma (y')^2 \sec x & -\dfrac{1}{b'} \Sigma (y')^2 \sec^2 x \end{vmatrix}$$

The computational procedure is to first calculate a' and b' from Eq. (10) and then iterate with Eq. (14) to the required accuracy. This technique is especially convenient with a digital computer. The probable error of the data points is

$$PE_y = 0.6745 \sqrt{\frac{\Sigma (y - ab^{-\sec x})^2}{N - 2}} \tag{15}$$

The probable errors of the constants are found by expanding Eq. (14) in terms of the coefficients of the data points

and summing the probable errors (Ref. 4, p. 229),

$$PE_a = \frac{PE_y}{|\Delta b'|} \sqrt{\Sigma (y')^2 [\sec x \Sigma (y')^2 \sec x - \Sigma (y')^2 \sec^2 x]^2}$$

$$PE_b = \frac{PE_y}{|\Delta a'|} \sqrt{\Sigma (y')^2 [\sec x \Sigma (y')^2 - \Sigma (y')^2 \sec x]^2} \tag{16}$$

Eqs. (10), (14)–(16) have been programmed for the 7094 computer and are being used in the evaluation of the Moon radiometric observations.

c. *Antenna calibration.* The equivalent antenna temperature of a radio source, referred to the output of the waveguide switch in the radiometer (Figs. 8 and 9, is given by

$$T_A = \frac{1/L}{4\pi} \int_{radio\,source} T_s (\Omega) G (\Omega) d\Omega \tag{17}$$

where L is the loss between the antenna aperture and the output of the switch, $T_s (\Omega)$ is the equivalent noise temperature of the source, and $G (\Omega)$ is the directivity function of the antenna. If the directivity is expressed

$$G (\Omega) = G_0 f (\Omega) \tag{18}$$

where G_0 is the peak directivity and $f (\Omega)$ is the antenna pattern, normalized to unity in the direction of peak directivity, then

$$T_A = \frac{G_0/L}{4\pi} \int_{radio\,source} T_s (\Omega) f (\Omega) d \tag{19}$$

An equivalent source temperature T'_s is defined by the relation

$$T'_s \int_{radio\,source} f (\Omega) d\Omega = \int_{radio\,source} T_s (\Omega) f (\Omega) d\Omega \tag{20}$$

Insertion of Eq. (20) into Eq. (3) yields

$$T_A = \left\{ \frac{(G_0/L)}{4\pi} \int_{radio\,source} f (\Omega) d\Omega \right\} T'_s \tag{21}$$
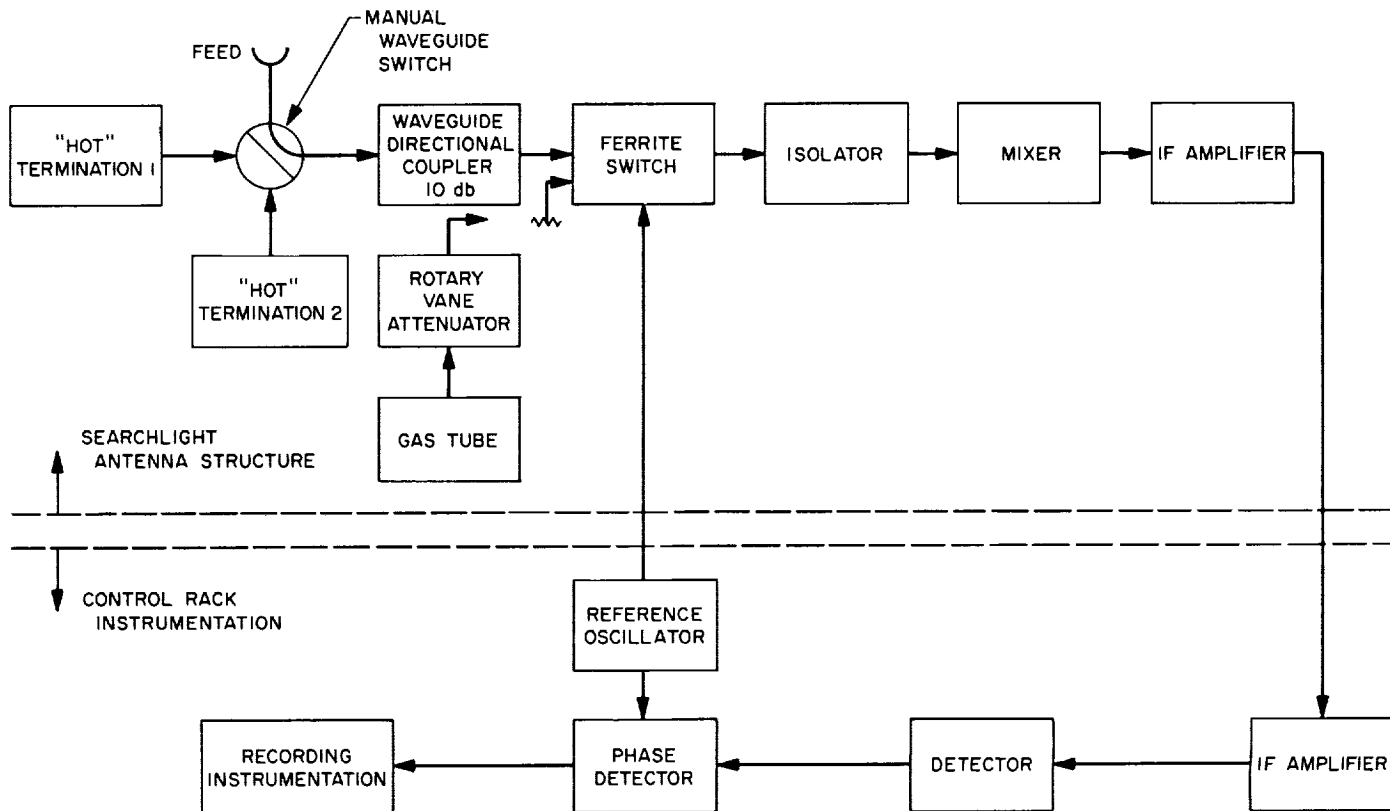
Fig. 8. Millimeter radiometer
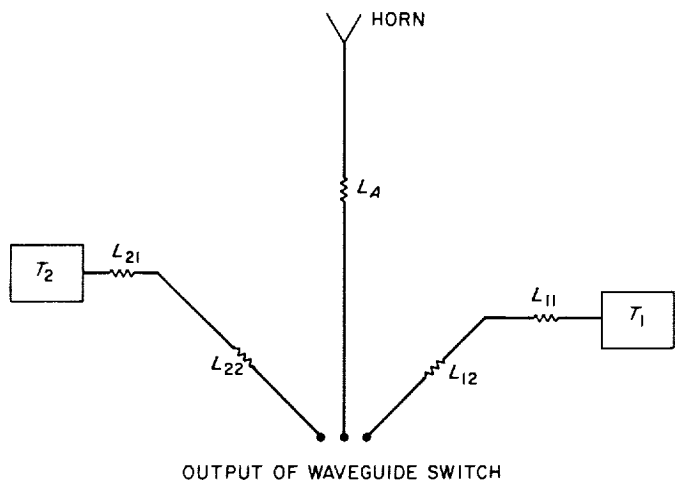
Fig. 9. 90-Gc radiometer waveguide losses

with a calibrated attenuator to obtain (G_0/L). For the 60-in. searchlight antenna used in the eclipse and lunation experiments, the beam correction factor was 0.4695 for a source with a 15-min semidiameter.

2. Simultaneous Lobing Radiometric Tracking System, *B. Seidel and G. S. Levy*

The S-band systems for the DSIF 85-ft and future 210-ft antennas will use simultaneous lobing angle tracking feed systems. A radiometer, which could be used with the tracking feed, would be a useful device for angle pointing and gain calibrations of the antenna systems using radio star sources. This procedure will become quite important when the 210-ft antenna is completed, as it will not be possible to utilize a collimation tower for calibration.

As stated in *SPS 37-35*, Vol. IV, p. 271, on Sept. 17, 1965, tracking data were taken on Omega Nebula when Omega Nebula was at a declination very close to that of the Mars *Mariner*, and the sidereal hour angle (HA) was about 3 hrs behind the *Mariner*. This source, therefore,

The bracketed quantity is known as the beam-correction factor. This factor has been determined experimentally on a long-distance antenna range by measuring $f(\Omega)$ directly and integrating it over the angle subtended by the source, and using standard gain horns in conjunction

provided not only repeatability data and tracking jitter data, but also provided a comparison against the *Mariner* tracking data taken on the same day. These data have now been reduced and are presented in Fig. 10. Though the tracking of the *Mariner IV* appears relatively rough, the mean of the tracking data can be compared to the tracking data taken on Omega Nebula. This comparison yields a difference, or apparent boresight shift, of about 0.065 deg in HA. At present, the reason for the apparent boresight shift is unknown but is most likely due to the

uncertainty in the position of the radio center of Omega Nebula; this is being investigated.

On October 5 and 6, the system was operated in an experiment designed to eliminate the necessity of a collimation tower when phasing the receiver for automatic tracking of a spacecraft.

The experiment consisted of using the simultaneous lobing radiometer, the radio source Crab Nebula, a signal
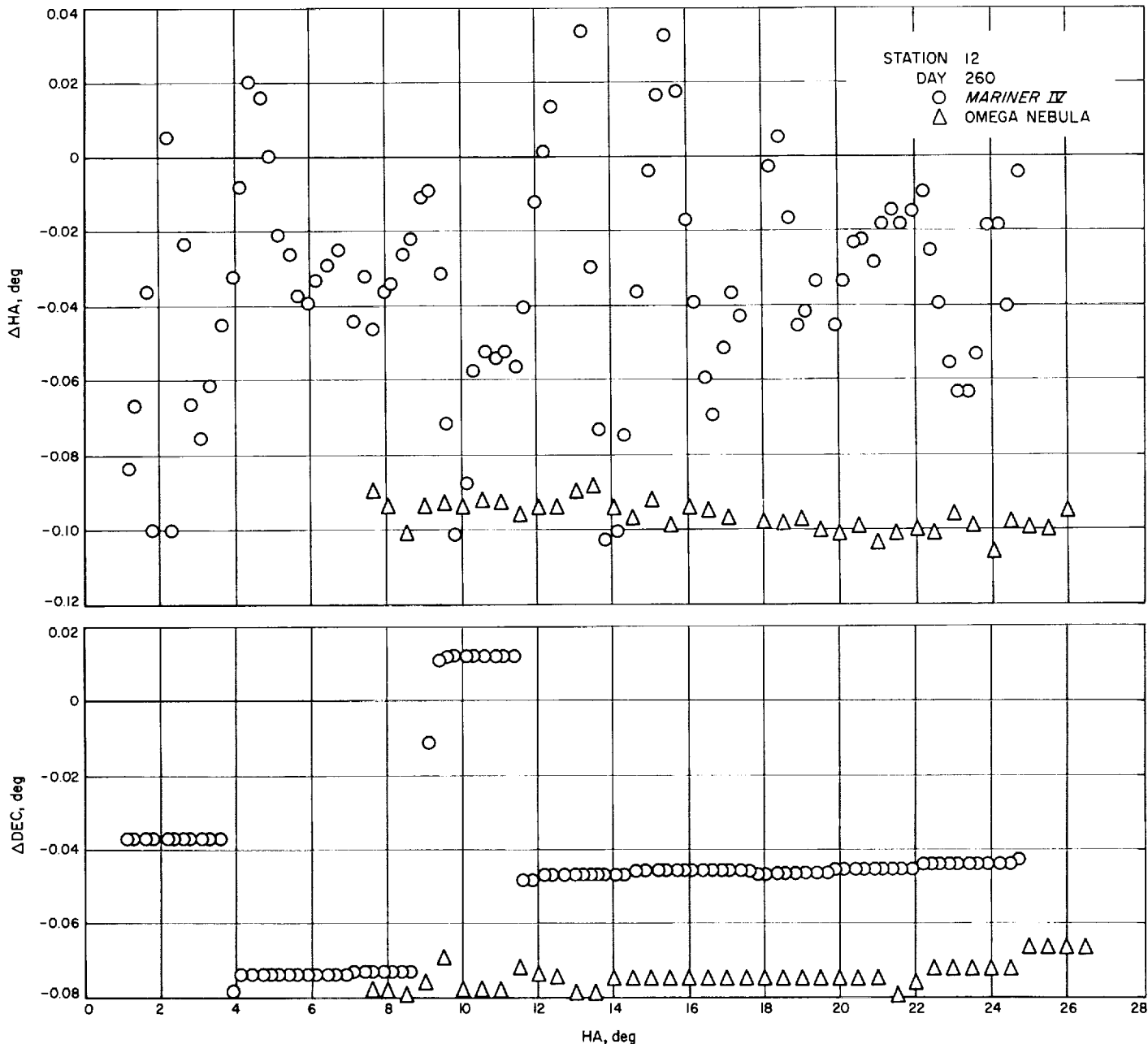
Fig. 10. HA and Dec pointing error vs HA
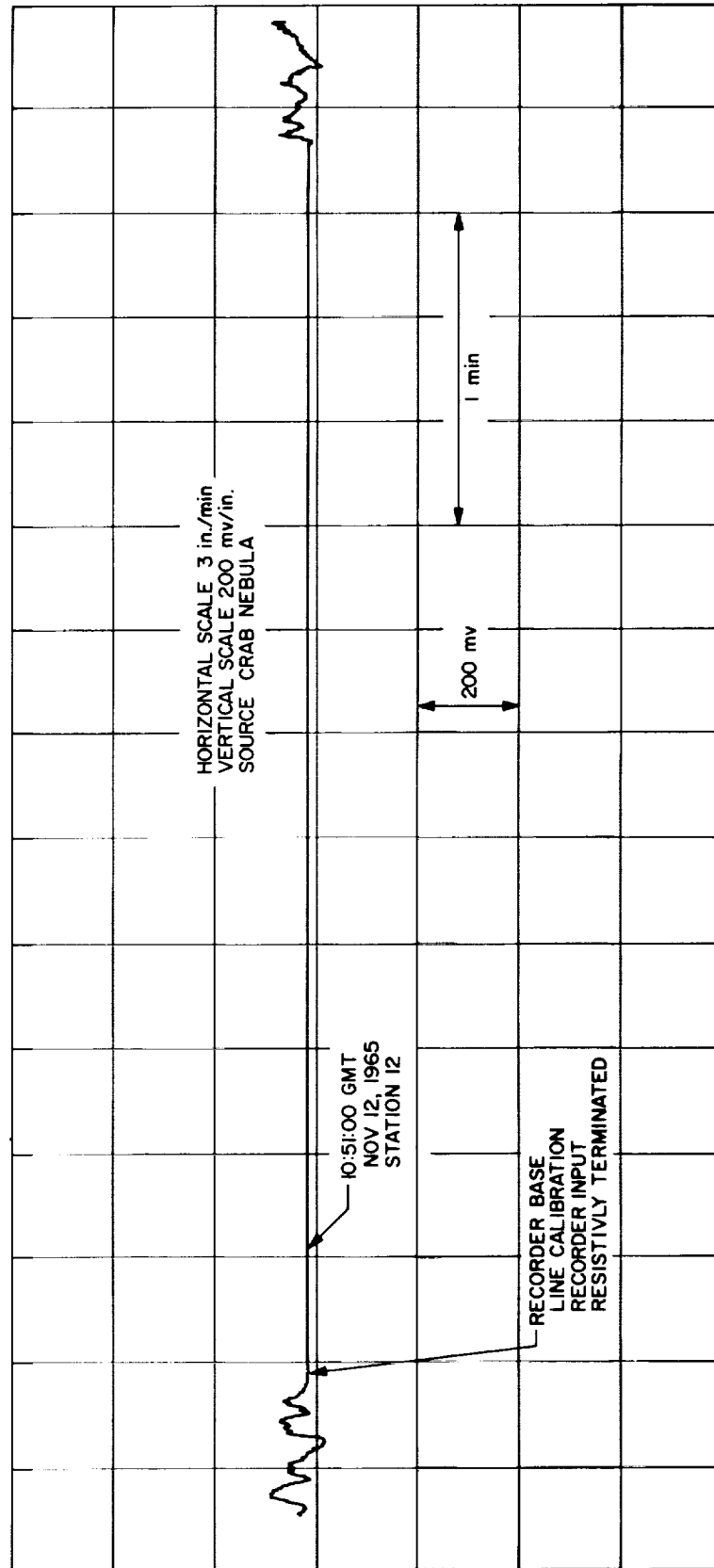
Fig. 11. Calibration of recorder base line

HORIZONTAL SCALE 3 in./min
VERTICAL SCALE 200 mv/in.
SOURCE CRAB NEBULA

1 min

200 mv

10:51:00 GMT
NOV 12, 1965
STATION 12

RECORDER BASE
LINE CALIBRATION
RECORDER INPUT
RESISTIVLY TERMINATED

generator to phase the receiver, and of checking the boresight and phasing against the collimation tower. The experiment was performed on the HA axis of the Echo Station's 85-ft diameter antenna. The results of this experiment are shown in Table 3.

Table 3. Phasing of HA channel of the receiver using a radio source

| October, 1965 | Radio source | |
	Phase shifter setting, deg	Collimation tower boresight HA coordinate
5	341	272.062
6	17.5	272.050

The phasing and HA coordinate obtained in the conventional manner are shown in Table 4.

Table 4. Phasing of HA channel of the receiver using the collimation tower

| October, 1965 | Collimation tower | |
	Phase shifter setting, deg	Collimation tower boresight HA coordinate
5	310	272.058
6	358	272.054

The average difference between the phase shifter settings obtained from the collimation tower and the radio source is approximately 25 deg. A test run against the collimation tower indicates that the null depth of the angle error feed (HA axis) is great enough to cause no boresight shift, within the resolution of the encoders, for this amount of phase misalignment. It can, therefore, be stated that the primary effect of this amount of phase misalignment will be an increase in tracking jitter. This increase in tracking jitter, on a theoretical basis, is approximately 10% for a phase misalignment of 25 deg.

Another question that arose involves the effect of the electro-mechanical dynamics of the antenna-servo system on the tracking of the source. That is, do the dynamics of the system cause the antenna to lag the source? This question was answered by connecting a strip chart recorder to record the error voltages being sent from the simultaneous lobing radiometer to the servo system. First the input to the recorder was terminated so that a base line could be established (Fig. 11). The antenna was then moved ahead of the source in HA and the brakes locked. The source was then allowed to drift through the beam
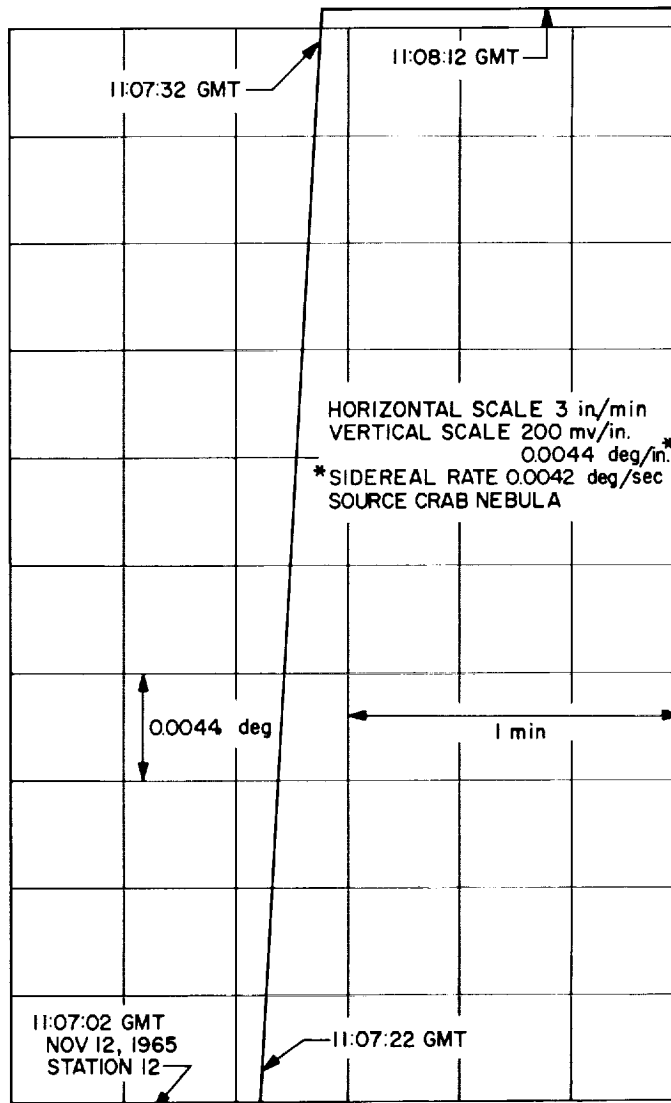
Fig. 12. Calibration of servo error voltage in terms of degrees

of the antenna (Fig. 12) while the recorder was used to record the HA error voltage, thereby calibrating the chart in terms of angular degrees. The antenna was then allowed to automatically track the source (Crab Nebula) in HA while error voltage recordings were taken (Fig. 13). Fig. 13 clearly shows that the mean electro-mechanical dynamical errors of the antenna-servo system are less than 0.002 deg.

It is felt that the above results clearly demonstrate the feasibility of the system, but it is also felt that further testing of the system is required, especially in the area of phase alignment, before the system can be pronounced operational.
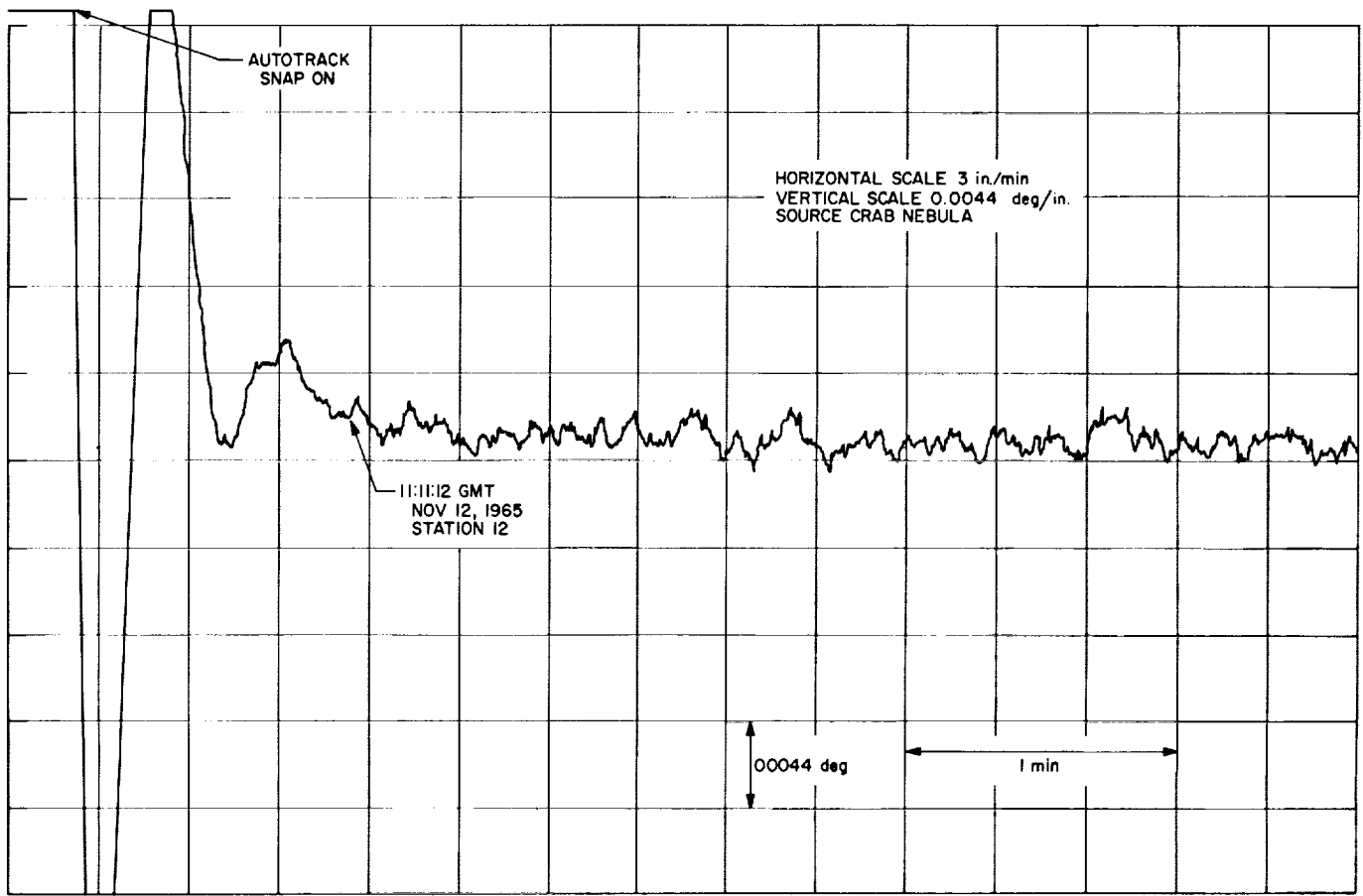
Fig. 13. Angle error recording of automatic tracking of Crab Nebula (HA axis)

3. Antenna Temperature Analysis

T. Otoshi and C. T. Stelzried

a. Summary. The equations to calculate the gain, efficiency, and antenna temperature of a circularly symmetrical antenna are presented in a form suitable for a digital computer program. The parabolic antennas and low noise circular feed horns, such as those presently in operation in the DSIF, have circular symmetry.

The computer program uses data obtained from experimental E- and H-plane antenna patterns and from sky and ground brightness temperature curves. The probable errors for determining gain and antenna temperature are also calculated.

b. Antenna gain. A. Ludwig (Ref. 5) pointed out that the total power pattern of an antenna with complete circular symmetry can be described in terms of two

selected patterns. For a linearly polarized antenna, the two required patterns are the E- and H-plane patterns. For an RCP or LCP antenna, the two required patterns are the receive patterns taken with the illuminator, illuminating first in RCP and then in LCP.

The following discussion presents a derivation of the equations for gain, efficiency, and antenna temperature of a linearly polarized and circularly symmetric antenna in terms of E- and H-plane patterns. The case for circularly polarized antenna is not analyzed in this article.

Following the analysis given by Ludwig in *SPS 37-26*, Vol. IV, pp. 200–207, let E_0 (R, ϕ, θ) represent the far field electric field pattern of any antenna located at the origin of the spherical coordinate system (R, ϕ, θ), as shown in Fig. 14. R is defined as the distance of E_0 from the origin, ϕ is the azimuthal angle in radians, θ is the polar angle in radians, and a_R, a_ϕ, a_θ are the associated unit vectors.
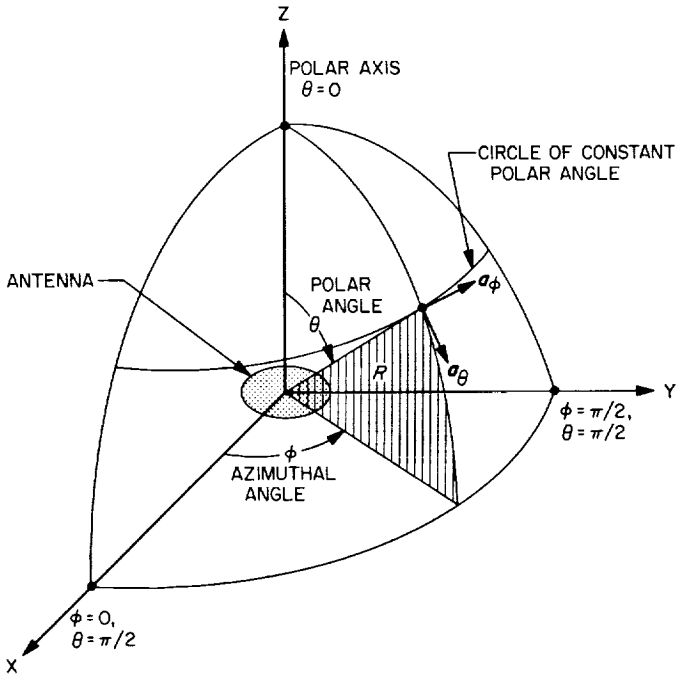
Fig. 14. The (R, ϕ, θ) spherical coordinate system

For this analysis, assume that the antenna has complete circular symmetry and is excited by the dominant or any $m = 1$ cylindrical mode.[2] Then,

$$E_0(R, \phi, \theta) = \frac{e^{-j\kappa R}}{R} [A_1(\theta) \sin \phi \, a_\theta + B_1(\theta) \cos \phi \, a_\phi]$$

(1)

where

$$\kappa = \frac{2\pi}{\lambda}$$

$$A_1(\theta) = |A_1(\theta)| e^{j\Phi_{A_1}(\theta)}$$

$$B_1(\theta) = |B_1(\theta)| e^{j\Phi_{B_1}(\theta)}$$

The subscript 1 is used to identify the $m = 1$ cylindrical mode. The terms $|A_1(\theta)|$, $|B_1(\theta)|$, $\Phi_{A_1}(\theta)$, and $\Phi_{B_1}(\theta)$ will be defined in the following: Note that if $\phi = 0$, Eq. (1) becomes

$$E_0(R, 0, \theta) = \frac{|B_1(\theta)|}{R} e^{-j[\kappa R - \Phi_{B_1}(\theta)]} a_\phi$$

(2)

and

$$|E_0(R, 0, \theta)| = \frac{|B_1(\theta)|}{R}$$

(3)

---

[2]The modes of a cylindrical waveguide are given in terms of TE_{mn} and TM_{mn}. For a more general expression of far field expansion, refer to discussion on spherical wave functions given in SPS 37-24, Vol. IV, pp. 150–154 and SPS 37-26, Vol. IV, pp. 197–200.

A study of Eq. (2) and Fig. 14 will reveal that $|B_1(\theta)|/R$ is the H-plane amplitude pattern measured as a function of θ at a distance R from the origin. The term $\Phi_{B_1}(\theta)$ in Eq. (2) is the H-plane phase pattern. If $\phi = \pi/2$, Eq. (1) becomes

$$E_0(R, \pi/2, \theta) = \frac{|A_1(\theta)|}{R} e^{-j[\kappa R - \Phi_{A_1}(\theta)]} a_\theta$$

(4)

Then

$$|E_0(R, \pi/2, \theta)| = \frac{|A_1(\theta)|}{R}$$

(5)

The latter equation is the E-plane amplitude pattern as measured at a distance R from the origin. The term $\Phi_{A_1}(\theta)$ in Eq. (4) is the E-plane phase pattern.

The complete expression for the far field electric field amplitude obtained from Eq. (1) is

$$|E_0(R, \phi, \theta)| = \frac{1}{R} [|A_1(\theta)|^2 \sin^2 \phi + |B_1(\theta)|^2 \cos^2 \phi]^{1/2}$$

(6)

The total radiated power is

$$P_T = \frac{1}{2} \left(\frac{\varepsilon}{\mu}\right)^{1/2} R^2 \int_0^\pi \int_0^{2\pi} |E_0(R, \phi, \theta)|^2 \sin \theta \, d\phi \, d\theta$$

(7)

Substitution of Eq. (6) into Eq. (7) and integration over ϕ gives

$$P_T = \frac{\pi}{2} \left(\frac{\varepsilon}{\mu}\right)^{1/2} \int_0^\pi [|A_1(\theta)|^2 + |B_1(\theta)|^2] \sin \theta \, d\theta$$

(8)

Let $P(\theta, \phi) = $ power/unit solid angle in any direction (θ, ϕ).

$$P(\theta, \phi) = \frac{1}{2} \left(\frac{\varepsilon}{\mu}\right)^{1/2} R^2 |E_0(R, \phi, \theta)|^2$$

(9)

Substitution of Eq. (6) into Eq. (9) gives

$$P(\theta, \phi) = \frac{1}{2} \left(\frac{\varepsilon}{\mu}\right)^{1/2} [|A_1(\theta)|^2 \sin^2 \phi + |B_1(\theta)|^2 \cos^2 \phi]$$

(10)

From definition of gain function in any direction and substitutions of Eq. (8) and (10) gives

$$G(\theta, \phi) = \left[\frac{P(\theta, \phi)}{\frac{P_T}{4\pi}} \right] = \frac{4[\,|A_1(\theta)|^2 \sin^2 \phi + |B_1(\theta)|^2 \cos^2 \phi\,]}{\int_0^\pi [\,|A_1(\theta)|^2 + |B_1(\theta)|^2\,] \sin \theta \, d\theta} \tag{11}$$

Special example. Assume that the angles, at which the maximum power per unit solid angle occurs, are in one of the principal planes (E or H) so that $\phi = 0$ or $\pi/2$ in Eq. (11). Let

$p_1(\theta) =$ One of the principal plane power patterns normalized by the maximum power per unit solid angle.

$p_2(\theta) =$ Second principal plane pattern normalized by the maximum power per unit solid angle.

Then for the special case

$$G_M = \frac{4}{\int_0^\pi [p_1(\theta) + p_2(\theta)] \sin \theta \, d\theta} \tag{12}$$

To evaluate the denominator by a numerical integration technique, let the integral between $\theta = 0$ and π be represented by the area of n number of rectangles. At the ith rectangle, the width of the rectangle is

$$\Delta \theta_i = \frac{\theta_{i+1} - \theta_{i-1}}{2}$$

and the height of the rectangle at the ith subregion is given by

$$(p_{1i} + p_{2i}) \sin \theta_i$$

where $i = 1, 2, \cdots, n$. For the end points, $i = 1$ and $i = n$.

$$\sin \theta_1 = \sin 0 = 0$$

$$\sin \theta_n = \sin \pi = 0$$

Then Eq. (12) becomes approximately[3]

$$G_M \approx \frac{8}{\sum_{i=2}^{n-1} [(p_{1i} + p_{2i}) (\sin \theta_i) (\theta_{i+1} - \theta_{i-1})]} \tag{13}$$

[3]Even though G_M, as defined here, is for a special case and the maximum gain may not necessarily occur in one of the principal E or H planes, the same expression will be used for developing subsequent equations; and, therefore, G_M may be considered to be a reference gain for the general case.

and in decibels

$$(G_M)_{\text{db}} = 10 \log_{10} G_M \tag{14}$$

c. Beam efficiency. Beam efficiency as defined here will refer to the fractional power contained in the annular solid angle between $\theta = 0$ and any angle θ_i.

$$\eta_B = \left\{ \frac{\int_0^{\theta_i} \int_0^{2\pi} P(\theta, \phi) \sin \theta \, d\phi \, d\theta}{P_T} \right\} \tag{15}$$

Substituting Eqs. (10) and (8) into Eq. (15) and integrating over ϕ gives

$$\eta_B = \left\{ \frac{\int_0^{\theta_i} [\,|A_1(\theta)|^2 + |B_1(\theta)|^2\,] \sin \theta \, d\theta}{\int_0^\pi [\,|A_1(\theta)|^2 + |B_1(\theta)|^2\,] \sin \theta \, d\theta} \right\} \tag{16}$$

In terms of normalized patterns,

$$\eta_B = \left\{ \frac{\int_0^{\theta_i} [p_1(\theta) + p_2(\theta)] \sin \theta \, d\theta}{\int_0^\pi [p_1(\theta) + p_2(\theta)] \sin \theta \, d\theta} \right\} \tag{17}$$

and the numerical evaluation of Eq. (17) is performed by the approximation

$$\eta_B \approx \left\{ \frac{\sum_{i=2}^{i} (p_{1i} + p_{2i}) (\sin \theta_i) (\theta_{i+1} - \theta_{i-1})}{\sum_{i=2}^{n-1} (p_{1i} + p_{2i}) (\sin \theta_i) (\theta_{i+1} - \theta_{i-1})} \right\} \tag{18}$$

The fractional power contained in only the ith annular solid angle segment is given as

$$\eta_{Bi} \approx \left\{ \frac{(p_{1i} + p_{2i}) (\sin \theta_i) (\theta_{i+1} - \theta_{i-1})}{\sum_{i=2}^{n-1} [(p_{1i} + p_{2i}) (\sin \theta_i) (\theta_{i+1} - \theta_{i-1})]} \right\} \tag{19}$$

d. Antenna temperature. From the definition of antenna temperature

$$T_A = \left[\frac{\int_0^\pi \int_0^{2\pi} T_b(\theta, \phi) P(\theta, \phi) \sin \theta \, d\phi \, d\theta}{P_T} \right] \tag{20}$$

Assume the brightness temperature function $T_b(\theta, \phi)$ of the environment surrounding the antenna has no variation along the ϕ coordinate. This will be true for an antenna pointed at zenith over flat Earth. Then substitution in terms of the normalized power patterns and integrating over ϕ gives

$$T_A = \left[\frac{\int_0^\pi T_b(\theta) [p_1(\theta) + p_2(\theta)] \sin\theta \, d\theta}{\int_0^\pi [p_1(\theta) + p_2(\theta)] \sin\theta \, d\theta} \right] \quad (21)$$

and for numerical evaluation of Eq. (21), the following approximate equation is given

$$T_A \approx \left\{ \frac{\sum_{i=2}^{n-1} [T_{bi}(p_{1i} + p_{2i})(\sin\theta_i)(\theta_{i+1} - \theta_{i-1})]}{\sum_{i=2}^{n-1} [(p_{1i} + p_{2i})(\sin\theta_i)(\theta_{i+1} - \theta_{i-1})]} \right\} \quad (22)$$

where T_{b_i} is the average brightness temperature of the ith annular solid angle segment.

Note also that Eq. (22) may be written as

$$T_A \approx \sum_{i=2}^{n-1} (T_{bi}\eta_{Bi}) \quad (23)$$

where n_{Bi} was defined by Eq. (19).

e. Probable error analyses. In practice an antenna pattern recording is usually made on a decibel scale. A sample recording for antenna patterns in decibels may be seen in Fig. 15. The following terms are defined.

F_{1i} = Value of the ith data point on pattern 1, as read relative to zero db of the recording scale. The values of any F_{1i} points must be negative db, with the exception of any zero db values.

F_{2i} = Value of the ith data point in db on pattern 2, as read relative to zero db of the recording scale.

$F_{M1,2}$ = The least negative data point value in db on patterns 1 and 2, respectively.

Using definitions made previously, the normalized power ratio for the ith data point is expressed as

$$p_{1i} = 10^{[(F_{1i} - F_{M1})/10]} \quad (24)$$
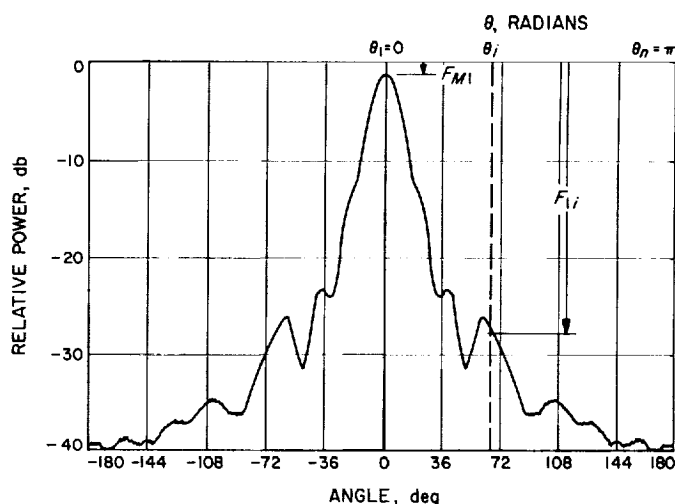
Fig. 15. Sample pattern recording for an antenna with circular symmetry

and

$$p_{2i} = 10^{[(F_{2i} - F_{M2})/10]} \quad (25)$$

The total probable error in determining the antenna gain, from experimental antenna patterns and assumed brightness temperatures, is a function of the probable errors of all the independent variables.

The equations for numerical evaluation of antenna gain and antenna gain in db were respectively given in Eqs. (13) and (14). The probable errors for these two expressions are related by

$$PE_{(G_M)_{db}} = (PE_{G_M})\left[\frac{\partial (G_M)_{db}}{\partial G_M} \right]$$
$$= (PE_{G_M})\left[\frac{1}{\left(G_M \frac{\ln 10}{10} \right)} \right] \quad (26)$$

From the expression of G_M in Eq. (13), let

$$\sum_{i=2}^{n-1} (p_{1i} + p_{2i})(\sin\theta_i)(\theta_{i+1} - \theta_{i-1}) = \Sigma_D$$

Then

$$G_M = \frac{8}{\Sigma_D}$$

If u_i represent any of the independent variables of G_M, then

$$\frac{\partial G_M}{\partial u_i} = \frac{-8}{\Sigma_D^2}\left(\frac{\partial \Sigma_D}{\partial u_i}\right) = \frac{-G_M^2}{8}\left(\frac{\partial \Sigma_D}{\partial u_i}\right) \tag{27}$$

and

$$\frac{\partial \Sigma_D}{\partial u_i} = \frac{\partial}{\partial u_i}\left\{\begin{array}{l} \cdots \\ + (p_{1(i-1)} + p_{2(i-1)})(\sin\theta_{i-1})(\theta_i - \theta_{i-2}) \\ + (p_{1i} + p_{2i})(\sin\theta_i)(\theta_{i+1} - \theta_{i-1}) \\ + (p_{1(i+1)} + p_{2(i+1)})(\sin\theta_{i+1})(\theta_{i+2} - \theta_i) + \\ \cdots \end{array}\right\} \tag{28}$$

The above two equations are useful for evaluating the total probable error of G_M which is expressed as follows, taking account of probable errors of p_{1i} and p_{2i} in terms of the probable errors of F_{1i}, F_{2i}, F_{M1}, and F_{M2}.

$$PE_{G_M} = \sqrt{\sum_{i=2}^{n-1}\left\{\begin{array}{l}\left[PE_{F_{1i}}\left(\frac{\partial p_{1i}}{\partial F_{1i}}\right)\left(\frac{\partial G_M}{\partial p_{1i}}\right)\right]^2 + \left[PE_{F_{M1}}\left(\frac{\partial p_{1i}}{\partial F_{M1}}\right)\left(\frac{\partial G_M}{\partial p_{1i}}\right)\right]^2 \\ + \left[PE_{F_{2i}}\left(\frac{\partial p_{2i}}{\partial F_{2i}}\right)\left(\frac{\partial G_M}{\partial p_{2i}}\right)\right]^2 + \left[PE_{F_{M2}}\left(\frac{\partial p_{2i}}{\partial F_{M2}}\right)\left(\frac{\partial G_M}{\partial p_{2i}}\right)\right]^2 \\ + \left[PE_{\theta_i}\left(\frac{\partial G_M}{\partial \theta_i}\right)\right]^2\end{array}\right\}} \tag{29}$$

where

PE_{θ_i} = probable error of the ith measurement of θ.

$PE_{F_{M1,2}}$ = probable errors of $F_{M1,2}$

$PE_{F_{1i}}$ = probable error of the ith measurement of F_1.

$PE_{F_{2i}}$ = probable error of the ith measurement of F_2.

$PE_{F_{1i}}$ and $PE_{F_{2i}}$ can be expressed in terms of a constant probable error, PE_a (round-off error, recorder jitter, etc.), and a system nonlinearity error proportional to the recorder signal level, as shown in the following:

$$PE_{F_{1i}}^2 = PE_a^2 + (bF_{1i})^2 \tag{30}$$

$$PE_{F_{2i}}^2 = PE_a^2 + (bF_{2i})^2 \tag{31}$$

After performing the required partial differentiations and making the above substitutions into Eqs. (29) and (26), the following expression is obtained for the total probable error of the antenna gain in db.

$$PE_{(G_M)_{db}} = \frac{G_M}{8} \sqrt{\sum_{i=2}^{n-1} \left\{ \begin{array}{l} [PE_{\bar{a}}^2 + PE_{F_{M1}}^2 + b^2 F_{1i}^2][p_{1i}(\sin\theta_i)(\theta_{i+1} - \theta_{i-1})]^2 \\ + [PE_{\bar{a}}^2 + PE_{F_{M2}}^2 + b^2 F_{2i}^2][p_{2i}(\sin\theta_i)(\theta_{i+1} - \theta_{i-1})]^2 \\ + \left(\frac{10}{\ln 10}\right)^2 PE_{\theta_i}^2 \left[\begin{array}{l} (p_{1(i-1)} + p_{2(i-1)})(\sin\theta_{i-1}) + (p_{1i} + p_{2i})(\cos\theta_i)(\theta_{i+1} - \theta_{i-1}) \\ - (p_{1(i+1)} + p_{2(i+1)})(\sin\theta_{i+1}) \end{array} \right]^2 \end{array} \right\}}$$

(32)

$$\sqrt{\sum_{i=2}^{n-1} \left\{ \begin{array}{l} [PE_{T_{bi}}^2][(p_{1i} + p_{2i})(\sin\theta_i)(\theta_{i+1} - \theta_{i-1})]^2 \\ + \left(\frac{\ln 10}{10}\right)^2 [PE_{\bar{a}}^2 + PE_{F_{M1}}^2 + b^2 F_{1i}^2][(T_{bi} - T_A)(p_{1i}\sin\theta_i)(\theta_{i+1} - \theta_{i-1})]^2 \\ + \left(\frac{\ln 10}{10}\right)^2 [PE_{\bar{a}}^2 + PE_{F_{M2}}^2 + b^2 F_{2i}^2][(T_{bi} - T_A)(p_{2i}\sin\theta_i)(\theta_{i+1} - \theta_{i-1})]^2 \\ + [PE_{\theta_i}^2] \left[\begin{array}{l} (T_{b(i-1)} - T_A)(p_{1(i-1)} + p_{2(i-1)})(\sin\theta_{i-1}) \\ + (T_{bi} - T_A)(p_{1i} + p_{2i})(\cos\theta_i)(\theta_{i+1} - \theta_{i-1}) \\ - (T_{b(i+1)} - T_A)(p_{1(i+1)} + p_{2(i+1)})(\sin\theta_{i+1}) \end{array} \right]^2 \end{array} \right\}}$$

(33)

In a similar manner, the probable error of determining antenna temperature from Eq. (22) is given in the following:

$$PE_{T_A} = \frac{G_M}{8}$$

where

G_M and T_A are determined respectively from Eqs. (13) and (22).

$PE_{T_{bi}}$ = probable error of the ith value of the brightness temperature.

Note that if $T_{bi} = T_A$ in Eq. (33), the error terms due to the patterns go to zero. For example, an antenna completely surrounded by black walls of an enclosure at uniform temperature T_b will have an antenna temperature of T_b. Pattern errors will not result in antenna temperature errors for this case.

Account has not been taken in these error expressions for the dynamic range of the patterns. The limits of this source of error will be found by computing the parameters in two ways. The parameters are first solved with the patterns as taken and then resolved using those portions of the patterns at the lower dynamic limit replaced with an assumed greater (say an additional 40 db) limit.

A program is presently being written for the IBM 7094 computer for computations of $(G_M)_{db}$, n_B, and T_A and their probable errors. Plans are being made to extend the analysis to the most general polarization case, which would be elliptical polarization.

4. Error Analysis of CW Signal Power Calibration with Thermal Noise Standards, M. S. Reid[1] and C. T. Stelzried

a. Summary. A complete error analysis of the calibration by thermal loads of the CW received signal power from a spacecraft is presented. The analysis, which considers only random errors, is in three parts:

(1) The accuracy of the calibration of the test transmitter power level.

(2) The spacecraft CW power measurement accuracy, which includes the accuracy of the AGC curve calibrations.

(3) A comparison of the CW received spacecraft power level at the three Goldstone stations where the experimental measurements were made. This includes the analysis of the probable errors associated with the antenna gain measurements.

A graph of CW signal power measurement resolution vs signal level for various values of the time constant and bandwidth parameters is presented. The analysis also includes a graph of the variation of antenna boresight error with source diameter, for two values of beamwidth. An expression for normalized system temperature probable error with the antenna pointed at a source is derived, and is shown to be a measure of the probable error in the measured value of source temperature.

b. Introduction. An attempt to improve the accuracy of the calibration of the CW received signal power in the DSN is presently underway. A convenient measure of the spacecraft received power level is the receiver AGC voltage, which is calibrated for absolute received power, defined at the receiver input, with a calibrated test transmitter. The method, equipment and results are discussed elsewhere (e.g., *SPS 37-35,* Vol. III, p. 58, and *SPS 37-36,* Vol. III, p. 44). A partial error analysis of the test transmitter calibration has been published (*SPS 37-32,* Vol. IV, p. 246). A complete error analysis of the spacecraft power measurement falls into the following parts:

(1) The accuracy of the calibration of the test transmitter power level.

(2) The spacecraft CW power measurement accuracy, which includes the accuracy of the AGC curve calibrations.

(3) A comparison of received spacecraft power level at the three Goldstone stations where this experiment was performed. This includes the analysis of the probable errors associated with the antenna gain measurements.

c. The calibration of the test-transmitter power level. A preliminary error analysis has been performed on the equation which calibrates the test-transmitter power level (*SPS 37-32,* Vol. IV, p. 246). This analysis is considered here in greater detail. The equation is

$$P_{si} = (Y-1) kT_s B \frac{1}{g(f_s)} \qquad (1)$$

where the terms have been defined in *SPS 37-32,* Vol. IV, p. 244 and *SPS 37-35,* Vol. III, p. 58. If the probable error of the measurement of P_{si} is $PE_{P_{si}}$ and k is assumed to be exact, then

$$(PE_{P_{si}})^2 = \left(\frac{\partial P_{si}}{\partial Y}\right)^2 (PE_Y)^2 + \left(\frac{\partial P_{si}}{\partial T_s}\right)^2 (PE_{T_s})^2$$
$$+ \left(\frac{\partial P_{si}}{\partial B}\right)^2 (PE_B)^2 + \left(\frac{\partial P_{si}}{\partial g(f_s)}\right)^2 (PE_{g(f_s)})^2 \qquad (2)$$

which may be written as

$$\left(\frac{PE_{P_{si}}}{P_{si}}\right)^2 = \left(\frac{PE_Y}{Y}\right)^2 \left[1 + \frac{kT_s B}{P_{si} g(f_s)}\right]^2 + \left(\frac{PE_{T_s}}{T_s}\right)^2$$
$$+ \left(\frac{PE_B}{B}\right)^2 + \left(\frac{PE_{g(f_s)}}{g(f_s)}\right)^2 \qquad (3)$$

Each term on the right hand side of Eq. (3) is now considered separately.

(PE_Y/Y). The term PE_Y, the probable error in the Y-factor measurement, is a function of:

(1) The nonlinearity of the precision attenuator (*SPS 37-35,* Vol. III, p. 58).

(2) The test transmitter input signal level instability $\Delta P_{si}/P_{si}$.

(3) A radiometer gain instability $[(1/\tau B) + (\Delta G/G_0)^2]^{1/2}$ where τ = post detector time constant and G = receiver gain (Ref. 6).

(4) The resetability of the attenuator and null indicator for repeated Y-factor measurements.

[1] On leave of absence from the National Institute for Telecommunications Research, Johannesburg, South Africa.

From the manufacturer's specifications (1) may be written $a_1 Y_{db}$ where a_1 is a constant and Y_{db} is the measured Y-factor in db. (4) is a constant and may be written as a_2. Then

$$\left(\frac{PE_Y}{Y}\right)^2 = (a_1 Y_{db})^2 + a_2^2 + \left(\frac{1}{\tau B}\right) + \left(\frac{\Delta G}{G_0}\right)^2 + \left(\frac{\Delta P_{si}}{P_{si}}\right)^2 \tag{4}$$

(PE_{T_s}/T_s). The system temperature T_s, when connected to the antenna, is given by

$$T_s = \frac{T_0 + T_R}{Y_{0A}} \tag{5}$$

where

T_0 = temperature of ambient load

T_R = receiver temperature

Y_{0A} = Y-factor (ratio) switching between ambient load and antenna at zenith

This Y-factor method of system temperature measurement was chosen because a simple and quick test was required, which could be performed by station personnel without interfering with the normal station precalibration routine. An ambient load was chosen, as this is the most convenient device from an operational point of view as well as the most accurate for the given station conditions.

From Eq. (5):

$$(PE_{T_s})^2 = \left(\frac{\partial T_s}{\partial T_0}\right)^2 (PE_{T_0})^2 + \left(\frac{\partial T_s}{\partial T_R}\right)^2 (PE_{T_R})^2 + \left(\frac{\partial T_s}{\partial Y_{0A}}\right)^2 (PE_{Y_{0A}})^2 \tag{6}$$

After the differentiation has been performed and the equation normalized, Eq. (6) becomes

$$\left(\frac{PE_{T_s}}{T_s}\right)^2 = \left(\frac{PE_{T_0}}{T_0}\right)^2 \left[1 - \frac{T_R}{T_s Y_{0A}}\right]^2 + \left(\frac{PE_{T_R}}{T_R}\right)^2 \left[1 - \frac{T_0}{T_s Y_{0A}}\right]^2 + \left(\frac{PE_{Y_{0A}}}{Y_{0A}}\right)^2 \tag{7}$$

The probable error in this Y-factor measurement is similar to the probable error of the Y-factor in Eq. (3), except that here there is no input signal. Thus

$$\left(\frac{PE_{Y_{0A}}}{Y_{0A}}\right)^2 = (a_1 Y_{0A(db)})^2 + a_2^2 + \frac{1}{\tau B} + \left(\frac{\Delta G}{G_0}\right)^2 \tag{8}$$

and Eq. (7) becomes

$$\left(\frac{PE_{T_s}}{T_s}\right)^2 = \left(\frac{PE_{T_0}}{T_0}\right)^2 \left[1 - \frac{T_R}{T_s Y_{0A}}\right]^2 + \left(\frac{PE_{T_R}}{T_R}\right)^2 \left[1 - \frac{T_0}{T_s Y_{0A}}\right]^2 + (a_1 Y_{0A(db)})^2 + a_2^2 + \frac{1}{\tau B} + \left(\frac{\Delta G}{G_0}\right)^2 \tag{9}$$

If the temperature of the load is considered variable and written T_L instead of T_0, then a graph of Eq. (9) may be drawn. This is shown in Fig. 16, where the variation of normalized system temperature probable error has been drawn. The following values have been used for the various parameters

$$\frac{1}{\tau B} = 10^{-5} \qquad \left(\frac{\Delta G}{G_0}\right) = 0.005 \text{ db}$$

$$a_1 = 0.00354 \text{ db} \qquad a_2 = 0.003 \text{ db}$$

With the given station conditions the receiver temperature was known to 1°K. In this case the graph shows that the normalized system temperature probable error is a minimum for load temperatures between about 100 and 450°K, and is equal to approximately 0.008. This indicates an error of 0.4°K in a 50°K system temperature.

If, however, the receiver temperature is more accurately known $(PE_{T_R} \approx 0.2°K)$, that is, if it has been determined with precision cryogenic loads, then the normalized system temperature probable error is a minimum for load temperatures of about 77°K.

(PE_B/B). In order to derive an expression for the normalized bandwidth probable error the method of measuring bandwidth must be considered. Measurements of attenuation and frequency are made over the band and the resulting curve integrated by a trapezoidal process to yield the total bandwidth. This is expressed by Eq. (10). The terms are defined in Fig. 17.

$$B = \tfrac{1}{2} y_1 (f_2 - f_1) + \tfrac{1}{2} y_n (f_n - f_{n-1}) + \tfrac{1}{2} \sum_{i=2}^{n} y_i (f_{i+1} - f_{i-1}) \tag{10}$$
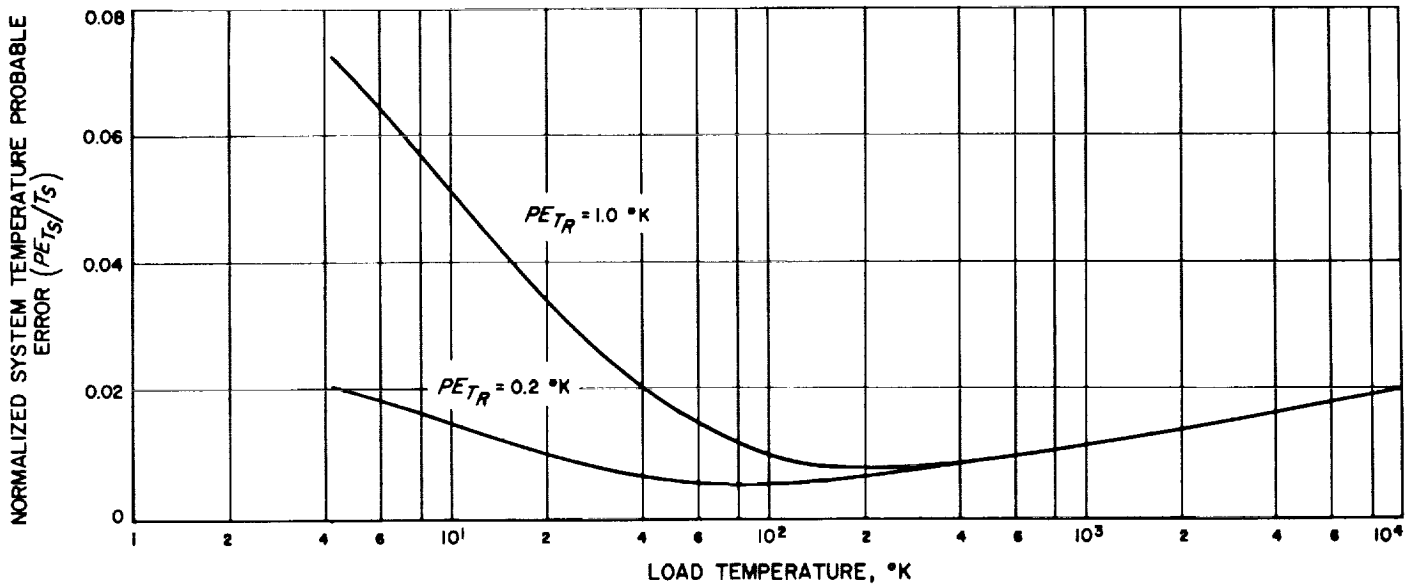
Fig. 16. System temperature probable error vs load temperature
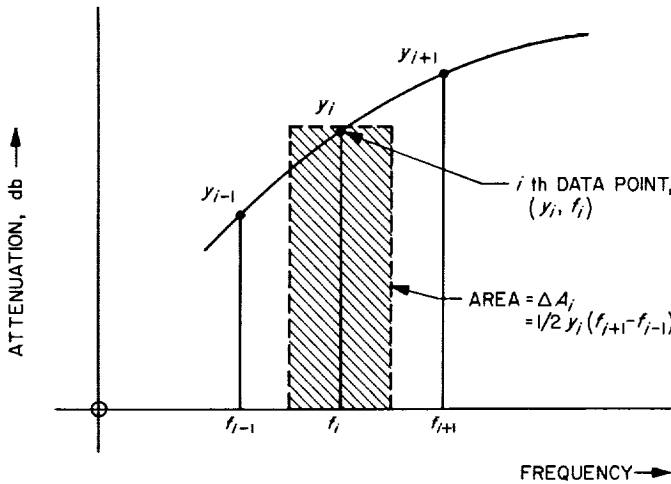
Fig. 17. The measurement of bandwidth
by trapezoidal integration

The probable error PE_B is derived from Eq. (10) thus

$$(PE_B)^2 = (PE_{y_1})^2 \left(\frac{\partial B}{\partial y_1}\right)^2 + \cdots + (PE_{y_i})^2 \left(\frac{\partial B}{\partial y_i}\right)^2$$
$$+ \cdots + (PE_{y_n})^2 \left(\frac{\partial B}{\partial y_n}\right)^2 + (PE_{f_1})^2 \left(\frac{\partial B}{\partial f_1}\right)^2$$
$$+ \cdots + (PE_{f_i})^2 \left(\frac{\partial B}{\partial f_i}\right)^2$$
$$+ \cdots + (PE_{f_n})^2 \left(\frac{\partial B}{\partial f_n}\right)^2 \Bigg\} \quad (11)$$

Therefore

$$\left(\frac{PE_B}{B}\right) = \frac{1}{B} \Bigg\{ [(a_1 Y_{1(db)})^2 + a_2^2]\, y_1^2 \left(\frac{f_2 - f_1}{2}\right)^2$$
$$+ [(a_1 Y_{n(db)})^2 + a_2^2]\, y_n^2 \left(\frac{f_n - f_{n-1}}{2}\right)^2$$
$$+ \tfrac{1}{4} \sum_{i=2}^{n} [(a_1 Y_{i(db)})^2 + a_2^2]\, y_i^2 (f_{i+1} - f_{i-1})^2$$
$$+ \left(\frac{PE_f}{2}\right)^2 \left[y_1^2 + y_n^2 + \sum_{i=2}^{n} (y_{i-1} - y_{i+1})^2 \right] \Bigg\}^{1/2}$$
$$(12)$$

The bandwidth B and normalized probable error in bandwidth PE_B/B of each narrow band filter were determined by Eqs. (10) and (12), which have been programmed for a 1620 computer. For example, approximately 210 data points were measured on the Pioneer Station's narrow band filter, and the computer output gave the following results

$$B = 11.569 \text{ kc} \qquad \frac{PE_B}{B} = 0.00026$$

It is known from experience, however, that due to effects such as crystal aging, etc., not taken into consideration by the error analysis, a practical value for PE_B/B is 0.0026.

$PE_{g(f_s)}/g(f_s)$. It is not practical to reduce this term any further.

Fig. 18 shows a graph of the normalized probable error $PE_{P_{si}}/P_{si}$ against P_{si} in dbm for two values of bandwidth and time constant

$$B_1 = 10\,\text{kc} \qquad \tau_1 = 0.1\,\text{sec}$$

$$B_2 = 1\,\text{kc} \qquad \tau_2 = 1.0\,\text{sec}$$

and for the other constants

$$k = 1.38 \times 10^{-23}$$

$$a_1 = 0.00354\,\text{db}$$

$$a_2 = 0.003\,\text{db}$$

$$\left\{ \left(\frac{\Delta G}{G_0}\right)^2 + \left(\frac{\Delta P_{si}}{P_{si}}\right)^2 \right\}^{\frac{1}{2}} = 0.01\,\text{db}$$

$$T_s = 45°\text{K}$$

$$\frac{PE_B}{B} = 0.0026$$

$$g(f_s) = 1$$

$$\frac{PE_{g(f_s)}}{g(f_s)} = 0.003$$

$$Y = 20\,\text{db}$$

$$T_R = 11°\text{K}$$

$$T_0 = 247°\text{K}$$

$$Y_{0.1} = 8.5\,\text{db}$$

$$PE_{T_R} = 1°\text{K}$$

$$PE_{T_0} = 0.1°\text{K}$$

The figure shows that maximum resolution at low power levels is obtained by narrowing the bandwidth and increasing the postdetector time constant. From the graph $PE_{P_{si}}/P_{si} = 0.022$ for $\tau = 1.0$ sec, $B = 10$ kc and $P_{si} = -120$ dbm; i.e.,

$$\left(\frac{PE_{P_{si}}}{P_{si}}\right)_{\substack{test \\ transmitter}} = 0.022$$

or approximately 2% or 0.1 db. These normalized probable errors relate to a single measurement only and could be improved by making several Y-factor measurements, for example, and averaging the results.
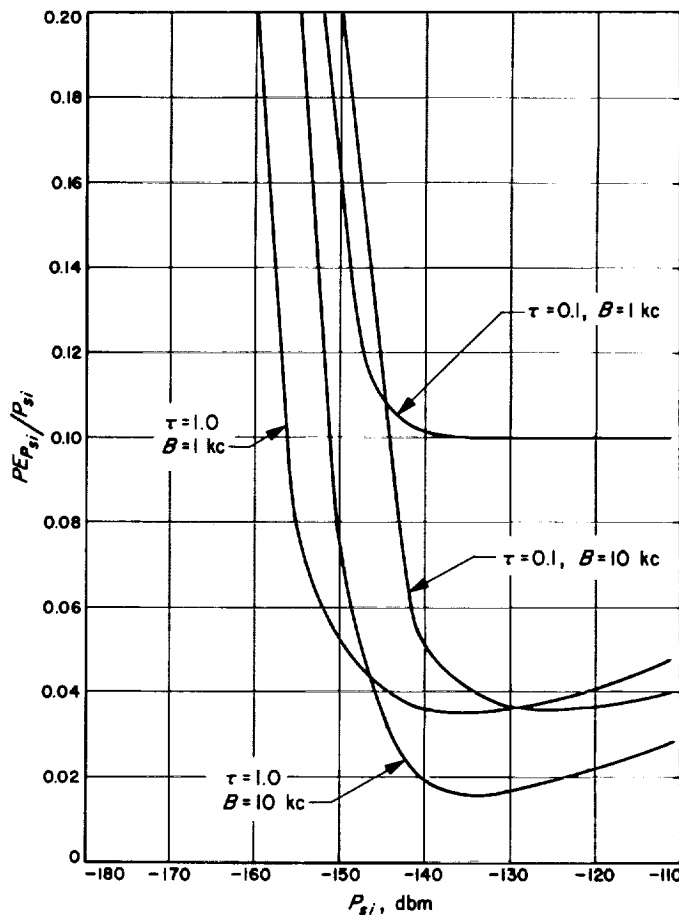
Fig. 18. CW signal power measurement resolution vs signal level

d. The spacecraft CW power measurement accuracy. The accuracy of the transfer of the calibrated test transmitter power levels to the spacecraft power level through the AGC curves involves:

(1) The receiving system linearity, RF to IF.

(2) The linearity and calibration of the variable attenuator in the test transmitter.

(3) The calibration of the step attenuator in the test transmitter, if applicable.

These errors, (1), (2), and (3), may be written as a_3, a_4, and a_5, respectively. Other errors associated with spacecraft power measurement are AGC voltage indicator jitter a_6; AGC graph calibration error a_7; and antenna to spacecraft pointing error a_8. Furthermore, a system gain change that occurs after the AGC and test transmitter calibrations, but before the spacecraft is acquired, may also be a source of error. If this is written a_9, then the

normalized probable error in spacecraft power measurement is

$$\left(\frac{PE_{P_{si}}}{P_{si}}\right)_{spacecraft} = \left\{\left(\frac{PE_{P_{si}}}{P_{si}}\right)^2_{\substack{test \\ transmitter}} + \sum_{i=3}^{9}(a_i)^2\right\}^{\frac{1}{2}} \quad (13)$$

assuming:

$$a_3 = a_5 = 0.02 \text{ db}$$
$$a_4 = 0.04 \text{ db}$$
$$a_6 = 0.01 \text{ db}$$
$$a_7 = 0$$
$$a_8 = 0.0125 \text{ db}$$
$$a_9 = 0.25 \text{ db}$$

then
$$\sum_{i=3}^{9}(a_i)^2 = 0.00366$$

and

$$\left(\frac{PE_{P_{si}}}{P_{si}}\right)_{spacecraft} = \{0.000400 + 0.00366\}^{\frac{1}{2}}$$

$$= 0.064 \text{ or approximately } 0.3 \text{ db}.$$

e. Error analysis of antenna gain measurements. In order to compare the CW signal received power levels at different stations the antenna efficiency of each station must be taken into account (*SPS 37-36*, Vol. III, p. 44). The requirements for this test were similar to those for the system temperature measurement: a simple, quick test which would not interrupt normal station operation to any great extent. Once again a Y-factor method was chosen. Each station tracked a radio source and the antenna efficiency was derived from Eq. (14)

$$\eta = \frac{\text{Measured source temperature}}{\text{Theoretical source temperature}}$$

$$= \frac{T_0 + T_R}{T_{st}}\left\{\frac{1}{Y_1} - \frac{1}{Y_2}\right\} \quad (14)$$

where

T_{st} = Theoretical source temperature, °K

T_R = Receiver temperature, °K

T_0 = Temperature of ambient load, °K

Y_1 = Y-factor (ratio) switching between ambient load and antenna on the radio source

Y_2 = Y-factor (ratio) switching between ambient load and antenna off the radio source

Two radio sources, Omega and Taurus A (Ref. 7), were chosen and each station tracked these almost nightly, over a period of several weeks.

If PE_η is the probable error of the antenna efficiency, then from Eq. (14)

$$(PE_\eta)^2 = \left(\frac{\partial\eta}{\partial T_0}\right)^2(PE_{T_0})^2 + \left(\frac{\partial\eta}{\partial T_R}\right)^2(PE_{T_R})^2 + \left(\frac{\partial\eta}{\partial Y_1}\right)^2(PE_{Y_1})^2$$
$$+ \left(\frac{\partial\eta}{\partial Y_2}\right)^2(PE_{Y_2})^2 + \left(\frac{\partial\eta}{\partial T_{st}}\right)^2(PE_{T_{st}})^2 \quad (15)$$

From Eqs. (14) and (15)

$$\left(\frac{PE_\eta}{\eta}\right)^2 = \left(\frac{PE_{T_0}}{T_0}\right)^2\left[\frac{T_0}{T_0 + T_R}\right]^2 + \left(\frac{PE_{T_R}}{T_R}\right)^2\left[\frac{T_R}{T_0 + T_R}\right]^2$$
$$+ \left(\frac{PE_{Y_1}}{Y_1}\right)^2\left[\frac{Y_2}{Y_2 - Y_1}\right]^2$$
$$+ \left(\frac{PE_{Y_2}}{Y_2}\right)^2\left[\frac{Y_1}{Y_2 - Y_1}\right]^2 + \left(\frac{PE_{T_{st}}}{T_{st}}\right)^2 \quad (16)$$

Y_2 is given by

$$Y_2 = \frac{T_0 + T_R}{T_{sA}} \quad (17)$$

where T_{sA} = system temperature with the source outside the antenna beam. The error in the measurement of Y_2 is then given by

$$\left(\frac{PE_{Y_2}}{Y_2}\right)^2 = \left(\frac{PE_{T_{sA}}}{T_{sA}}\right)^2 + \frac{1}{\tau B} + \left(\frac{\Delta G}{G_0}\right)^2 + (a_1 Y_{2(db)})^2 + a_2^2 \quad (18)$$

Y_1 is given by

$$Y_1 = \frac{T_0 + T_R}{T_{sA} + T_s} = \frac{T_0 + T_R}{T_{ss}} \quad (19)$$

where T_s is the measured source temperature and T_{ss} is the system temperature with the antenna on the radio source.

The error associated with the Y_1 measurement may be written

$$\left(\frac{PE_{Y_1}}{Y_1}\right)^2 = \left(\frac{PE_{T_{ss}}}{T_{ss}}\right)^2 + \frac{1}{\tau B} + \left(\frac{\Delta G}{G_0}\right)^2 + (a_1 Y_{1(db)})^2 + a_2^2 \quad (20)$$

Eqs. (19) and (20) relate to the accuracy of the Y-factor measurement and items that affect that accuracy. Thus, terms such as PE_{T_0}, PE_{T_R} [and $PE_{T_{sA}}$ from Eq. (18)] do not enter Eqs. (19) and (20) since any change in these terms while the Y-factor is being measured will be small and may, therefore, be neglected.

An expression must now be derived for $PE_{T_{ss}}$. This inaccuracy arises from a tracking error on the radio source. As a first approximation a circular radio source of diameter D and uniform brightness distribution will be considered. Fig. 19 is a diagram of the system geometry. Since this is symmetrical about the source center 0 the misalignment x of the antenna beam on the source need only be considered along one axis. From the geometry of the figure

$$\left(\frac{D}{2}\right)^2 = x^2 + \Psi^2 - 2x\Psi \cos(\pi - \phi)$$

$$\therefore \Psi = -x\cos\phi \pm \frac{D}{2}\left\{1 - \left(\frac{2x}{D}\right)^2 (1 - \cos^2\phi)\right\}^{1/2} \quad (21)$$

Since Ψ must be a positive number, the negative sign in Eq. (21) is ignored and with the assumption that x is small Eq. (21) becomes

$$\Psi = \frac{D}{2}\left\{1 - \tfrac{1}{2}\left(\frac{2x}{D}\right)^2 \sin^2\phi - \left(\frac{2x}{D}\right)\cos\phi\right\} \quad (22)$$

T_{ss} may be written (Ref. 8)

$$T_{ss} = K' \int\!\!\int_\Omega T_D(\psi,\phi) f(\psi,\phi)\, d\Omega \quad (\Omega = \text{all space})$$

$$(23)$$

where

T_D = source or disc temperature

K' = a constant

If the antenna main lobe is assumed Gaussian then

$$f(\psi,\phi) = e^{-(\psi/\alpha)^2} \quad (24)$$

where $\quad \alpha = 0.6\,\psi_H$

and $\quad \psi_H$ = the half power points on the main lobe

also $\quad dA = r^2 \sin\psi d\psi d\phi \quad (25)$

where $\quad r$ = distance from antenna to radio source

$\quad$ = constant for small x
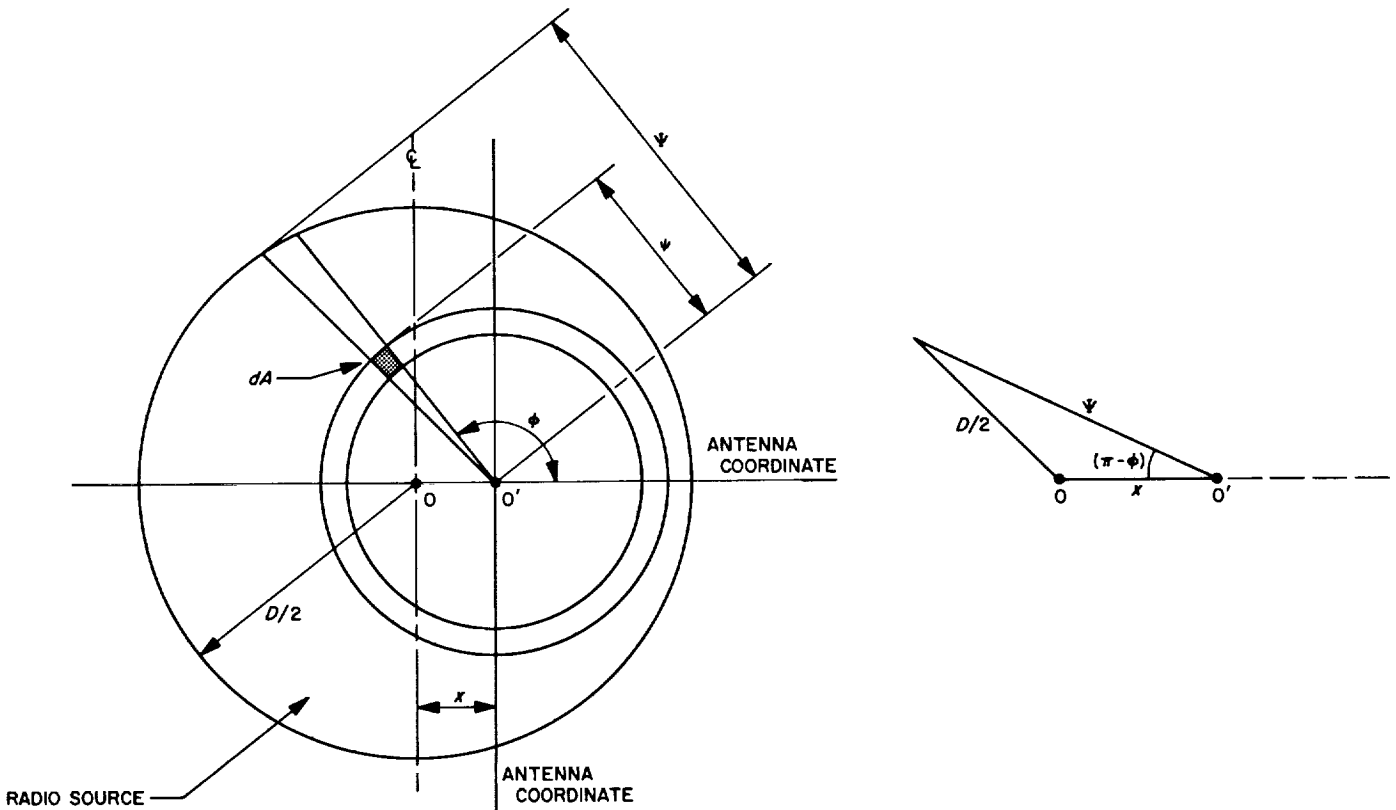
Fig. 19. Geometry of antenna misalignment

Substituting Eqs. (24) and (25) into Eq. (23) and assuming T_D constant

$$T_{ss} = kT_D \int_0^{2\pi} d\phi \int_0^\Psi e^{-(\psi/\alpha)^2} \sin \psi d\psi \qquad (26)$$

For small beamwidth $\sin \psi \approx \psi$ and using Eq. (22), Eq. (26) becomes

$$T_{ss} = \tfrac{1}{2}kT_D\alpha^2 \int_0^{2\pi} \{1 - e^{-(\Psi^2/\alpha)^2}\} d\phi$$

$$= \tfrac{1}{2}kT_D\alpha^2 \int_0^{2\pi} \{1 - e^{-(D/2\alpha)^2 [f(x,\phi)]}\} d\phi$$

where

$$f(x,\phi) = 1 - \left(\frac{2x}{D}\right)^2 \sin^2 \phi - 2\left(\frac{2x}{D}\right)\cos \phi$$

$$+ \left(\frac{2x}{D}\right)^3 \sin^2 \phi \cos \phi$$

$$+ \tfrac{1}{4}\left(\frac{2x}{D}\right)^4 \sin^4 \phi + \left(\frac{2x}{D}\right)^2 \cos^2 \phi$$

Substituting and expanding

$$T_{ss} = \tfrac{1}{2}kT_D\alpha^2 \int_0^{2\pi} 1 - e^{-(D/2\alpha)^2}$$

$$\left\{1 - \left(\frac{D}{2\alpha}\right)^2\left[-\left(\frac{2x}{D}\right)^2 \sin^2 \phi - 2\left(\frac{2x}{D}\right)\cos \phi\right.\right.$$

$$+ \left(\frac{2x}{D}\right)^3 \sin^2 \phi \cos \phi + \tfrac{1}{4}\left(\frac{2x}{D}\right)^4 \sin^4 \phi$$

$$+ \left(\frac{2x}{D}\right)^2 \cos^2 \phi\right] + \tfrac{1}{2}\left(\frac{D}{2\alpha}\right)^4$$

$$\times \left[-\left(\frac{2x}{D}\right)^2 \sin^2 \phi - 2\left(\frac{2x}{D}\right)\cos \phi\right.$$

$$+ \left(\frac{2x}{D}\right)^3 \sin^2 \phi \cos \phi + \tfrac{1}{4}\left(\frac{2x}{D}\right)^4 \sin^4 \phi$$

$$\left.\left.+ \left(\frac{2x}{D}\right)^2 \cos^2 \phi\right]^2\right\} d\phi \qquad (27)$$

All terms in Eq. (27) containing odd powers of $\cos \phi$ become zero through the integration. Performing the integration term by term and after some simplification Eq. (27) reduces to

$$T_{ss} = kT_D\alpha^2\pi \left\{1 - e^{-(D/2\alpha)^2}\left[1 + \left(\frac{D}{2\alpha}\right)^4\right.\right.$$

$$\times \left(\frac{2x}{D}\right)^2 - \frac{3}{32}\left(\frac{D}{2\alpha}\right)^2\left(\frac{2x}{D}\right)^4$$

$$\left.\left.- \frac{1}{32}\left(\frac{D}{2\alpha}\right)^4\left(\frac{2x}{D}\right)^6 + \frac{35}{4096}\left(\frac{D}{2\alpha}\right)^4\left(\frac{2x}{D}\right)^8\right]\right\}$$
$$(28)$$

Since (x/D) is small, $(x/D)^4$ and higher powers may be neglected. Eq. (28) then reduces to

$$T_{ss} = kT_D\alpha^2\pi \left\{1 - e^{-(D/2\alpha)^2}\left[1 + \left(\frac{D}{2\alpha}\right)^4\left(\frac{2x}{D}\right)^2\right]\right\}$$
$$(29)$$

from which may be deduced

$$\left(\frac{PE_{T_{ss}}}{T_{ss}}\right) = \left(\frac{PE_x}{x}\right)\left\{\frac{-2\left(\frac{D}{2\alpha}\right)^4\left(\frac{2x}{D}\right)^2}{e^{(D/2\alpha)^2} - \left(\frac{D}{2\alpha}\right)^4\left(\frac{2x}{D}\right)^2 - 1}\right\} \quad (30)$$

It may be shown from Eq. (30) that

$$\text{as } \frac{D}{2\alpha} \to 0, \left(\frac{PE_{T_{ss}}}{T_{ss}}\right) \to \frac{-4\left(\frac{2x}{D}\right)^2}{1 - 2\left(\frac{2x}{D}\right)^2}$$

where the negative sign in the numerator indicates that the error due to the antenna misalignment always reduces the measured value of source temperature. It may also be shown from Eq. (30) that

$$\text{as } \frac{D}{2\alpha} \to \infty, \left(\frac{PE_{T_{ss}}}{T_{ss}}\right) \to 0$$

Expressions must now be found for PE_x and x which will reduce Eq. (30) to definable terms or to terms that can be accurately estimated.

From Eq. (29)

$$\frac{T_{ss}}{T_{ss(0)}} = \frac{1 - e^{-(D/2\alpha)^2}\left[1 + \left(\frac{D}{2\alpha}\right)^4\left(\frac{2x}{D}\right)^2\right]}{1 - e^{-(D/2\alpha)^2}} \qquad (31)$$

where $T_{ss(0)}$ is the system temperature on the source with no antenna misalignment.

Solving Eq. (31) for x

$$x = \pm \left\{ \frac{\left[e^{(D/2\alpha)^2} - 1\right]\left[1 - \dfrac{T_{ss}}{T_{ss(0)}}\right]}{\left(\dfrac{D}{2\alpha}\right)^4 \left(\dfrac{2}{D}\right)^2} \right\}^{1/2} \quad (32)$$

This is a general expression relating any antenna misalignment x, with the corresponding normalized system temperature, $T_{ss}/T_{ss(0)} = t_{ss}$. In the boresighting process the antenna is slewed amounts X_1, X_2, along a diameter, on either side of the source center. Thus

$$X_{1,2} = \pm \left\{ \frac{\left[e^{(D/2\alpha)^2} - 1\right]\left[1 - t_{ss(X_1, X_2)}\right]}{\left(\dfrac{D}{2\alpha}\right)^4 \left(\dfrac{2}{D}\right)^2} \right\}^{1/2} \quad (33)$$

where $t_{ss(X_1, X_2)}$ is the normalized reduction in $T_{ss(0)}$ associated with $X_{1,2}$, when the antenna is slewed. Nominally $t_{ss(X_1)} = t_{ss(X_2)}$ and usually this is 0.5. Since $x = X_1 - X_2$, and choosing X_2 as reference (i.e., $PE_{X_2} = 0$)

$$PE_x = PE_{X_1} \quad (34)$$

and from Eq. (34)

$$PE_x = PE_{t_{ss(X_1)}} \left(\frac{\partial X_1}{\partial t_{ss(X_1)}} \right) \quad (35)$$

Performing the differentiation and dropping the subscript 1

$$PE_x = \left[\frac{PE_{t_{ss(X)}}}{t_{ss(X)}} \right] \left(\frac{t_{ss(X)}}{2} \right) \left\{ \frac{1}{\left(\dfrac{D}{2\alpha}\right)^2 \left(\dfrac{2}{D}\right)} \right\} \left\{ \frac{e^{(D/2\alpha)^2} - 1}{1 - t_{ss(X)}} \right\}^{1/2} \quad (36)$$

$PE_{t_{ss(X)}}/t_{ss(X)}$ may be converted to db by

$$\left(\frac{PE_{t_{ss(X)}}}{t_{ss(X)}} \right)_{ratio} = PE_{t_{ss(X)}\,(db)} \left(\frac{1}{10} \ln 10 \right)$$

so that

$$\left[\frac{PE_x}{PE_{t_{ss(X)}\,(db)}} \right] = \left(\frac{\ln 10}{10} \right) \left(\frac{t_{ss(X)}}{2} \right) \left\{ \frac{1}{\left(\dfrac{D}{2\alpha}\right)^2 \left(\dfrac{2}{D}\right)} \right\}$$
$$\times \left\{ \frac{e^{(D/2\alpha)^2} - 1}{1 - t_{ss(X)}} \right\} \quad (37)$$

Eq. (37) is the general boresight error equation which relates $PE_x/PE_{t_{ss(X)}}$ to the source diameter, the antenna beamwidth and the "depth" of the boresight,

$$t_{ss(X)} = \frac{T_{ss(X)}}{T_{ss(0)}}$$

The variation of $PE_x/PE_{t_{ss(X)}}$ in deg/db with D in minutes is shown in Fig. 20 for two beamwidths (0.3 and 0.1 deg to the half-power points) and two boresight depths ($t_{ss(X)} = 0.5$ and 0.9). The approximate angular diameters of Taurus A and Omega at 2295 Mc are indicated. The graph shows that for a source diameter of 8 min and for a standard boresight ($t_{ss(X)} = 0.5$) $PE_x/PE_{t_{ss(X)}} = 0.015$ (0.3 deg beam).

Thus, if the null indicator resetability error is 0.1 db in the boresighting process then the angular probable error of pointing on the source is 0.0015 deg. It must be noted that this error is due only to the inaccuracy of the boresight and does not take into account other inaccuracies such as errors from the servo system, etc. In practice the pointing accuracy is approximately 0.01 deg (Ref. 9).

An expression has now been found for PE_x (Eq. 36). The normal error analysis method would proceed by substituting Eq. (36) into Eq. (30), setting $x = 0$ and solving for system temperature probable error. This would yield the system temperature probable error at the point where the antenna had been nominally boresighted to the source center. The expressions in Eqs. (30) and (36), however, are such that $PE_{T_{ss}}/T_{ss}$ becomes indeterminate when x is set equal to zero. This is due to the slope of the $PE_{T_{ss}}/T_{ss}$ vs x curve being equal to zero when x equals zero. This slope remains zero for small variations about $x = 0$ due to the approximations made in the integration of the antenna equation (Eq. 23). This problem is avoided by the use of incrementals in place of the more usual
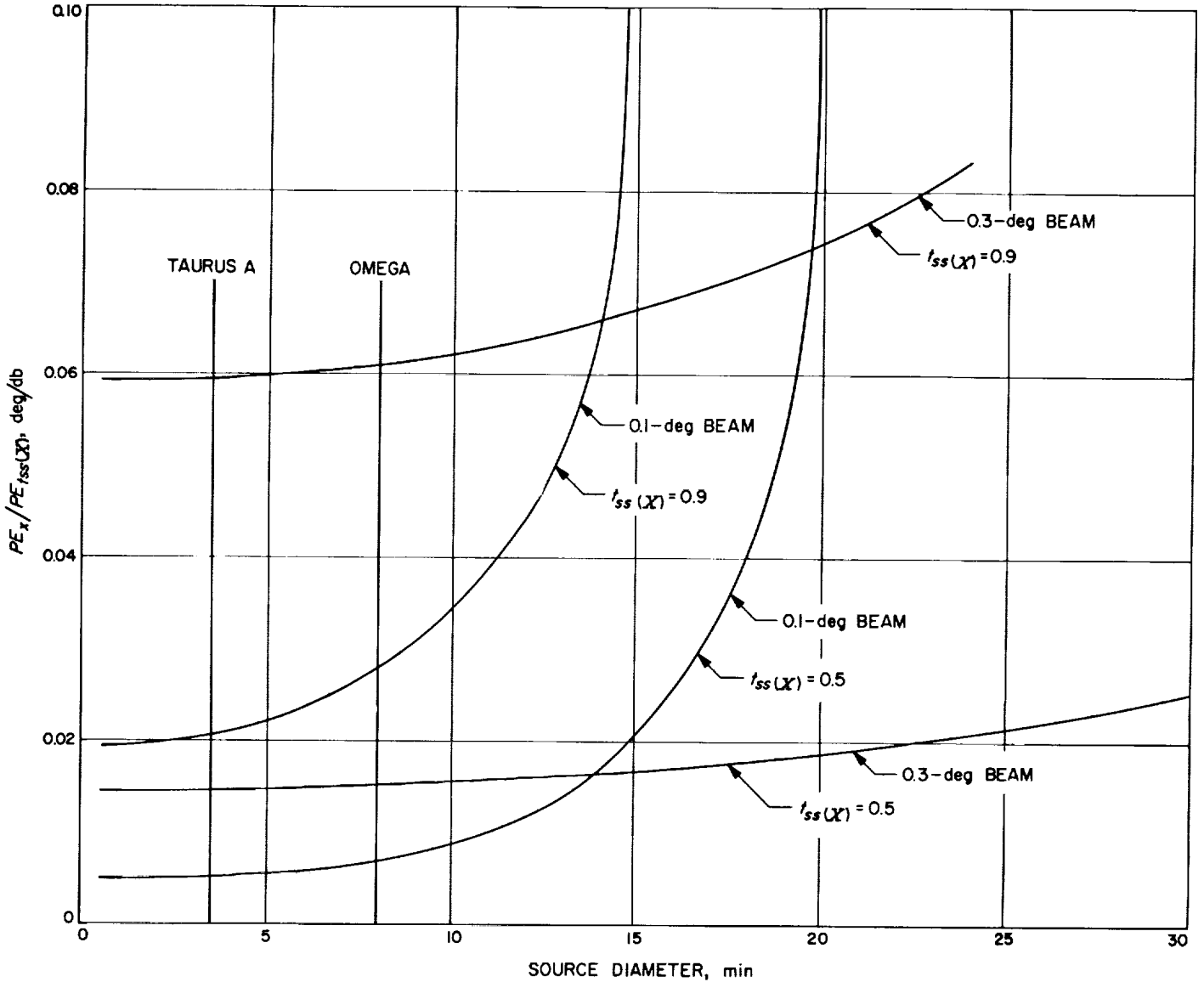
Fig. 20. Antenna boresight error vs source angular diameter

differentials. In the general case the normalized system temperature on the source may be written, from Eq. (31),

$$t_{ss} = \frac{T_{ss}}{T_{ss(0)}} = 1 - \left\{ \frac{e^{-(D/2\alpha)^2}\left(\frac{D}{2\alpha}\right)^4\left(\frac{2x}{D}\right)^2}{1 - e^{-(D/2\alpha)^2}} \right\} \quad (38)$$

where the subscript 0 refers to the point $x = 0$; and

$$\frac{PE_{T_{ss(0)}}}{T_{ss(0)}} = \left. \frac{T_{ss(x+\Delta x)} - T_{ss(x)}}{T_{ss(x)}} \right|_{\substack{x=0 \\ \Delta x = PE_x}} \quad (39)$$

Substituting from the general expression Eq. (29), Eq. (39) becomes

$$\frac{PE_{T_{ss(0)}}}{T_{ss(0)}} = \left[\frac{\left(\frac{D}{2\alpha}\right)^4\left(\frac{2}{D}\right)^2 e^{-(D/2\alpha)^2}}{e^{-(D/2\alpha)^2} - 1} \right]^{(PE_x)^2} \quad (40)$$

Eq. (40) is a slowly varying function with angular source diameter. For a source 8 min in diameter, a 0.3 deg beamwidth and a 0.5 boresight, $PE_{T_{ss(0)}}/T_{ss(0)}$ is approximately 0.0028; i.e., the probable error of a measured source temperature of 50°K is 0.14°K. This probable error arises only from the boresighting error. The expression for $PE_{T_{ss(0)}}/T_{ss(0)}$ given by Eq. (40), together with Eqs. (16), (18), and (20), forms the defining expression for

276

the normalized probable error in antenna efficiency. For the above conditions the given value of $PE_{T_{ss}(0)}/T_{ss}(0)$ is negligible compared with the other terms defining antenna efficiency. Assuming $PE_{T_{ss}(0)}/T_{ss}(0) = 0.0028$, neglecting $PE_{T_{st}}/T_{st}$, and with the values given in sections c and d for the other parameters $PE_\eta/\eta = 0.013$.

The best estimate of spacecraft received power is achieved by comparing or averaging measurements made at more than one station and taking the antenna efficiency of each station into account. For this purpose the probable error of the theoretical source temperature may be neglected and the term $PE_{T_{st}}/T_{st}$ in Eq. (16) is reduced to zero. The probable error of the antenna efficiency is a measure of the accuracy of the averaging process and, therefore, provides a check for bias errors in the final result.

The normalized probable error of the best estimate of spacecraft received power PE_{sp}/sp is given by a combination of Eqs. (13) and (16). Thus

$$\left(\frac{PE_{sp}}{sp}\right) = \left\{\left(\frac{PE_{P_{si}}}{P_{si}}\right)^2_{spacecraft} + \left(\frac{PE_\eta}{\eta}\right)^2\right\}^{\frac{1}{2}} \quad (41)$$

Eq. (41) is a summary of all random errors in the calibration by thermal loads of the CW received signal power from a spacecraft.

References

1. Potter, P. D., "A New Horn Antenna with Suppressed Sidelobes and Equal Beam-widths," *Microwave Journal*, pp. 71–78, June 1963.

2. Yeh, Cavour, *Excitation of Higher Order Modes by a Step Discontinuity of a Circular Waveguide*, Technical Report No. 32-496, Jet Propulsion Laboratory, Pasadena, 1964.

3. Nagelberg, E. R., Shefer, J., "Mode Conversion in Circular Waveguides," *The Bell System Technical Journal*, pp. 1321–1338, Sept. 1965.

4. Worthing, A. G., and Gaffner, J., *Treatment of Experimental Data*, 9th Printing, John Wiley & Sons, New York, 1943.

5. Ludwig, A., private communication, Jet Propulsion Laboratory, Pasadena, 1965.

6. Tuiri, M. E., "Radio Astronomy Receivers," IEEE Transactions, Vol. MIL-B, Nos. 3 and 4, p. 267, July–October 1964.

7. Jet Propulsion Laboratory, Antenna Acceptance Test, Procedure No. DZM-1074-TP, p. 38, April 29, 1965.

8. Ko, H. C., *On the Analysis of Radio Astronomical Observations Made with High-Resolution Radio Telescope Antennas*, RF Project, No. 844, Report No. 1, Ohio State University, Columbus, p. 17, February 1961.

9. Levy, G. S., private communication, Jet Propulsion Laboratory, Pasadena, 1965.

XX. Spacecraft Radio

A. A Similarity Principle for Multipacting Discharges

R. Woo

1. Introduction

In SPS 37-35, Vol. IV, p. 282, a similarity principle for multipacting discharges was described. It was shown that a physical explanation resulted for a controversial assumption made by previous workers in the analysis of the parallel plates case. This article continues the discussion of applications of this similarity principle.

2. One-sided Multipacting in Parallel Plates Geometry

If a dc voltage is applied across the parallel plates in addition to the rf voltage, multipacting involving only one electrode may result. This case is known as one-sided multipacting and has been studied quite extensively by Vance (Ref. 1). Scaling relations for this case will be derived through application of the similarity principle. Some comments on Vance's analysis will also be made.

With the addition of a dc voltage, the equation of motion is

$$\frac{d^2s}{dt^2} = -\frac{e}{m}\frac{V_{dc}}{d}(1 + \gamma \sin \omega t) \qquad (1)$$

where γ is V_{rf}/V_{dc} and V_{dc}, is the dc voltage. The boundary conditions are

Initial: at $\omega t = \phi_0$, $s = 0$ and $\frac{ds}{dt} = v_i$

Final: at $\omega t = \phi_1$, $s = 0$ and $\frac{ds}{dt} = v_f$ (2)

The phase synchronization for one-sided multipacting is

$$\phi_1 = \phi_0 + 2n\pi \qquad (3)$$

where $n = 1, 2, 3, \cdots$. As in the case of two-sided multipacting, we are faced with two types of boundaries of the one-sided multipacting region—one determined from

phase considerations and the other from the minimum arrival energy requirement.

a. *Phase-controlled boundary.* Normalization of Eqs. (1) and (2) according to $x = s/d$ and $\phi = \omega t$ results in phase similar solutions $x(\phi)$ if the quantities $V_{rf}/(fd)^2$, $V_{dc}/(fd)^2$ and v_f/v_i remain invariant. Normalization of velocity v with respect to the initial velocity allows plotting of the solutions in the phase plane of Fig. 1. For comparison, the two-sided multipacting case is also shown. All similar solutions are represented by only one plot. It should also be pointed out that, whereas Vance assumed v_i independent of fd, the similarity principle requires that v_i be proportional to fd.

b. *Minimum electron arrival energy boundary.* When fd, in addition to V_{rf} and V_{dc}, is invariant we have the

solutions similar in kinetic energy as well. When fd varies, the results of the similarity principle do not apply. However, some conclusions for scaling may still be drawn. It is reasonable to assume that corresponding points are assigned similar values of ϕ_0. The v_f/v_i ratio k may also be considered constant, since v_i and v_f are approximately constant. From the phase similarity results, we know that $V_{rf}/(fd)^2$, $V_{dc}/(fd)^2$ and v_f/fd can be expressed in terms of ϕ_0 and k. Therefore, it can be concluded that V_{rf}/fd and V_{dc}/fd are invariant for the minimum arrival energy boundary.

Fig. 1. Phase plane plot in parallel plates geometry: (a) two-sided multipacting; (b) one-sided multipacting
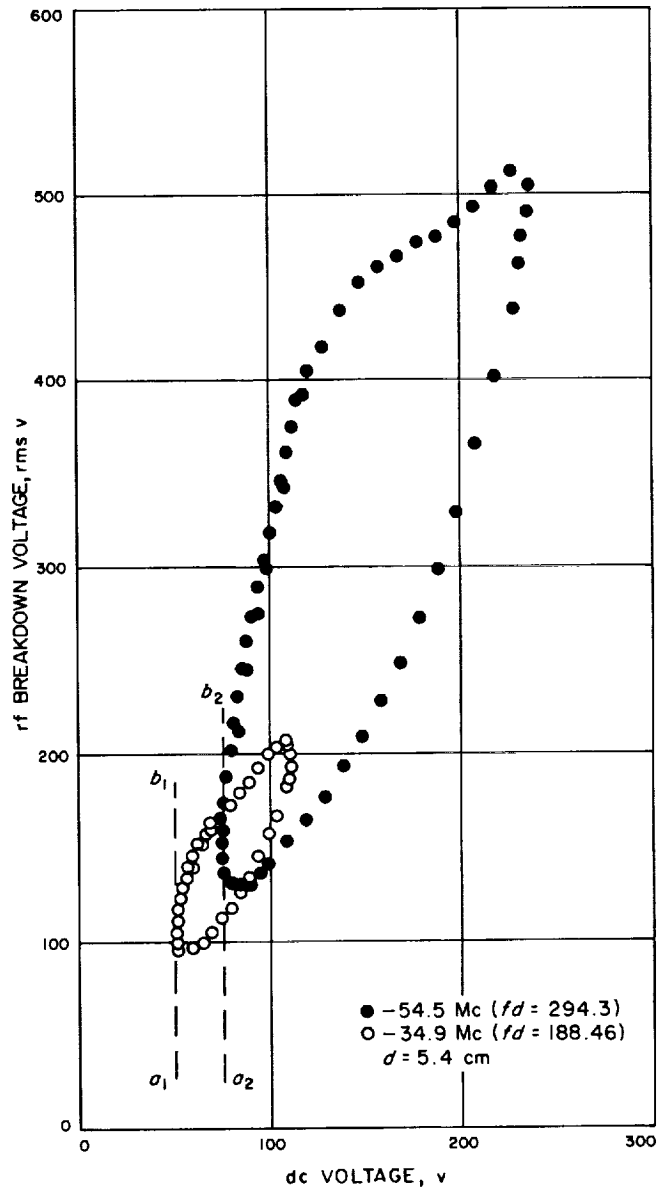
Fig. 2. One-sided multipacting experimental data of Vance (Ref. 1)

Fig. 2 shows Vance's closed experimental one-sided multipacting regions obtained at 34.9 and 54.5 Mc. a_1b_1 and a_2b_2 represent the minimum energy boundaries for 34.9 and 54.5 Mc, respectively. As shown by Vance, the electron gains all its energy from the dc field and the condition for the minimum energy boundary is given by

$$V_{dc} \geq v_{min}\left(\frac{k+1}{k}\right)\omega d\,\frac{m}{e}\frac{1}{2\pi} \qquad (4)$$

where v_{min} is the magnitude of the required minimum electron velocity for secondary emission.

It should be re-emphasized that the assumption of a constant k is satisfactory because both v_i and v_f may be considered approximately constant. Vance used a value of 2.9 for k in Eq. (4), and it should be pointed out that this represents a minimum electron arrival energy of only 33.6 ev. Since this is somewhat low, larger values of k are to be expected.

For dc voltages greater than a_1b_1 for 34.9 Mc or a_2b_2 for 54.5 Mc, the kinetic energy of the electrons is sufficient, and phase considerations determine the boundary. Even though the upper limit of the closed multipacting region is determined from the maximum displacement condition, this can be considered a special phase limit condition so that the phase similarity relations still apply. A check of Vance's data shows that the correspondence of all scaling relations is quite satisfactory.

Zager and Tishin's data are reproduced in Fig. 3. In their data, no attempts were made to isolate the one-sided

and two-sided multipacting cases, and this accounts for the extension of the multipacting regions to zero dc voltages. Good scaling correspondence is also demonstrated.

The experimental data taken by Vance at 54.5 Mc are reproduced by themselves in Fig. 4. AB was determined in the same manner as by Vance—through the condition that the electron return to the emitting electrode at the end of one rf cycle. If Eq. (1) is integrated twice to

Fig. 4. One-sided multipacting experimental
data of Vance (Ref. 1)

Fig. 3. Experimental data of Zager and Tishin (Ref. 2)
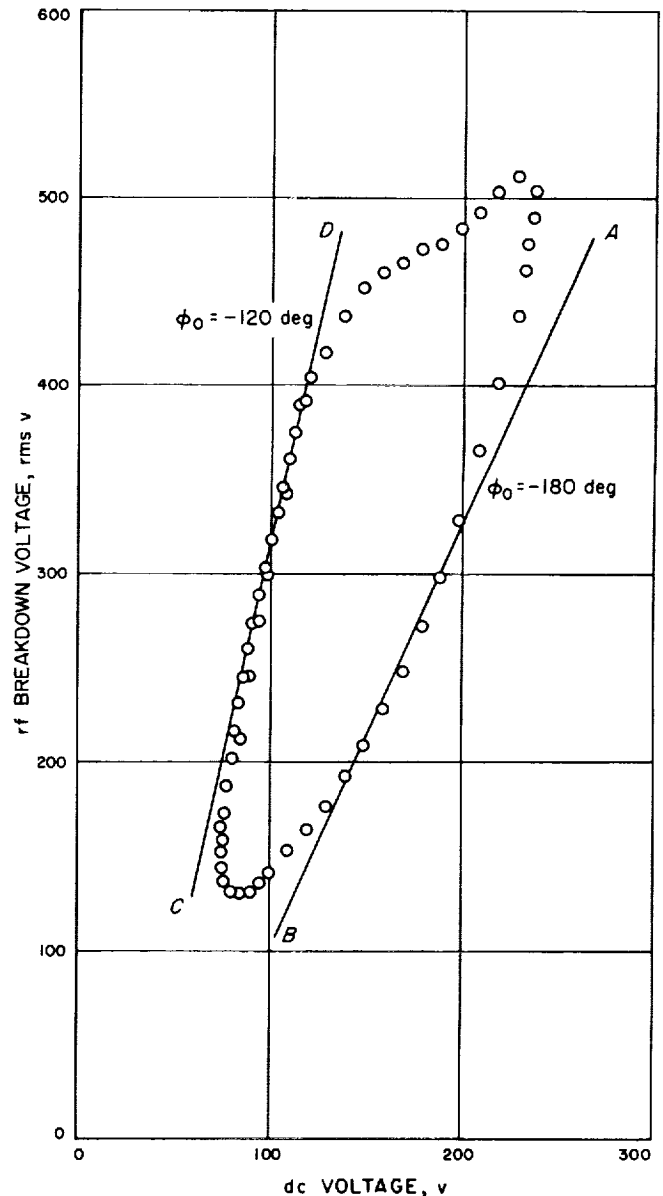
obtain the expression for $s(t)$, and if this is evaluated for $s = 0$ and $\phi = \phi_0 + 2\pi$, we get

$$V_{rf} = \frac{m}{e}\,\omega d \frac{v_i}{\cos \phi_0} - \frac{\pi}{\cos \phi_0} V_{dc} \qquad (5)$$

$\phi_0 = -180$ deg minimizes Eq. (5) and it is with this value substituted in Eq. (5) that AB is drawn. At this point we wish to add CD which corresponds to $\phi_0 = -120$ deg and to point out the good correlation. This should not be too surprising. Since the maximum displacement condition determines the upper limit of the closed multipacting region, while the minimum energy requirement determines the "lower limit," the side boundaries of the multipacting region must be determined from phase relations.

Points satisfying Eq. (5) have the sole property of possessing the same initial velocity in addition to identical emission and arrival phase angles. The assumption of a constant initial velocity does not violate the similarity principle since these similar phase relations hold for the multipacting region of a single physical configuration taken at one particular frequency. In contrast, the similarity principle applies to the case where either one or

both of physical dimension and frequency vary. The emission and arrival phase angles are similar while the initial velocity is proportional to fd. As a result, the trajectory solutions are similar also.

3. Multipacting in the Presence of a Steady Magnetic Field

For multipacting in the presence of a steady magnetic field, the additional phase similarity condition is that B_{dc}/f remain invariant. Of course, if kinetic energy is to be the same, fd must also be invariant.

The particular case of a steady transverse magnetic field was analyzed by Deb and Goswami (Ref. 3) employing the parameter α, the ratio of the cyclotron frequency to the frequency of the applied electric field. For small values of α compared to unity, they found $V_{rf}/(fd)^2 = F(\phi_0, k, \alpha)$ where ϕ_0 is the initial emission phase angle, k is v_f/v_i; and F is functional notation. Of course this is exactly what the similarity principle predicts since α is proportional to B_{dc}/f.

It can thus be seen that the addition of a small steady transverse magnetic field simply shifts the plot of Fig. 5 to higher voltages and lower frequencies. This behavior coincides qualitatively to the experimental results of Huber, Ozaki and Kleider (Ref. 4). When α is unity, one-sided multipacting becomes a possibility, and Brown (Ref. 5) studied this case. However, no appropriate data appear to be available.

4. Multipacting in Coaxial Geometries

Let us consider the coaxial case of Fig. 6 in which the electron is traveling only in the radial direction. It must be emphasized that this is for simplicity only and that
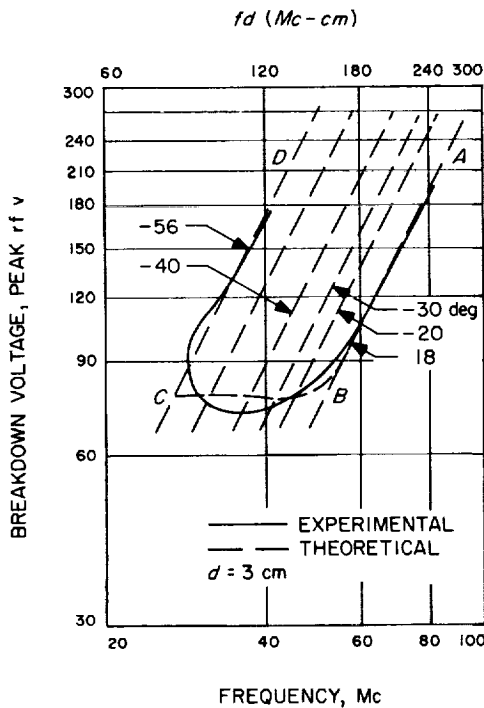
Fig. 5. Two-sided multipacting breakdown region of Hatch and Williams (Ref. 4). $k = 3.0$; energy$_{min} = 60$ ev
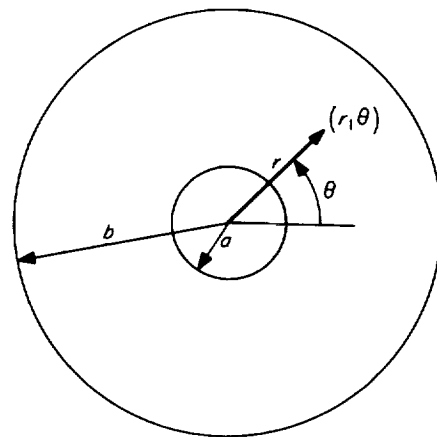
Fig. 6. Coaxial geometry

the end results in the two dimensional case are the same. If both rf and dc voltages are considered, the equation of motion is given by:

$$\frac{d^2r}{dt^2} = \frac{e}{m} \frac{V_{dc}}{r \ln \frac{b}{a}} (1 + \gamma \sin \omega t) \qquad (6)$$

where r is the radial displacement; b and a are the radii of the outer and inner electrodes, respectively, and γ is V_{rf}/V_{dc}. This is a nonlinear differential equation and so far no analytic solution seems available. Although Bol (Ref. 6) studied some electron trajectory plots on the analog computer, no definite information was gained. Under these conditions the similarity principle can be significantly useful.

We normalize as follows:

$$x = \frac{r}{a} \qquad (7)$$

$$\phi = \omega t \qquad (8)$$

Since both one-sided and two-sided multipacting may occur, boundary conditions for both cases must be considered. We take the two-sided case first. Because of the nonuniformity of the fields, the boundary conditions for electrons traveling in opposing directions are different. If the initial velocity for an electron traveling from a to b is v_1, and if the velocity is normalized as $y = dr/dt/v_1$, the boundary conditions are

Initial: at $\phi = \phi_0$, $x = 1$ and $y = 1$

Final: at $\phi = \phi_1$, $x = \dfrac{b}{a}$ and $y = k_1$ (9)

If the initial velocity for an electron traveling from b to a is v_2 and if velocity is normalized as $y = dr/dt/v_2$, the boundary conditions are

Initial: at $\phi = \phi_1$, $x = \dfrac{b}{a}$ and $y = 1$

Final: at $\phi = \phi_2$, $x = 1$ and $y = k_2$ (10)

It is important to note that v_1 and v_2 need not be the same. The phase condition for two-sided multipacting is

$$\phi_2 - \phi_0 = 2n\pi \qquad (11)$$

For similar boundary conditions, k_1 and k_2 must be invariant.

For one-sided multipacting at the outer electrode, the boundary conditions are

Initial: at $\phi = \phi_0$, $x = \dfrac{b}{a}$ and $y = 1$

Final: at $\phi = \phi_3$, $x = \dfrac{b}{a}$ and $y = k_3$ (12)

while for one-sided multipacting at the inner electrode we have

Initial: at $\phi = \phi_0$, $x = 1$ and $y = 1$

Final: at $\phi = \phi_3$, $x = 1$ and $y = k_4$ (13)

In both cases y is velocity normalized to the initial emission velocity. The phase condition for one-sided multipacting is

$$\phi_3 - \phi_0 = 2n\pi \qquad (14)$$

where $n = 1, 2, 3, \cdots$. It is obvious that for similar boundary conditions k_3 or k_4, whichever may be the case, must be invariant.

With only an rf voltage V_{rf} applied, both one-sided and two-sided multipacting may occur. Obviously, one-sided multipacting becomes more probable for larger b/a ratios. In terms of the experimental parameters, phase similar situations occur when $V_{rf}/(fa)^2$ is invariant while similarity in kinetic energy results when fd is also invariant. To determine the minimum energy boundary, some approximate calculations may be carried out via numerical integration, assuming the respective k's in the boundary conditions are constant.

With the addition of a dc voltage V_{dc}, both one-sided and two-sided multipacting may still occur. For phase similarity we have $V_{rf}/(fa)^2$ and $V_{dc}/(fa)^2$ invariant; while for similarity in kinetic energy fa must be invariant. If we follow the same arguments used in the parallel plates case, we can conclude that V_{rf}/fa and V_{dc}/fa are invariant for the minimum energy boundary.

As can be seen from the above results, the similarity principle offers a large amount of multipacting information for the coaxial case. Of considerable interest to

those interested in coaxial transmission lines is the fact that geometrically-similar transmission lines possess a similar characteristic impedance. It should also be pointed out that the similarity principle serves as a valuable tool when electron trajectory solutions are being sought through numerical integration on a computer, since similar solutions need only be solved once.

As previously mentioned, the one-dimensional case was taken for simplicity and the results are identical for the two-dimensional case. It is interesting to derive the equations describing two-dimensional electron motion in the coaxial case.

The coordinates are defined in Fig. 6. The equation of motion is given by

$$m \frac{d}{dt}(v_r \mathbf{i}_r + v_\theta \mathbf{i}_\theta) = e \frac{V_{dc}}{r \ln \frac{b}{a}} (1 + \gamma \sin \omega t) \mathbf{i}_r$$

where

$$v_r = \frac{dr}{dt}, \qquad v_\theta = r \frac{d\theta}{dt} \qquad \text{and } \gamma = \frac{v_{rf}}{V_{dc}} \qquad (15)$$

v_r and v_θ are the respective r and θ components of velocity v while $\mathbf{i}_r$ and $\mathbf{i}_\theta$ are the respective unit vectors in the r and θ directions. Since

$$\frac{d\mathbf{i}_r}{dt} = \frac{\partial \mathbf{i}_r}{\partial r} \frac{dr}{dt} + \frac{\partial \mathbf{i}_r}{\partial \theta} \frac{d\theta}{dt}$$

$$\frac{d\mathbf{i}_\theta}{dt} = \frac{\partial \mathbf{i}_\theta}{\partial r} \frac{dr}{dt} + \frac{\partial \mathbf{i}_\theta}{\partial \theta} \frac{d\theta}{dt} \qquad (16)$$

and since we know that all partials of unit vectors are zero except

$$\frac{\partial \mathbf{i}_r}{\partial \theta} = \mathbf{i}_\theta$$

$$\frac{\partial \mathbf{i}_\theta}{\partial \theta} = -\mathbf{i}_r \qquad (17)$$

we get the following simultaneous differentional equations

$$\frac{d_r^2}{dt^2} - \frac{C^2}{r^3} = \frac{e}{m} \frac{V_{dc}}{r \ln \frac{b}{a}} (1 + \gamma \sin \omega t) \qquad (18)$$

and

$$r^2 \frac{d\theta}{dt} = C \qquad (19)$$

where C is a constant. Eq. (19) is just a statement of the principle of conservation of angular momentum. If the initial velocity does not have a θ component, Eq. (19) is zero and Eq. (18) reduces to Eq. (6).

5. Conclusions

The results of the similarity principle for multipacting discharges can be summarized as follows. First of all, a physical explanation is rendered for the success of the constant v_f/v_i assumption made by previous workers. Secondly, it is shown that the results of the different multipacting cases can be unified under the similarity principle. Thirdly, the application of the similarity principle to geometries other than parallel plates yields much information otherwise unavailable. Obviously all configurations have certainly not been covered in this article. Lastly, the scaling results serve as an indispensible tool when electron trajectories have to be determined through numerical integration on the computer.

References

1. Vance, E. F., "One-Sided Multifactor Discharge Modes," *Journal of Applied Physics*, Vol. 34, p. 3237, 1963.

2. Zager, B. A., and Tishin, V. G., *Žurnal Tehničeskoĭ Fiziki*, Vol. 34, p. 297, 1964. (Translation, "Multipactor Discharge and Ways of Suppressing It," *Soviet Physics—Technical Physics*, Vol. 9, No. 2, p. 234, August 1964.)

References (Cont'd)

3. Deb, S., and Goswami, S. N., "Breakdown in High Frequency Electrodeless Discharge at Low Pressure in the Presence of a Steady Transverse Magnetic Field," *British Journal of Applied Physics*, Vol. 15, p. 1501, 1964.

4. Huber, E. L., Ozaki, H. T., and Kleider, A., "Effect of Magnetic Field on High Frequency Electrodeless Discharge at Low Pressure," *Bulletin of the American Physics Society*, Vol. 2, p. 86, 1957. (See also Hughes Aircraft Co. Research Laboratories Technical Memorandum No. 471, April 1, 1957.

5. Brown, S. C., *Basic Data of Plasma Physics*, John Wiley and Sons, Inc., New York, p. 214, 1959.

6. Bol, K., "The Multipactor Effect in Klystrons," *Institute of Radio Engineers National Convention Record*, Vol. 2, Pt. 3, p. 151, 1954.

XXI. Communications System Research: Information Processing

A. Run-Length Encoding [1]

S. W. Golomb

1. Summary

The literature in statistical communication theory generally contains a significant shift in viewpoint between the discrete and the continuous case. In the latter context, a particular distribution is assumed almost from the outset, and most of the theorems refer to such things as the "white gaussian noisy channel," or other equally specific assumptions. For the discrete case, on the other hand, the results are rarely evaluated in terms of specific distributions. The present remarks are intended as a step in this direction, viz, the explicit form which Huffman coding assumes when applied to the geometric distribution. It would also be appropriate to have explicit answers for the binominal distribution, the Poisson distribution, etc. These results have practical import for the transmission of experiments from a spacecraft to Earth.

2. A Context for the Problem

Suppose one is playing a game of chance. Each game consists of a sequence of favorable events (probability

p), terminated by the first occurrence of an unfavorable event (probability $q = 1 - p$). More specifically, say the game is roulette, and the unfavorable event is the occurrence of 0, which has a probability of $q = \frac{1}{37}$. One is concerned about communicating the game's description back to Earth. This is the situation in many space telemetry situations.

A binary channel is available, but a "fee" is charged for each bit sent. The problem is how to encode the vicissitudes of the wheel so as to place the least strain on the communication channel. It is easily seen that, for the case $p = q = \frac{1}{2}$, the best that can be done is to use 0 and 1 to represent the two possible outcomes. However, the case at hand involves $p >> q$, for which the "direct coding" method is shockingly inefficient.

A suggestion is to encode the run lengths between successive unfavorable events. In general, the probability of a run length of n is $p^n q$, for $n = 0, 1, 2, 3, \cdots$, which is the familiar geometric distribution (Ref. 1).

3. The Encoding Procedure

If the list of possible outcomes were finite, we could list them with their probabilities, and apply Huffman coding (Refs. 2 or 3 p. 77 et seq.). However, with an

infinite list, it is clear that we cannot start at the bottom and work our way up. Fortunately, the fact that the probabilities follow a distribution law furnishes a short cut, as follows:

Let $m = -\log 2/\log p$. (That is, $p^m = \frac{1}{2}$.) The results will be most readily applicable for those p such that m is an integer (viz., $p = 0.5 \cdots, p = 0.707 \cdots, p = 0.794 \cdots, p = 0.849 \cdots, p = 0.873 \cdots$, etc.). The resulting coding scheme is especially simple when m is a power of 2, but any integer m is a favorable case.

If $p^m = \frac{1}{2}$, then a run length $n + m$ is only half as likely as a run length of n. (The respective probabilities are $p^{m+n} q = \frac{1}{2} p^n q$ and $p^n q$.) Thus, we would expect the codeword for run-length $n + m$ to be one bit longer than the codeword for run-length n. This argument, although nonrigorous, leads to the correct conclusion that there should be m codewords of each possible word-length, except for the shortest wordlengths, which are not used at all if $m > 1$, and possibly one transitional wordlength which is used fewer than m times. Knowing this answer, there is a rigorous proof by mathematical induction. The dictionaries for the first several values of m are as shown in Table 1, where $G(n)$ is used to designate $p^n q$.

In general, let k be the smallest positive integer such that $2^k \geq 2m$. Then the corresponding code dictionary contains exactly m words of every word length $\geq k$, as well as $2^{k-1} - m$ words of length $k - 1$. (The simplification which occurs for m, a power of 2, is that the collection of words of length $k - 1$ is empty.) This result is obtained by seeing how much "signal space" is used up by having m words of every length $\geq k$. This consumes

$$\frac{m}{2^k} + \frac{m}{2^{k+1}} + \frac{m}{2^{k+2}} + \cdots = \frac{m}{2^{k-1}}$$

leaving $1 - m/2^{k-1} = (2^{k-1} - m)/2^{k-1}$ unused, which means that $2^{k-1} - m$ words of length $k - 1$ may be adjoined.

4. Further Examples

We will consider the cases $m = 14$ and $m = 16$, to illustrate what happens when m is not a power of 2 and when m is a power of 2, respectively. The dictionaries in these two cases are shown in Table 2. In the case $m = 14$, we find $k = 5$, and $2^{k-1} - m = 2$, so that there are two codewords of length 4, followed by fourteen codewords of lengths five, six, seven, etc. On the other hand, since $m = 16$ is a power of 2, the corresponding dictionary contains exactly sixteen words of every wordlength starting with length 5.

Table 1. Run-length dictionaries for small m

$m = 1$			$m = 2$		
n	$G(n)$	Codeword	n	$G(n)$	Codeword
0	1/2	0	0	0.293	00
1	1/4	10	1	0.207	01
2	1/8	110	2	0.116	100
3	1/16	1110	3	0.104	101
4	1/32	11110	4	0.073	1100
5	1/64	111110	5	0.051	1101
6	1/128	1111110	6	0.036	11100
7	1/256	11111110	7	0.025	11101
8	1/512	111111110	8	0.018	111100
9	1/1024	1111111110	9	0.013	111101
10	1/2048	11111111110	10	0.009	1111100

$m = 3$			$m = 4$		
n	$G(n)$	Codeword	n	$G(n)$	Codeword
0	0.206	00	0	0.151	000
1	0.164	010	1	0.128	001
2	0.130	011	2	0.109	010
3	0.103	100	3	0.092	011
4	0.081	1010	4	0.078	1000
5	0.064	1011	5	0.066	1001
6	0.051	1100	6	0.056	1010
7	0.041	11010	7	0.048	1011
8	0.032	11011	8	0.040	11000
9	0.026	11100	9	0.034	11001
10	0.021	111010	10	0.029	11010

Table 2. Run-length dictionaries for $m = 14$ and $m = 16$

$m = 14$				$m = 16$			
n	Codeword	n	Codeword	n	Codeword	n	Codeword
0	0000	24	101100	0	00000	24	101000
1	0001	25	101101	1	00001	25	101001
2	00100	26	101110	2	00010	26	101010
3	00101	27	101111	3	00011	27	101011
4	00110	28	110000	4	00100	28	101100
5	00111	29	110001	5	00101	29	101101
6	01000	30	1100100	6	00110	30	101110
7	01001	31	1100101	7	00111	31	101111
8	01010	32	1100110	8	01000	32	1100000
9	01011	33	1100111	9	01001	33	1100001
10	01100	34	1101000	10	01010	34	1100010
11	01101	35	1101001	11	01011	35	1100011
12	01110	36	1101010	12	01100	36	11001000
13	01111	37	1101011	13	01101	37	1100101
14	10000	38	1101100	14	01110	38	1100110
15	10001	39	1101101	15	01111	39	1100111
16	100100	40	1101110	16	100000	40	1101000
17	100101	41	1101111	17	100001	41	1101001
18	100110	42	1110000	18	100010	42	1101010
19	100111	43	1110001	19	100011	43	1101011
20	101000	44	11100100	20	100100	44	1101100
21	101001	45	11100101	21	100101	45	1101101
22	101010	46	11100110	22	100110	46	1101110
23	101011	47	11100111	23	100111	47	1101111

In a practical situation, if $m = - \log 2/\log p$ is not an integer, then the best dictionary will oscillate between $[m]$ words of a given length and $[m] + 1$ words of another length. (Here $[m]$ denotes the greatest integer $\leq m$.) For large m, however, there is very little penalty for picking the nearest integer when designing the code. Very often, the underlying probabilities are not known accurately enough to justify picking a nonintegral value of m. (For example, saying $p = 0.95$ on the basis of statistical evidence may involve as large a round-off error as saying $m = 14$.) For the "player" of Section 2, the approximation $m = 25$ corresponds closely to $q = 1/37$.

5. Decoding

The dictionaries in Table 2 exhibit striking patterns which suggest that a rather simple decoding procedure might be employed. For the case $m = 16$, the following rule for decoding is adequate:

Start at the beginning (left end) of the word, and count the number of 1's preceding the first 0. Let this number be $A \geq 0$. Then the word consists of $A + 5$ bits. Let the last 5 bits be regarded as the ordinary binary representation of the integer $R, 0 \leq R \leq 15$. Then the correct decoding of the word is $16A + R$. This simple decoding reveals an equally simple method of encoding. To encode the number N, we divide N by 16 to get $N = 16A + R$, and write A 1's followed by the 5-bit binary representation of R.

The case $m = 14$ is only slightly more complicated. Suppose a word starts in A 1's, and the next three bits are not all 0's. Then we consider the word to consist of $A + 5$ bits altogether. Let the last 5 bits be the binary representation of the integer R. Then the correct decoding of the codeword is $14A + R - 2$. On the other hand, if the initial A 1's are followed by three or more 0's, we regard the codeword as consisting of a total of $A + 4$ bits. Letting the last 4 bits be the binary representation of an integer R', the correct decoding in this case is $14A + R'$. This procedure also can be inverted to describe direct encoding from ordinary numbers to codewords.

6. Implementation

Observe that although run length coding is a big improvement over no coding at all, it is less than 100% efficient for the mission at hand. In fact, there is a method (the "M.I.T." method) which is 100% efficient. However, a hasty briefing on this method convinces us that it is unimplementable, because it requires infinite computing capability. The run length system is to be employed after

all. As it turns out, however, the experimenter has bribed the "croupier," and the "Unfavorable Case" occurs only half as often as expected. Fortunately, the coding procedure is such that the cost of communicating has also decreased as a result!

It is appropriate to mention that there really is a method, invented by Elias and Shannon (see Ref. 3, p. 61), which is 100% efficient for communicating events from a $p{:}q$ distribution. Moreover, the assertion that "infinite computing capability" is required is a gross overstatement. Nevertheless, we quite possibly made the correct practical decision. We shall leave it to the reader to judge.

B. Detection of Mixtures of Many Data Sources
I. Eisenberger

1. Summary

In this article we consider the distribution of a mixture of m normal distributions $(m = 3, 4, 5, \cdots)$ and find conditions for which the distribution is polymodal. Thus, inferences on possible data sources can be made from received spacecraft telemetry.

2. Principal Results

If a random variable is such that, with probability $p_i, 0 < p_i < 1$, it comes from a normal distribution with density $N(\mu_i, \sigma_i)$, $i = 1, 2, \cdots, m$,

$$\sum_{i=1}^{m} p_i = 1$$

its density function is given by

$$f(x; p_i) = \sum_{i=1}^{m} p_i N(\mu_i, \sigma_i)$$

With respect to modality, the following results are derived:

(1a) If $\mu_i = \mu$, $i = 1, 2, \cdots, m$, $f(x;p_i)$ is unimodal for all sets $\{p_i\}$.

(2a) Assuming without loss of generality that $\mu_{i+1} \geq \mu_i$, for $i = 1, 2, \cdots, m - 1$, and that there is at least one value of i such that $\mu_{i+1} > \mu_i$, a sufficient condition that $f(x;p_i)$ be unimodal for all sets $\{p_i\}$ is

that

$$(\mu_m - \mu_1)^2 < \frac{27\sigma_1^2\sigma_m^2}{4(\sigma_1^2 + \sigma_m^2)}$$

and

$$\max\left[\sigma_1^2(\mu_m - \mu_1)(\mu_m - \mu_1)^2, \frac{4\sigma_i^2}{27}(\mu_i - \mu_1)^3\right]$$

$$< \sigma_1^2\sigma_i^2(\mu_i - \mu_1) - \frac{4\sigma_i^2}{27}(\mu_i - \mu_1)^3$$

for $i = 2, 3, \cdots, m - 1$

(3a) A sufficient condition that there exist sets $\{p_i\}$ for which $f(x;p_i)$ is polymodal is that

$$(\mu_m - \mu_1)^2 > \frac{8\sigma_1^2\sigma_m^2}{\sigma_1^2 + \sigma_m^2}$$

(4a) For every set of values of μ_i and σ_i, sets $\{p_i\}$ exist for which $f(x; p_i)$ is unimodal.

A distribution is called parapolymodal if it is unimodal, but its density function has more than 2 points of inflection. With respect to paramodality, the following results are derived:

(1b) If $\mu_i = \mu$, $i = 1, 2, \cdots, m$, and assuming without loss of generality that $\sigma_{i+1} \geqq \sigma_i$, $i = 1, 2, \cdots$, $m - 1$ and that there are at least two values of i such that $\sigma_{i+1} > \sigma_i$, a necessary and sufficient condition that there exist sets $\{p_i\}$ for which $f(x;p_i)$ is parapolymodal is that

$$\sigma_1 < \sigma_m(3^{1/2} - 2^{1/2})$$

(2b) If the means are not all equal, but $f(x;p_i)$ is unimodal for all sets $\{p_i\}$, the condition given in (1b) is sufficient that there exist sets $\{p_i\}$ for which $f(x;p_i)$ is parapolymodal.

(3b) For fixed values of μ_i and σ_i, if there exist sets $\{p_i\}$ such that $f(x;p_i)$ is polymodal, then there exist sets $\{p_i\}$ for which $f(x;p_i)$ is parapolymodal.

(4b) For every set of values of μ_i and σ_i, sets $\{p_i\}$ exist for which $f(x;p_i)$ is neither polymodal nor parapolymodal.

3. Introduction

With respect to spacecraft experiments, it is often reasonable for the experimenter to assume that a distribution, although not known exactly, nevertheless should exhibit certain general characteristics. When these characteristics do not appear in the received telemetry, the experimenter may well suspect that the population from which he is sampling is not the one he thought it was. For example, the assumption that the distribution of particle energies is unimodal is a reasonable one when the particles all come from the same source. However, since the distribution of particle energies of a mixture of two or more kinds of particles is very likely to be polymodal, polymodality suggests strongly the presence of more than one kind of particle in the population. Consequently, the modal properties of a mixture of two normal distributions were previously investigated (SPS 37-18, Vol. IV, pp. 178–180 and Ref. 4).

As previously noted (SPS 37-35, Vol. IV, p. 331), however, it may not be necessary to demand polymodality before one should suspect a mixture of two or more populations. The number of points of inflection that a density function possesses may often be a sufficient criterion. A positive, continuous density function, ranging from $+\infty$ to $-\infty$, has at least two points of inflection. If it has only two, the distribution is unimodal. Since the converse is not true, it may be possible to detect a mixture of two or more distributions when the number of points of inflection exceeds two, even when the distribution is unimodal. Thus, it also becomes important to consider ways in which this type of distribution can arise. The analysis in SPS 37-35, Vol. IV, p. 331 (as well as in SPS 37-18, Vol. IV, pp. 178–180 and Ref. 4) was restricted to the case of a mixture of two normal distributions. In this article, however, we take up the general case of a mixture of $m(> 2)$ normal distributions.

4. Theoretical Discussion of Results (1a)–(4a)

a. Proof of (1a). Putting $f'(x; p_i) = 0$, one obtains

$$f'(x; p_i) = -\sum_{i=1}^m \frac{p_i(x - \mu_i)}{\sigma_i^3(2\pi)^{1/2}} \exp{-\frac{\frac{1}{2}(x - \mu_i)^2}{\sigma_i^2}} = 0 \quad (1)$$

Result (1a) follows immediately from Eq. (1), since if $\mu_i = \mu$ for all i, the only root is $x = \mu$.

b. Proof of (2a). We can now assume that $\mu_{i+1} \geqq \mu_i$, $i = 1, 2, \cdots m - 1$, and there is at least one value of i such that $\mu_{i+1} > \mu_i$. Since $x = \mu_1$ is not a root of $f'(x;p_i) = 0$, one can divide Eq. (1) by

$$\frac{(x - \mu_1)p_2}{\sigma^3(2\pi)^{1/2}} \exp\left[-\frac{\frac{1}{2}(x - \mu_1)^2}{\sigma_1^2}\right]$$

After rearranging, one obtains

$$g(x) = \frac{\mu_2 - x}{x - \mu_1} h_1(x) + \sum_{i=3}^{m} \frac{p_i (\mu_i - x)\sigma_2^3}{p_2 (x - \mu_1)\sigma_1^3} h_{i-1}(x) = \frac{p_1\sigma_2^3}{p_2\sigma_1^3}$$

where

$$h_i(x) = \exp\left[-\frac{\tfrac{1}{2}(x - \mu_{i+1})^2}{\sigma_{i+1}^2} + \frac{(x - \mu_1)^2}{\sigma_1^2} \right]$$

$$i = 1, 2, \cdots, m-1$$

Now let the ratios p_i/p_2, $i = 3, 4, \cdots, m$, be fixed but arbitrary; that is, let $p_i/p_2 = k_i$, $i = 3, 4, \cdots, m$. For any value of p_1, $0 < p_1 < 1$, one has

$$p_2 = \frac{1 - p_1}{1 + \sum\limits_{i=3}^{m} k_i}, \qquad \frac{p_1\sigma_2^3}{p_2\sigma_1^3} = \frac{p_1\sigma_2^3\left(1 + \sum\limits_{i=3}^{m} k_i\right)}{(1 - p_1)\sigma_1^3}$$

Thus, it is seen that as p_1 varies from zero to one,

$$\frac{p_1\sigma_2^3}{p_2\sigma_1^3} > 0$$

and this term takes on all finite positive values exactly once on the interval $0 < p_1 < 1$ for all fixed values of σ_1 and σ_2. Moreover,

$$p_i = \frac{k_i(1 - p_1)}{1 + \sum\limits_{i=3}^{m} k_i}$$

$i = 3, 4, \cdots, m$, so that each value of p_1 determines a unique set $\{p_i\}$ for each set of fixed values $k_3, k_4, \cdots, k_m$. Thus, for any fixed set of k_i, each value of x for which $g(x) > 0$ is a root of the equation $g(x) = p_1\sigma_2^3/p_2\sigma_1^3$ for a unique set $\{p_i\}$, and, hence, is a root of $f'(x; p_i) = 0$ for the same set $\{p_i\}$. For $x \gtrsim \mu_m$ and $x < \mu_1$, $g(x) < 0$, so that one is interested only in values of x on the interval $\mu_1 < x < \mu_m$. On this interval, $g(x) \to \infty$ as $x \to \mu_1$, and for some x_0, (depending on the fixed values of the set $\{k_i\}$) such that $\mu_2 < x_0 < \mu_m$, $g(x_0) = 0$. Therefore, since $g(x)$ is continuous on $\mu_1 < x < x_0$, $g(x)$ takes on all positive finite values at least once on this interval. Moreover, if $g(x)$ is monotone decreasing on this interval for all possible sets of k_i, all positive values will be attained exactly once, so that $f(x; p_i)$ will have a single maximum and, hence, will be unimodal for all possible sets $\{p_i\}$. Since decreasing monotonicity is implied by $g'(x) < 0$ on $\mu_1 < x < x_0$, and x_0 can be made arbitrarily close to μ_m by a proper choice of the ratios $\{k_i\}$, conditions for which this relation is satisfied will now be investigated for $\mu_1 < x < \mu_m$.

For $\mu_1 < x < \mu_m$,

$$g'(x) = \frac{h_1(x)}{\sigma_1^2\sigma_2^2 (x - \mu_1)^2} \left[\sigma_1^2 (x - \mu_1)(\mu_2 - x)^2 + \sigma_2^2 (x - \mu_1)^2 (\mu_2 - x) - \sigma_1^2\sigma_2^2 (\mu_2 - \mu_1) \right]$$

$$+ \sum_{i=3}^{m} \frac{k_i h_{i-1}(x)\sigma_2^3}{\sigma_1^2\sigma_i^5 (x - \mu_1)^2} \left[\sigma_1^2 (x - \mu_1)(\mu_i - x)^2 + \sigma_i^2 (x - \mu_1)^2 (\mu_i - x) - \sigma_1^2\sigma_i^2 (\mu_i - \mu_1) \right]$$

$$< \frac{h_1(x)}{\sigma_1^2\sigma_2^2 (x - \mu_1)^2} \left\{ \max\left[\sigma_1^2 (\mu_m - \mu_1)(\mu_m - \mu_2)^2, \frac{4\sigma_1^2}{27} (\mu_2 - \mu_1)^3 \right] + \frac{4\sigma_2^2 (\mu_2 - \mu_1)^3}{27} - \sigma_1^2\sigma_2^2 (\mu_2 - \mu_1) \right\}$$

$$+ \sum_{i=3}^{m-1} \frac{k_i h_{i-1}(x)\sigma_2^3}{\sigma_1^2\sigma_i^5 (x - \mu_1)^2} \left\{ \max\left[\sigma_1^2 (\mu_m - \mu_1)(\mu_m - \mu_i)^2, \frac{4\sigma_1^2}{27} (\mu_i - \mu_1)^3 \right] + \frac{4\sigma_i^2}{27} (\mu_i - \mu_1)^3 - \sigma_1^2\sigma_i^2 (\mu_i - \mu_1) \right\}$$

$$+ \frac{k_m h_{m-1}(x)\sigma_2^3}{\sigma_1^2\sigma_m^5 (x - \mu_1)^2} \left[\frac{4}{27} (\sigma_1^2 + \sigma_m^2)(\mu_m - \mu_1)^3 - \sigma_1^2\sigma_m^2 (\mu_m - \mu_1) \right] < 0$$

if

$$(\mu_m - \mu_1)^2 < \frac{27\sigma_1^2\sigma_m^2}{4(\sigma_1^2 + \sigma_m^2)},$$

and

$$\max\left[\sigma_1^2(\mu_m - \mu_1)(\mu_m - \mu_1)^2, \frac{4\sigma_1^2}{27}(\mu_i - \mu_1)^3\right] < \sigma_1^2\sigma_i^2(\mu_i - \mu_1) - \frac{4\sigma_1^2}{27}(\mu_i - \mu_1)^3$$

for $i = 2, 3, \cdots, m - 1$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (2)

Thus, for values of μ_i and σ_i satisfying Ineq. (2), $g(x)$ decreases monotonically on $\mu_i < x < \mu_m$ for every set $\{k_i\}$. Then for each set $\{p_i\}$, there exists only one value of x for which $f'(x; p_i) = 0$. This must be a maximum since $f(x; p_i) \to 0$ as $x \to \pm\infty$. This proves (2a).

c. Proof of (3a). However, for $x = \mu_1 + \mu_m/2$

$$g'\left(\frac{\mu_1 + \mu_m}{2}\right) = \frac{h_1\left(\frac{\mu_1 + \mu_m}{2}\right)}{\sigma_1^2\sigma_2^2\left(\frac{\mu_1 + \mu_m}{2}\right)^2}\left[\sigma_1^2\left(\frac{\mu_m - \mu_1}{2}\right)\left(\frac{2\mu_2 - \mu_1 - \mu_m}{2}\right)^2 + \sigma_2^2\left(\frac{\mu_m - \mu_1}{2}\right)^2\left(\frac{2\mu_2 - \mu_1 - \mu_m}{2}\right) - \sigma_1^2\sigma_2^2(\mu_2 - \mu_1)\right]$$

$$+ \sum_{i=3}^{m-1}\frac{k_ih_{i-1}\left(\frac{\mu_1 + \mu_m}{2}\right)\sigma_2^3}{\sigma_1^2\sigma_i^5\left(\frac{\mu_m - \mu_1}{2}\right)^2}\left[\sigma_i^2\left(\frac{\mu_m - \mu_1}{2}\right)\left(\frac{2\mu_i - \mu_1 - \mu_m}{2}\right)^2\right.$$

$$\left. + \sigma_i^3\left(\frac{\mu_m - \mu_1}{2}\right)^2\left(\frac{2\mu_i - \mu_1 - \mu_m}{2}\right) - \sigma_1^2\sigma_i^2(\mu_i - \mu_1)\right]$$

$$+ \frac{k_mh_{m-1}\left(\frac{\mu_1 + \mu_m}{2}\right)\sigma_2^3}{\sigma_1^2\sigma_m^5\left(\frac{\mu_m - \mu_1}{2}\right)^2}\left[\frac{\sigma_1^2 + \sigma_m^2}{8}(\mu_m - \mu_1)^3 - \sigma_1^2\sigma_2^2(\mu_m - \mu_1)\right]$$

Now suppose that

$$\frac{\sigma_1^2 + \sigma_m^2}{8}(\mu_m - \mu_1)^3 - \sigma_1^2\sigma_2^2(\mu_m - \mu_1) = \varepsilon > 0$$

and that k_m is sufficiently large so that $(\mu_i + \mu_m)/2 < x_0$ and that the sum of the first $m - 1$ terms of $g'[(\mu_1 + \mu_m)/2]$ is less in absolute value than

$$\frac{k_mh_{m-1}\left(\frac{\mu_1 + \mu_m}{2}\right)\sigma_2^3}{\sigma_1^2\sigma_m^5\left(\frac{\mu_m - \mu_1}{2}\right)^2}\,\varepsilon.$$

Then $g'[(\mu_1 + \mu_m)/2] > 0$. This means that, since the k_i can be chosen arbitrarily, for values of μ_i and σ_i satisfying the inequality

$$(\mu_m - \mu_1)^2 > \frac{8\sigma_1^2\sigma_m^2}{\sigma_1^2 + \sigma_m^2}$$

there exist sets $\{k_i\}$ such that for an interval about $x = (\mu_1 + \mu_m)/2$, $g(x)$ is increasing. For each x_2 in this interval there is an $x_1 < x_2$, an $x_3 > x_2$ and a set $\{p_i\}$, such that

$$g(x_1) = g(x_2) = g(x_3) = \frac{p_1 \sigma_2^3}{p_2 \sigma_1^3}$$

Consequently, for this set $\{p_i\}$, $f'(x_1; p_i) = f'(x_2; p_i) = f'(x_3; p_i) = 0$.

Now suppose that for two of the values x_1, x_2, x_3, say x_1 and x_3, $f(x_1; p_i)$ and $f(x_3; p_i)$ are not extreme values of $f(x; p_i)$. Then $f''(x_1; p_i) = f''(x_3; p_i) = 0$ and $f''(x_2; p_i) < 0$ [assuming $f(x; p)$ unimodal]. From $f'(x_1; p_i)$, one always has

$$1 + \sum_{i=2}^{m} \frac{p_i (x_1 - \mu_i) \sigma_1^3}{p_1 (x_1 - \mu_1) \sigma_i^3} h_{i-1}(x_1)$$

$$= 1 + \frac{(1 - p_1)(x_1 - \mu_2) \sigma_1^3}{p_1 \left(1 + \sum_{i=3}^{m} k_i\right)(x_1 - \mu_1) \sigma_2^3} h_1(x_1)$$

$$+ \sum_{j=3}^{m} \frac{k_j (1 - p_1) \sigma_1^3 (x_1 - \mu_j)}{p_1 \left(1 + \sum_{i=3}^{m} k_i\right)(x_1 - \mu_1) \sigma_j^3} h_{j-1}(x_1) = 0 \tag{3}$$

From $f''(x_1; p_i) = 0$, one also has

$$1 + \sum_{i=2}^{m} \frac{p_i \sigma_1^5 [(x_1 - \mu_i)^2 - \sigma_i^2]}{p_1 \sigma_i^5 [(x_1 - \mu_1)^2 - \sigma_1^2]} h_{i-1}(x_1)$$

$$= 1 + \frac{(1 - p_1) \sigma_1^5 [(x_1 - \mu_2)^2 - \sigma_2^2]}{\sigma_2^5 p_1 \left(1 + \sum_{i=3}^{m} k_i\right)[(x_1 - \mu_1)^2 - \sigma_1^2]} h_1(x_1)$$

$$+ \sum_{j=3}^{m} \frac{k_j (1 - p_1) \sigma_1^5 [(x_1 - \mu_j)^2 - \sigma_j^2]}{p_1 \left(1 + \sum_{i=3}^{m} \sigma_j^5\right)[(x_1 - \mu_1)^2 - \sigma_1^2]} h_{j-1}(x_1)$$

$$= 0 \tag{4}$$

Equating Eqs. (3) and (4) one finds that a necessary condition for $f'(x_1; p_i) = f''(x_1; p_i) = 0$ is that

$$\left[\frac{\sigma_1^2 [(x_1 - \mu_2)^2 - \sigma_2^2]}{\sigma_2^5 [(x_1 - \mu_1)^2 - \sigma_1^2]} - \frac{(x_1 - \mu_2)}{(x_1 - \mu_1) \sigma_2^3}\right] h_1(x_1)$$

$$+ \sum_{j=3}^{m} \left\{\frac{k_j \sigma_1^2 [(x_1 - \mu_j)^2 - \sigma_j^2]}{\sigma_j [(x_1 - \mu_1)^2 - \sigma_j^2]} - \frac{k_j (x_1 - \mu_j)}{(x_1 - \mu_1) \sigma_j^3}\right\} h_{j-1}(x_1) = 0 \tag{5}$$

Since the coefficients of the $h_i(x_1)$ in Eq. (5) are rational functions of x_1, independent of sets $\{p_i\}$, there exists an interval about x_1 and sets $\{p_i\}$ (depending on x) such that for all values x_α in this interval $f'(x_\alpha; p_i) = 0 \neq f''(x_\alpha; p_i)$. And, if these values of x_α are chosen sufficiently close to x_1, for the corresponding values x_β about x_2 and x_λ about x_3 such that $f'(x_\beta; p_i) = f'(x_\lambda; p_i) = 0$, one has, due to continuity, $f''(x_\beta; p_i) \neq 0$, $f''(x_\lambda; p_i) \neq 0$. Thus, for these sets $\{p_i\}$, $f(x; p_i)$ has at least three extreme values and, hence, is polymodal. This proves (3a).

One can also arrive at Result (3a) intuitively. If one chooses any $m - 2$ of the m values of p_i sufficiently small, the contribution to $f(x; p_i)$ of the normal distributions associated with these p_i can be made so negligible that $f(x; p_i)$ will be essentially a mixture of two distributions. One can then apply a result derived in SPS 37-18, Vol. IV, pp. 178–180 and in Ref. 4, which states that for a mixture of two normal distributions with densities $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$, there exist values of p for which $f(x; p)$ is bimodal if $(\mu_2 - \mu_1)^2 > 8\sigma_1^2\sigma_2^2/(\sigma_1^2 + \sigma_2^2)$. In fact, Result (3a) can be restated more generally, as follows:

(3a'). A sufficient condition that there exist sets $\{p_i\}$ for which $f(x; p_i)$ is polymodal is that for some i, j, one has

$$(\mu_i - \mu_j)^2 > \frac{8\sigma_i^2\sigma_j^2}{\sigma_i^2 + \sigma_j^2}$$

The proof of (3a') would proceed along the same lines as that of (3a) and will be omitted.

d. Proof of (4a). Since $g(x)$ is finite and never zero on $\mu_1 < x < x_0 < \mu_m$ and $g(x_0) = 0$, $g(x) \to \infty$ as $x \to \mu_1$, regardless of the behavior of $g(x)$ in the interior of the interval $\mu_1 < x < x_0$; there is a half interval of values of x about μ_1 and a half interval of values of x about x_0, both half intervals within $\mu_1 < x < x_0$, such that for each value of x within these half intervals $g(x)$ is monotone decreasing. Hence, for the sets $\{p_i\}$ associated with each such value of x, there is only one root of $f'(x; p_i) = 0$, and $f(x; p_i)$ is thus unimodal. This proves (4a).

5. Theoretical Discussion of Results (1b)–(4b)

a. Proof of (1b). Since a point of inflection of $f(x; p_i)$ corresponds to an extreme value of $f'(x; p_i)$, the zeros of $f''(x; p_i)$ will be investigated. Putting $f''(x; p_i) = 0$, one obtains

$$f''(x; p_i) = \sum_{i=1}^{m} \frac{p_i}{\sigma_i^5 (2\pi)^{1/2}} [(x - \mu_i)^2 - \sigma_i^2] \exp\left[\frac{-\frac{1}{2}(x - \mu_i)^2}{\sigma_i^2}\right] \tag{6}$$

Now let $\mu_i = \mu$, $i = 1, 2, \cdots, m$. If $\sigma_i = \sigma$ for all i, $f(x; p_i)$ is a normal density function for all sets $\{p_i\}$ and is neither polymodal nor parapolymodal. Hence, we can assume that $\sigma_{i+1} \geqslant \sigma_i$, $i = 1, 2, \cdots, m$ and that there are at least two values of i such that $\sigma_{i+1} > \sigma_i$. Since $x = \mu \pm \sigma_1$ is not a root of $f''(x; p_i) = 0$; one can divide Eq. (6) by

$$\frac{p_2\left[(x - \mu)^2 - \sigma_1^2\right]}{\sigma_2^5 (2\pi)^{1/2}} \exp\left[\frac{-\frac{1}{2}(x - \mu)^2}{\sigma_1^2}\right]$$

After rearranging, one obtains

$$g_1(y) = \frac{y^2 - \sigma_2^2}{y^2 - \sigma_1^2} h_1(y) + \sum_{i=3}^{m} \frac{\sigma_2^5 p_i (y^2 - \sigma_i)^2}{\sigma_i^5 p_2 (y^2 - \sigma_1^2)} h_{i-1}(y)$$

$$= -\frac{\sigma_2^5 p_1}{\sigma_1^5 p_2} \tag{7}$$

where

$$y = x - \mu$$

and

$$h_i(y) = \exp\left[\frac{1}{2} y^2 \left(\frac{1}{\sigma_i^2} - \frac{1}{\sigma_{i+1}^2}\right)\right], \quad i = 1, 2, \cdots, m - 1$$

As in Section 3, let $p_i/p_2 = k_i$ be fixed but arbitrary for $i = 3, 4, \cdots, m$. Then one can conclude that for any fixed set $\{k_i\}$ each value of y for which $g_1(y) < 0$ is a root of the equation

$$g_1(y) = \frac{-\sigma_2^5 p_1}{\sigma_1^5 p_2}$$

for a unique set $\{p_i\}$ is a root of $f''(x; p_i) = 0$ for the same set $\{p_i\}$. For $y \geqslant \sigma_m$, $y \leqslant -\sigma_m$, and $-\sigma_1 < y < \sigma_1$, $g_1(y) > 0$, so that one is interested only in values of y on the intervals $\sigma_1 < y < \sigma_m$ and $-\sigma_m < y < -\sigma_1$. For $\sigma_1 < y < \sigma_m$, $g_1(y) \to -\infty$ as $y \to \sigma_1^+$ and for some y_0 (depending on the fixed values of the set $\{k_i\}$) such that $\sigma_2 < y_0 < \sigma_m$, $g(y_0) = 0$. Therefore, $g_1(y)$ takes on all negative finite values on $\sigma_1 < y < y_0$, and if $g_1(y)$ is monotone increasing on this interval for all possible sets $\{k_i\}$, all negative values will be attained exactly once. Consequently, $f'(x; p_i)$ will have only one extreme value for all sets $\{p_i\}$. $f(x; p_i)$ will then have only one point of inflection on the interval $\sigma_1 < y < \sigma_m$. Similarly, for $-\sigma_m < y < \sigma_1$, $g_1(y) \to -\infty$ as $y \to -\sigma_1$ and $g_1(-y_0) = 0$. If $g_1(y)$ is monotone decreasing on $-y_0 < y < -\sigma_1$, $f(x; p_i)$ will have only one point of inflection on $-\sigma_m < y < -\sigma_1$ for all possible sets $\{p_i\}$. Thus, if $g_1(y)$ is monotone increasing on $\sigma_1 < y < y_0$ and monotone decreasing

$-y_0 < y < -\sigma_1$ for all possible sets $\{k_i\}$, $f(x; p_i)$ will have only two points of inflection for all possible sets $\{p_i\}$ and will be neither polymodal nor parapolymodal. Since increasing monotonicity is implied by $g'(y) > 0$ on $\sigma_1 < y < y_0$, and y_0 can be made arbitrarily close to σ_m by a proper choice of the ratios k_i, conditions for which this relation is satisfied will now be investigated for $\sigma_1 < x < \sigma_m$. Due to the symmetry of $g_1(y)$ about $y = 0$, the results obtained for $\sigma_1 < y < \sigma_m$ will also hold for $-\sigma_m < y < -\sigma_1$.

For $\sigma_1 < y < \sigma_m$,

$g_1'(y)$

$$= \frac{y h_1(y) (\sigma_2^2 - \sigma_1^2)}{\sigma_1^2 \sigma_2^2 (y^2 - \sigma_1^2)^2} \left[2\sigma_1^2 \sigma_2^2 + (y^2 - \sigma_2^2)(y^2 - \sigma_1^2)\right]$$

$$+ \sum_{i=3}^{m} \frac{k_i \sigma_2^5 y h_{i-1}(y)(\sigma_i^2 - \sigma_1^2)}{\sigma_1^2 \sigma_i^7 (y^2 - \sigma_1^2)^2} \left[2\sigma_1^2 \sigma_i^2 + (y^2 - \sigma_i^2)(y^2 - \sigma_1^2)\right]$$

$$> \frac{y h_1(y) (\sigma_2^2 - \sigma_1^2)}{\sigma_1^2 \sigma_2^2 (y^2 - \sigma_1^2)^2} \left[2\sigma_1^2 \sigma_2^2 - \frac{(\sigma_2^2 - \sigma_1^2)^2}{4}\right]$$

$$+ \sum_{i=3}^{m} \frac{k_i y \sigma_2^5 h_{i-1}(y)(\sigma_i^2 - \sigma_1^2)}{\sigma_1^2 \sigma_i^7 (y^2 - \sigma_1^2)^2} \left[2\sigma_1^2 \sigma_i^2 - \frac{(\sigma_i^2 - \sigma_1^2)^2}{4}\right] > 0$$

if

$$\sigma_1^2 > \frac{(\sigma_i^2 - \sigma_1^2)^2}{8\sigma_i^2} \quad \text{for } i = 2, 3, \cdots, m \tag{8}$$

However, since $\sigma_m \geqslant \sigma_i$ for $i = 2, 3, \cdots, m - 1$, it follows that

$$\frac{\sigma_m^2 - \sigma_1^2}{\sigma_m} > \frac{\sigma_i^2 - \sigma_1^2}{\sigma_i}, \quad i = 2, 3, \cdots, m - 1$$

so that if $\sigma_1^2 > (\sigma_m^2 - \sigma_1^2)^2/8\sigma_m^2$, then $\sigma_1^2 > (\sigma_i^2 - \sigma_1^2)^2/8\sigma_i^2$, $i = 2, 3, \cdots, m$. Hence, condition 8 can be stated as $g'(y) > 0$, if $\sigma_1^2 > (\sigma_m^2 - \sigma_1^2)^2/8\sigma_m^2$, which is equivalent to

$$g'(y) > 0 \quad \text{if} \quad \sigma_1 > \sigma_m (3^{1/2} - 2^{1/2}) \tag{9}$$

Thus, for values of σ_1 and σ_m satisfying Ineq. (9), $g_1(y)$ increases monotonically on $\sigma_1 < y < y_0$ and decreases monotonically on $-y_0 < y < -\sigma_1$ for all possible sets $\{p_i\}$. Hence, $f(x; p_i)$ has only two points of inflection for all $\{p_i\}$ and is not parapolymodal.

For y^2 equal to $(\sigma_1^2 + \sigma_m^2)/2$, we have

$$g_1' \sqrt{\frac{\sigma_1^2 + \sigma_m^2}{2}} = \frac{\sqrt{\dfrac{\sigma_1^2 + \sigma_m^2}{2}} h_1 \sqrt{\dfrac{\sigma_1^2 + \sigma_m^2}{2}} (\sigma_2^2 - \sigma_1^2)}{\sigma_1^2 \sigma_2^2 \left(\dfrac{\sigma_m^2 - \sigma_1^2}{2}\right)^2} \left[2\sigma_1^2\sigma_2^2 - \frac{(\sigma_m^2 - \sigma_1^2)(2\sigma_2^2 - \sigma_1^2 - \sigma_m^2)}{4}\right]$$

$$+ \sum_{i=3}^{m-1} \frac{k_i \sigma_2^5 \sqrt{\dfrac{\sigma_1^2 + \sigma_m^2}{2}} h_{i-1} \sqrt{\dfrac{\sigma_1^2 + \sigma_m^2}{2}} (\sigma_m^2 - \sigma_i^2)}{\sigma_1^2 \sigma_i^7 \left(\dfrac{\sigma_m^2 - \sigma_1^2}{2}\right)^2} \left[2\sigma_1^2\sigma_i^2 - \frac{(\sigma_m^2 - \sigma_1^2)(2\sigma_i^2 - \sigma_1^2 - \sigma_m^2)}{4}\right]$$

$$+ \frac{k_m \sigma_2^5 \sqrt{\dfrac{\sigma_1^2 + \sigma_m^2}{2}} h_{m-1} \sqrt{\dfrac{\sigma_1^2 + \sigma_m^2}{2}} (\sigma_1^2 - \sigma_1^2)}{\sigma_1^2 \sigma_m^7 \left(\dfrac{\sigma_m^2 - \sigma_1^2}{2}\right)^2} \left[2\sigma_1^2\sigma_m^2 - \frac{(\sigma_m^2 - \sigma_1^2)^2}{4}\right]$$

Now suppose that $2\sigma_1^2\sigma_m^2 - (\sigma_m^2 - \sigma_1^2)^2/4 = \varepsilon < 0$ and k_m is sufficiently large so that $(\sigma_1^2 + \sigma_m^2)/2 < y_0^2$ and the sum of the first $m-1$ terms of $g_1'[(\sigma_1^2 + \sigma_m^2)/2]$ is less in absolute value than

$$\frac{k_m \sigma_2^5 \sqrt{\dfrac{\sigma_1^2 + \sigma_m^2}{2}} h_{m-1} \left(\sqrt{\dfrac{\sigma_1^2 + \sigma_m^2}{2}}\right) (\sigma_m^2 - \sigma_1^2)}{\sigma_2^2 \sigma_m^7 \left(\dfrac{\sigma_m^2 - \sigma_1^2}{2}\right)^2} \varepsilon$$

Then $g_1'[(\sigma_1^2 + \sigma_m^2)/2] < 0$.

This means that since the k_i can be chosen arbitrarily, for values of σ_i satisfying the inequality

$$\sigma_1^2 < \frac{(\sigma_m^2 - \sigma_1^2)^2}{8\sigma_m^2}$$

or, equivalently

$$\sigma_1 < \sigma_m (3^{1/2} - 2^{1/2}) \qquad (10)$$

there exist sets $\{k_i\}$ such that for an interval about $y = [(\sigma_1^2 + \sigma_m^2)/2]^{1/2}$, $g_1'(y)$ is decreasing. For each y_2 in this interval there is a $y_1 < y_2$, a $y_3 > y_3$ and a set $\{p_i\}$ such that $g(y_1) = g(y_2) = g(y_3) = -p_1\sigma_2^5/p_2\sigma_1^5$

Consequently, for this set $\{p_i\}$, $f''(x_1'; p_i) = f''(x_2; p_i) = f''(x_3; p_i) = 0$, where $x_i = y_i + \mu$, $i = 1, 2, 3$. By means of an analysis similar to that given in Section 3, it can be shown that if two of the values y_1, y_2, y_3 are not extreme values of $f'(x; p_i)$, there is, nevertheless, an interval of values about y_2 such that for y_α in this interval, there is a $y_\beta < y_\alpha$, a $y_\lambda > y_\alpha$ such that $f''(x_\beta; p_i) = f''(x_\lambda, p_i) = 0$

and $f''(x_\beta; p_i) \neq 0$ $f''(x_\lambda; p_i) \neq 0$. Thus, one can conclude that when $\mu_i = \mu$ for all i, $f(x; p_i)$ has either two points of inflection or six or more points of inflection, depending on whether σ_1 and σ_m satisfy Eqs. (9) or (10). Since it has been shown that when $\mu_i = \mu$ for all i, $f(x; p_i)$ is unimodal for all sets $\{p_i\}$, $f(x; p_i)$ is parapolymodal when it has six or more points of inflection. This proves (1b).

b. Proof of (2b). It has been shown that if the difference between μ_1 and each of the remaining μ_i is sufficiently small, $f(x; p_i)$ is unimodal for all sets $\{p_i\}$, but if the difference between μ_1 and μ_m is sufficiently large, the modality of $f(x; p_i)$ depends on the set $\{p_i\}$. However, it is also obvious that for a fixed set $\{p_i\}$ and fixed σ_i, as $(\mu_m - \mu_1)^2$ increases, regardless of the behavior of μ_2, $\mu_3, \cdots, \mu_{m-1}$, $f(x; p_i)$ tends continually toward polymodality, since $f(x; p_i)$ will eventually exhibit at least two concentrations of probability. When $f(x; p_i)$ is parapolymodal for $(\mu_m - \mu_1)^2 = 0$, there are $4k + 2$ points of inflection for some $k = 1, 2, \cdots, m - 1$. Hence, $f'(x; p_i) \to \infty$ as $x \to \pm\infty$; $f'(x; p_i)$ is positive and has $k + 1$ maxima and k minima when x is less than the mode, and negative with k maxima and $k + 1$ minima for x greater than the mode. As $f(x; p_i)$ tends to polymodality as $(\mu_m - \mu_1)^2$ increases with the set $\{p_i\}$ and the σ_i fixed, for some critical value of $(\mu_m - \mu_1)^2$, either one of the abscissa of the positive minima or one of the abscissa of the negative maxima changes sign. When this sign change occurs, $f(x; p_i)$ has at least three extreme values and is polymodal. However, as $(\mu_m - \mu_1)^2$ increases from zero to the critical value, there are always at least four extreme values and one zero of $f'(x; p_i)$, which means that if $f(x; p_i)$ is parapolymodal when $(\mu_m - \mu_1) = 0$, $f(x; p_i)$ remains parapolymodal as $(\mu_m - \mu_1)^2$ increases until it becomes polymodal. This proves (2b).

c. Proof of (3b). Now suppose for fixed values of μ_i and σ_i and a given set $\{p_i\}$, $f(x; p_i)$ has $k \geqslant 2$ modes. This means that each p_i, $i = 2, 3, \cdots, m - 1$ can then be taken sufficiently small such that:

(1) There exist values of p_1 and p_m for which $f(x; p_i)$ is bimodal.

(2) Holding each p_i, $i = 2, 3, \cdots, m - 1$ at the above sufficiently small values, there are two critical values of p_1, say u and v, such that for $v \ll p_1 < 1$ or $0 \ll p_1 < u$, $f(x; p_i)$ is unimodal, and for $u < p_1 < v$, $f(x; p_i)$ is bimodal.

When $f(x; p_i)$ is bimodal, $f'(x; p_i) \to 0$ as $x \to \pm \infty$, and $f'(x; p_i)$ has three distinct zeros and has at least two positive maxima and at least two negative minima. As p_1 increases to v or decreases to u, two adjacent zeros of $f'(x; p_i)$ move toward each other, and coincide when p_1

equals one of the critical values. When this occurs, then either,

(1) $f'(x; p_i)$ has at least one positive maximum, at least two negative minima and a maximum between two negative minima which coincides with a zero, or

(2) $f'(x; p_i)$ has at least two positive maxima, at least one negative minimum and a minimum between two positive maxima which coincides with a zero. But this means that for x_0 equal to the maximum (minimum) coinciding with a zero, $f(x_0; p_i)$ is not an extreme value, but is a point of inflection with zero slope. Thus, $f(x; p_i)$ has an extreme value which is a maximum and at least four points of inflection and, hence, is parapolymodal. Therefore, there is an interval of values of $p_1 \geqslant v$ and an interval of values of $p_1 \leqslant u$ for which $f(x; p_i)$ is parapolymodal. This proves (3b).
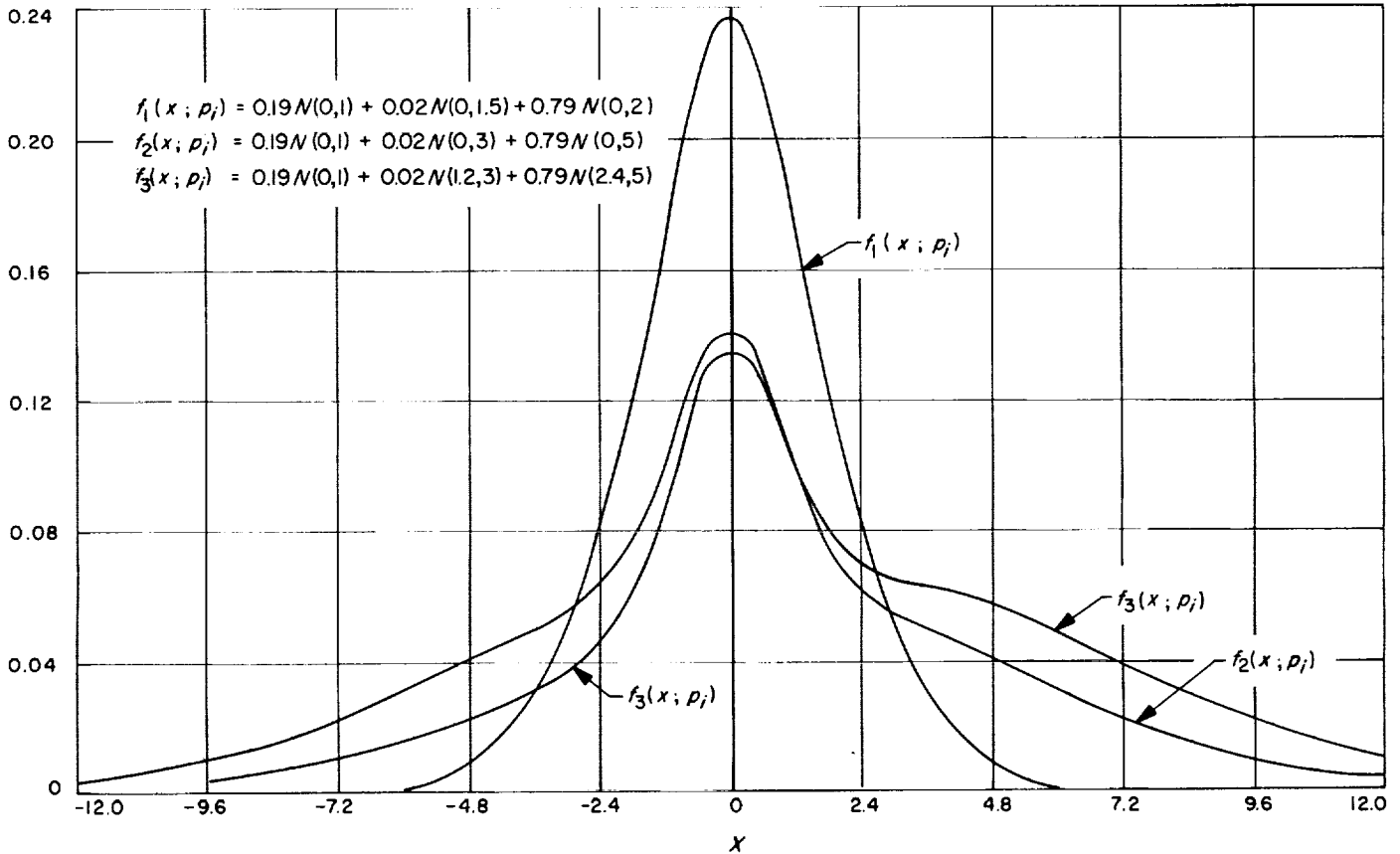
Fig. 1. Parapolymodal and nonparapolymodal distributions for $\mu_1 = \mu_i$, $i = 2, 3$, and the effect on paramodality of increasing $(\mu_j - \mu_i)^2$, when polymodality is impossible

d. Proof of (4b). The general form of Eq. (7) is given by

$$g_2(x) = \frac{(x - \mu_2)^2 - \sigma_2^2}{(x - \mu_1)^2 - \sigma_1^2} h_1^*(x)$$

$$+ \sum_{i=3}^{m} \frac{\sigma_2^5 p_i \left[(x - \mu_i)^2 - \sigma_i^2\right]}{\sigma_i^5 p_2 \left[(x - \mu_1)^2 - \sigma_1^2\right]} h_{i-1}^*(x) = -\frac{\sigma_2^5 p_1}{\sigma_1^5 p_2}$$

where

$$h_i^*(x) = \exp\left[\frac{-\frac{1}{2}(x - u_{i+1})^2}{\sigma_{i+1}^*} + \frac{(x - \mu_1)^2}{\sigma_1^2}\right],$$

$$i = 1, 2, \cdots, m - 1$$

In general, when $\mu_1 + \sigma_1 \neq \mu_i \pm \sigma_i$ or $\mu_1 - \sigma_1 \neq \mu_i \pm \sigma_i$ for all $i \geq 2$, for a fixed set $\{k_i\}$, there are two intervals for which $g_2(x) < 0$ and continuous on each open interval. At one endpoint of each interval, $g_2(x) = 0$ and $g_2(x) \to -\infty$ as x approaches the other endpoint from within the interval. Thus, regardless of the behavior of $g_2(x)$ in the interior of each interval, there is a half interval of values of x about each endpoint within each of the

two intervals, corresponding to sets $\{p_i\}$, for which the inverse function $g_2^{-1}(x)$ is double-valued. For each such set $\{p_i\}$, $f(x; p)$ has only two points of inflection and, hence, is neither polymodal nor parapolymodal. For the special case where $\mu_1 + \sigma_1 = \mu_i \pm \sigma_i$ or $\mu_1 - \sigma_1 = \mu_i \pm \sigma_i$ for all $i \geq 2$, one root of $f''(x; p_i)$ for all sets $\{p_i\}$ is $\mu_i \pm \sigma_i$, and since there is now only one interval for which $g_2(x) = 0$, which has the same properties as the two intervals in the general case, sets of $\{p_i\}$ exist for which $f(x; p_i)$ has only one additional point of inflection. Thus, in all cases, sets of $\{p_i\}$ exist for which $f(x; p_i)$ is neither polymodal nor parapolymodal. This proves (4b).

6. Examples

In Fig. 1, $f_1(x; p_i)$ and $f_2(x; p_i)$ are both unimodal for all sets $\{p_i\}$, since $\mu_1 = \mu_2 = \mu_3$. $f_3(x; p_i)$ is also unimodal for all sets $\{p_i\}$ since

$$(\mu_3 - \mu_1)^2 = 5.76 < \frac{27\sigma_1^2\sigma_3^2}{4(\sigma_1^2 + \sigma_3^2)} = 6.49$$
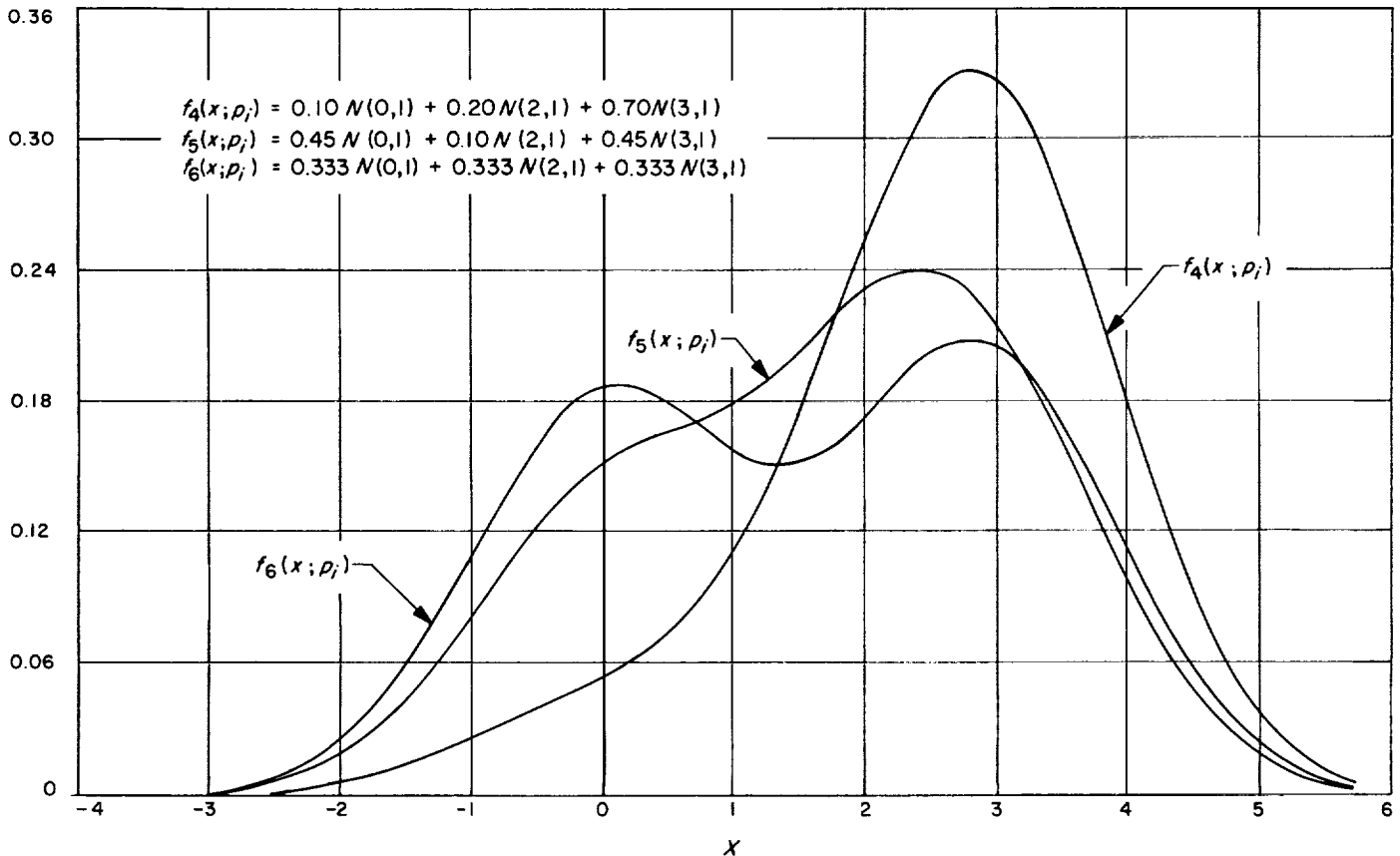
Fig. 2. The effect of the set $\{p_i\}$ on modality and paramodality when polymodality is possible

and

$$\max\left[\sigma_1^2(\mu_3 - \mu_1)(\mu_3 - \mu_2)^2, \frac{4\sigma_1^3}{27}(\mu_2 - \mu_1)^3\right] = 3.456$$

$$< \sigma_1^2\sigma_2^2(\mu_2 - \mu_1) - \frac{4\sigma_2^3}{27}(\mu_2 - \mu_1)^3 = 8.50$$

$f_1(x; p_i)$ cannot be parapolymodal since in this case, $\sigma_1 = 1 > \sigma_3(3^{1/2} - 2^{1/2}) = 0.636$. However, both $f_2(x; p_i)$ and $f_3(x; p_i)$ are parapolymodal, but it is readily seen that this property is more pronounced in $f_3(x; p_i)$ than in $f_2(x; p_i)$ due to the inequality of the means in $f_3(x; p_i)$.

The density functions in Fig. 2 are all mixtures of the same three normal distributions with different sets $\{p_i\}$. $f_4(x; p_i)$ is neither polymodal nor parapolymodal $f_5(x; p_i)$ is parapolymodal; and $f_6(x; p_i)$ is bimodal. Although $f_5(x; p_i)$ is unimodal, nevertheless the paramodality it exhibits is sufficiently pronounced to enable one to suspect that the distribution it represents is a mixture.

Fig. 3 illustrates the effect of increasing $(\mu_j - \mu_i)^2$. The same set of proportions (0.40, 0.24, 0.36) is associated with each density function and $\sigma_1 = \sigma_2 = \sigma_3 = 1$ in each case; yet, the density functions progress from unimodality with two points of inflection to paramodality, to bimodality and, finally, to trimodality.

$f_7(x; p_i) = 0.40N(0,1) + 0.24\ N(0.5,1) + 0.36N(0.75,1)$
$f_8(x; p_i) = 0.40N(0,1) + 0.24\ N(1,1) + 0.36\ N(2.75,1)$
$f_9(x; p_i) = 0.40N(0,1) + 0.24\ N(2,1) + 0.36\ N(3.75,1)$
$f_{10}(x; p_i) = 0.40N(0,1) + 0.24\ N(3,1) + 0.36\ N(5.75,1)$
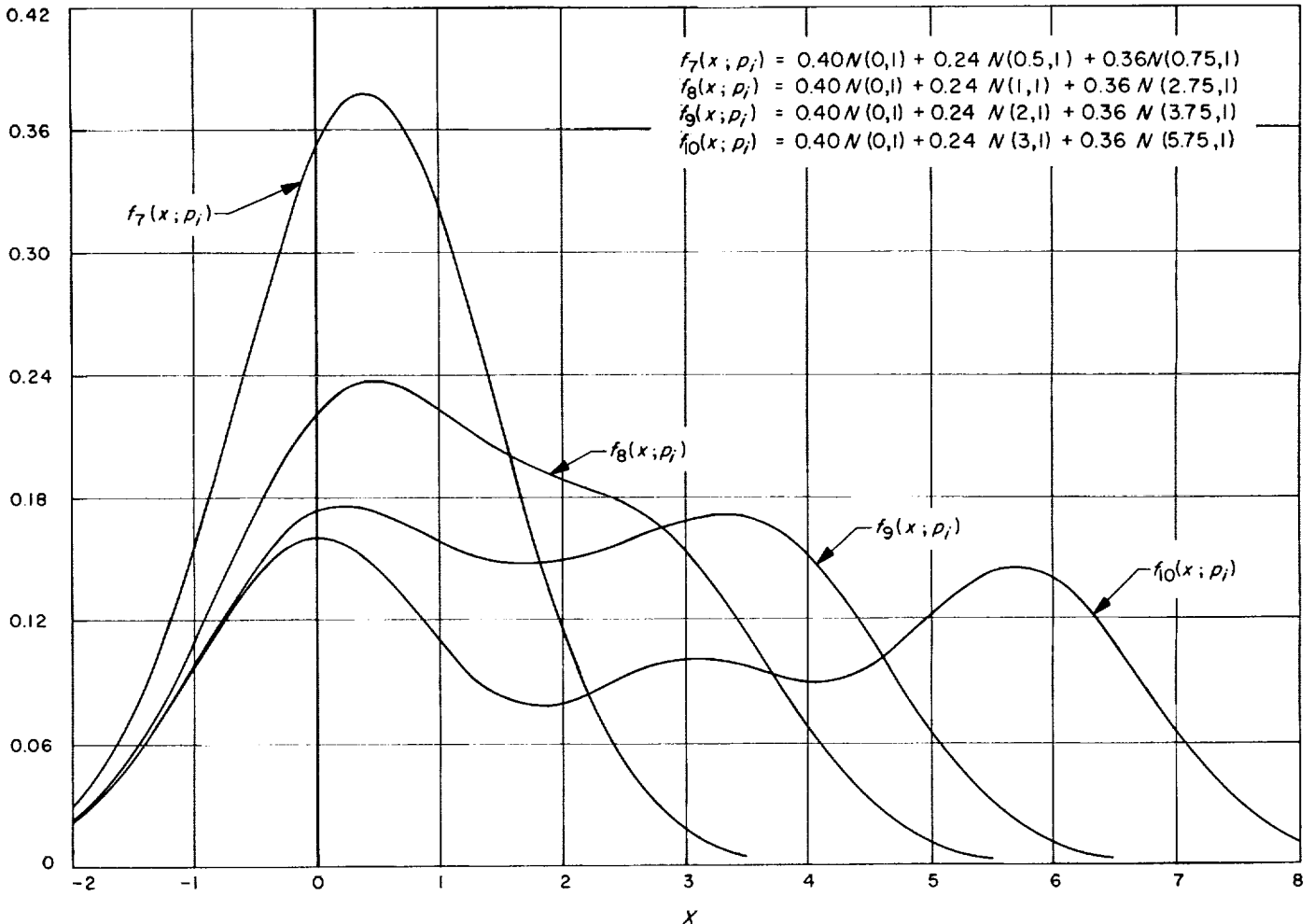
Fig. 3. The effect on modality and paramodality of increasing $(\mu_j - \mu_i)^2$

References

1. Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. I, John Wiley & Sons, New York, p. 174, 1950.

2. Huffman, D., "A Method for the Construction of Minimum Redundancy Codes," *Proc. IRE*, Vol. 40, No. 10, pp. 1098–1101, September, 1952.

3. Abramson, N., *Information Theory and Coding*, McGraw-Hill Book Co., New York, 1963.

4. *Technometrics*, Vol. 6, No. 4, pp. 357–363, November 1964.

XXII. Communications Systems Research: Detection Theory

A. Improvements to be Realized Through the Use of Block-Coded Communication Systems

W. C. Lindsey

As the need for higher data rates and improved performance in modern-day telemetry and command systems continues to increase, the communications engineer is faced with the problem of designing systems which satisfy more difficult constraints. Viterbi has shown that block-coding can be used to improve telemetry performance significantly over the more usual bit-by-bit detection scheme (Ref. 1).

In the discussion to follow it is assumed that a "clear" carrier signal is necessary in all systems for tracking purposes, and that the telemetry waveform is recovered by means of a narrow band tracking filter (a phase-locked loop). If conditions are such that there is noise in this tracking loop, the telemetry performance will be degraded. It is the purpose of this note to present the results of an extension of Viterbi's analysis to the case where this tracking noise is no longer negligible. The result, therefore, gives the practical improvement which block-coding can achieve relative to the uncoded system.

We consider the block-coded case first. Such coding consists of the transmission of digital data by means of an orthogonal dictionary which contains $N = 2^n$ code words (binary sequences of ± 1's), where n is the number of bits of information associated with each word. We assume that each word exists for a known time duration of $T = nT_b$ seconds, where $T_b = \mathscr{R}^{-1}$ is the time-per-bit, the reciprocal of the data rate. Briefly, the transmitter emits the waveform

$$\zeta(t) = (2P)^{1/2} \sin\left[\omega t + (cos^{-1} m) z_k(t)\right]$$

where

$$z_k(t) = x_k(t) \, S(t); \qquad k = 1, 2, \cdots N = 2^n$$

code word $\longrightarrow$ $\longleftarrow$ subcarrier

is the appropriate code word which amplitude modulates a unit power square-wave subcarrier (Ref. 2). Since $S(t)$ is a sequence of ± 1's, $z_k(t)$ is also a sequence of ± 1's. The subcarrier fundamental frequency is chosen such that the significant spectral components of the data-modulated subcarrier do not interfere with the receiver carrier-tracking process. The receiver observes the signal, corrupted by noise, which may be written

$$\psi(t) = (2m^2 P)^{1/2} \sin(\omega_1 t + \theta)$$

carrier component $\longrightarrow$

$$+ \left[2(1 - m^2) P\right]^{1/2} z_k(t) \cos(\omega_1 t + \theta + n(t).$$

$\longleftarrow$ data component

We assume that $n(t)$ is a white Gaussian process whose single-sided spectral density is N_0 (w/cps). The carrier component, tracked by means of a phase-locked loop, produces a local estimate of the carrier, say $r(t) = 2^{1/2} \cos(\omega_1 t + \hat{\theta}_1)$, which is then used to demodulate the information coherently. To do this, $y(t) = \psi(t) r(t)$ is presented to a maximum likelihood detector (a set of $N = 2^n$ correlators). The correlator outputs are sampled at the conclusion of the signaling interval (assumed known) to determine the *a posteriori* most probable code word. We further assume that the phase-error of the tracking filter, $\theta - \theta_1$, is essentially constant for at least a duration of T seconds.

We now wish to compare this block-coded system with an uncoded binary system in which the data, say $x_1(t) = -x_2(t)$, $0 \le t \le T_b$ from all data sources are time-multiplexed into a single pulse-code modulated (PCM) bit-stream and the results used to biphase-modulate a square wave subcarrier. In this mechanization (Ref. 2), the subcarrier is obtained by appropriately combining a pseudo-noise (PN) sequence with a unit power square wave, say $S(t)$. The transmitted waveform for this system is

$$\zeta(t) = (2P)^{1/2} \{\sin \omega_1 t + (\cos^{-1} m) [x_k(t) \cdot PN \cdot S(t)] \}$$

The receiver tracks the noise-corrupted carrier component just as indicated before but demodulates the data using only one cross-correlator. The cross-correlator output is sampled at the conclusion of the signaling interval and makes a maximum likelihood decision to determine the most probable value of x_k.

Figs. 1 and 2 indicate the maximum improvement in signal-to-noise ratio, i.e., PT_b/N_0, which may be expected from block-coding when the modulation index is chosen to minimize the probability of error (SPS 37-34, Vol. IV, p. 242). These figures are plots (for $n = 5, 7, 8$) of the improvement η (in db) vs the data-rate-to-carrier-tracking-loop bandwidth ratio,[1] say $\delta = \mathcal{R}/W_L$. We note that, for $\delta = 0$, the phase-locked loop estimate of the received carrier phase θ becomes uniformly distributed while $\delta = \infty$ corresponds to perfect carrier tracking, i.e., zero carrier phase-jitter.

Thus the improvement to be realized using a single-channel block-coded system instead of the single-channel PCM/PSK/PM system is clearly depicted. The parameter η is thus defined as

$$\eta = \frac{(PT_b/N_0)_{n=1}}{(PT_b/N_0)_n}$$

---

[1]The loop bandwidth is defined by $W_L = (1/2\pi j) \int_{-j\infty}^{j\infty} |H(s)|^2 \, ds$ where $H(s)$ is the transfer function of the carrier-tracking loop as determined by linear phase-lock loop theory.
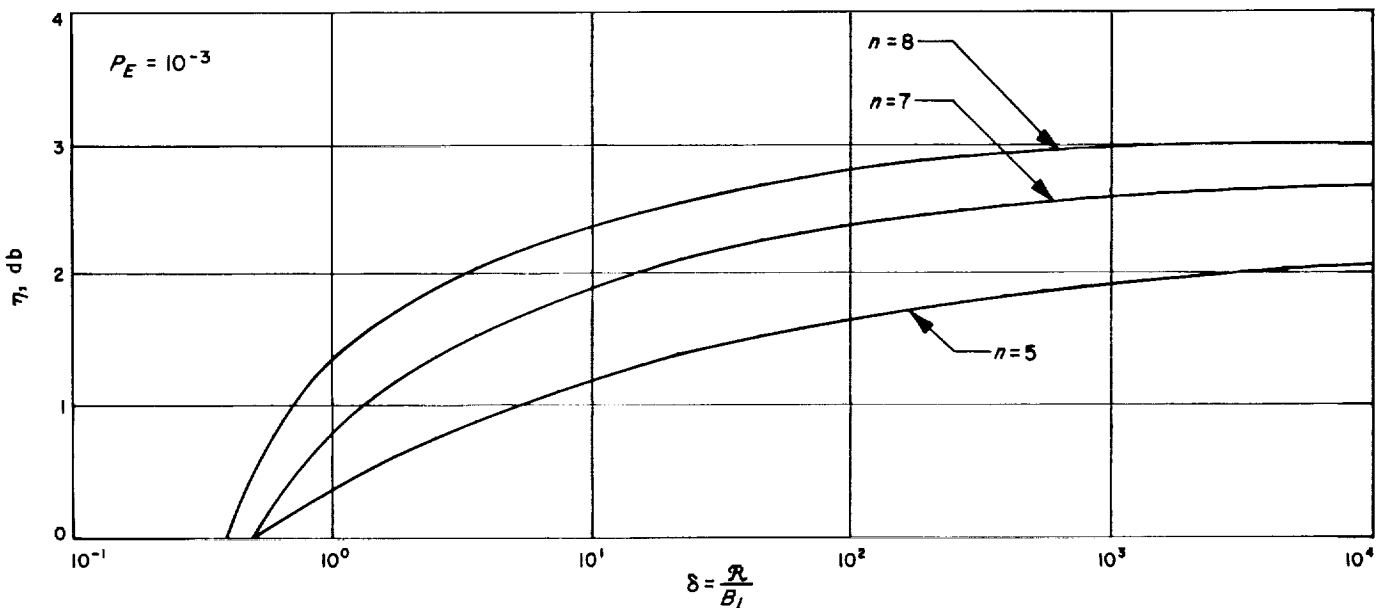
Fig. 1. Relative improvement in PT_b/N_0 using orthogonal coding, as a function of the parameter $\delta = \mathcal{R}/B_L$, for $P_E = 10^{-3}$
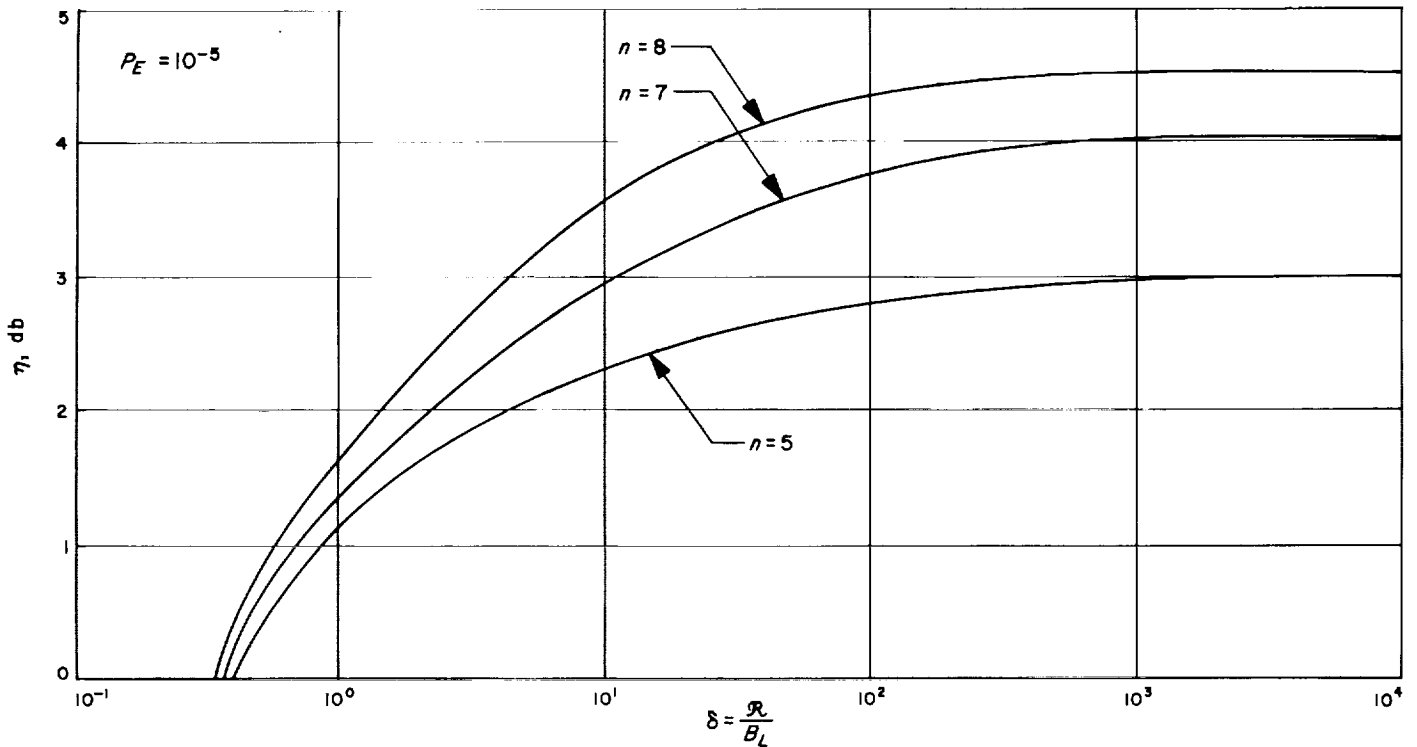
Fig. 2. Relative improvement in PT_b/N_0 using orthogonal coding as a function of the parameter $\delta = \mathcal{R}/B_L$, for $P_E = 10^{-5}$

Fig. 1 is plotted for an error rate of 10^{-3} (typical for telemetry systems) while Fig. 2 is plotted for an error rate of 10^{-5} (typical for command systems). Notice the asymptotes approached as the carrier reference phase-jitter approaches zero, i.e., as δ approaches infinity. This provides a check with presently tabulated results (Ref. 1) for perfect phase synchronization.

We conclude from these results that, for a given n (i.e., the number of information bits coded into each word), the improvement to be realized through the use of block codes depends on the parameter δ = (Data Rate/Carrier Tracking-Loop Bandwidth) to a significant degree in the regions of current design interest. Values of $\delta < 10$ make block coding seem hardly worth the effort.

Finally, we can extend these results both to the case of a bi-orthogonal code dictionary and to the case of incoherent detection. For $n \geq 5$ the performances of the bi-orthogonal system (SPS 37-34, Vol. IV, p. 242) and the incoherent system (Ref. 3) are approximately equal to the system we have studied, hence, Figs. 1 and 2 present an upper bound on the improvements which can be realized for those cases.

B. Word Synchronization Over the Binary Symmetric Channel

J. J. Stiffler

1. Introduction

In SPS 37-35, Vol. IV, p. 349, we considered the amount of time needed to establish symbol synchronization for various communication systems by utilizing a maximum likelihood detection procedure. In this article we apply the same technique to the problem of word synchronization over the binary symmetric channel. We shall consider three situations:

(1) A synchronizing sequence is transmitted repetitively during a synchronization mode. When synchronization is obtained, the communication mode begins, but until that time no data are transmitted.

(2) A known synchronizing pattern is periodically transmitted, interspersed with data. Synchronization involves the recognition of the pattern in question.

(3) The information is encoded into comma-free words, and the comma-free property is used to establish synchronization.

The maximum likelihood detectors can be readily determined in each of these cases. In situation (1), the optimum detector simply determines the Hamming distance between each of the phases of a locally generated replica of the sync sequence and the received sequence. After a suitable number of repetitions of the sync sequence have been received, the phase corresponding to the minimum distance is used as the estimation of the true sync position. In the second situation, the optimum synchronizer is somewhat more complicated, but the procedure for case (1) is nearly optimum, and practical considerations would favor its use over the strictly optimum approach. The third case also involves essentially the same approach, except that the locally generated synchronizing pattern is replaced by that word in the dictionary closest (in the Hamming distance sense) to the received sequence of bits. Again this procedure is suboptimum, but it is a reasonable, practical approximation to the optimum approach.

In the remainder of this article we derive estimations as to the number of observables which will be needed to establish synchronization with satisfactory reliability in each of these three situations.

2. Repetitive n-Symbol Synchronizing Sequence

Let d_ν be the Hamming distance between the synchronizing sequence and the sequence obtained by permuting it cyclically ν positions to the right. The observables

$$N_\nu = \sum_{j=1}^{M} n_j (\nu)$$

would, in the absence of errors, equal $M d_\nu$ for all $\nu = 0, 1, \cdots, N - 1$. ($n_j (\nu)$ represents the Hamming distance between the jth repetition of the νth phase shift of the received sequence and the locally generated sequence.)

If the receiver signal is subject to a total of e errors in this interval (of M repetitions), the probability distribution of N_ν is easily determined:

$$P (N_\nu | e \text{ errors}) = Pr \left(\frac{M d_\nu - N_\nu + e}{2} \right.$$

errors in the $M d_\nu$ positions of disagreement and

$$\frac{N_\nu - M d_\nu + e}{2}$$

errors in the $M (n - d_\nu)$ positions of agreement $\Bigg)$

$$= \begin{cases} \dfrac{\dbinom{M d_\nu}{\dfrac{e + M d_\nu - N_\nu}{2}} \dbinom{M (n - d_\nu)}{\dfrac{e - (M d_\nu - N_\nu)}{2}}}{\dbinom{M n}{e}} & \max [0, M d_\nu - e] \leq M d_\nu + e, \\[20pt] & \text{and } N_\nu \text{ same parity as } M d_\nu + e \\[10pt] 0 & \text{otherwise} \end{cases}$$

(1)

Consequently,

$$P (N_\nu) = \sum_{e=0}^{Mn} P (N_\nu | e \text{ errors}) P(e \text{ errors})$$

$$= \sum_{e=0}^{Mn} \dbinom{M d_\nu}{\dfrac{e + M d_\nu - N_\nu}{2}} \dbinom{M (n - d_\nu)}{\dfrac{e - (M d_\nu - N_\nu)}{2}} p^e (1 - p)^{Mn - e}.$$

(2)

This is true for all v, $v = 0, 1, \cdots, N - 1$, where it is observed that $d_0 = 0$. The probability of a correct decision after M observations of each phase is

$$P_s = Pr\{N_0 < N_1, N_0 < N_2, \cdots, N_0 < N_{N-1}\}$$

$$\geqq 1 - \sum_{i=1}^{N-1} Pr\{N_0 \geqq N_i\} \qquad (3)$$

When the different phases are observed in parallel, we note that the number of errors e in the received sequence is the same for all phases. Thus,

$$Pr\{N_0 \geqq N_i\} = \sum_{e=0}^{Mn} Pr\{N_0 \geqq N_i|e\} Pr(e)$$

$$= \sum_{e=0}^{Mn} Pr\{N_i \leqq e|e\} Pr(e)$$

$$= \sum_{e=0}^{Mn} \sum_{N_i=0}^{e} Pr\{N_i|e\} Pr(e)$$

$$= \sum_{e=0}^{Mn} \sum_{j=0}^{e} \binom{Md_i}{\dfrac{e + Md_i - j}{2}}$$

$$\times \binom{M(n - d_i)}{\dfrac{e - (Md_i - j)}{2}} p_e (1 - p)^{Mn-e}$$

$$\qquad (4)$$

where the binomial coefficient $\binom{N}{i}$ is zero unless i is a non-negative integer.

Eqs. (3) and (4) provide an upper bound on the amount of time needed to establish synchronization. It is sometimes convenient to approximate this bound by using the Central Limit Theorem. The number of errors observed at the vth epoch is just $k_1 + Md_v - k_2$, where k_1 is the number of errors in the $M(n - d_v)$ positions of agreement and k_2 the number in the Md_v positions of disagreement. But

$$\eta_v = E\{k_1 + Md_v - k_2\} = Md_v + M(n - d_v)p - Md_v p$$

$$\qquad (5)$$

while

$$\sigma_v^2 = VAR\{k_1 + Md_v - k_2\}$$

$$= VAR\{k_1\} + VAR\{k_2\} = MN p(1 - p). \qquad (6)$$

Consequently, if $d_v \geqq d_\mu$ for all $v \neq 0$, then

$$Pe \leqq (N - 1) Pr\{N_0 - N_\mu \geqq 0\} \leqq \frac{N-1}{(2\pi)^{1/2} \sigma}$$

$$\times \int_0^\infty \exp\left\{\frac{-(y - (\eta_0 - \eta_\mu))^2}{8\sigma^2}\right\} dy$$

$$= \frac{(N-1)}{2}\left\{1 - erf\frac{M^{1/2}d_\mu(1 - 2p)}{2(2)^{1/2}[np(1-p)]^{1/2}}\right\}$$

$$\leqq N - 1\frac{\exp\left\{-\dfrac{Md_\mu^2(1-2p)^2}{8np(1-p)}\right\}}{(2\pi)^{1/2}\dfrac{M^{1/2}d_\mu(1-2p)}{2[np(1-p)]^{1/2}}}$$

$$= \frac{\exp\left\{-\dfrac{Md_\mu^2(1-2p)^2}{8np(1-p)} + \log_e(N-1)\right\}}{(2\pi)^{1/2}\dfrac{M^{1/2}d_\mu(1-2p)}{2[np(1-p)]^{1/2}}}$$

$$\qquad (7)$$

where use is made of the upper bound on the variance σ_x^2 of the difference between two random variables: $(\sigma_x^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2) \leqq (\sigma_1 + \sigma_2)^2$. The probability of a synchronization error will be small if

$$M > \frac{8np(1-p)\log_e(N-1)}{d_\mu^2(1-2p)^2} \qquad (8)$$

3. Synchronizing Sequences Interspersed With Data

Eq. (3) is still applicable, but now $Pr\{N_0 \geqq N_i\}$ must be redetermined. Let the 0th and ith phases have Mn_i symbols in common yielding a distance, in the absence of errors, of Md_i.

Then

$$Pr\{N_0 \geqq N_i\} = \sum_{N_0=0}^{Mn} \sum_{N_i=0}^{N_0} Pr\{N_i\} Pr\{N_0\}$$

$$= \sum_{N_0=0}^{Mn} \sum_{e=0}^{Mn_i} \binom{M(n - n_i)}{N_0 - e}\binom{Mn_i}{e}$$

$$p^{N_0}(1 - p)^{Mn-N_0}$$

$$\times \sum_{j=0}^{N_0} \sum_{e_1=0}^{e} \frac{\binom{Md_i}{e_1}\binom{M(n_i - d_i)}{e - e_1}}{\binom{Mn_i}{e}}$$

$$\times \binom{M(n - n_i)}{j - e + 2e_1 - d_i}\left(\frac{1}{2}\right)^{j-e+2e_1-d_i} \qquad (9)$$

If $n_i = 0$, this reduces to

$$Pr\{N_0 \geqq N_i \,|\, n_i = 0\} = \sum_{N_0 = 0}^{Mn} \binom{Mn}{N_0}$$

$$p^{N_0}(1-p)^{Mn-N_0} \sum_{j=0}^{N_0} \binom{Mn}{j}\left(\frac{1}{2}\right)^j \quad (10)$$

Presumably, the synchronizing sequence will be chosen so that

$$Pr\{N_0 \geqq N_i \,|\, n_i > 0\} \leqq Pr\{N_0 \geqq N_i \,|\, n_i = 0\} \quad (11)$$

(e.g., a Barker-type sequence could be used.) Then

$$P_e \leqq (N-1)\,Pr\{N_0 \geqq N_i \,|\, n_i = 0\} \quad (12)$$

Again, the Central Limit Theorem may be used to obtain an approximation to Eq. (10). Here we note that

$$E(N_0) = Mnp$$

$$VAR(N_0) = Mnp(1-p)$$

$$E(N_i \,|\, n_i = 0) = \frac{Mn}{2}$$

$$VAR(N_i \,|\, n_i = 0) = \frac{Mn}{4} \quad (13)$$

Since N_0 and N_i are statistically independent when $n_i = 0$, it follows that

$$Pr\{N_0 \geqq N_i \,|\, n_i = 0\} \approx \frac{1}{2}\left\{1 - erf\frac{(Mn)^{1/2}(1-2p)}{(2)^{1/2}(1+4p(1-p))^{1/2}}\right\}$$

$$\leqq \frac{\exp\left\{-\dfrac{Mn(1-2p)^2}{2[1+4p(1-p)]}\right\}}{(2\pi)^{1/2}\dfrac{(Mn)^{1/2}(1-2p)}{(1+4p(1-p))^{1/2}}} \quad (14)$$

Since $P_e \leqq (N-1)\,Pr\{N_0 \geqq N_i \,|\, n_i = 0\}$, the probability of a synchronization error will be small if

$$M > \frac{2[1+4p(1-p)]\log_e(N-1)}{n(1-2p)^2} \quad (15)$$

4. Random Sequences of Code Words

We can immediately establish an upper bound on the probability of a sync error by noting that, if there are 2^k words and if Mn-tuples are observed at each phase, there are 2^{kM} possible decodings. The probability that this minimum distance N_i accumulated at the ith phase is less

than that distance N_0 at the 0th (true) phase is certainly bounded by

$$Pr\{N_i \leqq N_0\} \leqq \sum_j Pr\{N_i^j \leqq N_0\} \leqq 2^{kM}\,Pr\{N_i^* \leqq N_0\} \quad (16)$$

where N_i^j is the distance between the jth sequence of code words and the received sequence, and N_i^* corresponds to that sequence of code words which has the minimum error-free distance from the received sequence. This follows from the observation that when $p < \frac{1}{2}$, the probability that $N_i^j \leqq N_0$ is a monotonically decreasing function of the error-free distance between the two sequences. If the dictionary is comma-free with index d, this error-free distance is at least Md regardless of the received sequence. Consequently,

$$Pe \leqq (N-1)\max_i Pr\{N_i \leqq N_0\} \leqq (N-1)\,2^{kM}$$

$$\times \sum_{e=0}^{Mn} \sum_{j=0}^{e} \frac{\binom{Md}{\frac{e+Md-j}{2}}\binom{M(n-d)}{\frac{e-(Md-j)}{2}}}{2} p^e(1-p)^{Mn-e} \quad (17)$$

Applying the Central Limit Theorem (Laplace-DeMoivre Theorem), we find that [cf. Eq. (7)]

$$Pe < \frac{\exp\left\{kM\log_e 2 + \log_e(N-1) - \dfrac{Md^2(1-2p)^2}{8\,np(1-p)}\right\}}{(2\pi)^{1/2}\dfrac{M^{1/2}d(1-2p)}{2[np(1-p)]^{1/2}}} \quad (18)$$

If

$$M > \frac{8\,np(1-p)\log_e(N-1)}{d^2(1-2p)^2 - 8\,np(1-p)\,k\log_e 2} \quad (19)$$

and the index of comma freedom is bounded by

$$d > n\frac{\left[8\,p(1-p)\dfrac{k}{n}\log_e 2\right]^{1/2}}{1-2p} \quad (20)$$

then we can expect the probability of a sync error to be small. Since this is an upper bound on the error probability, the converse of this statement need not be true.

Another approach to estimating the sync error probability is possible when it is specified that the dictionary being used is capable of correcting up to e errors. We

use the fact that no two n-tuples x and y in any coset of an e-error correcting code can have the property that

$$|x - y| \leq |x| + |y| \leq 2e + 1 \quad (|z| \text{ denotes the Hamming weight of } z).$$

Suppose that the distance between the jth received n-tuple at the ith phase and the nearest code word would be, in the absence of errors, d_{ij}. Errors will change this distance to m_{ij}. If m_{ij} is not greater than e, the distance n_{ij} between this n-tuple and the nearest code word is m_{ij}; if m_{ij} is greater than e, this distance is at least $2e + 1 - m_{ij}$. Note that $n_{ij} = 2e + 1 - m_{ij}$ for all $e < m_{ij} < 2e + 1$ only if the code is close-packed. Equating n_{ij} to min $\{m_{ij}, 2e + 1 - m_{ij}\}$ consequently provides a conservative estimate of the true minimum distance. (We can ignore the event $m_{ij} > 2e + 1$ since this will presumably occur with very low probability.) We thus conclude that

$$P(N_i = A) = \sum_{n_{i1}} \sum_{n_{i2}} \cdots \sum_{n_{iM}} P(n_{i1} = A_1) P(n_{i2} = A_2) \cdots$$
$$\times P(n_{iM} = A - A_1 - A_2 \cdots - A_{M-1}) \quad (21)$$

where

$$P(n_{ij} = A_j) = P(n_{ij} = A_j \mid m_{ij} \leq e) P(m_{ij} \leq e)$$
$$+ P(n_{ij} = A_j \mid m_{ij} > e) P(m_{ij} > e)$$
$$= P(m_{ij} = A_j \mid m_{ij} \leq e) P(m_{ij} \leq e)$$
$$+ P(m_{ij} = 2e + 1 - A_j \mid m_{ij} > e) P(m_{ij} > e)$$
$$= \begin{cases} P(m_{ij} = A_j) + P(m_{ij} = 2e + 1 - A_j) & 0 \leq A_j \leq e \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

and where

$$P\{m_{ij} = B_j\} = \sum_{i=0}^{d_{ij}} \binom{d_{ij}}{i} \binom{n - d_{ij}}{B_j + i - d_{ij}} p^{B_j + 2i - d_{ij}} (1 - p)^{n - B_j - 2i + d_{ij}} \quad (23)$$

Unfortunately it is difficult to carry this analysis much further without some additional qualifications. It is not even possible to state that the worst situation is obtained when d_{ij} attains its minimum possible value.

Again, however, we may obtain an approximate analysis using the Gaussian approximation to the distribution in question. We observe that

$$m_{ij} = e + \tfrac{1}{2} - E|n_{ij} - (e + \tfrac{1}{2})|,$$

so that

$$E(m_{ij}) = e + \tfrac{1}{2} - E|n_{ij} - (e + \tfrac{1}{2})|$$

It is easily verified[2] that $E(|x|) \leq E^{1/2}(x^2)$ for any random variable x. Thus,

$$E(m_{ij}) \geq e + \tfrac{1}{2} - [E(n_{ij} - (e + \tfrac{1}{2}))^2]^{1/2}$$
$$= e + \tfrac{1}{2} - [E(n_{ij} - (e + \tfrac{1}{2}))^2 + VAR(n_{ij})]^{1/2}$$
$$= e + \tfrac{1}{2} - [((e + \tfrac{1}{2} - np) - d_{ij}(1 - 2p))^2 + np(1 - p)]^{1/2} \quad (24)$$

Moreover,

$$VAR(m_{ij}) = np(1 - p) \quad (25)$$

The lower bound is also valid for the true phase $(d_{ij} \equiv 0)$, but to obtain an upper bound in this case we simply observe that

$$E(n_{0j}) \leq np \quad (26)$$

Since the variance is finite we conclude that a sufficient condition for synchronization is that

$$E(m_{0j}) < E(m_{ij}), \quad i = 1, 2, \cdots, N-1 \quad (27)$$

and hence, from Eqs. (24) and (26), that

$$\Delta_e^2 > (\Delta_e - d_{ij}(1 - 2p))^2 + \sigma^2 \quad (28)$$

[2]For a proof note that

$$0 \leq \Sigma (u + |x|v)^2 p(x) = u^2 \Sigma p(x) + 2uv \Sigma |x| p(x) + v^2 \Sigma x^2 p(x)$$
$$= u^2 + 2uv E(|x|) + v^2 E(x^2) \text{ for any real } u \text{ and } v.$$

Setting $u = E(|x|)$ and $v = -1$ we conclude that $E(x^2) \geq E^2(|x|)$, which was to be shown.

where $\Delta_e = e + \frac{1}{2} - np$ and $\sigma^2 = np\,(1 - p)$. This, in turn, is satisfied if

$$\frac{\Delta_e}{1 - 2p}\left[1 - \left(1 - \frac{\sigma^2}{\Delta_e^2}\right)^{\frac{1}{2}}\right] < d_{ij} < \frac{\Delta_e}{1 - 2p}$$
$$\times \left[1 + \left(1 - \frac{\sigma^2}{\Delta_e^2}\right)^{\frac{1}{2}}\right] \qquad (29)$$

(It is assumed here that $\Delta_e > 0$ and that $p < \frac{1}{2}$. If this were not true the synchronous error rate would be intolerably high.) We also observe that the most difficult situation, so far as synchronization is concerned, occurs when $E(m_{ij})$ attains its minimum value. The bound (28) is a minimum either for minimum d_{ij} or maximum d_{ij}. But $d_0 \leq e$ for close-packed[3] e-error correcting codes with the index of comma-freedom d_0. The minimum value of the bound (28) is achieved for d_0 if

$$e\,(1 - 2p) - (e + \frac{1}{2} - np) \leq (e + \frac{1}{2} - np) - d_0\,(1 - 2p) \qquad (30)$$

or if

$$d_0 \leq d_1 = \frac{2\Delta_e}{1 - 2p} - e \qquad (31)$$

If this inequality does not hold, the worst case occurs when $d_{ij} = e$.

We observe that if the synchronous error probability is to be reasonably small $\Delta_e = h\sigma$ for some h on the order of, say, five. In this case, condition Eq. (29) for synchronizability reduces to approximately,

$$\frac{\sigma}{2h\,(1 - 2p)} = \frac{\Delta_e}{(1 - 2p)\,(2h^2)}$$
$$< d_{ij} < \frac{2\Delta_e}{1 - 2p} = d_1 + e \qquad (32)$$

Note that if $d_0 \geq d_1 > 0$, synchronization is always possible; if $d_1 \leq 0$, the present analysis does not assure the synchronizability of the code in question. In either case, the apparent upper bound (30) on the permitted distance d_{ij} is vacuous. If this upper bound can be exceeded, synchronization is presumably not possible for any index of comma freedom.

---

[3]If the code is not close-packed it is possible for d_{ij} to be greater than e; however, since the distance to all code vectors must then be greater than e, the expected final distance to the nearest vector must be increased over that in the close-packed case. Consequently, this argument establishes a bound for nonclose-packed codes, as well.

We conclude that synchronization is possible if

$$\frac{2\Delta_e}{1 - 2p} > e \qquad (33)$$

and

$$d_0 > \frac{\Delta_e}{(1 - 2p)\,(2h^2)} > \frac{e}{4h^2} \qquad (34)$$

For large n, the synchronous word error probability is approximately

$$P_e \approx \frac{1}{2}\left(1 - erf\,\frac{\Delta_e}{(2\sigma)^{\frac{1}{2}}}\right) = \frac{1}{2}\left[1 - erf\left(\frac{h}{2^{\frac{1}{2}}}\right)\right] \qquad (35)$$

As n is increased, e must grow at least as rapidly as $np + h\,(npq)^{\frac{1}{2}}$. But from Eq. (33), Δ_e must also grow this rapidly. This can be true only if h is proportional to n. We infer, therefore, that an error-correcting code is synchronizable if the word error probability is proportional to $(2\pi an)^{-\frac{1}{2}}\exp\{-an/2\}$ for some constant a. In this case, from the inequality (34), d_0 does not need to grow with n; the comma-free codes of index one, discussed in Ref. 4, may be sufficient for all n whenever these conditions occur. Even if the Ineq. (29) is not satisfied, it may still be possible to synchronize, since only a sufficient condition has been established here. In particular, if d_0 satisfies Eq. (20), the analysis leading to that result is applicable, regardless of the error-correcting capabilities of the code. We observe, however, that if the code is efficient (i.e., nearly close-packed) then the condition (20) will generally be impossible to satisfy, except for low error probabilities (as is the case here). This suggests that efficient self-synchronization is feasible with these codes only when they are to be operated at relatively low error probabilities. This is independent of the question as to whether it is possible to obtain codes with an index of comma-freedom satisfying the Ineq. (29). It must be remembered, though, that the random selection of code words will normally insure that d_{ij} is usually at neither of its extremes. Synchronization will still probably be possible even if the Ineq. (29) does not hold. It will presumably take considerably longer, however, and, at any rate, cannot be estimated without a detailed investigation of the specific dictionary being used. Moreover, if h does not increase at least as rapidly as n, we can conclude that for large n

$$Pr\,\{e - n_{0i} > n\epsilon\}$$
$$\times \frac{1}{2}\left\{1 - erf\left(\frac{(n)^{\frac{1}{2}}\,\epsilon}{(2pq)^{\frac{1}{2}}} - \frac{h}{2^{\frac{1}{2}}}\right)\right\} \qquad (36)$$

becomes arbitrarily small for any $\epsilon > 0$. The expected number of errors in synchronous operation thus becomes concentrated about e, the maximum number of errors that can be observed with a close-packed code regardless of the transmitted sequence. This suggests, therefore, that for efficient error-correcting codes a condition not unlike that of Eq. (33) is, at least asymptotically, necessary as well as sufficient.

The number of observations needed when the Ineqs. (33) and (34) are satisfied is easily bounded in the standard way:

$$P_e = Pr\ \{N_0 < N_1 \text{ or } N_0 < N_2 \text{ or } \cdots \text{ or } N_0 < N_{N-1}\}$$

$$\leq (N-1)\,Pr\,\{N_0 - N_1 < 0\}$$

$$= \frac{(N-1)}{2}\left(1 - erf\,\frac{\mu}{(2\sigma)^{1/2}}\right)$$

$$\geq \frac{\exp\left\{\log_e(N-1) - \frac{\mu^2}{2\sigma^2}\right\}}{(2\pi)^{1/2}\,\mu/\sigma} \tag{37}$$

where

$$\mu = E\,(N_i - N_0) = M\,[E\,(n_{ij}) - E\,(n_{0j})]$$

$$\geq M\,\{\Delta_e - [(\Delta_e + d_0(1 - 2p))^2 + np\,(1-p)]^{1/2}\}$$

$$= M\Delta_e\left\{1 - \left[\left(1 - \frac{d_0(1-2p)}{\Delta_e}\right)^2 + \frac{1}{h^2}\right]^{1/2}\right\} \tag{38}$$

and $\sigma^2 \leq 4\,Mnp\,(1-p)$. The terms Δ_e and h are as previously defined. Note that under the conditions described above for large n, d_0/Δ_e and $1/h^2$ are presumably small, so that

$$\mu \approx M\,[d_0(1-2p) - \Delta_e/2h^2] \tag{39}$$

An upper bound on the number of words which must be observed, before synchronization is reliably obtained, is afforded by observing the term

C. Extension of Koerner's Theorem to Arbitrary Tensor Product Code

P. Stanek

In SPS 37-17, Vol. IV, p. 71, Koerner discusses a decoding technique for the n-bit orthogonal codes of Reed-Muller. His result appears as an instance of a more general theorem.

Briefly, let $y\,(t)$ be the received signal and for each code word $x_i\,(t)$, define

$$c_i = \int_0^{nT} y\,(t)\,x_i\,(t)\,dt, \qquad T = \text{bit transmission time.} \tag{1}$$

Then the most likely word transmitted is x_{i_0} where $c_{i_0} \geq c_i$ for all i. With 2^n code words, a system based on this technique alone requires 2^n parallel correlators, somewhat too many for even small values of n.

Now the ith code word is an L-tuple of ± 1's, each component of duration τ, so that $\tau = nT/L$. Hence Eq. (1) can be rewritten as

$$c_i = \sum_{k=1}^{L} \int_{(k-1)\tau}^{k\tau} y\,(t)\,x_i\,(t)\,dt \tag{2}$$

and on the interval from $(k-1)\,\tau$ to $k\tau$, $x_i\,(t)$ has a constant value, say a_{ki}.

$$y_k = \int_{(k-1)\tau}^{k\tau} y\,(t)\,dt, \qquad k = 1, \cdots, L \tag{3}$$

then Eq. (2) finally becomes

$$c_i = \sum_{k=1}^{L} y_k\,a_{ki} \tag{4}$$

The vector $c = (c_1, \cdots, c_L)$ is the image of the vector $y = (y_1, \cdots, y_L)$ under the linear transformation with matrix $A = (a_{ij})$. This matrix A will be called the code matrix.

$$\frac{\mu^2}{2\sigma^2} - \log_e(N-1) \approx \frac{M\Delta_e^2\left\{1 - \left[1 - \left(\frac{d_0(1-2p)}{\Delta_e}\right)^2 + \frac{1}{h^2}\right]^{1/2}\right\}^2}{8\,np\,(1-p)} - \log_e(N-1)$$

$$\approx \frac{Mh^2\,[d_0(1-2p) - \Delta_e/2h^2]^2 - 8\Delta_e^2\log_e(N-1)}{8\Delta_e^2} \tag{40}$$

For orthogonal codes of maximum length, $L = 2^n$, and an unsophisticated decoder uses $2^n (2^n - 1)$ additions or subtractions to compute all components of the vector $c = yA$. If T_0 is the computer add time and T_{w} is the least time required to decode a word, then we have shown

$$T_w \leq 2^n (2^n - 1) T_0 \qquad (5)$$

Let the code matrix A be m by m.

Theorem. If $A_1, \cdots, A_v$ are p_1 by $p_1, \cdots, p_v$ by p_v square matrices ± 1's and if $A = A_1 \otimes \cdots \otimes A_v$, then T_w can be reduced to no more than

$$[m(p_1 - 1) + m(p_2 - 1) + \cdots + m(p_v - 1)] T_0$$

$$= \left[\sum_{i=1}^{J} p_i - v \right] m T_0$$

Proof. If $v = 1$, the conclusion is Eq. (5). If $v = 2$, notice that $A_1 \otimes A_2 = (A_1 \otimes I)(I \otimes A_2)$, for suitable identity matrices I. Then

$$c = yA = y(A_1 \otimes I)(I \otimes A_2) \qquad (6)$$

which involves $m(p_1 - 1)$ additions or subtractions for the first factor and $m(p_2 - 1)$ for the second. The theorem is proved for $v = 2$. Also, the matrix $I \otimes A_2$ appears as

$$I \otimes A_2 = \begin{pmatrix} A_2 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ & & & \\ 0 & 0 & \cdots & A_2 \end{pmatrix} \qquad (7)$$

where 0 is a p_1 by p_1 matrix of 0's. This is a good position for induction and the theorem follows. For Hadamard matrices A_1 and A_2, $A_1 \otimes A_2$ is another Hadamard matrix (see Ref. 1, p. 55) and so the hypothesis is not unrealistic.

Corollary. (Koerner's Theorem.) For n-bit Reed-Muller codes, $T_w \leq 2^n n T_0$.

In this case, the code matrix A is an n-fold tensor product of the 2 by 2 matrix

$$H = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad (8)$$

and so $T_w \leq mnT_0 = 2^n n T_0$.

It should be pointed out that there is no guaranty that the bounds obtained on T_w are the best possible (i.e., there

may exist more efficient encoding methods). However, if the code matrix A can be factored as a tensor product, this is always advisable since the form

$$[m(k_1 - 1) + \cdots + m(k_v - 1)] T_0 \qquad (9)$$

where $m = k_1 \cdots k_v$, monotonically decreases as v increases. Naturally each $k_i \geq 2$; hence, we see that the method affords its maximum efficiency to the Reed-Muller codes.

D. Results from the Two Station Space Diversity Experiment

W. F. Gillmore, Jr.

1. Introduction

The purpose of this experiment is to explore the possibility of improving the signal-to-noise ratio by combining signals from a number of tracking stations. Although there are many possible ways of doing this, the one studied here seems to be the most promising. Rather than phasing the antennas directly into an array, video outputs from the individual station receivers are combined by simple addition.

Telemetry from *Mariner IV* was used as the test signal for this experiment. At the time of the experiment it had passed the planet Mars and was more than 120 million miles away. This provided a signal which can be considered typical for a deep space probe. Two similar tracking stations, Echo and Pioneer, were used for this experiment. Although there is no unobstructed space path between the stations, a microwave link is available for sending video signals over the hills between them.

The data for the diversity experiment were taken on July 29. Video telemetry from both the Echo and Pioneer tracking stations was sent to the nearby Venus Station over microwave links. Considerable care was taken to ensure that both data and calibration signals were properly recorded. While the two stations were tracking *Mariner*, the doppler from each station was being automatically counted and sent to JPL at 60-sec intervals. These datum points were stored in a standard manner on a disk file by 7040 computers. Later, the data were logged on magnetic tape for permanent storage. The 60-sec

doppler counts are differenced to find a doppler difference frequency at 60-sec intervals throughout the experiment.

Results from this experiment have been made more interesting and useful by processing the data in several ways. Signal-to-noise ratio measurements of the *Mariner* telemetry signals were made independently on a signal-to-noise ratio estimator and on a multichannel correlator. The doppler difference frequency was found and compared with the difference frequency computed from a good ephemeris of *Mariner IV*. A brief summary of all these results is given in the following Section.

2. Summary of Results

Typical results from the telemetry part of the experiment are summarized in Table 1. For convenience, all of the entries in this table have the same units, cycles per second. They represent total signal power divided by noise power spectral density. Since they are all essentially signal-to-noise ratios, they are expressed in db.

Table 1. Summary of telemetry results

Measurement	Signal-to-noise ratio estimator, db	Multichannel autocorrelator, db
(1) Echo channel	22.8	24.4
(4) Pioneer channel	21.3	23.2
(2) Sum channel	24.6	27.0
Computed sum	25.0	26.8

Each of the numbers in the first three rows of Table 1 represent about 20 min of data processing time. This gives all of the measurements roughly the same theoretical accuracy. Each measurement should be within 0.1 db of the correct value.

Numbers in the second column of Table 1 were obtained on the signal-to-noise estimator (*SPS 37-27*, Vol. IV, pp. 169–184), while numbers in the third column were obtained on the multichannel autocorrelator (*SPS 37-28*, Vol. IV, pp. 219–222). The last row of the table is included for comparison purposes. It shows what the signal-to-noise ratio of the sum channel should be, if the measured signal-to-noise ratios in the first two rows were exact. With perfect measurements rows three and four of the table would be identical.

Fig. 3 shows the results from the doppler part of the experiment. Sixty-second doppler counts were made at
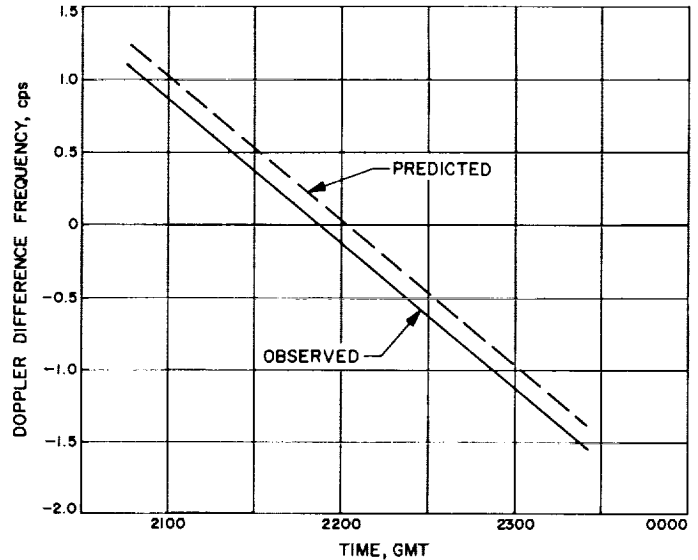
Fig. 3. Summary of doppler results

both the Echo and Pioneer tracking stations. The difference between these counts is the S-band doppler difference frequency, which is plotted in Fig. 3 as a function of time.

A theoretical second curve is plotted in Fig. 3. This curve shows the doppler difference frequency, as it was predicted by the JPL orbit determination programs. The predicted curve was based on the best orbit available for the *Mariner* spacecraft.

The main purpose for plotting the curves shown in Fig. 3 was to determine how well a computer could control the phasing of antennas from ephemeris data. In order for RF phasing to give any gain at all it must be accurate to better than ±60 deg. With the frequency error between the two curves in Fig. 3 at almost 0.2 cps, the computer could not predict phase for more than about 1 sec.

Errors, such as those displayed in Fig. 3, can sometimes be corrected fairly well by shifting the theoretical curve parallel to itself. If this is done for the predicted curve, the error becomes less than 0.02 cps over a 2-hr period. Under these conditions, if the correct phase and frequency can be found, initially, the computer predictions can be used for about 4 hr.

3. The Telemetry Experiment

Telemetry signals were taken from the same point within the receivers at both the Pioneer and Echo tracking stations. The connection was made at the output of

a video isolation amplifier immediately following the 10-Mc telemetry phase detector. Both of these signals were at convenient voltage levels and bandwidths for transmission to the Venus Station over microwave links.

At the Venus Station both signals were taken from the microwave links and fed to an Ampex FR-1200 tape recorder. Separate tracks were used for each of the telemetry signals, and a third track was used to record a standard NASA time code. Each tape contains about 1 hr of data, and relatively little time was lost by changing reels.

Data tapes were played back through a special device for summing the two signals. A block diagram of this is shown in Fig. 4. The two band-pass filters were each set for a passband from 0.25-400 cps. Since output levels from the amplifiers in Fig. 4 were not readily adjustable, they were left with the same gain throughout the entire experiment. Outputs from channels 1 and 4 contain the video telemetry output from the Echo and Pioneer tracking stations respectively. Channel 2 contains the sum of channels 1 and 4.
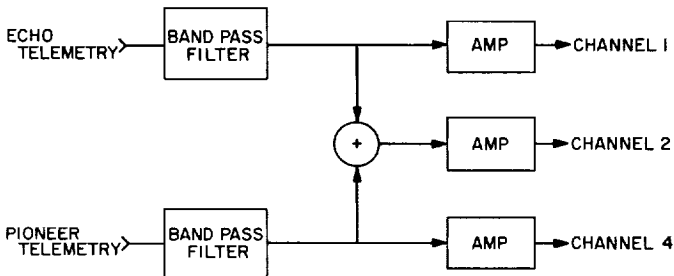
Fig. 4. Special summing device for video telemetry signals

All data processing was done at the output of the special summing device of Fig. 4. Results from the signal-to-noise estimator required that the data tape be played back three times, once for each channel. The multichannel auto-correlator was able to process the three channels simultaneously with a single tape playback.

Results from the signal-to-noise ratio estimator were printed out directly by an SDS-920 computer. They include the mean, standard deviation, their ratio, and a number proportional to the result given in Table 1. Arbitrary gain constants in the receiver will affect both the mean and standard deviation, but the final signal-to-noise ratio is supposed to remain unchanged. It is useful to have the standard deviation available because it is a

number which is directly proportional to the square root of the power spectral density of the noise.

When two signals are added together, their RMS signal voltages add directly, as shown in Eq. (1). Noise voltages,

$$D_2 = D_1 + D_4 \tag{1}$$

however, add orthogonally as shown in Eq. (2). From these two equations

$$N_2 = (N_1^2 + N_4^2)^{1/2} \tag{2}$$

the signal-to-noise ratio ρ_2 of the combined signals can be predicted. Eq. (3) gives

$$\rho_2 = \frac{S_2}{N_2} = \frac{S_1 + S_4}{(N_1^2 + N_4^2)^{1/2}} = \frac{S_1}{(N_1^2 + N_4^2)^{1/2}} + \frac{S_4}{(N_1^2 + N_4^2)^{1/2}}$$

$$= \frac{S_1}{N_1 [1 + (N_4/N_1)^2]^{1/2}} + \frac{S_4}{N_4 [(N_1/N_4)^2 + 1]^{1/2}}$$

$$= \frac{\rho_1}{[1 + (N_4/N_1)^2]^{1/2}} + \frac{\rho_4}{[(N_1/N_4)^2 + 1]^{1/2}} \tag{3}$$

the desired result. The ratios N_4/N_1 and N_1/N_4, as well as the signal-to-noise ratios ρ_1 and ρ_4, can be calculated directly from the computer print out. Note that ρ_2 cannot show a 3-db improvement over both ρ_1 and ρ_4 unless $\rho_1 = \rho_4$.

Fig. 5 shows typical power spectral density curves for channels 1, 2 and 4. Vertical lines indicate the areas which were taken to be synchronizing signal, data signal, and noise (Ref. 5). These spectra were obtained from the correlator output. Signal-power-to-noise-power-spectral-density ratios were obtained from the area under the first data sideband. This ratio was multiplied by $\pi^2/8$ to account for the power in the higher order sidebands. The results are given in the last column of Table 1.

The calculated value given in Table 1 for the multichannel autocorrelator was found from Eq. (3). The ratio between N_1 and N_4 can be the same as that determined previously, or it can be found from RMS voltmeter readings. Voltmeter readings taken in channels 1 and 4 before the mixer of Fig. 4 are almost entirely determined by broad-band white noise. Since the bandwidths of the two channels are the same, the ratio of the voltmeter readings gives N_1/N_4 directly. It is satisfying to find that both methods of calculations give the same predicted results for the sum channel to within less than 0.1 db.
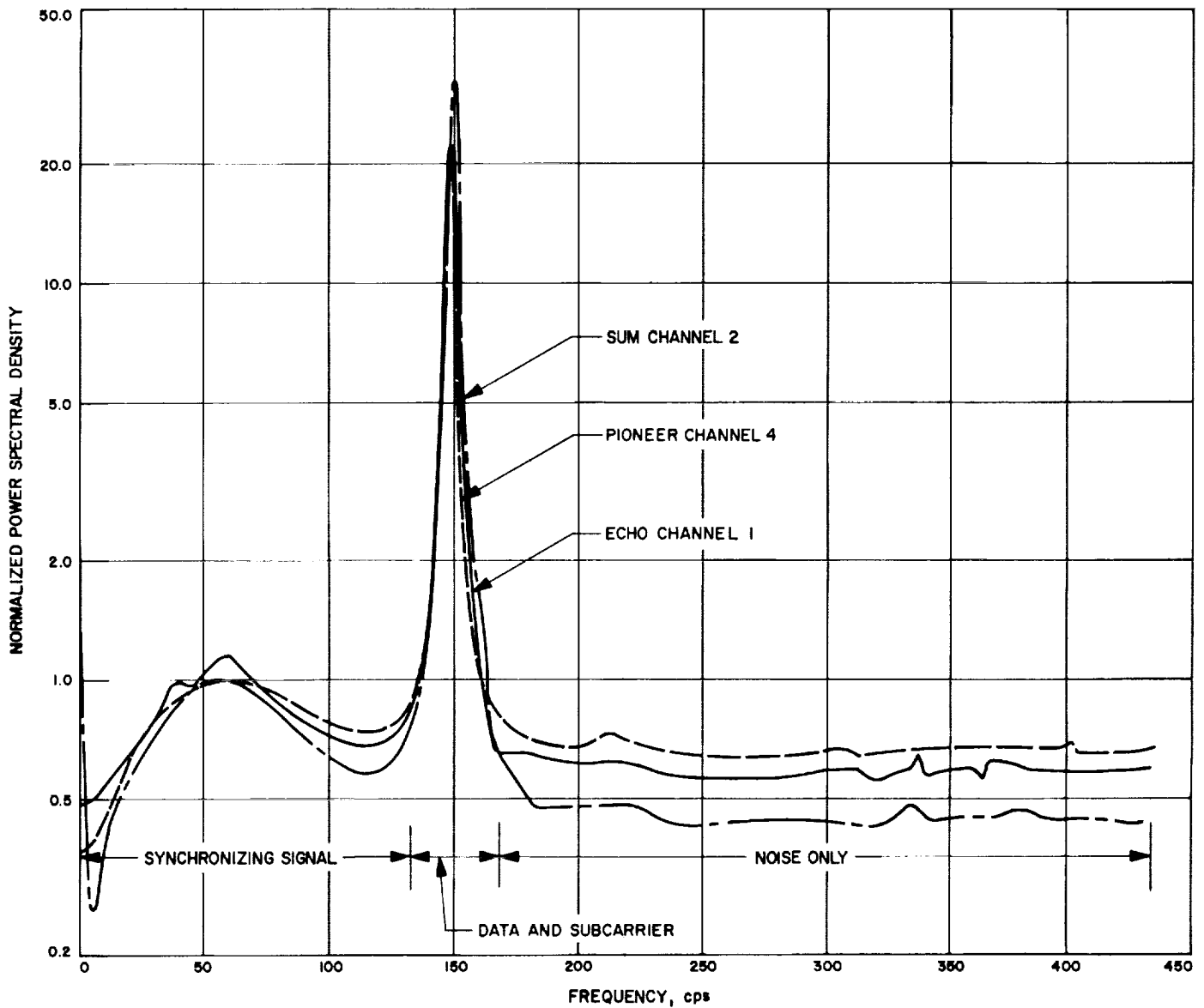
Fig. 5. Power spectral density curves made on the multichannel autocorrelator: Echo channel 1; Pioneer channel 4; sum channel 2

There are some general conclusions which can be drawn from processing the telemetry. Of the two methods, the signal-to-noise ratio estimator was the most convenient to use because a computer program was already available to perform the calculations. It also gave a number proportional to the strength of the noise directly on the computer print out. On the other hand, the results obtained by the signal-to-noise ratio estimator were consistently low. This may be related to the fact that it was very difficult to make the machine lock onto the data subcarrier. One of the difficulties with the machine is its

limited dynamic range, and the *Mariner* signals were apparently very close to its threshold.

The multichannel autocorrelator had the advantage of processing all of the channels simultaneously. Although it was a little difficult to use, without writing a new set of computer programs, its results seem to be quite consistent and accurate. There is no lost data time because it is not necessary for the machine to lock onto the data subcarrier. It also has a larger dynamic range. Unfortunately, this particular autocorrelator was not able to

measure the absolute level of the incoming signal or noise directly.

4. The Doppler Experiment

During a preliminary run of this experiment high frequency doppler signals were taken from both the Echo and Pioneer tracking stations. The signal used in each case is the one known as 1-Mc biased doppler. At the Echo Station this signal was taken directly from the receiver while the corresponding signal from the Pioneer Station was received at the Echo Station via the microwave link. Both signals were fed to the special detector device shown in Fig. 6.
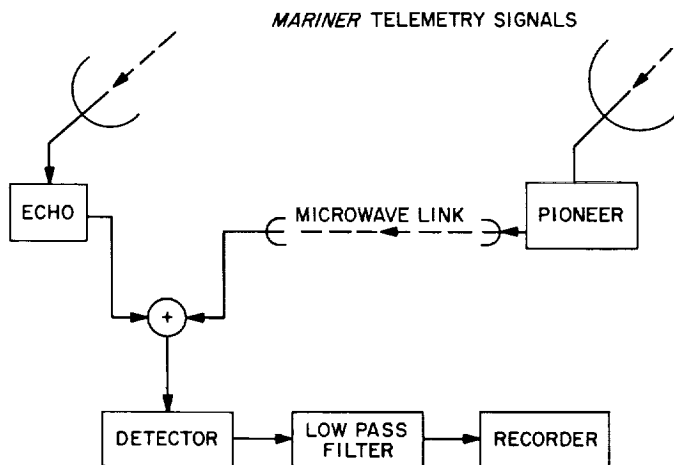
Fig. 6. Envelope detector for observing doppler difference frequency between Echo and Pioneer tracking stations

Each of the 1-Mc biased doppler signals was sinusoidal, with a frequency equal to 1 Mc plus the doppler frequency. Although the 1-Mc bias frequency was the same for both signals, it was not phase coherent. Also, a considerable amount of broad band noise was present from the two receivers. The lowpass filter shown in Fig. 6 proved to be quite essential, and its high frequency cutoff was set as low as 10 cps. Fig. 7 shows a typical waveform at the recorder input. It consists of the doppler difference frequency plus some added noise.

In the final run of the experiment a somewhat different method was used to obtain the doppler difference frequency. Each station has a doppler counter which can be used to count the unbiased doppler frequency directly. For convenience, the counting interval was set at 60 sec.
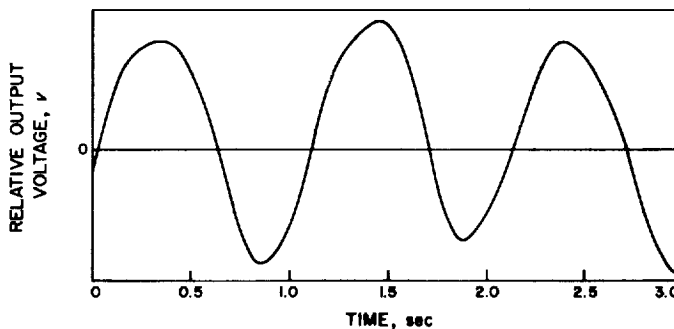
Fig. 7. Typical waveform of doppler difference signal after low pass filtering

The station clocks were carefully synchronized to ensure that the counting would be done during simultaneous time intervals. By differencing the two counter readings, the required doppler difference frequency is obtained.

Although the second method is not quite as accurate as the first method, it is much easier to execute. The counter readings for each 60-sec count were sent digitally to the laboratory on teletype lines. As the datum points were received they were logged in a standard manner by IBM 7040 computers. The data were first stored in a disk file and then transferred to magnetic tape for permanent storage. Later, special cards were made for each of the datum points. All of these data served as input to the 7094 orbit determination program.

After all of the data were taken, the orbit program was run to determine the best possible orbit for Mariner. Coordinates for the two tracking stations were the best which could be found from a land survey between stations and from tracking data throughout the entire Mariner flight. Output from the program gave the predicted doppler frequency and a residual for each station. The predicted doppler difference frequency P is computed from Eq. (4), where P_P and P_E

$$P = P_P - P_E \qquad (4)$$

are the predicted doppler frequencies for the Pioneer and Echo Stations respectively. Eq. (5)

$$0 = P + R_P - R_E \qquad (5)$$

gives the observed doppler difference frequency 0 in terms of P, and the residuals R_P and R_E at Pioneer and

Echo respectively. All of the residuals are defined as observed, minus predicted values.

It can be seen from the results of Fig. 3 that there is a considerable error in the predicted curve. This error is not entirely unexpected because the program was constructed to minimize errors in doppler predictions at three tracking stations spaced around the world, rather than the difference between two closely spaced stations. With this much error in the orbit program, it was useless to try to track phase at the two stations and look for residuals in phase directly. The data show that the computer can predict phase for about 1 sec, at most, before the error exceeds 1 rad.

A more optimistic appraisal of the results in Fig. 3 can be made if the predicted curve is shifted closer to the measured curve by moving it parallel to itself. When this is done, it leaves an error of less than 0.02 cps after a 2-hr operating period. At this rate it would take over 4 hr before the phase error exceeds 1 rad, provided that the initial frequency and phase were found exactly.

5. Conclusions

From analysis of the *Mariner IV* telemetry data, it can be reasonably concluded that there is little or no difficulty in achieving the theoretical gain in signal-to-noise ratio when multiple tracking stations are used. Both the signal-to-noise estimator and the multichannel autocorrelator are capable of giving good results for the necessary measurements.

One conclusion which can be drawn from the doppler experiment is that the present orbit determination program would not be sufficient to control the phasing of an array of antennas. Furthermore, even if the same frequency reference were used at all stations, propagation uncertainties might cause serious phasing problems.

The final conclusion seems to be that the easiest way to make signals add together reliably is to do the adding at the demodulated level, after the carrier has been removed. An extra advantage of this is that no compensation is necessary for the unequal movement of the phase centers of different antennas.

References

1. Golomb, S., et al., *Digital Communications with Space Applications*, Prentice-Hall EE Series, Englewood Cliffs, New Jersey, 1964.

2. Springett, J. C., *Telemetry and Command Techniques for Planetary Spacecraft*, Technical Report No. 32-495, Jet Propulsion Laboratory, Pasadena, California, January 15, 1964.

3. Lindsey, W. C., "Coded Noncoherent Communications," *IEEE Transactions on Space Electronics and Telemetry*, Vol. Set-11, No. 1, pp. 6–13, March 1965.

4. Stiffler, J. J., "Comma-Free Error-Correcting Codes," *IEEE Transactions on Information Theory*, Vol. IT-11, No. 1, pp. 107–112, January 1965.

5. Springett, J. C., *Telemetry and Command Techniques for Planetary Spacecraft*, Technical Report No. 32-495, Jet Propulsion Laboratory, Pasadena, January 14, 1965.

XXIII. Communications Systems Research: Mathematical Research

A. Variance of Estimates of Spectral Densities

E. C. Posner

1. Summary

This paper gives a computational method for finding lower bounds on the variance of estimates of spectral densities obtained by an autoregressive filter technique. To be found is the sum of the squares of the elements of a certain linear recurring sequence. A system of linear equations is obtained which yields this sum as a quadratic form in the initial conditions of the recursion. The equations are solved exactly for two step recursions. The same procedure will work on filters whose transfer function has both numerator and denominator.

2. Introduction

Suppose we start with $n+1$ measured correlation coefficients $\widehat{R}_j$, $0 \leq j \leq n$, which are unbiased estimators of the first $n+1$ correlation coefficients R_j of a sampled

stationary Gaussian process. The differences $\widehat{R}_j - R_j$ are $n+1$ mutually independent normal random variables of mean 0, variance σ^2. The true spectrum

$$P(f) = \sum_{j=-\infty}^{\infty} R_j \exp(2\pi i j f)$$

is often estimated (Ref. 1) by

$$\widetilde{P}(f) = \sum_{j=-n}^{n} \widehat{R}_i \exp(2\pi i j f),$$

which in effect assumes that the R_j for $|j| > n$ are zero. The technique used here is rather to estimate all the remaining R_j from $R_0, R_1, \cdots, R_n$, using the fact that a polynomial filter has been used. The estimator $\widehat{P}(f)$ is obtained as

$$\widehat{P}(f) = \sum_{j=-\infty}^{\infty} \widehat{R}_i \exp(2\pi i j f). \tag{1}$$

where the $\widehat{R}_j$ for $|j| > n$ are obtained as linear combinations of the $\widehat{R}_j$, $0 \leq j \leq n$ ($R_j = R_{-j}$, $\widehat{R}_j = \widehat{R}_{-j}$ in this notation). Thus, we use for the "missing" $\widehat{R}_j$ these linear combinations.

In practice, the filter itself is estimated from the data; by assuming a filter known in advance, we shall obtain here a lower bound for the variance of $\hat{P}(f)$. Since in truth

$$P(f) = \sum_{j=-\infty}^{\infty} R_j \exp(2\pi i j f) , \qquad (2)$$

the actual integrated square error in $\hat{P}(f)$ is given by the random variable

$$\tau^2 = \int_0^1 [P(f) - \hat{P}(f)]^2 \, df . \qquad (3)$$

It then follows quite straightforwardly that τ^2 is given by

$$\tau^2 = \sum_{j=-\infty}^{\infty} (R_j - \hat{R}_j)^2 . \qquad (4)$$

3. Determining the Missing $\hat{R}_j$

We need a formula for the $\hat{R}_j$, $|j| > n$. We shall first derive

$$R_{l+n+1} = \sum_{j=0}^{n} a_j R_{l+j} , \text{all } l \geq 0 , \qquad (5)$$

for certain constants a_j. The sampled process x_i satisfies (Ref. 1)

$$x_{l+n+1} = \sum_{j=0}^{n} a_j x_{l+j} + n_{l+n+1} , \text{for all } l . \qquad (6)$$

The a_j are constants derivable from the filter, and the n_j, $-\infty < j < \infty$, are mutually independent identically distributed normal random variables.

By stationarity,

$$R_j = \sum E(x_j x_0), \ -\infty < j < \infty . \qquad (7)$$

Hence

$$R_{l+n+1} = E(x_{l+n+1} x_0) = E\left[\left(\sum_{j=0}^{n} a_j x_{l+j} \, x_0 \right) \right] , \qquad (8)$$

and so

$$R_{l+n+1} = \sum_{j=0}^{n} a_j E(x_{l+j} x_0) , \qquad (9)$$

which proves Eq. (5). The required estimates for the missing $\hat{R}$'s can now be taken as the quantities

$$\hat{R}_{l+n+1} = \sum_{j=0}^{n} a_j \hat{R}_{l+j} , \ -\infty < l < \infty . \qquad (10)$$

The $\hat{R}_j$ chosen this way are unbiased estimators of the R_j.

Now consider the polynomial

$$g(z) = z^{n+1} - \sum_{j=0}^{n} a_j z^j , \qquad (11)$$

the reciprocal of the z-transform of the sample-filter response. Realizability of such a filter means in effect (Ref. 1) that every root of $g(z) = 0$ is less than 1 in absolute value. As a consequence (Ref. 2), the $\hat{R}_j$ approach 0 geometrically as $|j| \to \infty$, so all assumptions on convergence used above are justified. We note that $S_j = R_j - \hat{R}_j$ satisfies the same recursion (5) or (10) that $\{R_j\}$ and $\{\hat{R}_j\}$ satisfy. We are given $S_0, S_1, \cdots, S_n$, and the other S_j are determined from these $n+1$.

The problem of this paper is to find

$$S^2 = \sum_{j=0}^{\infty} S_j^2 , \qquad (12)$$

giving $S_0, \cdots, S_n$ and the fact that the S_j satisfy

$$S_{l+n+1} = \sum_{j=0}^{n} a_j S_{l+j} , \text{all } l \qquad (13)$$

(we have changed notation slightly to avoid summing those S_j of negative indices). Knowledge of this S^2-sum will then give us the actual integrated square error τ^2 as a random variable. We also find.

$$s^2 = E(\tau^2) , \qquad (14)$$

the expected L_2 − error, in a later section.

4. S^2 as a Quadratic Form

This section proves that S^2 is a quadratic form q in the S_i, $0 \leq i \leq n$ (q will certainly be positive definite, i.e., $q(S_0, \cdots, S_n) > 0$ unless $S_0 = \cdots = S_n = 0$). Furthermore, equations determining the coefficients are given.

The S_i are given by (Ref. 2)

$$S_i = \sum_{j=0}^{n} c_j u_j^i , \text{all } i , \qquad (15)$$

where the u_j are the roots of $g(s) = 0$ (this is only true if $g(s)$ has distinct roots; if not, a slight modification of the argument has to be given). The c_j are (possibly complex) constants to be determined from $S_0, S_1, \cdots, S_n$. Squaring Eq. (15), we find, for all i

$$S_i^2 = \sum_{j,k} c_j c_k (u_j u_k)^i . \qquad (16)$$

Hence

$$\sum_{i=0}^{\infty} S_i^2 = \sum_{j,k} c_j c_k \sum_{i=0}^{\infty} (u_j u_k)^i . \qquad (17)$$

But

$$\sum_{i=0}^{\infty} (u_j u_k)^i = (1 - u_j u_k)^{-1}.$$ (18)

Hence

$$\sum_{i=0}^{\infty} S_i^2 = \sum_{j,k} \frac{1}{(1 - u_j u_k)} c_j c_k.$$ (19)

The question now hinges on the dependence of the c_j on $S_0, S_1, \cdots, S_n$. From Eq. (15), we see that the c_j are linear combinations of the S_i, $0 \leq i \leq n$, since the equations

$$S_i = \sum_{j=0}^{n} c_j u_j^i \ , 0 \leq i \leq n,$$ (20)

have a unique solution because of Van der Monde's Determinant Theorem (Ref. 2). Hence, $c_j c_k$ is a sum of products of pairs of the first $(n+1)$ of the S_i. Since the $u_j u_k$ are merely constants determined by the a_i, we have proved that S^2, that is, $\displaystyle\sum_{i=0}^{\infty} S_i^2$, is indeed a quadratic form in the S_i, $0 \leq i \leq n$.

The coefficients of this form can be found by the above argument, but this would involve finding the roots of $g(s) = 0$. Since we want to avoid root-finding, we proceed differently.

First notice that Eq. (19), coupled with Eq. (20), indicates that the coefficients of the quadratic form $S^2 = q(S_0, \cdots, S_n)$ are rational functions of the coefficients a_i of the polynomial $g(s)$. For these coefficients, as a simple *gedanken* calculation shows, are rational symmetric functions of the roots u_j. So we look for the coefficients of the quadratic form q as rational functions of the a_i.

The idea will be to use the recursion (13) for the S_i to obtain certain conditions that q must satisfy. These conditions in turn will uniquely determine the coefficients of each $S_i S_j$ occurring in q as rational functions of the a_i, $0 \leq i \leq n$.

5. Finding the Coefficients of the Quadratic Form q

Our key observation is based on the identity

$$q(S_0, S_1, \cdots, S_n) = S_0^2 + q(S_1, S_2, \cdots, S_{n+1}),$$ (21)

where

$$S_{n+1} = \sum_{i=0}^{n} a_i S_i.$$ (22)

These equations merely express the fact that

$$\sum_{i=0}^{\infty} S_i^2 = S_0^2 + \sum_{i=1}^{\infty} S_i^2.$$

In the second sum, we have the first sum, but $(S_0, S_1, \cdots, S_n)$ is replaced by $(S_1, S_2, \cdots, S_{n+1})$ as the initial condition. Since the recursion given by Eq. (5) remains the same, we conclude that

$$\sum_{i=1}^{\infty} S_i^2 = q(S_1, S_2, \cdots, S_{n+1}).$$ (23)

This proves Eq. (21). We now prove that Eq. (21) determines the quadratic form q uniquely, and furthermore we show how the coefficients of q are determined.

Let

$$q(S_0, \cdots, S_n) = \sum_{i,j=0}^{n} q_{ij} S_i S_j \ , q_{ij} = q_{ji}.$$ (24)

The coefficients q_{ij} of the form are to be determined from Eq. (21). Eqs. (21) and (22) imply that

$$\sum_{i,j=0}^{n} q_{ij} S_i S_j = S_0^2 + \sum_{i,j=1}^{n} q_{i-1,j-1} S_i S_j + 2 \sum_{i=1}^{n} \sum_{j=0}^{n} q_{i-1,n} a_j S_i S_j$$
$$+ q_{nn} \sum_{i,j=0}^{n} a_i a_j S_i S_j.$$ (25)

Eq. (25) merely arises by formal substitution of Eq. (24) into Eq. (21).

Since Eq. (25) is to be an identity in quadratic forms, we can equate coefficients of $S_i S_j$ on both sides. We arrive at the sought-after system of linear equations for the $(n+1)(n+2)/2$ unknowns q_{ij}, $0 \leq i,j \leq n$:

$$\begin{cases} q_{00} = 1 + a_0^2 q_{nn}; \\ q_{0i} = a_0 q_{i-1,n} + a_0 a_i q_{nn}, 1 \leq i \leq n; \\ q_{ij} = q_{i-1,j-1} + a_j q_{i-1,n} + a_i q_{j-1,n} + a_i a_j q_{nn}, 1 \leq i,j \leq n. \end{cases}$$ (26)

If we define $q_{ij} = 0$ if i or j is less than 0, we can simplify Eqs. (26) to

$$\begin{cases} q_{00} = 1 + a_0^2 q_{nn}; \\ q_{ij} = q_{i-1,j-1} + a_j q_{i-1,n} + a_i q_{j-1,n} + a_i a_j q_{nn}, \begin{cases} 0 \leq i,j \leq n, \\ (i,j) \neq (0,0). \end{cases} \end{cases}$$ (27)

While a proof that Eqs. (27) have a unique solution can be given algebraically, it is more instructive to give an analytic proof of this fact. First note that we have proved that the equations have at least one solution, since the

quadratic form $q = \sum_{i=0}^{\infty} S_i^2$ has been proved to exist. To prove that this solution is the only solution, note that Eqs. (27) imply, and are in fact equivalent to, Eq. (21); that is how Eq. (27) was derived. Thus if Eqs. (27) had another solution, there would be a quadratic form $\widetilde{q}(S_0, \cdots, S_n)$ with

$$\widetilde{q}(S_0, \cdots, S_n) = S_0^2 + \widetilde{q}(S_1, \cdots, S_{n+1}), \qquad (28)$$

whenever S_{n+1} satisfies Eq. (22). We would then conclude that if the S_i for $i > n$ are defined by the recursion (5), then

$$\widetilde{q}(S_0, \cdots, S_n) = \sum_{i=0}^{N} S_i^2 + \widetilde{q}(S_{N+1}, \cdots, S_{N+n+1}),$$

$$(29)$$

for all $N \geq n$. For we merely keep feeding Eq. (28) back into itself on the right side to go from N to $N+1$.

If $S_0, S_1, \cdots, S_n$ are arbitrary real numbers, we have already observed that $S_i \to 0$ as $i \to \infty$, since the largest root of $g(s) = 0$ is less than 1 in absolute value. Thus, $\widetilde{q}(S_{N+1}, \cdots, S_{N+n+1}) \to 0$ as $N \to \infty$. We conclude

$$\widetilde{q}(S_0, \cdots, S_n) = \sum_{i=0}^{\infty} S_i^2. \qquad (30)$$

Since $\widetilde{q}(S_0, \cdots, S_n)$ has been proved equal to $q(S_0, \cdots, S_n)$ for all real $(S_0, \cdots, S_n)$, $q = \widetilde{q}$, that is, the coefficients of these two forms are the same. Or in other words, Eqs. (27) have a unique solution for the q_{ij}.

One thus proceeds to solve Eqs. (27) for the desired q_{ij}. While the number of unknowns appears staggering (for $n = 50$, there are 1326 unknowns), the special form of Eq. (28) makes machine solution feasible.

6. The Case n = 1

We shall solve Eq. (28) for the case $n = 1$, the case of a quadratic recursion corresponding to a second-order filter. We have

$$\begin{cases} q_{00} = 1 + a_0^2 \, q_{11}\,; \\ q_{01} = a_0 q_{01} + a_0 a_1 q_{11}\,; \\ q_{11} = q_{00} + 2a_1 \, q_{01} + a_1^2 \, q_{11}\,. \end{cases} \qquad (31)$$

Hence

$$q_{01} = \frac{a_0 a_1 q_{11}}{1 - a_0} \qquad (32)$$

(the magnitude of a_0, being the absolute value of the product of the roots of $g(s) = 0$, is less than unity).

We use Eq. (32) and the first of Eqs. (31), substituted into the last equation of (31), to find q_{11}:

$$q_{11} = \frac{1}{1 - a_0^2 - a_1^2 - \left(\dfrac{2a_0 a_1^2}{1 - a_0}\right)}. \qquad (33)$$

Then q_{00} is found as $1 + a_0^2 q_{11}$; similarly Eq. (32) gives q_{01}.

This answer (33) is not especially illuminating; however, it was checked with the answer found by the algebraic method suggested in Sect. 4. Since the quadratic equation $g(s) = 0$ can be solved explicitly, the coefficients can be found in closed form. The two answers, as required, agree.

7. Expected Error

The quantities $S_0, \cdots, S_n$ are random variables, so $q(S_0, \cdots, S_n)$ is not really known when measurements are made. Hence, one is really interested in the expected total square error $s^2 = E(\tau^2)$, where τ^2 is given by Eq. (4). That is,

$$s^2 = E\big(q(S_0, \cdots, S_n)\big). \qquad (34)$$

Since the S_i are independent and identically distributed normal random variables of mean 0 and variance σ^2, then $E(S_i S_j) = 0$ whenever $0 \leq i \neq j \leq n$. The cross-terms in q thus vanish when the expectation is taken, and we conclude

$$s^2 = \sigma^2 \sum_{i=0}^{n} q_{ii}, \qquad (35)$$

since $E(S_i^2) = \sigma^2$, $0 \leq i \leq n$. The quantity $Tr(q) = \sum_{i=0}^{n} q_{ii}$ is called the *trace* of the form q, i.e., the sum of the diagonal elements of the $n \times n$ real symmetric matrix (q_{ij}).

It may be possible to solve Eqs. (27) for $Tr(q)$ without finding all the q_{ij}; however, we have been unable to find such a simplification. For the case $n = 2$, the results of the previous section gives the somewhat uninformative expression

$$s^2 = \sigma^2(1 + a_0^2 q_{11} + q_{11})$$

$$= \sigma^2 \left\{ \frac{2(1 - a_0) - a_1^2(1 + a_0)}{(1 - a_0^2)(1 - a_0) - a_1^2(1 + a_0)} \right\}. \qquad (36)$$

One can verify that the denominator on the right side of Eq. (36) is always positive when $g(s) = 0$ has roots inside the unit circle of the complex plane. One then concludes that

$$\frac{2(1 - a_0) - a_1^2(1 + a_0)}{(1 - a_0^2)(1 - a_0) - a_1^2(1 + a_0)} > 1. \qquad (37)$$

Thus, $s^2 > \sigma^2$. This provides a check on our method, since $q_{00} > 1$ from Eq. (31).

B. Joint Asymptotic Distributions

E. Rodemich

1. Summary

Necessary and sufficient conditions are given for the existence of a joint asymptotic distribution for the maxima of samples of a pair of non-independent random variables. The class of all joint asymptotic distributions is described. A simple estimator for determining such a distribution out of a one-parameter family is given. This method is useful in the extreme-value theory of error-probability estimation (Ref. 3) when used with an "out-of-lock" detector as well as a bit detector.

2. Introduction

Consider a sequence of independent pairs of random variables, (X_j, Y_j), $j = 1, 2, \cdots$, each with the same joint distribution

$$F(x, y) = \Pr \{X_j < x, Y_j < y\}.$$

If the marginal distributions are of "exponential type," the following is known: if $\alpha_n, \beta_n, \gamma_n, \delta_n$ are constants such that

$$U_n = \alpha_n \left[\max (X_1, X_2, \cdots, X_n) - \beta_n\right],$$

$$V_n = \gamma_n \left[\max (Y_1, Y_2, \cdots, Y_n) - \delta_n\right]$$

are random variables with means approaching γ (Euler's constant) and variances approaching $\pi^2/6$ as $n \to \infty$, then

$$\lim_{n \to \infty} \Pr \{U_n < x\} = \exp (-e^{-x})$$

$$\lim_{n \to \infty} \Pr \{V_n < y\} = \exp (-e^{-y})$$

(Ref. 4, p. 166). It will be shown in Theorem 1 that under certain conditions on $F(x, y)$, the joint distribution of U_n and V_n also tends to a limit:

$$\lim_{n \to \infty} \Pr \{U_n < x, V_n < y\} = \exp \left[-(e^{-x} + e^{-y}) w (x - y)\right].$$

The set of all asymptotic distributions of this type is given by Theorem 2. The results are easily transformed to other kinds of marginal distributions, such as the uniform distribution on (0, 1). In the application of this theory, the X_j would be the output of a bit detector in a spacecraft command receiver and the Y_j would be the output of an "out-of-lock" indicator; such a command system has been proposed for *Voyager*.

The last section of this paper considers the problem of estimating the limit distribution by sampling. In practice, a large fixed n might be used, for which

$$\Pr \{U_n < x, V_n < y\} \cong \exp \left[-(e^{-x} + e^{-y}) w (x - y)\right].$$

If a form for the unknown function $w(z)$ is assumed, depending on a certain number of unknown parameters, these parameters may be estimated by various methods from N independent samples of the random variables. The estimate which is asymptotically the best, as $N \to \infty$, is the maximum likelihood estimate (Ref. 5, p. 500). This method, however, requires the solution of complicated equations, with complexity going to infinity with N. A simple estimate is described for one-parameter families of distributions, depending on $\Pr \{|U_n - V_n| < a\}$, where a is a positive constant. Similar estimates may be devised for families with more than one parameter.

These estimates would be used to estimate the probability of getting a correct command into the receiver; for a command of length L is read in correctly if and only if all L bits were detected correctly and the out-of-lock indication never occurred on the L bits. Thus, we need to know $\Pr (U_L < x, V_L < y)$, where, if any U exceeds x, a bit error is made, whereas if any V exceeds y, out-of-lock is detected and the command is rejected.

3. Theorems on Joint Asymptotic Distributions

Theorem 1. Let (X_j, Y_j), $j = 1, 2, \cdots$ be a sequence of independent pairs of random variables such that each pair has the joint cumulative distribution function $F(x,y)$ with marginal distributions of exponential type. Let

$$X_n^* = \max_{1 < j < n} X_j, \quad Y_n^* = \max_{1 < j < n} Y_j.$$

The normalized variables

$$U_n = \alpha_n (X_n^* - \beta_n), V_n = \gamma_n (Y_n^* - \delta_n),$$

with means approaching γ and variances approaching $\pi^2/6$, have a limiting joint distribution as $n \to \infty$, if and only if

$$w(z) = 1 - (e^{-z} + 1)^{-1} \lim_{n \to \infty} n \cdot \Pr \left\{ X_1 > \frac{z}{\alpha_n} + \beta_n, Y_1 > \delta_n \right\}$$

exists for all z. The limit cumulative distribution function of (U_n, V_n) is

$$\lim_{n \to \infty} \Pr \{U_n < x, V_n < y\} = \exp \left[-(e^{-x} + e^{-y}) w (x - y) \right].$$

Proof. We have

$$\Pr \{U_n < x, V_n < y\} = \Pr \left\{ X_n^* < \frac{x}{\alpha_n} + \beta_n, Y_n^* < \frac{y}{\gamma_n} + \delta_n \right\}$$

$$= \Pr \left\{ X_1 < \frac{x}{\alpha_n} + \beta_n, Y_1 < \frac{y}{\gamma_n} + \delta_n \right\}^n$$

$$= \left[1 - \Pr \left\{ X_1 \geq \frac{x}{\alpha_n} + \beta_n \right\} - \Pr \left\{ Y_1 \geq \frac{y}{\gamma_n} + \delta_n \right\} \right.$$

$$\left. + \Pr \left\{ X_1 \geq \frac{x}{\alpha_n} + \beta_n, Y_1 \geq \frac{y}{\gamma_n} + \delta_n \right\} \right]^n$$

By the normalization of U_n, V_n,

$$\Pr \left\{ X_1 \geq \frac{x}{\alpha_n} + \beta_n \right\} = \frac{1}{n} e^{-x} + o \left(\frac{1}{n} \right),$$

$$\Pr \left\{ Y_1 \geq \frac{y}{\gamma_n} + \delta_n \right\} = \frac{1}{n} e^{-y} + o \left(\frac{1}{n} \right).$$

Hence

$$\lim_{n \to \infty} \Pr \{U_n < x, V_n < y\} = \lim_{n \to \infty} \left[1 - \frac{1}{n} (e^{-x} + e^{-y}) + \Pr \left\{ X_1 \geq \frac{x}{\alpha_n} + \beta_n, Y_1 \geq \frac{y}{\gamma_n} + \delta_n \right\} \right]^n$$

$$= \exp \left[-(e^{-x} + e^{-y}) + \lim_{n \to \infty} n \Pr \left\{ X_1 \geq \frac{x}{\alpha_n} + \beta_n, Y_1 \geq \frac{y}{\gamma_n} + \delta_n \right\} \right],$$

where, if one of the three limits exists, they all exist. If

$$n' = \text{integral part of } (ne^y) = ne^{y_1},$$

then

$$y_1 = y + o(1).$$

By the lemma to be given below, applied first to Y_1, then X_1,

$$\lim_{n \to \infty} n \Pr \left\{ X_1 \geq \frac{x}{\alpha_n} + \beta_n, Y_1 \geq \frac{y}{\gamma_n} + \delta_n \right\}$$

$$= \lim_{n \to \infty} n \Pr \left\{ X_1 \geq \frac{x}{\alpha_n} + \beta_n, Y_1 \geq \delta_{n'} \right\}$$

$$= \lim_{n \to \infty} n \Pr \left\{ X_1 \geq \frac{x - y}{\alpha_{n'}} + \beta_{n'}, Y_1 \geq \delta_{n'} \right\}.$$

For fixed y, $n' \sim n e^y$. Hence this limit equals

$$e^{-y} \lim_{n \to \infty} n' \Pr \left\{ X_1 \geq \frac{x - y}{\alpha_{n'}} + \beta_{n'}, Y_1 \geq \delta_{n'} \right\}$$

The values of n' form a sequence of integers tending to $+ \infty$. Hence this last limit can be replaced by

$$e^{-y} \lim_{n \to \infty} n \Pr \left\{ X_1 \geq \frac{x - y}{\alpha_{n'}} + \beta_{n'}, Y_1 \geq \delta_{n'} \right\},$$

assuming that this limit exists. If it does not exist, then the original limit does not exist for all x, y.

Now we have

$$\lim_{n \to \infty} \Pr \{ U_n > x, V_n > y \} = \exp \left\{ - (e^{-x} + e^{-y}) \left[1 - \frac{1}{e^{-(x-y)} + 1} \lim_{n \to \infty} n \Pr \left\{ X_1 \geq \frac{x - y}{\alpha_n} + \beta_n, Y_1 \geq \delta_n \right\} \right] \right\},$$

which is the desired result.

Lemma. Under the hypotheses of Theorem 1, if x is fixed and x', n' depend on n, such that

$$x + \log n = x' + \log n' + o(1),$$

$$\log n - \log n' = 0(1),$$

then for any sequence $\{ y_n \}$,

$$\lim_{n \to \infty} n \Pr \left\{ X_1 \geq \frac{x}{\alpha_n} + \beta_n, Y_1 \geq Y_n \right\}$$

$$= \lim_{n \to \infty} n \Pr \left\{ X_1 \geq \frac{x'}{\alpha_{n'}} + \beta_{n'}, Y_1 \geq y_n \right\}$$

Proof. If $G(x)$ is the marginal distribution of X_1,

$$\left| n \Pr \left\{ X_1 \geq \frac{x}{\alpha_n} + \beta_n, Y_1 \geq Y_n \right\} - n \Pr \left\{ X_1 \geq \frac{x'}{\alpha_{n'}} + \beta_{n'}, Y_1 \geq Y_n \right\} \right| \leq n \left| G \left(\frac{x}{\alpha_n} + \beta_n \right) - G \left(\frac{x'}{\alpha_{n'}} + \beta_{n'} \right) \right|.$$

We have

$$G \left(\frac{x}{\alpha_n} + \beta_n \right) = 1 - \frac{1}{n} e^{-x} + 0 \left(\frac{1}{n} \right),$$

$$G \left(\frac{x'}{\alpha_{n'}} + \beta_{n'} \right) = 1 - \frac{1}{n'} e^{-x'} + 0 \left(\frac{1}{n'} \right).$$

Hence

$$G \left(\frac{x}{\alpha_n} + \beta_n \right) - G \left(\frac{x'}{\alpha_{n'}} + \beta_{n'} \right) = \frac{1}{n} \left(-1 + \exp \left[o(1) \right] + o \left(\frac{1}{n} \right) + o \left(\frac{1}{n'} \right) = o \left(\frac{1}{n} \right),$$

and the above expression tends to zero.

Theorem 2. An asymptotic joint distribution function $F(x, y)$ with marginal distributions $e^{-e^{-x}}, e^{-e^{-y}}$ has the form

$$F(x, y) = \exp \left[- (e^{-x} + e^{-y}) w (x - y) \right] \tag{1}$$

where, for some non-negative measure μ on $(0, \infty)$ with

$$\int_0^\infty d\mu(t) \leq 1,$$

$$\int_0^\infty t \, d\mu(t) \leq 1,$$

$$w(z) = 1 - \frac{1}{e^z + 1}\left[e^z \int_0^{e^{-z}} t \, d\mu(t) + \int_{e^{-z}}^\infty d\mu(t) \right]. \quad (2)$$

Conversely any such measure yields an asymptotic joint distribution function.

Proof. An asymptotic distribution function is its own asymptote. Hence by Theorem 1, it has the form (1), with

$$\lim_{z \to \pm\infty} w(z) = 1. \quad (3)$$

The formula given above for $w(z)$ is a result of the condition that $F(x, y)$ shall be a distribution function. Clearly any distribution function of the form (1) is asymptotic, since

$$F(x + \log n, y + \log n)^n = F(x, y).$$

It may be assumed that $F(x, y)$, hence also $w(z)$, is a C^2 function, since the general case may be obtained from this by taking limits. Then $F(x, y)$ must satisfy the conditions

$$\frac{\partial F}{\partial x}, \; \frac{\partial F}{\partial y}, \; \frac{\partial^2 F}{\partial x \partial y} \geq 0.$$

This implies the following conditions on $w(z)$:

$$(1 + e^z) \, w'(z) - w(z) \leq 0,$$
$$(1 + e^z) \, w'(z) + e^z w(z) \geq 0, \quad (4)$$
$$(1 + e^z) \, w''(z) - (1 - e^z) \, w'(z) \geq 0.$$

Let

$$v(z) = 1 - \frac{1}{e^z + 1}\left[e^z \int_0^{e^{-z}} t \, d\mu(t) + \int_{e^{-z}}^\infty d\mu(t) \right],$$

where $\mu(t)$ is a bounded, non-decreasing C^1 function. Then

$$(1 + e^z)v'(z) - v(z) = -1 + \int_{e^{-z}}^\infty d\mu(t),$$

$$(1 + e^z)v'(z) + e^z v(z) = e^z \left[1 - \int_0^{e^{-z}} t \, d\mu(t) \right], \quad (5)$$

$$(1 + e^z)v''(z) - (1 - e^z)v'(z) = e^{-z} \mu'(e^{-z}).$$

Put $\quad e^{-z}\mu'(z) = (1 + e^z)w''(z) - (1 - e^z)w'(z).$

Integrating,

$$\mu(e^{-z}) = C_1 - (1 + e^z) \, w'(z) + w(z),$$

where C_1 is a constant. By (3) and (4), $\mu(t)$ is non-decreasing, with $C_1 \leq \mu(t) \leq C_1 + 1$.

From (4)

$$\frac{d}{dz}\left[(1 + e^z)w'(z) - w(z) \right]$$

$$= (1 + e^z)w''(z) - (1 - e^z) \, w'(z) \geq 0.$$

Thus the first expression in (4) is non-decreasing. Suppose that it has the value A at $z = z_1$. Then for $z < z_1$,

$$(1 + e^z)w'(z) - w(z) \leq A,$$

$$\frac{d}{dz}\left[(1 + e^{-z})w(z) \right] \geq -Ae^{-z},$$

$$(1 + e^{-z})w(z) \geq -Ae^{-z} + Ae^{-z_1} - (1 + e^{-z_1})w(z_1).$$

As $z \to -\infty$, $w(z) \to 1$. Hence $A \geq -1$, and the first expression in (4) can be bounded on both sides:

$$-1 \leq (1 + e^z)w'(z) - w(z) \leq 0.$$

Thus $\quad \displaystyle\lim_{z \to +\infty} (1 + e^{-z})w'(z) = \lim_{z \to +\infty} e^{-z}w(z) = 0.$

Similarly, $\quad \displaystyle\lim_{z \to -\infty} (1 + e^z)w'(z) = 0.$

By integrating by parts,

$$v(z) = 1 - \frac{1}{e^z + 1}\left\{ \mu(+\infty) - e^z \int_0^{e^{-z}} \mu(t) \, dt \right\}$$

$$= 1 - \frac{1}{e^z + 1}\left\{ C_1 \right.$$

$$+ \lim_{\zeta \to -\infty} \left[-(1 + e^\zeta)w'(\zeta) + w(\zeta) \right] - C_1 + e^z$$

$$\times \int_z^\infty \left[(1 + e^{-\zeta})w'(\zeta) - e^{-\zeta}w(\zeta) \right] d\zeta \Big\}$$

$$= 1 - \frac{1}{e^z + 1}\left\{ 1 + e^z \cdot (1 + e^{-\zeta})w(\zeta) \Big|_z^\infty \right\}$$

$$= 1 - \frac{1}{e^z + 1}\left\{ 1 + e^z \left[1 - (1 + e^{-z})w(z) \right] \right\}$$

$$= w(z).$$

Hence the formula (2) is valid. Comparing (4) with (5), we see that

$$\int_0^\infty d\mu(t) \leq 1,$$

$$\int_0^\infty t\, d\mu(t) \leq 1. \tag{6}$$

Conversely, if $w(z)$ is given by (2), and $\mu(t)$ is a non-decreasing C^1 function satisfying (6), (4) is valid, so that $F(x, y)$ is a distribution function.

Theorem 2a. If the distribution function $F(x, y)$ of Theorem 2 is symmetric, $w(z)$ is an even function, given for positive z by

$$w(z) = 1 - \int_1^{e^{-z}} \frac{t\, d\mu_1(t)}{1 + t} - \frac{1}{1 + e^z}\int_{e^{-z}}^1 d\mu_1(t),$$

where $\mu_1(t)$ can be any non-negative measure on $(0, 1]$ of mass ≤ 1.

Proof. $F(x, y)$ is symmetric if $w(z)$ is an even function. This leads to the condition $t\, d\mu(t) = d\mu(1/t)$. Write the integrals in (2) in terms of integrals over $(0,1]$. This leads to the above formula with $d\mu_1(t) = (1 + t)d\mu(t)$. We have

$$\int_0^\infty t d\mu(t) = \int_0^\infty d\mu(t) = \int_0^1 d\mu_1(t).$$

Hence the inequalities on $\mu(t)$ are equivalent to $\int_0^1 d\mu_1(t) < 1$.

Example (i). If the mass of μ_1 is 1, concentrated at $t = 1$,

$$w(z) = \frac{e^{|z|}}{e^{|z|} + 1}.$$

Then $F(x, y) = \exp\left[-e^{-\min(x, y)}\right]$.
Hence $X = Y$ with probability 1.

Example (ii). If the mass of μ_1 is 1, concentrated at $t = e^{-z_1}$, $z_1 > 0$,

$$w(z) = \begin{cases} \dfrac{e^{z_1}}{1 + e^{z_1}}, & |z| \leq z_1, \\[2ex] \dfrac{e^z}{1 + e^{|z|}}, & |z| \geq z_1. \end{cases}$$

Putting $e^{-q} = e^{z_1}/(1 + e^{z_1})$,

$$F(x, y) = \begin{cases} e^{-e^{-(x+q)}}e^{-e^{-(y+q)}}, & |x - y| \leq z_1 \\[1ex] \exp\left[-e^{-\min(x, y)}\right], & |x - y| \geq z_1 \end{cases}$$

Within the strip $|x - y| < z_1$, the distribution is the same as in the independent case, with each variable translated by q. $|x - y| = z_1$ with positive probability, and $|x - y| > z_1$ with zero probability.

Example (iii). If

$$d\mu_1 = \frac{8c}{(1 + t)^2}\, dt, c \leq \frac{1}{4},$$

$$w(z) = 1 - \frac{4ce^z}{(e^z + 1)^2} = 1 - c\operatorname{sech}^2\frac{z}{2}.$$

This is probably the simplest class of $w(z)$ which leads to a continuous probability density depending on a single parameter.

4. Estimation of a Symmetric Asymptotic Distribution

Suppose that a pair of random variables X, Y have a joint distribution of the asymptotic type of Theorem 1:

$$\Pr\{X < x, Y < y\} = \exp\left[-(e^{-x} + e^{-y})\, w(x - y)\right].$$

For any $w(z) \neq 1$, there is an associated family of possible functions $w(z)$, depending linearly on a parameter c:

$$w(z) = 1 - c\, u(z), \tag{7}$$

obtained by varying the normalization of the measure μ of Theorem 2. On the basis of a set of independent measurements of X and Y, a "best" value for c may be obtained in various ways. Assuming that $w(z)$ is actually given by (7), for some c, the value for c determined from sampling, $\hat{c}$, is a random variable whose probability distribution depends on c.

If $\hat{c}$ is determined from a large number N of samples, $\hat{c} \rightarrow c$ most rapidly when $\hat{c}$ is the maximum likelihood estimate. Then the variance of $\hat{c}$ is asymptotically equal to

$$\frac{1}{N}\left\{-E\left[\frac{\partial^2}{\partial^2 c}(\log F_{xy})\right]\right\}^{-1}$$

(Ref. 6, p. 236), assuming that $F(x, y)$ is sufficiently differentiable. $N \cdot \operatorname{Var}(\hat{c})$ was computed numerically as a function of c for the distribution of Example (iii) above. This is shown in Fig. 1.
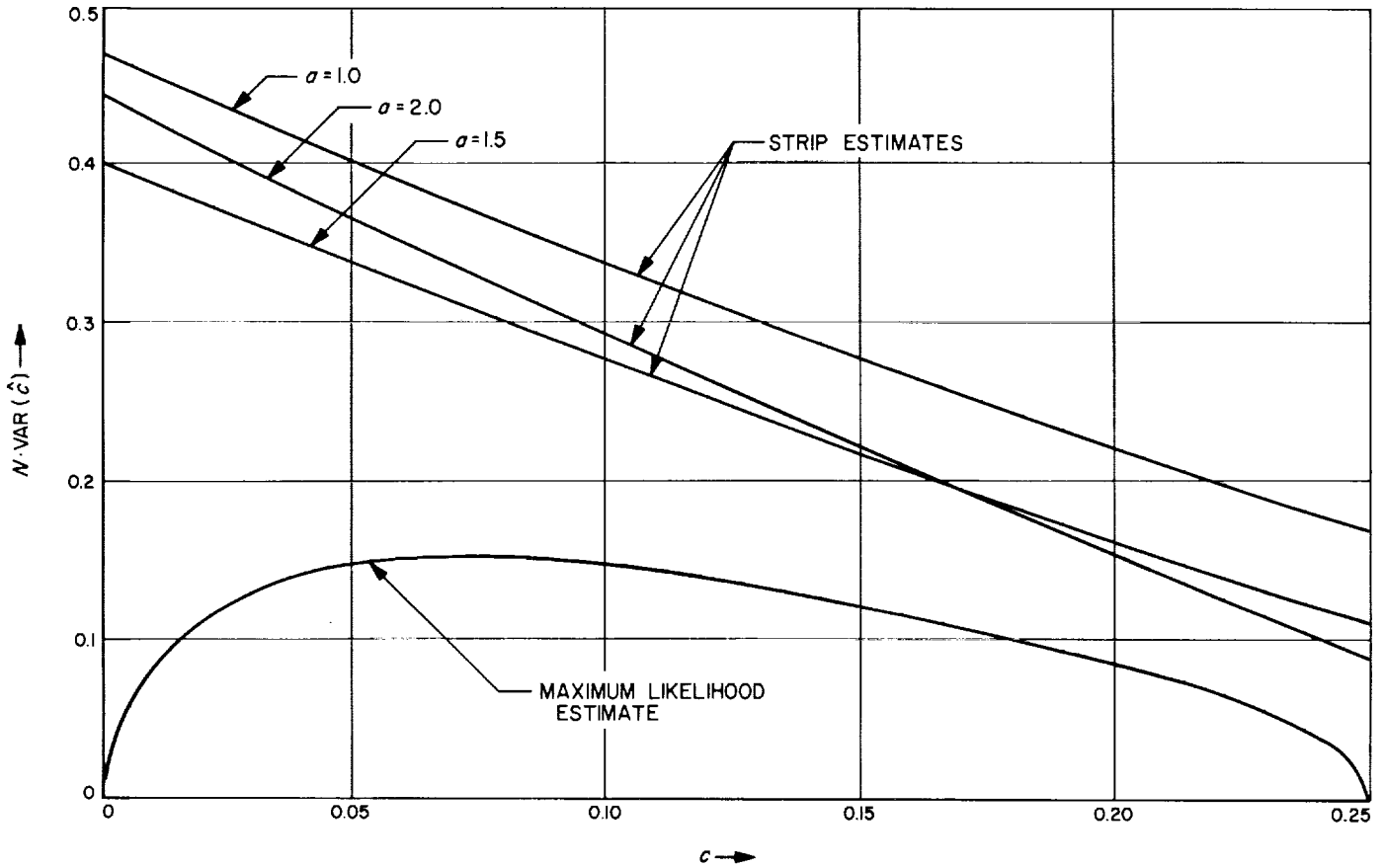
Fig. 1. Variance of estimators for the parameter c for the family of distributions (iii)

Another method for finding $\hat{c}$ depends on the expression for the probability that $|X - Y| < a$, for some positive constant a. We have

$$\Pr \{b < X - Y < a\}$$

$$= \int_{-\infty}^{\infty} dy \int_{y+b}^{y+a} d_x F_\nu(x, y)$$

$$= \int_{-\infty}^{\infty} dy \left[F_\nu(y + a, y) - F_\nu(y + b, y) \right]$$

$$= \int_{-\infty}^{\infty} dy \, e^{-y} \left[w(z) + (e^{-z} + 1)w'(z) \right] e^{-e^{-y} (e^{-z}+1) w(z)} \Big|_{z_-b}^{a}$$

$$= \left(\frac{1}{e^{-z} + 1} + \frac{w'(z)}{w(z)} \right) \Big|_{z_-b}^{a} .$$

In particular,

$$\Pr \{|x - y| < a\} = \frac{e^a - 1}{e^a + 1} + 2 \frac{w'(a)}{w(a)} ,$$

if $w(z)$ is an even function. If $v_N(a)$ is the number of times $|X - Y| < a$ in N samples, $N^{-1} v_N(a)$ is an approximation to this probability. Thus we are led to the estimate $\hat{c}$ given by

$$N^{-1} v_N(a) = \frac{e^a - 1}{e^a + 1} - \frac{2\hat{c}u'(a)}{1 - \hat{c}u(a)} .$$

By the law of large numbers, this tends to $\Pr\{|X - Y| < a\}$ with probability 1. Hence $\hat{c} \to c$. Putting

$$\xi_N = N^{-1} v_N(a),$$

we have

$$\hat{c} = \frac{\tanh \frac{1}{2} a - \xi_N}{2u'(a) + u(a) \cdot \left(\tanh \frac{1}{2} a - \xi_N \right)} = g(\xi_N) .$$

ξ_N has mean $p = \Pr\{|x - y| < a\}$, and for large N it is approximately normal, with variance $(1/N)p(1 - p)$. Hence

$$E(\hat{c}) = E\left[g(p) + (\xi_N - p)g'(p) + \tfrac{1}{2}(\xi_N - p)^2 g''(p)\right]$$
$$+ o\left(\frac{1}{N}\right)$$
$$= c + \frac{1}{2N}p(1 - p)g''(p) + o\left(\frac{1}{N}\right),$$

$$E\{(\hat{c} - c)^2\} = E\left[(\xi_N - p)^2 g'(p)^2\right] + o\left(\frac{1}{N}\right)$$
$$= \frac{1}{N}p(1 - p)g'(p)^2 + o\left(\frac{1}{N}\right)$$

It follows that as $N \to \infty$,

$$N \, \mathrm{Var}(\hat{c}) \sim p(1 - p)g'(p)^2 = \frac{1}{4}p(1 - p)\left[\frac{w(a)^2}{w'(a)}\right]^2.$$

For various values of a, this is plotted as a function of c in Fig. 1, for

$$w(z) = 1 - \frac{4ce^z}{(1 + e^z)^2}.$$

Notice that the variance does not depend very critically on the value of a. For $1.5 < a < 2$, it is approximately twice the variance of the maximum likelihood estimate over much of the range of c. Hence, this estimator is a good one to use to avoid having to solve the likelihood equations.

C. Independent Bivariate Extreme-Value Distributions

T. Pitcher

1. Summary

Let $(x_1, x_2, \cdots)$ be independent random variables with distribution function $G(x)$ and let $(y_1, y_2, \cdots)$ be independent random variables with distribution function $H(y)$. Let the joint distribution of x_i and y_i be $F(x, y)$. Then the distribution of $X_n = \max(x_1, \cdots, x_n)$ is $G^n(x)$, the distribution of $Y_n = \max(y_1, \cdots, y_n)$ is $H^n(y)$, and the joint distribution of X_n, Y_n is $F^n(x, y)$. We wish to determine when, and in what sense, X_n and Y_n become

independent as $n \to \infty$. Such knowledge would be useful in applying extreme-value theory to *Voyager* type command detectors; see the preceding article by Rodemich.

For a joint distribution $F(x, y)$ with margins $G(x)$ and $H(y)$, Sibuya (Ref. 7) has introduced the function $\Omega_F(s,t)$ defined on $[0 \le s \le 1, 0 \le t \le 1]$ by

$$\Omega_F(G(x), H(y)) = \frac{F(x, y)}{G(x)\,H(y)}.$$

Then $(\Omega_F - 1)$ is a measure of the difference between F and the corresponding independent distribution for x and y. A precise theorem is proved in this direction.

2. Properties of Ω_F

We have

$$\Omega_{F^n}(G^n(x), H^n(y)) = \frac{F^n(x, y)}{G^n(x)\,H^n(y)} = \Omega_F^n(G(x), H(y))$$

so that

$$\Omega_{F^n}(s, t) = \Omega_F(s^{1/n}, t^{1/n}).$$

If U and V are strictly monotone functions and $\widetilde{F}(x, y) = F(U^{-1}(x), V^{-1}(y))$ then $\widetilde{F}$ has the margins $\widetilde{G}(x) = G(U^{-1}(x))$ and $\widetilde{H}(y) = H(V^{-1}(y))$, and

$$\Omega_F(G(x), H(y)) = \Omega_{\widetilde{F}}(G(U^{-1}(U(x))), H(V^{-1}(V(y))))$$
$$= \frac{\widetilde{F}(U(x), V(y))}{\widetilde{G}(U(x))\widetilde{H}(V(y))} = \frac{F(x, y)}{G(x)H(y)}$$
$$= \Omega_F(G(x), H(y)).$$

Hence the function Ω_F is invariant under the transformation $\widetilde{x} = U(x), \widetilde{y} = V(y)$. In particular by using the transformation $\widetilde{x} = G(x), \widetilde{y} = H(y)$, we can reduce the problem to the case in which the margins are uniformly distributed on $[0, 1]$ if G and H are strictly monotone. The corresponding transformation for densities is given by

$$\frac{\partial^2 \widetilde{F}}{\partial x \partial y}(x, y) = \frac{\partial^2 F}{\partial x \partial y}(G^{-1}(x), H^{-1}(y))/G'(G^{-1}(x))\,H'(H^{-1}(y)),$$

assuming that the derivatives exist.

3. Conditions for Independence

In the following we shall assume, unless otherwise stated, that the above transformation has been performed and hence that the marginal distributions are uniform.

Let

$$P(x,y) = 1 + F(x,y) - x - y.$$

Sibuya (Ref. 7) has proved that if

$$P(1-\epsilon, 1-\epsilon) = o(\epsilon)$$

then

$$\lim_{n \to \infty} \Omega_{F^n}(s,t) = 1 \text{ for all } s \text{ and } t.$$

We are therefore motivated to prove the following theorem:

Theorem 1. Suppose $F(x,y) = \int_0^x \int_0^y f(u,v) \, du \, dv$. If, for some $a < 1$ either

$$\int_a^1 \int_a^1 \frac{f(x,y)}{(1-x)^{1/2}(1-y)^{1/2}} \, dx \, dy < \infty \qquad (1)$$

or

$$\int_a^1 \int_a^1 f^2(x,y) \, dx \, dy < \infty \qquad (2)$$

then $P(1-\epsilon, 1-\epsilon) = o(\epsilon)$; hence, asymptotic independence holds.

Proof. In the first case

$$P(1-\epsilon, 1-\epsilon) \le \epsilon \int_{1-\epsilon}^1 \int_{1-\epsilon}^1 \frac{f(x,y)}{(1-x)^{1/2}(1-y)^{1/2}} \, dx \, dy = o(\epsilon) \,,$$

and in the second case

$$P(1-\epsilon, 1-\epsilon) = \int_{1-\epsilon}^1 \int_{1-\epsilon}^1 f(x,y) \, dx \, dy$$

$$< \left[\int_{1-\epsilon}^1 \int_{1-\epsilon}^1 f^2(x,y) dx \, dy \right]^{1/2} \left[\int_{1-\epsilon}^1 \int_{1-\epsilon}^1 dx \, dy \right]^{1/2}$$

$$= \epsilon \left[\int_{1-\epsilon}^1 \int_{1-\epsilon}^1 f^2(x,y) dx \, dy \right] = o(\epsilon) \,.$$

We then have in the general case

Theorem 2. In the general case (i.e., without assuming $G(x) = x$, $H(y) = y$) $\Omega_{F^n} \to 1$ if for some A and B with $G(A) < 1$ and $H(B) < 1$ either

$$\int_A^\infty \int_B^\infty \frac{f(x,y)}{(1-G(x))^{1/2}(1-H(y))^{1/2}} \, dx \, dy < \infty \qquad (1)$$

or

$$\int_A^\infty \int_B^\infty \frac{f^2(x,y)}{g(x) \, h(y)} \, dx \, dy < \infty \qquad (2)$$

where we have written

$$f(x,y) = \frac{\partial^2 F}{\partial x \partial y}(x,y) \,, \quad g(x) = \frac{dG(x)}{dx} \,, \quad h(y) = \frac{dH(y)}{dy} \,.$$

Proof. This follows from the theorem above on applying the transformation $\tilde{x} = G(x)$, $\tilde{y} = H(y)$.

4. Asymptotic Orthogonality

If the x_i and y_j are uniformly distributed, then the random variable.

$$\xi_n = (X_n)^n \,, \quad \eta_n = (Y_n)^n$$

are also uniformly distributed. Writing F_n for the joint distribution of (ξ_n, η_n) we have

$$F_n(s,t) = F^n\left(s^{\frac{1}{n}}, t^{\frac{1}{n}}\right) = st \, \Omega_{F^n}(s,t) \,.$$

Theorem 3. If $\Omega_{F^n}(s,t) \to 1$ for all s and t then ξ_n and η_n are asymptotically orthogonal.

Proof. ξ_n and η_n each have mean ½ and variance $\frac{1}{12}$ so it only remains to show that

$$\int_0^1 \int_0^1 xy \, dF_n(x,y) \to \frac{1}{4} = \int_0^1 \int_0^1 xy \, dx \, dy$$

and this is an immediate consequency of the continuity of xy and the fact that $F_n(x,y) \to xy$ for all x and y.

5. Uniform Marginals

Morgenstern (Ref. 8) and Plackett (Ref. 9) among others, have introduced one-parameter families of joint distributions. In the case of uniform margins Morgenstern's, is given by

$$F_a(x,y) = xy \, (1 + a(1-x)(1-y)) \qquad 0 \le a \le 1 ;$$

$a = 0$ is the independent case. Since

$$P_a(1-\epsilon, 1-\epsilon) = F_a(1-\epsilon, 1-\epsilon) - 1 + 2\epsilon$$

$$= (1-\epsilon)^2(1+a\epsilon^2) - 1 + 2\epsilon = 0(\epsilon^2) = o(\epsilon),$$

$\Omega_{F^n_a}(x,y) \to 1$ for all x and y and all a.

Plackett's family of distributions is defined for all $0 \le a \le \infty$ by

$$a = \frac{F_a(x,y)(1-x-y+F_a(x,y))}{(x-F_a(x,y)(y-F_a(x,y)))} \,.$$

Here $\alpha = 0$ corresponds to $x = 1-y$, $\alpha = \infty$ corresponds to $x=y$ and $\alpha=1$ gives the independent case. If $0 < \alpha < \infty$

$$P_\alpha(x,y) = 1 - x - y + F_\alpha(x,y)$$

$$= \frac{\alpha(x-F_\alpha(x,y))(y-F_\alpha(x,y))}{F_\alpha(x,y)}$$

so

$$P_\alpha(1-\epsilon, 1-\epsilon) = 0\left((1-\epsilon - H(1-\epsilon, 1-\epsilon))^2\right).$$

It is known (Ref. 10) that any distribution F with uniform margins must satisfy

$$\max(x+y-1, 0) \leq F(x,y) \leq \min(x,y).$$

In our case this implies

$$1 - 2\epsilon \leq F_\alpha(1-\epsilon, 1-\epsilon) \leq 1-\epsilon$$

or

$$\epsilon \leq 1-\epsilon - F_\alpha(1-\epsilon, 1-\epsilon) \leq 0$$

so that

$$P_\alpha(1-\epsilon, 1-\epsilon) = 0(\epsilon^2) = o(\epsilon).$$

Hence, Plackett's distributions also satisfy Sibuya's condition.

Unfortunately, data taken by Ashlock show that asymptotic independence does not hold in the command detector case. Hence, a method such as in the above paper is needed to find probabilities.

References

1. Hannan, E. J., *Time Series Analysis*, pp. 14–18, Meuthen, London, 1960.

2. Zierler, Neal, "Linear Recurring Sequences," *Journal of the Society for Industrial Applications of Mathematics*, Vol. 17: pp. 31–48, 1959.

3. Ashlock, J. C., and Posner, E. C., *Extreme-Value Theory Used for Bit Error Probability Estimation in the Ranger Command Detector*, TR 32-705, Jet Propulsion Laboratory, Pasadena, January 15, 1965.

4. Gumbel, E., *Statistics of Extremes*, Columbia University Press, New York, 1958.

5. Cramér, H., *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

6. Mood, A., and Graybill, F., *Introduction to the Theory of Statistics*, Second Edition, McGraw-Hill, New York, 1963.

7. Sibuya, M., "Bivariate Extreme Statistics I," *Annals of the Institute for Statistical Mathematics*, Tokyo, Vol. II: p. 195–210, 1960.

8. Morgenstern, D., "Einfache Beispiele Zweidimensionaler Verteilungen," *Mitteilungsblatt für Mathematische Statistik*, Vol. 8: p. 234–235, 1956.

9. Plackett, R. L., "A Class of Bivariate Distributions," *Journal of the American Statistical Association*, Vol. 60: p. 516–522, 1965.

10. Frechet, M., "Sur les tableux de corrélation dont les marges sout donnees," *Annales de l'Universite de Lyon*, Sect. A, Series 3, Vol. 14: p. 53–77, 1951.

XXIV. Communications Systems Research: Combinatorial Communications

A. A Problem on Rook Domains

E. Rodemich

1. Introduction

Suppose we have a k-dimensional array of non-negative integers, of length n in each direction. The rook domain $R(P)$ of a point P in the array with coordinates $(j_1, \cdots, j_k)$ is defined to be the set of all points Q whose coordinates $(j'_1, \cdots, j'_k)$ differ from those of P in, at most, one position. Denote the element of the array at P by $f(P)$. Then we have the following:

a. Conjecture. If the array has the property that whenever $f(P) = 0$,

$$\sum_{Q \in R(P)} f(Q) \geq n,$$

then

$$\sum f(P) \geq n^k/k \quad \text{(Ref. 1)}.$$

Various proofs of this are known for $k = 2$, and for general k with $n \leq k$.

There is a generalization of this conjecture. For $f(P) = 0$,

$$\sum_{Q \in R(P)} f(Q) \geq n,$$

is the sum over all one-dimensional coordinate planes through P of the sum of the elements in each plane. Define $S^j(P)$, for $1 \leq j \leq k - 1$, to be the sum of the elements in each j-dimensional coordinate plane through P. The number of such planes is

$$\binom{k}{j}.$$

b. Conjecture. For some j, $1 \leq j \leq k - 1$, if the array has the property that whenever $f(P) = 0$,

$$S^j(P) \geq \frac{n^j}{k}\binom{k}{j},$$

then $\sum f(P) \geq n^k/k$.

The following results are proved here:

c. Theorem 1. If $n \leq 2k$, and whenever $f(P) = 0$, $S^1(P) \geq n$, then $\sum f(P) \geq n^k/k$.

d. Theorem 2. For any n, k, if whenever $f(P) = 0$, $S^{k-1}(P) \geq n^{k-1}$, then $\sum f(P) \geq n^k/k$.

e. Theorem 3. For any n, k, and any j with $2 \leq j \leq k - 2$, if

$$S^j(P) \geq \frac{n^j}{k}\binom{k}{j}$$

for $f(P) = 0$, then $\sum f(P) > n^k/k \times (1 - 1/4k)$. If $j = 1$, $\sum f(P) > n^k/(k + \frac{1}{4})$.

Only a weakened form of the hypothesis is necessary for Theorem 1. For any n, k, j, we have the following progressively weaker hypotheses:

(i)   For $f(P) = 0$, $S^j(P) \geq \dfrac{n^j}{k}\dbinom{k}{j}$.

(ii) $\displaystyle\sum_{P, f(P)=0} S^j(P) \geq \frac{n^j}{k}\binom{k}{j} \sum_{P, f(P)=0} 1$.

(iii) For some j-dimensional direction, let $\sigma^j(P)$ denote the sum of elements in the j-plane in this direction through P.

Then

$$\sum_{P, f(P)=0} \sigma^j(P) \geq \frac{n^j}{k} \sum_{P, f(P)=0} 1.$$

Theorem 1 and the first part of Theorem 3 are proved under hypothesis (iii).

2. Proofs of Theorems 1–3

Proof of Theorem 1. We have

$$\sum_{P, f(P)=0}\left[\sigma^1(P) - \frac{n}{k}\right] \geq 0.$$

Let all the rows in the given direction be numbered. Let the jth row contain n_j non-zero elements, with sum s_j. Then the jth row contributes $n - n_j$ terms to the above sum. Hence

$$\sum_j (n - n_j)\left(s_j - \frac{n}{k}\right) \geq 0.$$

$n_j = 0$ implies $s_j = 0$. $s_j < 2$ implies $n_j \leq 1$. Hence

(i) if $s_j - n/k \geq 0$, $n_j \geq 1$, which implies $n - n_j \leq n - 1$, and $(s_j - n/k)(n - n_j) \leq (n - 1)(s_j - n/k)$;

(ii) if $s_j - n/k < 0$, since $n \leq 2k$, $s_j < 2$, which implies $n_j \leq 1$, $n - n_j \geq n - 1$, $(s_j - n/k)(n - n_j) \leq (n - 1) \times (s_j - n/k)$.

Every term in this sum falls under one of these cases. Hence

$$(n - 1)\sum_j\left(s_j - \frac{n}{k}\right) \geq \sum_j (n - n_j)\left(s_j - \frac{n}{k}\right) \geq 0,$$

and

$$\sum_j S_j \geq \frac{n}{k}\sum_j 1 = \frac{n^k}{k}.$$

Proof of Theorem 2. Let $\sigma_{l,m}$ be the sum of elements in the hyperplane $\{(j_1, \cdots, j_k), j_l = m\}$.

Pick the minimum hyperplane sum. The hypothesis is unchanged under a permutation of directions, and of coordinates in each direction. Hence we may assume that $\sigma_{1,1}$ is minimum.

Consider the n^{k-2} diagonals in this plane, each generated by displacing a fixed element $(1, j_2, \cdots, j_k)$:

$$\{(1, j_2 + t, j_3 + t, \cdots, j_k + t), \qquad t = 0, 1, \cdots, n - 1\},$$

where the coordinates are evaluated modulo n. Let s be the minimum sum of one of these diagonals. If $s \geq n/k$, then, $\sigma_{1,1} \geq n^{k-1}/k$, $\sum f(P) \geq n^k/k$. Hence we may assume $s < n/k$.

By permutation of the coordinates, we may assume that s is the sum of the elements at

$$\{(1, t, \cdots, t), \qquad t = 1, \cdots, n\},$$

and, if there are z zeros on this diagonal, they occupy the first z positions. Then we have

$$\sigma_{1,1} + \sum_{j=2}^{k} \sigma_{j,t} \geq n^{k-1}, \qquad t = 1, 2, \cdots, z,$$

$$z\sigma_{1,1} + \sum_{j=2}^{k}\sum_{t=1}^{z} \sigma_{j,t} \geq zn^{k-1}.$$

If $\sum f(P) = S$,

$$\sum_{t=1}^{z} \sigma_{j,t} = S - \sum_{t=z+1}^{n} \sigma_{j,t} \leq S - (n - z)\sigma_{1,1},$$

by the minimal property of $\sigma_{1,1}$. Hence

$$z\sigma_{1,1} + (k-1)[S - (n-z)\sigma_{1,1}] \geqq zn^{k-1},$$

$$(k-1)S + k\left[z - n\left(1 - \frac{1}{k}\right)\right]\sigma_{1,1} \geqq zn^{k-1}.$$

Since $s < n/k$, $z > n(1 - 1/k)$. Also, $\sigma_{1,1} \leqq S/n$. Hence

$$(k-1)S + k\left[z - n\left(1 - \frac{1}{k}\right)\right]S/n \geqq (k-1)S + k$$

$$\times \left[z - n\left(1 - \frac{1}{k}\right)\right]\sigma_{1,1} \geqq zn^{k-1},$$

or

$$kzS/n \geqq zn^{k-1},$$

$$S \geqq n^k/k.$$

A different proof of this result is given in *SPS 37-35*, Vol. IV, pp. 304–305.

Proof of Theorem 3. (i) Assume hypothesis (iii). Let the j-dimensional planes in the given direction be numbered from 1 to n^{k-j}, and let the rth plane have n_r non-zero elements of sum s_r. Then, as in the proof of Theorem 1,

$$\sum_{r=1}^{n^{k-j}} (n^j - n_r)\left(s_r - \frac{n^j}{k}\right) \geqq 0,$$

$$n^j \sum_r s_r \geqq \frac{n^j}{k} \sum_r (n^j - n_r) + \sum_r n_r s_r.$$

Let

$$\sum s_r = S, \sum n_r = N.$$

Then since $n_r \leqq s_r$, $N \leqq S$. Also,

$$\sum n_r s_r \geqq \sum n_r^2 \geqq \left(\sum n_r\right)^2/n^{k-j},$$

by Schwarz' inequality. Hence

$$n^j S \geqq \frac{n^j}{k}(n^k - N) + N^2/n^{k-j}.$$

The right side of this inequality is a quadratic in N which assumes its minimum at $N = n^k/(2k)$. Hence

$$n^j S \geqq \frac{n^j}{k}\left(n^k - \frac{n^k}{2k}\right) + \frac{n^{j+k}}{4k^2} = \frac{n^{j+k}}{k}\left(1 - \frac{1}{4k}\right),$$

$$\sum_P f(P) = S \geqq \frac{n^k}{k}\left(1 - \frac{1}{4k}\right).$$

Equality cannot hold here. If it did, we would have $s_r = n_r$ for all r, $S = N = n^k/(2k)$, which is impossible.

This inequality is slightly weaker than the result to be proved for $j = 1$.

(ii) Now let $j = 1$. For $f(P) = 0$, we have

$$S^1(P) = \sum_{l=1}^{k} \sigma^l(P) \geqq n,$$

where $\sigma^l(P)$ is the sum of the row through P in the lth direction. This implies

$$S^1(P) \geqq n + \sum_{l=1}^{k}(\sigma^l(P) - n)_+,$$

where the subscript $+$ denotes that the term in the sum is to be counted only if it is positive; for, if any terms in the sum are non-zero,

$$n + \sum_{l=1}^{k}(\sigma^l(P) - n)_+ \leqq \sum_{l=1}^{k} \sigma^l(P),$$

independently of the hypothesis.

Sum $S^1(P)$ over all points P of the array. Each row is counted n times, so

$$\sum_P S^1(P) = nkS,$$

where $S = \sum f(P)$. Break the sum up into the sum over zeros and the sum over non-zeros. If all the kn^{k-1} rows are numbered, and the rth row has n_r positive elements of sum s_r,

$$\sum_{f(P)=0} S^1(P) \geqq n \sum_{f(P)=0} 1 + \sum_r (n - n_r)(s_r - n)_+,$$

$$\sum_{f(P)\neq 0} S^1(P) = \sum_r n_r s_r.$$

Hence if N denotes the number of positive elements,

$$nkS \geqq n^{k+1} - nN + \sum_r (n - n_r)(s_r - n)_+ + \sum_r n_r s_r. \tag{1}$$

The integers n_r, s_r in Eq. (1) satisfy the relations

$$\sum_r s_r = kS \,, \tag{2}$$

$$\sum_r n_r = kN \,, \tag{3}$$

$$s_r \geqq n_r \geqq 0 \,, \qquad n_r \leqq n \,. \tag{4}$$

If we minimize S, subject only to the conditions of Eqs. (1)–(4), where the s_r and n_r are independent variables, the minimum will be no larger than the actual minimum of $\sum f(P)$.

First note that the right side of Eq. (1) may be rewritten as

$$n^{k+1} - nN + \sum_{s_r \geqq n} [ns_r - n(n - n_r)] + \sum_{s_r < n} n_r s_r \,.$$

By Eq. (4), this expression is not increased if an $s_r > n$ is decreased by 1 and an $s_r < n$ is increased by 1. This operation does not disturb Eqs. (2), (3), or (4). Hence by repeated application of this process, we reach a set of s_r's for which either $s_r \geqq n$ for all r, or $s_r \leqq n$ for all r. In the first case, we have $S \geqq n^k$. Thus we may assume that this does not occur.

Our minimum problem is now reduced to

$$nkS \geqq n^{k+1} - nN + \sum_r n_r s_r \,, \tag{1'}$$

$$\sum_r s_r = kS \,, \tag{2}$$

$$\sum_r n_r = kN \,, \tag{3}$$

$$n \geqq s_r \geqq n_r \geqq 0 \,. \tag{4'}$$

Eliminating N, we have

$$nkS \geqq n^{k+1} + \sum_r n_r \left(s_r - \frac{n}{k} \right) \,.$$

Varying the n_r's independently, subject to Eq. (4'), the right side is a minimum when

$$n_r = \begin{cases} 0, & s_r \geqq \dfrac{n}{k} \,, \\[2mm] s_r, & s_r < \dfrac{n}{k} \,. \end{cases}$$

Thus

$$nkS \geqq n^{k+1} + \sum_{s_r < n/k} s_r \left(s_r - \frac{n}{k} \right) \,,$$

$$kS = \sum s_r, \qquad 0 \leqq s_r \leqq n \,.$$

To reduce the problem to one which is practical to solve, replace the sums by integrals and s_r by an increasing function. This further reduces the minimum. Putting $m = kn^{k-1}$, we have

$$kS = \int_0^m s(t)\, dt, \qquad 0 \leqq s(t) \leqq n \,,$$

$$nkS \geqq n^{k+1} + \int_{s(t) < n/k} s(t) \left[s(t) - \frac{n}{k} \right] dt \,.$$

In the interval where $s(t) \geqq n/k$, $s(t)$ may be replaced by a step function taking the two values n/k and n. Then, on the interval on which $s(t) \leqq n/k$, $s(t)$ may be replaced by its average value μ, since the second integrand is a convex function. This reduces the problem to

$$kS = p\mu + (m - p)\, n \,, \tag{5}$$

$$nkS \geqq n^{k+1} + p\mu \left(\mu - \frac{n}{k} \right) \,, \tag{6}$$

$$0 \leqq p \leqq m \,,$$

$$0 \leqq \mu \leqq \frac{n}{k} \,,$$

with the independent variables μ, p. Expressing μ in terms of p,

$$\mu = \frac{1}{p} \left[kS - (m - p)\, n \right] \,,$$

$$nkS \geqq n^{k+1} + (kS - mn) \left(2n - \frac{n}{k} \right)$$

$$+ \frac{(kS - mn)^2}{p} + pn \left(n - \frac{n}{k} \right) \,. \tag{7}$$

From Eq. (6), if $\mu = 0$ or n/k, $S \geqq n^k/k$. Hence if S has a smaller minimum, this minimum value and the corresponding value of p give equality in Eq. (7). Extending the range of p to $(0, \infty)$ may further reduce

the minimum. The right side of Eq. (7) has its minimum for fixed S, and $p > 0$, at

$$p = \frac{mn - kS}{n\sqrt{1 - \frac{1}{k}}} = \frac{k(n^k - S)}{n\sqrt{1 - \frac{1}{k}}}.$$

Hence

$$nkS \geq n^{k+1} + k(S - n^k)\left(2n - \frac{n}{k}\right)$$
$$+ 2kn(n^k - S)\sqrt{1 - \frac{1}{k}},$$

which implies

$$S \geq 2n^k \frac{1 - \sqrt{1 - \frac{1}{k}}}{2 - \sqrt{1 - \frac{1}{k}}}.$$

We have

$$\frac{2 - \sqrt{1 - \frac{1}{k}}}{1 - \sqrt{1 - \frac{1}{k}}} = \frac{1 + \frac{1}{k} + \sqrt{1 - \frac{1}{k}}}{\frac{1}{k}}$$

$$< \frac{1 + \frac{1}{k} + 1 - \frac{1}{2k}}{\frac{1}{k}} = 2\left(k + \frac{1}{4}\right).$$

Therefore

$$S \geq \frac{n^k}{k + \frac{1}{4}}.$$

3. Cubes With a Restricted Number of Distinct Positive Elements

The first theorem below tells us that the conjectures are true for cubes which do not contain more than one distinct positive element. This suggests considering next, cubes containing only two distinct positive elements. Theorem 5, however, relates this case to a generalization of the original case, in which the positive elements may be any real numbers ≥ 1.

Note that the proofs of Theorems 1–4 carry through if the positive elements are any real numbers ≥ 1.

a. Theorem 4. Under hypothesis (iii), if all positive elements of the cube have the same value, then $\sum f(P) \geq n^k/k$. Another proof was found in *SPS 37-16*, Vol. IV, pp. 35–36. More generally, our proof is valid if the sum of a j-dimensional hyperplane and the number of non-zero elements in it vary together: i.e., $s_p \geq s_r$ implies $n_p \geq n_r$.

Proof: As in the proof of Theorem 3, we have

$$n^j \sum_r s_r \geq \frac{n^j}{k} \sum_r (n^j - n_r) + \sum_r n_r s_r.$$

The average value of s_r is

$$\bar{s} = \frac{\sum s_r}{\sum 1} = n^{j-k} \sum s_r.$$

Since n_r varies with s_r, there is a value $\bar{n}$ such that

$$s_r > \bar{s} \rightarrow n_r \geq \bar{n},$$

$$s_r < \bar{s} \rightarrow n_r \leq \bar{n}.$$

Then

$$\sum (n_r - \bar{n})(s_r - \bar{s}) \geq 0,$$

$$\sum n_r s_r \geq \bar{s} \sum n_r + \bar{n} \sum (s_r - \bar{s}) = \bar{s} \sum n_r,$$

or

$$n^{k-j} \sum n_r s_r \geq (\sum n_r)(\sum s_r).$$

Combining this with the above inequality,

$$n^k \sum s_r \geq \frac{n^k}{k} \sum (n^j - n_r) + (\sum n_r)(\sum s_r),$$

or

$$\sum (n^j - n_r)\left(\sum s_r - \frac{n^k}{k}\right) \geq 0.$$

The first factor on the left is non-negative. If it is zero, $\sum s_r \geq \sum n_r = n^k$. Otherwise,

$$\sum s_r \geq \frac{n^k}{k}.$$

b. Lemma. If $r = p/q$ is a rational number with $1 \leq r \leq M$, M an integer ≥ 2, then there is a k-cube of integers of side $N = q(M - 1)$ such that each entry is 1 or M, and the sum of each row is Nr.

Proof: Define t by

$$rN = tM + N - t.$$

Then

$$t = \frac{N(r-1)}{M-1} = p - q.$$

Hence t is an integer, and $0 \le t \le N$. Define the function $f(i_1, \cdots, i_k)$ on the k-cube $1 \le i_1, \cdots, i_k \le N$ by

$$f(i_1, \cdots, i_k) = \begin{cases} M & \text{if } \sum i_j \equiv 1, 2, \cdots, t \pmod{N}, \\ 1 & \text{otherwise}. \end{cases}$$

Then the array of values $f(i_1, \cdots, i_k)$ has the desired properties.

c. Theorem 5. Let R be a set of non-negative real numbers. Define

$$D_m(k, j, R) = \sup \frac{n^k}{\sum f(P)},$$

over all n, all k-dimensional cubical arrays of side n, with elements taken from R, such that hypothesis (m) of Section 1 $(m = \text{i, ii, iii})$ is satisfied. Then

$$D_m(k, j, \{x \mid x = 0 \text{ or } x \ge 1\}) = \sup_{M \ge 1} D_m(k, j, \{0, 1, M\}). \tag{9}$$

(In particular, the conjecture for cubes whose elements are zero or real numbers ≥ 1 is equivalent to the conjecture for all cubes whose elements consist of $0, 1$ and at most one other positive integer.)

Proof. For any given k, j, and any $\varepsilon > 0$, we will show if a cube of side n whose elements are either zero or real numbers ≥ 1 satisfies hypothesis (m), then for some N, M there is a cube of side Nn whose elements are $0, 1$, or M, which satisfies hypothesis (m), and if the sum of elements in each cube is S, S', respectively,

$$S' \le N^k S/(1-\varepsilon).$$

It follows that the right side of Eq. (9) is at least $1 - \varepsilon$ times the left side. Letting $\varepsilon \to 0$, we have Eq. (9), with ($\le$). The reverse inequality is true because $\{0, 1, M\}$ is a restriction of the set on the left. Hence Eq. (9) is true.

Let $f(i_1, \cdots, i_k)$, $1 \le i_1, \cdots, i_k \le n$, be the given array. Select an array of rational numbers $g(i_1, \cdots, i_k)$ such that if $f(i_1, \cdots, i_k) = 0$, $g(i_1, \cdots, i_k) = 0$, and if $f(i_1, \cdots, i_k) \ge 1$,

$$\frac{\varepsilon S}{1-\varepsilon} n^{-k} \ge g(i_1, \cdots, i_k) - f(i_1, \cdots, i_k) \ge 0.$$

Then the array of $g(i_1, \cdots, i_k)$ satisfies hypothesis (m), and

$$S_1 = \sum g(P) \le \frac{\varepsilon S}{1-\varepsilon} + \sum f(P) = \frac{S}{1-\varepsilon}.$$

Let M be an integer which is an upper bound for $g(P)$, and let q be the least common denominator of $\{g(P)\}$. If $N = q(M-1)$, by the lemma there corresponds to each positive $g(i_1, \cdots, i_k)$ a k-cube of side N:

$$\{h(i_1, \cdots, i_k; l_1, \cdots, l_k), 1 \le l_1, \cdots, l_k \le N\}$$

such that each element is 1 or M, and the sum of every row is $Ng(i_1, \cdots, i_k)$. This implies that the sum of any j-dimensional plane is $N^j g(i_1, \cdots, i_k)$.

We will define $f'(i_1, \cdots, i_k)$ for $1 \le i_1, \cdots, i_k \le Nn$. There is a unique way of putting

$$i_p = N(q_p - 1) + l_p, \qquad p = 1, \cdots, k,$$

with

$$1 \le q_p \le n,$$
$$1 \le l_p \le N.$$

If $g(q_1, \cdots, q_k) = 0$, let

$$f(i_1, \cdots, i_k) = 0.$$

If $g(q_1, \cdots, q_k) > 0$, let

$$f'(i_1, \cdots, i_k) = h(q_1, \cdots, q_k; l_1, \cdots, l_k).$$

Every zero in the array $\{f'(i_1, \cdots, i_k)\}$ corresponds to a zero in the array $\{g(i_1, \cdots, i_k)\}$, and the sum of the j-dimensional hyperplane in a given direction through it is a certain sum of sums of j-dimensional hyperplanes in N-sided subcubes. It is easily shown that the

j-dimensional sums in the new cube are N^j times corresponding sums in the cube $g(P)$. Hence the nN cube satisfies hypothesis (m). The sum of its elements is

$$S' = N^k S_1 \leqq \frac{N^k S}{(1 - \varepsilon)}.$$

B. Hadamard Matrices of Orders 116 and 232

L. D. Baumert

In this Report, a construction is given for an Hadamard matrix of order 116, which was the smallest previously unsolved case. Using this matrix one easily constructs an Hadamard matrix of order 232 which also was an unsolved case.

1. Introduction

An Hadamard matrix (Ref. 2) is a square matrix of *ones* and *minus ones* whose row (column) vectors are orthogonal. The order of an Hadamard matrix is necessarily 1, 2, or $4t$, with t a positive integer. It has been conjectured that if $n = 4t$, there exists an Hadamard matrix of order n. While numerous constructions have been given, for particular values of n the conjecture is still unresolved. The constructions of Refs. 2–10 exhaust the known values of n. The smallest unsolved cases were 116, 188, 232, 236, and 260, two of which are constructed here.

The connection between Hadamard matrices and the regular simplex, orthogonal and biorthogonal codes of space telemetry is explained in Ref. 7, and in *SPS 37-32*, Vol. IV, p. 277.

2. Hadamard Matrices of Orders 116 and 232

An Hadamard matrix of order 116 is given by

$$H = \begin{vmatrix} A & B & C & D \\ -B & A & -D & C \\ -C & D & A & -B \\ -D & -C & B & A \end{vmatrix}$$

where each of A, B, C, and D is a 29×29 symmetric circulant matrix. The first rows of A, B, C, D (using $+$ for $+1$ and $-$ for -1) are:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	+	+	−	−	+	−	−	+	−	+	+	+	−	+
B	+	+	+	+	−	+	+	−	+	−	−	−	+	+
C	+	+	+	−	−	−	+	+	−	−	+	−	+	−
D	+	−	+	−	−	−	+	+	−	−	+	−	+	+

15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
+	+	+	−	+	+	+	−	+	−	−	+	−	−	+
+	+	+	+	−	−	−	+	−	+	+	+	−	+	+
−	−	−	+	−	+	−	−	+	+	−	−	−	+	+
+	+	+	+	−	+	−	−	+	+	−	−	−	+	−

Calling this matrix H_{116}, one constructs an Hadamard matrix of order 232 by taking the tensor product with the Hadamard matrix of order 2, i.e., $H_{232} = H_2 \otimes H_{116}$.

This matrix was discovered on the SDS 930 computer delivered recently to JPL Section 331. The program was written partially to test the machine. Running time was less than 3 hr. Previous attempts to run a similar program on an IBM 7094 failed to yield such a matrix.

References

1. Golomb, S. W., and Posner, E. C., "Rook Domains, Latin Squares, Affine Planes, and Error-Distributing Codes," *IEEE Transactions of the Professional Technical Group on Information Theory*, Vol. IT-10, No. 3, pp. 196–208, July 1964.

2. Paley, R.E.A.C., "On Orthogonal Matrices," *Journal of Mathematics and Physics*, Vol. 12, pp. 311–320, 1933.

References (Cont'd)

3. Baumert, L. D., and Hall, M., Jr., "Hadamard Matrices of the Williamson Type," *Mathematics of Computation*, Vol. 19, pp. 442–447, 1965.

4. Brauer, A., "On a New Class of Hadamard Determinants," *Mathematische Zeitschrift*, Vol. 58, pp. 219–225, 1953.

5. Ehlich, H., "Neue Hadamard-Matrizen," *Archiv der Mathematik*, Vol. 16, pp. 34–36, 1965.

6. Goldberg, K., "Hadamard Matrices of Order Cube Plus One," Abstract 567-90, *Notices of the American Mathematical Society*, Vol. 7, p. 348, 1960.

7. Golomb, S. W., et. al., *Digital Communications with Space Applications*, Prentice-Hall Inc., New Jersey, 1964.

8. Stanton, R. G., and Sprott, D. A., "A Family of Difference Sets," *Canadian Journal of Mathematics*, Vol. 10, pp. 73–77, 1958.

9. Williamson, J., "Hadamard's Determinant Theorem and the Sum of Four Squares," *Duke Mathematical Journal*, Vol. 11, pp. 65–81, 1944.

10. Williamson, J., "Note on Hadamard's Determinant Theorem," Bulletin of the American Mathematical Society, Vol. 53, pp. 608–613, 1947.