THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 6, July–August 1977
Printed in U.S.A.

Reflection, Transmission, and Mode Conversion in a Corrugated Feed

By C. DRAGONE

(Manuscript received December 14, 1976)

Microwave antennas are often required to carry signals simultaneously over a broad range of frequencies—e.g., the combined TD-2 and TH common carrier bands encompass a total frequency ratio of about 1.8 to 1 as do the combined 18- and 30-GHz bands. To achieve these bandwidths, an efficient broadband feed horn is required. The corrugated (hybrid-mode) horn is a leading candidate, but it is not immune to some cross-polarization coupling, input reflection, and pattern asymmetry. These problems are introduced mainly by two phenomena: variation of the dominant mode shape with frequency and mode conversion along the horn taper and at waveguide transitions at the horn input. Simple formulas for computing the magnitude of these phenomena and their effects on return loss and radiation patterns are given.

I. INTRODUCTION

Corrugated feeds (also called hybrid-mode feeds) are widely used in reflector-type antennas because of their excellent radiation characteristics.[1–16] At the frequency ω_0 at which the surface reactance X_s of the corrugations becomes infinite, the radiation pattern of a properly designed feed is circularly symmetric, is free of cross-polarized components, and has low sidelobes. In principle these properties can be obtained over a frequency range of more than an octave. In fact, one can show that the

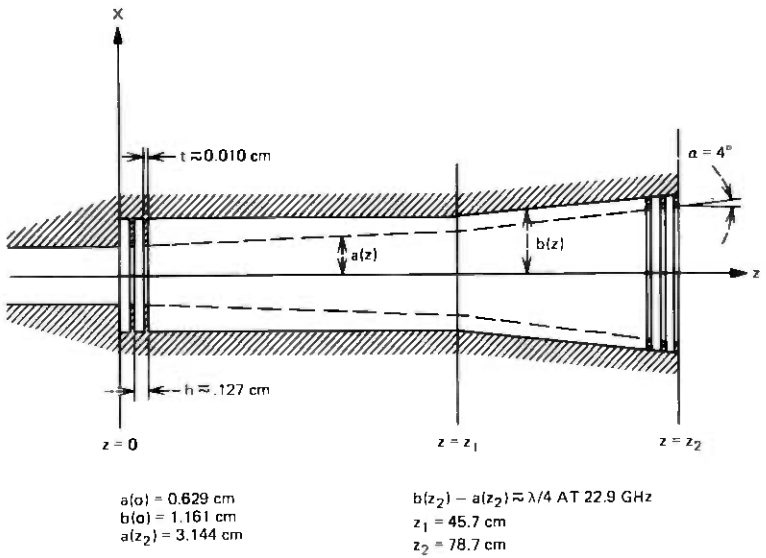

Fig. 1—Corrugated horn of Ref. 17.

field over an aperture illuminated by the fundamental mode has the remarkable property that in the limit, as

$$ka \to \infty, \quad (k = 2\pi/\lambda),$$

where a is the aperture radius, the field distribution becomes independent of the surface reactance X_s (provided $X_s \neq 0$). A corollary of this behavior is that a feed of sufficiently large aperture will have the above radiation characteristics over a wide range of frequencies provided only the fundamental mode is excited in the horn. To verify this behavior, an experiment, described in a companion article,[17] was made. A very long horn (see Fig. 1) was fabricated carefully, using a special fabrication technique to minimize geometrical imperfections in the corrugated walls, and the radiation characteristics were measured from 17 GHz to 35 GHz. From 17 GHz to 29 GHz the far field was found, as expected, to be essentially polarized in one direction. At frequencies above 29 GHz, however, a cross-polarized component was found to be caused by a certain undesirable mode, which will be called the HE'_{11}-mode. This mode was excited primarily at the input, where the corrugated waveguide was connected directly to a smooth waveguide, as shown in Fig. 1. A calculation, given in eqs. (102) and (103) of this article, was therefore made to determine the total amount of power converted from the TE_{11}-mode, incident at the input, to the HE'_{11}-mode.

A peculiarity of corrugated feeds is that there is some mode conversion even in a conical horn of constant surface reactance X_s (unless $X_s = 0$

or $X_s = \infty$). An evaluation of this effect is given in Section VII. However, in the experiment, the taper angle α (Fig. 1) was chosen sufficiently small ($\alpha \simeq 4°$) so that this effect was negligible.

The analysis starts in Section III with a derivation of the asymptotic properties for large ka of the modes of a corrugated waveguide. The results provide a simple and accurate representation of the modes in a feed aperture of more than a few wavelengths in diameter. Then, in Sections V and VI, a first-order derivation of the scattering parameters of a junction between two waveguides of slightly different characteristics is given. A simple relation [see, eqs. (83), (84), and (115) to (117)] is found between the scattering parameters and the coupling coefficients between the modes on the two sides of the junction. Each coefficient is given, except for a constant, by an integral of the type

$$\int\int_S (\mathbf{E}_1 \times \mathbf{H}_2^*) \cdot \mathbf{i}_z \, dxdy,$$

where S is the junction area, and $\mathbf{E}_1$ and $\mathbf{H}_2$ are the electric and magnetic vectors of the two modes, respectively. In Section IV, this surface integral is converted to a line integral, thus reducing the calculation of the coupling coefficients to a straightforward exercise. This result is useful also to calculate the far field of an aperture S illuminated by a mode $\mathbf{E}_1$, since the far field at a given observation point is, except for a constant, the coupling coefficient over S between $\mathbf{E}_1$ and the field $\mathbf{H}_2$ of a plane wave having the direction of the observation point. The far-field calculation is thus reduced to a contour integration.

The calculation of the scattering parameters is carried out in Sections IV to VI, using the above contour integral. It is found, for instance, that the input reflection of a corrugated feed connected to a smooth waveguide of the same diameter is simply given by the coefficient

$$\rho_1 = -\frac{\beta_1 - \beta_1'}{\beta_1 + \beta_1'},$$

β_1 and β_1' being the propagation constants in the two waveguides. An identical formula was derived by Brown[18] from the principle of conservation of momentum, but that derivation is not applicable to the present problem, which involves hybrid modes.

Finally, Section VII deals with the problem of spurious mode generation in a nonuniform waveguide whose parameters (radius and surface reactance) vary along the axis, as in Fig. 1. The differential scattering parameters that give, at any point in a nonuniform waveguide, the local coupling between the incident mode and the spurious modes are obtained from the analysis of Sections V and VI. By solving the differential equations specified by the above scattering parameters, we can thus determine the amplitudes of the spurious modes. An example is provided

of a first-order calculation of mode conversion in a conical waveguide such as the one in Fig. 1 for $z > z_1$. The result, eq. (154), is again quite simple.

II. PRELIMINARY CONSIDERATIONS

For a smooth waveguide, there is a simple relation between the propagation constant β of a mode and the waveguide diameter, but no such simple relation exists in the case of a corrugated waveguide [see eq. (20)]. For this reason, the properties of the corrugated waveguide modes cannot generally be determined as simply as in the case of a smooth waveguide. Also, the field configuration of each mode varies with waveguide diameter. There is, however, an important exception. When the radius a of the corrugated waveguide is sufficiently large, the propagation constant β for some of the modes is simply given by

$$\beta = \sqrt{(ka)^2 - u_{0m}^2}, \tag{1}$$

where u_{0m} is the mth zero of the Bessel function J_0 of order zero,

$$J_0(u_{0m}) = 0. \tag{2}$$

For all the other modes except one (for this special mode β is independent of a; see Appendix B) one has

$$\beta = \sqrt{(ka)^2 - u_{2m}^2}, \tag{3}$$

where u_{2m} is the mth root of the Bessel function of order two,

$$J_2(u_{2m}) = 0. \tag{4}$$

Equations (1) and (3) are valid provided $a \gg \lambda$, a condition which is satisfied to a good approximation by most feed apertures. Thus, the case

$$ka \gg 1 \tag{5}$$

is of considerable practical interest. One finds that as $ka \to \infty$, the properties of a mode become independent of the surface reactance X_s of the corrugated walls, except for the mode of Appendix B. Thus, the field distribution over the aperture of a feed illuminated by a single mode will be little affected by the surface reactance X_s (which varies fairly rapidly with frequency) provided ka is sufficiently large. This result, first pointed out by Thomas,[8] is very important for it implies that the aperture field distribution becomes frequency independent for large ka. The main purpose of this section is to determine the asymptotic behavior of the hybrid modes for large ka. It is shown that if $ka \neq \infty$ there is over the aperture of a feed a certain undesirable cross-polarized component, even if the aperture is illuminated by a single mode, unless of course X_s

$= \infty$. A simple expression for the amplitude of this component is given.

III. ASYMPTOTIC BEHAVIOR FOR LARGE *ka*

Consider a disk-loaded waveguide centered around the z-axis, as in Fig. 1, and assume its parameters a, b, and h are independent of z. Let r, ϕ, z be cylindrical coordinates defined by $x = r\cos\phi$ and $y = r\sin\phi$. The separation of the disks, which occupy the region $a < r < b$, is assumed to be much smaller than a wavelength λ,

$$kh \ll 1. \tag{6}$$

The region between two consecutive disks forms a radial line whose input reactance jX at $r = a$ is a function of the radial length $l = b - a$; for $ka \gg 1$, one has approximately

$$jX = jZ_0 \tan kl,$$

where $Z_0 = \sqrt{\mu_0/\epsilon_0}$. For a finite number of teeth per wavelength, the value of l must be corrected.* Because of condition (6) the effect of the disks can be accounted for adequately by introducing an effective surface reactance[5,12,19]

$$jX_s = jX\left(1 - \frac{t}{h}\right), \tag{7}$$

where t is the thickness of the disks, and by requiring that the field for $r < a$ satisfy the boundary conditions

$$\left.\begin{aligned} E_\phi &\simeq 0 \\ H_\phi &\simeq -\frac{E_z}{jX_s} \end{aligned}\right\} \text{ for } r = a, \tag{8}$$

where E_ϕ, H_ϕ, E_z are the ϕ and z components of the electric and magnetic field.

Let β be the propagation constant in the z direction,

$$\beta = k\cos\theta_1, \tag{9}$$

and assume θ_1 is real, so that $\beta < k$. The case where θ_1 is imaginary is considered in Appendix B. Assume the ϕ dependence of E_z is given by $\cos\phi$. Then, the field components of a mode that propagates in the z direction with propagation constant β are eiven by

$$E_z = AJ_1(\kappa r)\cos\phi\, e^{-j\beta z}, \tag{10}$$

* See Ref. 17 for the effect of a finite number of teeth per wavelength, which causes a reduction of the effective depth, l.

$$Z_0 H_z = B J_1(\kappa r) \sin\phi \, e^{-j\beta z}, \tag{11}$$

$$E_\phi = \frac{j}{\sin\theta_1}\left[B J_1'(\kappa r) + A \cos\theta_1 \frac{J_1(\kappa r)}{\kappa r}\right] \sin\phi \, e^{-j\beta z}, \tag{12}$$

$$Z_0 H_\phi = -\frac{j}{\sin\theta_1}\left[A J_1'(\kappa r) + B \cos\theta_1 \frac{J_1(\kappa r)}{\kappa r}\right] \cos\phi \, e^{-j\beta z}, \tag{13}$$

$$E_\rho = -\frac{j}{\sin\theta_1}\left[B \frac{J_1(\kappa r)}{\kappa r} + A \cos\theta_1 J_1'(\kappa r)\right] \cos\phi \, e^{-j\beta z}, \tag{14}$$

$$Z_0 H_\rho = -\frac{j}{\sin\theta_1}\left[A \frac{J_1(\kappa r)}{\kappa r} + B \cos\theta_1 J_1'(\kappa r)\right] \sin\phi \, e^{-j\beta z}, \tag{15}$$

$$(r \leqq a),$$

where

$$\kappa = k \sin\theta_1.$$

The boundary conditions (8) give[12]

$$\gamma = -\frac{u}{\cos\theta_1} \frac{J_1'(u)}{J_1(u)}, \tag{16}$$

$$\mathrm{y} = \frac{\cos\theta_1}{\sin\theta_1} \frac{1}{u} \frac{1}{\gamma} + \frac{1}{\sin\theta_1} \frac{J_1'(u)}{J_1(u)}, \tag{17}$$

where

$$u = ka \sin\theta_1,$$

$$\mathrm{y} = -\frac{Z_0}{X_s}, \tag{18}$$

and γ is the ratio between the TM and TE components of the hybrid mode,

$$\gamma = \frac{A}{B}. \tag{19}$$

By eliminating γ from eqs. (16) and (17), one obtains the eigenvalue equation

$$\frac{\mathrm{y}}{ka} = \frac{1}{u^2} \frac{J_1(u)}{u J_1'(u)} \left[\left(\frac{u J_1'(u)}{J_1(u)} \right)^2 - 1 + \frac{u^2}{(ka)^2} \right], \tag{20}$$

which is eq. (10) of Ref. 12.

The solutions of this equation are now studied for large ka. Both u and y are assumed to be finite. Then, in the limit as $ka \to \infty$, eq. (20) reduces to

$$\left(\frac{J_1'(u)}{J_1(u)} u \right)^2 - 1 = 0. \tag{21}$$

We distinguish two cases:

$$\frac{J_1'(u)u}{J_1(u)} = -1 \tag{22}$$

and

$$\frac{J_1'(u)u}{J_1(u)} = 1. \tag{23}$$

According to eq. (16) (with $\cos\theta_1 \approx 1$, since $\theta_1 \to 0$ as $ka \to \infty$), these two cases correspond respectively to

$$\gamma = 1 \tag{24}$$

and

$$\gamma = -1. \tag{25}$$

Using well known recurrence relations between the Bessel functions and their derivatives, and using conditions (22) and (23), we find

$$J_0(u) = 0 \tag{26}$$

and

$$J_2(u) = 0. \tag{27}$$

We conclude that for large ka, eq. (20) possesses the two sets of solutions

$$u \approx u_{0m} \ (m = 1, 2, \text{etc.}) \tag{28}$$

and

$$u \approx u_{2m} \ (m = 1, 2, \text{etc.}), \tag{29}$$

u_{0m} and u_{2m} being respectively the mth root of $J_0(u)$ and $J_2(u)$. Solutions (28) and (29) are characterized by $\gamma \approx 1$ and $\gamma \approx -1$; the corresponding modes will be designated,* respectively, HE_{1m} and HE'_{1m}.

Asymptotic series for u and γ in terms of

$$\frac{1}{ka}, \tag{30}$$

* This mode classification differs from the one by Clarricoats[12] and it was chosen for the following reason. Here, and in Ref. 17, we are interested in horns whose inner radius a varies gradually with z, while the wall susceptance y is approximately constant, as in Fig. 1, from z_2 to z_1. Consider therefore a mode propagating in Fig. 1 from z_2 towards z_1. Clarricoats' classification assigns in some cases a different name to this mode in different regions of the horn, even though there will be no discontinuous variation of the mode-field configuration, as it propagates in the horn. On the other hand, our classification based on the Bessel function roots u_{0m} and u_{2m}, assigns a single name everywhere in the horn. If instead the frequency is gradually changed the mode of a waveguide of given dimensions will retain the same name with Clarricoats' classification, whereas this is not always true with our classification. To understand better these considerations, see also Ref. 25.

are derived in Appendix A under the assumption $y \neq \infty$. For the HE_{1m} modes, characterized by $\gamma \to 1$ as $ka \to \infty$, it is found that

$$u = u_m = u_{0m}\left\{1 - \frac{1}{2}\frac{y}{ka} - \frac{1}{2}\left[1 - \frac{y^2}{4}(1 + u_{0m}^2)\right]\left(\frac{1}{ka}\right)^2 + \frac{y}{4}\left[1 - \frac{y^2}{12}(7u_{0m}^2 + 1)\right]\left(\frac{1}{ka}\right)^3 \cdots\right\} \quad (31)$$

and

$$\gamma = 1 - u_{0m}^2\left\{\frac{y}{2}\frac{1}{ka} - \frac{y^2}{8}(4 + u_{0m}^2)\left(\frac{1}{ka}\right)^2 - \frac{y}{2}\left[1 - \frac{1}{2}u_{0m}^2 - \frac{y^2}{4}(3u_{0m}^2 + 2)\right]\left(\frac{1}{ka}\right)^3 \cdots\right\}. \quad (32)$$

For the HE'_{1m}-modes, characterized by $\gamma \to -1$, u is given by

$$u'_m = u_{2m}\left\{1 - \frac{1}{2}\frac{y}{ka} \cdots\right\} \quad (33)$$

and

$$\gamma = -1 - u_{2m}^2\left\{\frac{1}{2}\frac{y}{ka} \cdots\right\}. \quad (34)$$

The x and y components of the electric field are now derived. First consider the HE_{1m} modes. One finds from eqs. (10) to (15), with $\cos\theta_1 = 1$ and γ given by eq. (32), that for large ka the transverse component of $\mathbf{E}$ is given by

$$\mathbf{E}_t \simeq -j\frac{ka}{u}A\left[J_0\left(\frac{r}{a}u\right)\mathbf{i}_x + \frac{1}{4}u^2\frac{y}{ka}J_2\left(\frac{r}{a}u\right)(\cos 2\phi\,\mathbf{i}_x + \sin 2\phi\,\mathbf{i}_y)\right], \quad (35)$$

omitting the factor $e^{-j\beta z}$. Amplitude A is determined by power P carried by the mode. From eq. (67) with du/dy given by eq. (92) and $\eta_i = A$

$$|A| \simeq \frac{1}{a}\sqrt{\frac{Z_0}{\pi}\frac{1}{a\beta}\frac{u^2}{ka}\frac{1}{J_1^2(u)}} \quad (36)$$

if $P = \frac{1}{2}$.

For the HE'_{1m} modes with $\gamma \simeq -1$, on the other hand,

$$\mathbf{E}_t \simeq j\frac{ka}{u}A\left[J_2\left(\frac{r}{a}u\right)(\cos 2\phi\,\mathbf{i}_x + \sin 2\phi\,\mathbf{i}_y) + \cdots\right], \quad (37)$$

where the dots represent terms that vanish as $ka \to \infty$. The amplitude $|A|$ for $P = \frac{1}{2}$ is still given by eq. (36).*

An important property of the field distribution (35) is that $E_y \to 0$ as $ka \to \infty$. Thus, in the limit as $ka \to \infty$, the field becomes polarized in one direction, *regardless of the value of the surface reactance* X_s (unless, of course $X_s = 0$). From eq. (35), the amplitude of E_y is porportional to the ratio

$$\frac{y}{ka}. \tag{38}$$

Therefore, in order that E_y be negligible over the aperture of a feed, it is sufficient that the aperture diameter be large and the thickness t of the disks (see Fig. 1) be small compared with their separation.† The far field of an aperture illuminated by the fundamental mode, the HE_{11} mode given by eq. (35) for $u = u_{01} = 2.4048$, is discussed in Ref. 17. From a comparison of the radiation patterns of E_x and E_y, we find that the ratio C^2 between the maximum value attained by $|E_y|^2$ and $|E_x|^2$ (which occurs on axis) is given by

$$C^2 = 0.14\left(\frac{y}{ka}\right)^2 = \frac{0.14}{\left[(1 - t/h)ka \tan\frac{\pi}{2}\frac{\omega}{\omega_0}\right]^2}, \tag{39}$$

where ω_0 denotes the frequency for which $y = 0$. One can easily verify using this formula that C^2 remains less than 0.000316 (−35 dB) over a frequency range $\omega_1 < \omega < 1.93\,\omega_1$, provided $ka > 10$ and $t/h < 0.1$.

Thus, good performance over a wide frequency range is possible, provided all the power incident at the input of the feed is converted to the HE_{11} mode. If, however, some of the input power is converted into some of the HE_{1m} modes, then, according to eq. (37), the field over the feed aperture will contain a cross-polarized component whose amplitude is essentially independent of the ratio y/ka. The resulting cross-polarized component of the far field is discussed in Ref. 17. If $\omega_1 < \omega < \omega_2$ denotes the frequency range over which only the fundamental mode (HE_{11}) propagates, it is pointed out in Ref. 17 that the largest value that ω_2/ω_1 can assume is 1.6839; this value is attained for $b/a = 1.8309$. Cutoff frequency formulas are derived in Appendix D.

In Appendix B, the properties of a surface-wave mode that can exist in a corrugated waveguide, in addition to the modes of eqs. (35) and (37), are briefly described.

---

* In eqs. (35) and (37), only the leading terms for the symmetrical, asymmetrical, and cross-polarized components are retained.

† Note that from eqs. (7), (8), and (18), y increases with t/h.

IV. COUPLING COEFFICIENT BETWEEN TWO MODES

Suppose the electric field $\mathbf{E}_1$ at the input of a corrugated waveguide is known, and we want to determine the resulting amplitude of one of the modes excited in the corrugated waveguide. We have to evaluate a surface integral of the form

$$\int\int_S (\mathbf{E}_1 \times \mathbf{H}_2^*) \cdot \mathbf{i}_z \, dxdy, \tag{40}$$

where $\mathbf{H}_2$ is the magnetic field of the mode whose amplitude is to be determined. This integral, identical to that involved in determining the far field radiated in a given direction by an aperture containing the field $\mathbf{E}_1$, is in general difficult to evaluate. However, in many cases, we can assume that

$$\frac{\partial \mathbf{E}_1}{\partial z} \approx -j\beta_1 \mathbf{E}_1, \tag{41}$$

where β_1 is a constant. This condition is approximately satisfied,* for instance, in the case of a feed aperture illuminated by a single mode propagating in the z direction with propagation constant β_1. We will show that the above surface integral can be reduced to a line integral which can be evaluated straightforwardly. We use the symbol $(\mathbf{E}_1, \mathbf{H}_2)$ for the integral (40), and call it the scalar product of the two modes $\mathbf{E}_1$ and $\mathbf{H}_2$.

If $\mathbf{E}_1, \mathbf{H}_1$ and $\mathbf{E}_2, \mathbf{H}_2$ are two solutions of Maxwell's equations, in free space,[20]

$$\nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2^* + \mathbf{E}_2^* \times \mathbf{H}_1) = 0 \tag{42}$$

in the absence of sources. Now, let the z dependence of the two solutions be given by

$$e^{-j\beta_1 z} \quad \text{and} \quad e^{-j\beta_2 z}. \tag{43}$$

Then, in eq. (42)

$$\nabla = \nabla_t - j(\beta_1 - \beta_2)\mathbf{i}_z, \tag{44}$$

where ∇_t is the transverse part of ∇. Therefore eq. (42) gives

$$\nabla_t(\mathbf{E}_1 \times \mathbf{H}_2^* + \mathbf{E}_2^* \times \mathbf{H}_1) = j(\beta_1 - \beta_2)[(\mathbf{E}_1 \times \mathbf{H}_2^*) \cdot \mathbf{i}_z + (\mathbf{E}_2^* \times \mathbf{H}_1) \cdot \mathbf{i}_z]. \tag{45}$$

Next, consider a new solution $\mathbf{E}_1'$, $\mathbf{H}_1'$ with propagation constant $\beta_1' = -\beta_1$ and with z components given by

$$E_{1z}' = -E_{1z} \tag{46}$$

* Of course, condition (41) is satisfied *exactly* by a mode in a cylindrical waveguide.

$$H'_{1z} = +H_{1z}. \tag{47}$$

Then the x, y components of $\mathbf{E}'_1$, $\mathbf{H}'_1$ simply coincide[21] with the x, y components of $\mathbf{E}_1$, $-\mathbf{H}_1$,

$$E'_{1x} = E_{1x}, \qquad E'_{1y} = E_{1y}, \tag{48}$$

$$H'_{1x} = -H_{1x}, \qquad H'_{1y} = -H_{1y}. \tag{49}$$

Therefore, replacing in eq. (45) β_1, $\mathbf{E}_1$, $\mathbf{H}_1$ with $-\beta_1$, $\mathbf{E}'_1$, $\mathbf{H}'_1$ and making use of eqs. (48) and (49), we obtain

$$\nabla_t \cdot (\mathbf{E}'_1 \times \mathbf{H}^*_2 + \mathbf{E}^*_2 \times \mathbf{H}'_1) = -j(\beta_1 + \beta_2)[(\mathbf{E}_1 \times \mathbf{H}^*_2) - (\mathbf{E}^*_2 \cdot \mathbf{H}_1)] \times \mathbf{i}_z. \tag{50}$$

By adding eq. (45) to eq. (50), we obtain

$$(\mathbf{E}_1 \times \mathbf{H}^*_2) \cdot \mathbf{i}_z = \nabla_t \cdot \mathbf{F}, \tag{51}$$

where

$$\mathbf{F} = \frac{j}{2(\beta_1 + \beta_2)} [\mathbf{E}'_1 \times \mathbf{H}^*_2 + \mathbf{E}^*_2 \times \mathbf{H}'_1] - \frac{j}{2(\beta_1 - \beta_2)} [(\mathbf{E}_1 \times \mathbf{H}^*_2) + (\mathbf{E}^*_2 \times \mathbf{H}_1)]. \tag{52}$$

We now integrate eq. (51) over a finite area S of the plane $z = 0$, making use of the divergence theorem,

$$\int\int_S (\mathbf{E}_1 \times \mathbf{H}^*_2) \cdot \mathbf{i}_z \, dxdy = \oint_C \mathbf{F} \cdot \mathbf{n} \, ds, \tag{53}$$

where C is the contour of S and $\mathbf{n}$ is the outward normal. To determine $\mathbf{F} \cdot \mathbf{n}$, let τ be a unit vector tangent to C,

$$\tau = \mathbf{i}_z \times \mathbf{n}. \tag{54}$$

Then, if $\mathbf{A}$ and $\mathbf{B}$ are two arbitrary vectors,

$$(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{n} = A_\tau B_z - A_z B_\tau, \tag{55}$$

where A_τ, B_τ are the components of $\mathbf{A}$ and $\mathbf{B}$ in the direction of τ. Therefore, from eq. (52), taking into account eqs. (46) to (49),

$$\mathbf{F} \cdot \mathbf{n} = -j \frac{\beta_2}{\beta_1^2 - \beta_2^2} [E_{1\tau} H^*_{2z} + E^*_{2\tau} H_{1z}] + j \frac{\beta_1}{\beta_1^2 - \beta_2^2} [E_{1z} H^*_{2\tau} + E^*_{2z} H_{1\tau}]. \tag{56}$$

Finally, from eqs. (53) and (56), we obtain the desired result,

$$(\mathbf{E}_1, \mathbf{H}_2) = \int\int_S (\mathbf{E}_1 \times \mathbf{H}_2^*) \cdot \mathbf{i}_z \, dxdy$$

$$= -j \frac{\beta_2}{\beta_1^2 - \beta_2^2} \oint_C (E_{1\tau} H_{2z}^* + E_{2\tau}^* H_{1z}) ds$$

$$+ j \frac{\beta_1}{\beta_1^2 - \beta_2^2} \oint_C (E_{1z} H_{2\tau}^* + E_{2z}^* H_{1\tau}) ds. \quad (57)$$

Thus, the scalar product (coupling coefficient) of two modes $\mathbf{E}_1$ and $\mathbf{E}_2$ can be determined straightforwardly from the values of $\mathbf{E}_1$ and $\mathbf{E}_2$ on the contour of the aperture S. This result has a number of applications. It can be used, as already pointed out, to determine the far field radiated by an aperture S with known field distribution $\mathbf{E}_1$, in which case $\mathbf{H}_2$ is the magnetic field* of a plane wave with propagation vector $\mathbf{k}$ and eq. (57) gives, except for a constant independent of $\mathbf{k}$, the field component radiated in the direction of $\mathbf{k}$ with the polarization of $\mathbf{H}_2$. In this article, we are interested in the special case where S is a circular area of radius a, in which case we can replace in eq. (57) τ with ϕ, since

$$\tau = \mathbf{i}_\phi.$$

If $\mathbf{E}_i$, $\mathbf{H}_i$ $(i = 1, 2)$ represents a mode of a corrugated waveguide of radius a, so that for $r = a$

$$E_{i\phi} = 0, \quad Z_0 H_{i\phi} = -j\mathrm{y}_i E_{iz}, \quad (58)$$

then eq. (57) simplifies to

$$(\mathbf{E}_1, \mathbf{H}_2) = -\frac{1}{Z_0} \frac{\beta_1}{\beta_1^2 - \beta_2^2} (\mathrm{y}_2 - \mathrm{y}_1) \oint_C E_{1z} E_{2z}^* \, ds. \quad (59)$$

Since the modes are characterized for $z = 0$ and $r = a$ by

$$E_{iz} = \eta_i J_1(u_i) \cos\phi, \quad (60)$$

where η_i is the coefficient A of the ith mode, then

$$(\mathbf{E}_1, \mathbf{H}_2) = -\frac{a\pi}{Z_0} \frac{\beta_1}{\beta_1^2 - \beta_2^2} (\mathrm{y}_2 - \mathrm{y}_1) \eta_1 \eta_2^* J_1(u_1) J_1(u_2). \quad (61)$$

Note that

$$a^2(\beta_1^2 - \beta_2^2) = u_2^2 - u_1^2. \quad (62)$$

If we assume

$$\mathrm{y}_2 = \mathrm{y}_1 + d\mathrm{y}, \quad (63)$$

$$u_2 = u_1 + du_1, \quad (64)$$

* There are two cases (two polarizations) that must be considered, for each value of $\mathbf{k}$.

from eq. (61) we obtain for the power carried by the mode $\mathbf{E}_1$

$$P_1 = \frac{1}{2}(\mathbf{E}_1, \mathbf{H}_1) = -\frac{\pi}{2Z_0}\frac{\beta_1 a^3}{2u_1}\frac{dy}{du_1}\eta_1^2 J_1^2(u_1). \tag{65}$$

The derivative dy/du_1, which appears in this expression, is calculated in Appendix C. In the following sections, we choose

$$(\mathbf{E}_i, \mathbf{H}_i) = 1, \tag{66}$$

in which case from eq. (65)

$$|\eta_i| = \sqrt{\frac{2Z_0}{a^2\pi}\frac{u_i}{\beta_i a}\left|\frac{du_i}{dy}\right|\frac{1}{J_1^2(u_i)}}. \tag{67}$$

These results are now applied to the problem of a junction between two different waveguides.

V. JUNCTION BETWEEN TWO WAVEGUIDES OF DIFFERENT SURFACE REACTANCE

Let two waveguides of different surface reactance, but the same diameter, be jointed at $z = 0$. Assume a single mode incident on the plane of the junction from the region $z < 0$ and let $\mathbf{E}_t$, $\mathbf{H}_t$ denote the transverse field components. To determine the amplitudes of the reflected and transmitted modes, we expand $\mathbf{E}_t$ and $\mathbf{H}_t$ on either side of the junction in an infinite series of modes, and then require continuity of $\mathbf{E}_t$ and $\mathbf{H}_t$ at the junction. A simple solution for the amplitudes of the scattered modes is then obtained assuming the difference in surface reactance is small. This result will be extended in Section VI to the more general case of two waveguides of slightly different diameter.

Let the transverse fields for $z < 0$ be represented by a superposition of the modes of the waveguide occupying the region $z < 0$,

$$\mathbf{E}_t = A_1\mathbf{e}_1 e^{-j\beta_1 z} + \sum_1^\infty R_i \mathbf{e}_i e^{j\beta_1 z}, \quad (z < 0) \tag{68}$$

$$\mathbf{H}_t = A_1\mathbf{h}_1 e^{-j\beta_1 z} - \sum_1^\infty R_i \mathbf{h}_i e^{j\beta_1 z}, \quad (z < 0), \tag{69}$$

where

$$A_1\mathbf{e}_1 e^{-j\beta_1 z}, \quad A_1\mathbf{h}_1 e^{-j\beta_1 z} \quad (\mathrm{Re}(\beta_1) > 0)$$

are the transverse field components of the incident mode, and

$$R_i\mathbf{e}_i e^{j\beta_i z}, \quad -R_i\mathbf{h}_i e^{j\beta_i z}$$

are those of the reflected modes.

Similarly, for $z > 0$,

$$\mathbf{E}_t = \sum_1^\infty T_i \mathbf{e}_i' e^{-j\beta_i' z}, \quad (z > 0) \tag{70}$$

$$\mathbf{H}_t = \sum_1^\infty T_i \mathbf{h}_i' e^{-j\beta_i' z}, \quad (z > 0), \tag{71}$$

where T_i are the amplitudes of the transmitted modes.

We assume that $\mathbf{e}_i$, $\mathbf{h}_i$ are normalized so that

$$(\mathbf{e}_i, \mathbf{h}_n) = \int\int_S (\mathbf{e}_1 \times \mathbf{h}_n^*) \cdot i_z \quad dxdy = \delta_{in}. \tag{72}$$

Similarly,

$$(\mathbf{e}_i', \mathbf{h}_n') = \delta_{in}. \tag{73}$$

Since $(\mathbf{e}_i, \mathbf{h}_i)$ represents twice the power carried by the ith mode $\mathbf{e}_i$, this power becomes imaginary if the mode is cutoff, in which case eq. (73) for $i = 1$ should be replaced with

$$(\mathbf{e}_i, \mathbf{h}_i) = j.$$

However, in this article the calculation of R_i, T_i is restricted to the modes that are not cut off by the two waveguides.

From eq. (61) with η_i, η_k given by eq. (67),

$$(\mathbf{e}_i, \mathbf{h}_n') = \frac{2}{a^2} \frac{\beta_i}{\beta_i^2 - \beta_n'^2} (\mathrm{y} - \mathrm{y}') \sqrt{\frac{u_i u_n'}{|\beta_i \beta_n' (dy/du_i)(dy'/du_n')|}}. \tag{74}$$

and

$$(\mathbf{e}_n', \mathbf{h}_i) = \frac{\beta_n'}{\beta_i} (\mathbf{e}_i, \mathbf{h}_n'), \tag{75}$$

where $(\mathbf{e}_i, \mathbf{h}_n')$ and $(\mathbf{e}_n', \mathbf{h}_i)$ are scalar products defined as in eq. (72) and

$$u_i^2 = (ka)^2 - (\beta_i a)^2, \tag{76}$$

a being the radius of the two waveguides. In eq. (74) 1/y and 1/y′ are the normalized surface reactances of the two waveguides.

Now assume y′ − y is very small and let

$$\delta \mathrm{y} = \mathrm{y}' - \mathrm{y}. \tag{77}$$

To determine R_i, T_i, we require continuity of $\mathbf{E}_t$ and $\mathbf{H}_t$ for $z = 0$,

$$A_1 \mathbf{e}_1 + \sum_1^\infty R_n \mathbf{e}_n = T_1 \mathbf{e}_1' + \sum_2^\infty T_n \mathbf{e}_n'. \tag{78}$$

$$A_1 \mathbf{h}_1 - \sum_1^\infty R_n \mathbf{h}_n = T_1 \mathbf{h}_1' + \sum_2^\infty T_n \mathbf{h}_n'. \tag{79}$$

Take the scalar product of the first equation with $\mathbf{h}_i'$ and of the second with $\mathbf{e}_i'$. One obtains, taking into account eq. (73) and assuming that the mode $\mathbf{e}_i'$ is not cutoff, so that β_i' is real,

$$A_1(\mathbf{e}_1, \mathbf{h}_i') + R_i(\mathbf{e}_i, \mathbf{h}_i') + \sum_{n \neq 1,i} R_n(\mathbf{e}_n, \mathbf{h}_i') = T_i, \tag{80}$$

$$A_1(\mathbf{e}_i', \mathbf{h}_1) - R_i(\mathbf{e}_i', \mathbf{h}_i) - \sum_{n \neq 1,i} R_n(\mathbf{e}_i', \mathbf{h}_n) = T_i'. \tag{81}$$

Now, assume for the moment that y, y′ $\neq \infty$. Furthermore assume *none of the modes under consideration is at cutoff. Then*

$$(\mathbf{e}_n, \mathbf{h}_1'), \quad (\mathbf{e}_i', h_n) \quad (i \neq n) \tag{82}$$

are small quantities of the same order of δy. Furthermore, as we show below, this is true also for R_n. It follows that the two sums involving R_n in eqs. (80) and (81) are of order higher than δy. Therefore, subtracting these two equations and neglecting terms of order higher than δy,

$$R_i = -A_1 \frac{(\mathbf{e}_1, \mathbf{h}_i') - (\mathbf{e}_i', \mathbf{h}_1)}{(\mathbf{e}_i, \mathbf{h}_i') + (\mathbf{e}_i', \mathbf{h}_i)}. \tag{83}$$

Adding eqs. (80) and (81), and neglecting terms of order higher than δy and solving for T_i, we obtain

$$T_i = A_1 \frac{(\mathbf{e}_1, \mathbf{h}_i') + (\mathbf{e}_i', \mathbf{h}_1)}{2}. \tag{84}$$

Using eqs. (74) and (75), we rewrite eqs. (83) and (84) in the form

$$R_i = -A_1 \frac{\beta_i - \beta_i'}{\beta_1 + \beta_i'} \sqrt{\frac{u_1 \beta_i}{u_i \beta_1}} \sqrt{\left| \frac{dy/du_i}{dy/du_1} \right|}, \tag{85}$$

$$T_i = A_1 \frac{1}{a^2} \frac{1}{\beta_1 - \beta_i'} (y - y') \sqrt{\frac{u_1 u_i'}{\beta_1 \beta_i'}} \frac{1}{|(dy/du_1)\,(dy'/du_i')|}. \tag{86}$$

The derivatives dy/du_1 and dy'/du_i' are derived in Appendix C.

It is interesting to note from eq. (85) that the reflection coefficient for the mode $i = 1$ is simply

$$\rho_1 = \frac{R_1}{A_1} = -\frac{\beta_1 - \beta_1'}{\beta_1 + \beta_1'}, \tag{87}$$

which coincides with a formula derived by Brown[18] from a principle of conservation of momentum. However, that derivation is not applicable to the present problem, which involves hybrid modes. Measurements of ρ_1 described in Ref. 17 show that this formula, although derived assuming y′ $\simeq$ y, is quite accurate even for relatively large differences between y and y′.

It is also interesting to note that the following interpretation can be given to eq. (86). If $\mathbf{E}_t$ for $z = 0$ were known, we could determine T_i simply using the formula

$$T_i = (\mathbf{E}_t, \mathbf{h}_i'), \quad (z = 0), \tag{88}$$

which follows from eq. (70), in view of the orthogonality relations (73). Now, if $y - y' \approx 0$, $\mathbf{E}_t$ does not differ much from $A_1\mathbf{e}_1$ and, therefore, we might be tempted to write in eq. (88) $\mathbf{E}_t \approx A_1\mathbf{e}_1$, in which case we would get

$$T_i \approx A_1(\mathbf{e}_1, \mathbf{h}_i'). \tag{89}$$

Alternatively, since from eq. (71) we also have

$$T_i = (\mathbf{e}_i', \mathbf{H}_t) \quad \text{for } z = 0, \tag{90}$$

we might be tempted to assume $\mathbf{H}_t \approx A_1\mathbf{h}_1$ for $z = 0$, in which case

$$T_i \approx A_1(\mathbf{e}_i', \mathbf{h}_1). \tag{91}$$

Neither of the two formulas is correct* even if $\delta y \approx 0$. However, according to eq. (84), a correct expression for small δy is obtained by taking the average of the two formulas. We now treat two special cases.

5.1 Limiting case $ka \gg 1$

Assume that both y and y′ are finite, but the radius a is very large,

$$ka \gg 1,$$

a condition which is often satisfied near the aperture of a feed. From eqs. (31) and (33)

$$\lim_{ka\to\infty} \frac{du}{dy} = -\frac{1}{2}\frac{u}{ka}. \tag{92}$$

Furthermore, for large ka,

$$\beta a \approx ka - \frac{1}{2}\frac{u^2}{ka}, \tag{93}$$

since $(\beta a)^2 = (ka)^2 - u^2$. Therefore,

$$a\beta_i' - a\beta_i \approx -\frac{1}{2}\frac{u_i'^2 - u_i^2}{ka} \approx \frac{1}{2}\left(\frac{u_i}{ka}\right)^2 \delta y, \tag{94}$$

since from eq. (92)

$$u_i' \approx u_i - \frac{1}{2}\frac{u}{ka}\,\delta y. \tag{95}$$

* They are often used, however.[22]

Using these results, from eqs. (85) and (86) we obtain

$$T_i = -A_1 \frac{u_1 u_i}{u_i^2 - u_1^2} \frac{(y' - y)}{ka}, \tag{96}$$

$$R_i = A_1 \frac{1}{4} \frac{u_i u_1}{(ka)^2} \frac{y' - y}{ka}. \tag{97}$$

One can show that these formulas are valid even if δy is not small, provided both

$$\frac{y'}{ka} \quad \text{and} \quad \frac{y}{ka}$$

are small.

An application of eq. (96) is considered in Section 7.1.

5.2 Case 1/y = 0

At the input of the feed of Fig. 1, the corrugated waveguide is connected to a smooth waveguide ($1/y = 0$) of the same diameter. We now wish to calculate the reflection and transmission coefficients of such a junction. Thus, assume $y \approx \infty$ for $z < 0$. For $y \approx \infty$, there are two types of modes: TE modes, in which case $\gamma \approx 0$, and TM modes, in which case $\gamma \approx \infty$. In the former case, from eq. (179) of Appendix C

$$\lim_{y \to \infty} \frac{dy}{du} = -y^2(u^2 - 1) \frac{kau}{(ka)^2 - u^2}, \quad (\gamma \approx 0). \tag{98}$$

In the latter case, from eq. (180)

$$\frac{dy}{du} \approx -\frac{u}{ka} y^2 \quad (\gamma \approx \infty). \tag{99}$$

Now let the incident mode be a TE_{11} mode. We distinguish two cases depending on whether the ith mode is a TM mode or a TE mode. In the former case, from eqs. (85) and (98)

$$\lim_{y \to \infty} R_i = -A_1 \frac{\beta_i - \beta_i'}{\beta_1 + \beta_i'} \sqrt{\frac{\beta_i \beta_1}{(u_1^2 - 1)}} \frac{1}{k}, \tag{100}$$

where u_1 is the first root of $J_1'(u_1) = 0$,

$$u_1 = 1.8411. \tag{101}$$

If, on the other hand, the ith mode is also a TE-mode, from eqs. (85) and (99),

$$\lim_{y \to \infty} R_i = -A_i \frac{\beta_i - \beta_i'}{\beta_1 + \beta_i'} \sqrt{\frac{\beta_1}{\beta_i} \frac{u_i^2 - 1}{u_1^2 - 1}}. \tag{102}$$

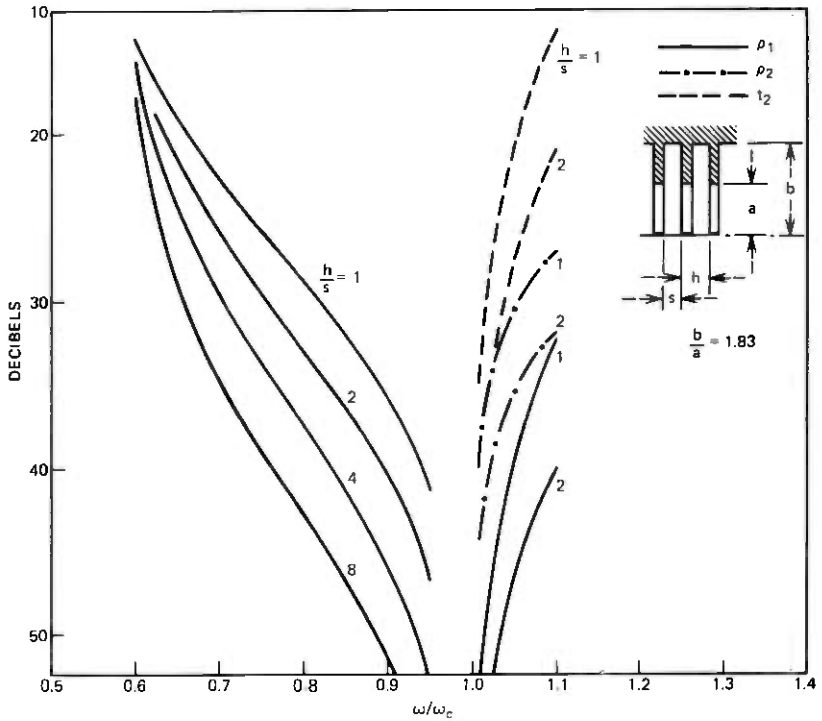

Fig. 2—Reflection, transmission, and coupling coefficients for input junction of Fig. 1.

From eq. (86) we obtain, using eq. (98),

$$\lim_{y\to\infty} T_i = A_1 \frac{1}{a\beta_1 - a\beta_i'} \sqrt{\frac{\beta_1 u_i'}{ka\beta_i'(u_1^2 - 1)}} \frac{1}{\sqrt{\left|\frac{dy'}{du_i'}\right|}}, \tag{103}$$

where dy'/du_i' can be determined using eq. (178), unless $y' \gg 1$, in which case we can use eq. (98) or (99) with y,u replaced by y', u_i'.

Equations (100) to (103) have been used to calculate the behavior of a junction with $b = 1.8309a$. Consideration has been restricted to the TE_{11} mode and the TM_{11} mode of the smooth waveguide, and the corresponding modes (HE_{11} and HE_{11}') of the corrugated waveguide. The results are shown in Fig. 2, where $i = 2$ refers to the TM_{11} mode (or the HE_{11}' mode),

$$\rho_1^2 = \left|\frac{R_1}{A_1}\right|^2 \tag{104}$$

is the input reflection,

$$\rho_2^2 = \left|\frac{R_2}{A_1}\right|^2 \tag{105}$$

gives the power converted into the TM_{11} mode, and

$$t_2^2 = \left|\frac{T_2}{A_1}\right|^2 \tag{106}$$

gives the power converted into the HE_{11}' mode. In Fig. 2, ω_c is the frequency at which $y' = \infty$. The corrugated guide at this frequency behaves like a smooth guide and, hence,

$$\rho_1 = \rho_2 = t_2 = 0. \tag{107}$$

The curves of Fig. 2 are useful in determining the practical bandwidth of the junction of Fig. 1.

VI. JUNCTION BETWEEN TWO WAVEGUIDES OF DIFFERENT DIAMETER

For some applications, to minimize the input reflection of a corrugated feed, it may be convenient to choose for the smooth waveguide a diameter different from that of the corrugated waveguide. In this section, the analysis of Section V is extended to the general case of a junction between two corrugated waveguides of different diameter. Let a and a' be the two diameters for $z > 0$ and $z < 0$, respectively, and assume again a single mode is incident on the junction, from the region $z < 0$.

If $\mathbf{E}_t$ for $z = 0$ were known, then the transmission coefficients T_i which appear in eqs. (70) and (71) could be determined at once using the formula*

$$T_i = \int\int_{S'} (\mathbf{E}_t \times \mathbf{h}_i^*) \cdot \mathbf{i}_z \, dS, \quad \text{for } z = 0, \tag{108}$$

which follows directly from eq. (70) in view of the orthogonality of the modes $\mathbf{e}_i'$, $\mathbf{h}_i'$ [see eq. (73)]. In eq. (108) S' denotes the circular area

$$0 < r < a'.$$

Now, for $z = 0$, $\mathbf{E}_t$ is given by eq. (68) inside the area

$$0 < r < a, \tag{109}$$

and it vanishes for $a < r < a'$. Therefore, eqs. (108) and (68) give

$$T_i = A_1(\mathbf{e}_1, \mathbf{h}_i') + \sum_{n=1}^{\infty} R_n(\mathbf{e}_n, \mathbf{h}_i'), \tag{110}$$

where

$$(\mathbf{e}_n, \mathbf{h}_i') = \int\int_S (\mathbf{e}_n \times \mathbf{h}_i'^*) \cdot \mathbf{i}_z \, dS, \tag{111}$$

S being the circular area (109), which corresponds to the waveguide of the region $z < 0$.

* Here we are only interested in calculating T_i and R_i for $\beta_i > 0$, $\beta_i' > 0$.

Equation (79), which was obtained by requiring continuity of $\mathbf{H}_t$ for $z = 0$, must be satisfied over the area S. By multiplying this equation with $\mathbf{e}_n$ and integrating over S, we obtain for $n \neq 1$

$$-R_n = \sum_1^\infty T_i(\mathbf{e}_n, \mathbf{h}_i') \quad (n \neq 1). \tag{112}$$

If the coefficients T_i in this relation are expressed in terms of the coefficients R_i using eq. (110), we get for $n \neq 1$

$$-R_n = (A_1 + R_1) \sum_{i=1}^{\infty} (\mathbf{e}_1, \mathbf{h}_i')(\mathbf{e}_n, \mathbf{h}_i') + R_n \sum_{i=1}^{\infty} (\mathbf{e}_n, \mathbf{h}_i')^2 + \sum_{s \neq n,1} R_s \sum_{i=1}^{\infty} (\mathbf{e}_s, \mathbf{h}_i')(\mathbf{e}_n, \mathbf{h}_i'). \tag{113}$$

For $n = 1$, the second sum of the right-hand side should be omitted and, furthermore, $-R_n$ should be replaced with $A_1 - R_1$.

We have thus obtained a system of equations in the unknowns R_1, R_2, etc. We solve* them in the limiting case where both $a' - a$ and $y' - y$ are very small, in which case

$$\left.\begin{aligned} (\mathbf{e}_n, \mathbf{h}_s') &\approx 0 \quad \text{for } n \neq s \\ R_i &\approx 0 \\ (\mathbf{e}_n, \mathbf{h}_k') - 1 &\approx 0 \end{aligned}\right\}, \tag{114}$$

and therefore the first two terms of the right-hand side of eq. (113) for $n \neq 1$ are respectively equal to

$$A_1[(\mathbf{e}_n, \mathbf{h}_1') + (\mathbf{e}_1, \mathbf{h}_n')]$$

and R_n. The last term can be neglected. Therefore, eq. (113) gives for $n \neq 1$

$$R_n \approx -\frac{1}{2}[(\mathbf{e}_1, \mathbf{h}_n') + (\mathbf{e}_n, \mathbf{h}_1')]A_1 \quad (n \neq 1). \tag{115}$$

Similarly, for $n = 1$,

$$R_1 \approx \frac{1}{2}[1 - (\mathbf{e}_1, \mathbf{h}_1')^2]A_1 \approx [1 - (\mathbf{e}_1, \mathbf{h}_1')]A_1. \tag{116}$$

The transmission coefficients can now be determined using eq. (110). We find for $n \neq 1$

$$T_n \approx \frac{1}{2}[(\mathbf{e}_1, \mathbf{h}_n') - (\mathbf{e}_n, \mathbf{h}_1')]A_1, \quad (n \neq 1), \tag{117}$$

which is a generalization of eq. (84).

* This derivation is not rigorous, for we neglect to examine the question of convergence of the summations in eq. (113). However, the validity of the results appears to be confirmed by the experimental results.

The coefficients $(\mathbf{e}_i, \mathbf{h}'_n)$ can be calculated using eq. (57). If $y' - y$ and $a' - a$ are very small, we can proceed as follows. The field components $\mathbf{e}_n$ and $\mathbf{h}_n$ of the nth mode are considered to be functions of the coordinates r,ϕ and of the two waveguide parameters a,y. Therefore,

$$\mathbf{h}'_n = \mathbf{h}_n(r, \phi, a', y') \simeq \mathbf{h}_n + \frac{\partial \mathbf{h}_n}{\partial y}\delta y + \frac{\partial \mathbf{h}_n}{\partial a}\delta a, \tag{118}$$

where $\mathbf{h}_n$, $\partial \mathbf{h}_n/\partial y$ and $\partial \mathbf{h}_n/\partial a$ are evaluated for $a' = a$, $y' = y$, and δy are δa denote $y' - y$ and $a' - a$. a similar relation can be written for $\mathbf{e}'_n$. It follows from eq. (118) that $(\mathbf{e}_i, \mathbf{h}'_n)$ for $i \neq n$ is a sum of two terms,

$$(\mathbf{e}_i, \mathbf{h}'_n) = \left(\mathbf{e}_i, \frac{\partial \mathbf{h}_n}{\partial y}\right)\delta y + \left(\mathbf{e}_i \frac{\partial \mathbf{h}_n}{\partial a}\right)\delta a, \tag{119}$$

since $(\mathbf{e}_i, \mathbf{h}_n) = 0$. The first term is simply the coefficient $(\mathbf{e}_i, \mathbf{h}'_n)$ calculated for $a' = a$; it corresponds to a junction between two waveguides of the same radius, but different surface reactance. The second term can be interpreted as the coefficient $(\mathbf{e}_i, \mathbf{h}'_n)$ relative to a junction between two waveguides having the same surface reactance but different radii a and a'. Since the term has already been treated in Section V, only the latter need be considered. If one sets

$$\delta \mathbf{h}_n = \frac{\partial \mathbf{h}_n}{\partial a}\delta a, \quad \delta \mathbf{e}_n = \frac{\partial \mathbf{e}_n}{\partial a}\delta a, \tag{120}$$

and if the ϕ variations of both modes are of the type considered in Section I, then, taking into account that $e_{i\phi} = 0$ for $r = a$, using eq. (57) we get

$$(\mathbf{e}_i, \delta \mathbf{h}_n) = \pi a \delta a \left[-j \frac{\beta_n}{\beta_i^2 - \beta_n^2} \left(e_{i\phi} \frac{\partial h^*_{nz}}{\partial a} + \frac{\partial e^*_{n\phi}}{\partial a} h_{iz} \right)_{\substack{\phi=90° \\ r=a}} \right.$$
$$\left. + j \frac{\beta_i}{\beta_i^2 - \beta_n^2} \left(e_{iz} \frac{\partial h^*_{n\phi}}{\partial a} + \frac{\partial e^*_{nz}}{\partial a} h_{i\phi} \right)_{\substack{\phi=0° \\ r=a}} \right]. \tag{121}$$

As an application, consider $y = \infty$, in which case $(\mathbf{e}_i, \mathbf{h}'_n)$ can be interpreted as the coefficient $(\mathbf{e}_i, \mathbf{h}'_n)$ relative to a junction between two smooth waveguides of radii a and $a' = a + \delta a$, respectively. Assume $\mathbf{e}_i$ is a TE mode and $\mathbf{e}_n$ is a TM mode, so that for $r = a$

$$e_{i\phi} = e_{iz} = \frac{\partial h^*_{nz}}{\partial a} = 0, \tag{122}$$

where the last term vanishes because $\mathbf{h}_n$ is a TM mode. Then eq. (121) gives

$$(\mathbf{e}_i, \delta \mathbf{h}_n) = \pi a \delta a \frac{1}{\beta_i^2 - \beta_n^2} \left[-j\beta_n \left(\frac{\partial e^*_{n\phi}}{\partial a} h_{iz} \right)_{\substack{\phi=90° \\ r=a}} \right.$$
$$\left. + j\beta_i \left(\frac{\partial e^*_{nz}}{\partial a} h_{i\phi} \right)_{\substack{\phi=0° \\ r=a}} \right]. \tag{123}$$

Similarly, interchanging $n \leftrightarrows i$ in eq. (121) and taking into account that for $r = a$

$$\frac{\partial e_{iz}}{\partial a} = e_{n\phi} = e_{nz} = h_{nz} = 0, \tag{124}$$

we obtain

$$(\mathbf{e}_n, \delta\mathbf{h}_i) = 0. \tag{125}$$

Now, the two modes are characterized by

$$h_{iz} = \frac{1}{Z_0}\eta_i J_1\left(\frac{r}{a}u_i\right)\sin\phi, \tag{126}$$

$$h_{i\phi} = -\frac{j}{Z_0}\eta_i \frac{\beta_i a}{u_i}\frac{J_1\left(\frac{r}{a}u_i\right)}{\frac{r}{a}u_i}\cos\phi, \tag{127}$$

and

$$e_{nz} = \eta_n J_1\left(\frac{r}{a}u_n\right)\cos\phi, \tag{128}$$

$$e_{n\phi} = j\eta_n \frac{\beta_n a}{u_n}\frac{J_1\left(\frac{r}{a}u_n\right)}{\frac{r}{a}u_n}\sin\phi, \tag{129}$$

where the amplitudes η_i and η_n are, because of the requirement (72), given by

$$\eta_i = \sqrt{\frac{2Z_0}{\pi a}}\frac{1}{a}\frac{u_i^2}{\sqrt{\beta_i}}\frac{1}{\sqrt{u_i^2 - 1}}\frac{1}{|J_1(u_i)|}\sqrt{\frac{1}{(ka)}}. \tag{130}$$

$$\eta_n = \sqrt{\frac{2Z_0}{\pi a}}\frac{1}{a}\frac{u_n}{\sqrt{\beta_n}}\frac{1}{|J_1(u_n)|}\sqrt{\frac{1}{(ka)}}. \tag{131}$$

From eqs. (123) and (126) to (131), taking into account that $J_1(u_n) = 0$, we obtain the final result

$$(\mathbf{e}_i, \delta\mathbf{h}_n) = 2\frac{ka}{\sqrt{a\beta_i a\beta_n}}\frac{1}{\sqrt{u_i^2 - 1}}\frac{\delta a}{a}. \tag{132}$$

Note that in deriving this relation it has been assumed that δa is sufficiently small so that

$$\frac{1}{\sqrt{\beta}} \approx \frac{1}{\sqrt{\beta_n'}}.$$

If this condition is not satisfied, we should replace β_n with β_n' in eq. (132).

Of special interest is the case where e_i and e_n represent the TE_{11} mode and the TM_{11} mode, respectively. In this case $u_i = 1.8411$ and, letting $i = 1$ and $n = 2$ for these two modes, we get

$$(\mathbf{e}_1, \delta \mathbf{h}_2) = 1.2937 \frac{ka}{\sqrt{a\beta_1 a\beta_2}} \frac{\delta a}{a}. \tag{133}$$

From eqs. (115), (117), (125), and (133) we then obtain for the conversion coefficients T_2 and R_2

$$T_2 \simeq -R_2 \simeq 0.646 \frac{ka}{\sqrt{a\beta_1 a\beta_2}} \frac{\delta a}{a} \times A_1, \tag{134}$$

where $|A_1|^2$, $|T_2|^2$, and $|R_2|^2$ represent the incident power, and the powers transmitted and reflected in the TM_{11} mode. We can verify* that T_2 is smaller by a factor of 2 than the conversion coefficient given in Ref. 22, which is due to the fact that the assumptions of Ref. 22 imply

$$T_n \simeq (\mathbf{e}_1, \mathbf{h}_n'), \tag{135}$$

rather than eq. (117).

Note that for $\beta_2 \to 0$, we have $a\beta_2' \to u_2\sqrt{\delta a/a}$, and therefore

$$T_2 \simeq 0.646 \sqrt{\frac{u_2}{a\beta_1}} \left(\frac{\delta a}{a}\right)^{3/4}. \tag{136}$$

Note T_2 remains finite even when the TM_{11} mode approaches cutoff in the first guide.

VII. MODE CONVERSION IN A NONUNIFORM WAVEGUIDE

Typically, a corrugated feed is made of one or more sections of nonuniform waveguide whose surface reactance and radius are functions of z. Since a nonuniform waveguide does not in general possess a natural mode of propagation, an incident mode will be scattered in forward and backward modes. This is true even for a conical waveguide of constant surface reactance (except when $y = 0$ or $y = \infty$). The analysis of Sections V and VI gives the differential scattering parameters which allow the local coupling into forward and backward modes to be determined at any point in a uniform waveguide. We can thus obtain a set of differential equations, whose coefficients are given by the above scattering parameters, and which can be solved, at least in principle, for the mode am-

* In Refs. 22 and 23, the TM_{11} mode was cut off to the left of the junction, and for this reason there is poor agreement between those measurements and eq. (134), which is not applicable in this case. However, numerical calculations by Masterman and Clarricoats agre well with eq. (134) at frequencies well above the cutoff frequency of the TM_{11} mode, as we may verify from Fig. 11 of Ref. 24.

plitudes. We confine ourselves to a first-order treatment assuming the total scattered power is much less than the incident power, since this is the most interesting case if the feed is well designed. It is convenient to assume for the moment that only y varies with z, in which case the waveguide can be approximated by a succession of junctions of the type considered in Section V. Let the HE_{11} mode be incident at the input ($z = z_1$). We wish to determine the resulting amplitude $T_2(z)$ of the HE'_{11} mode for $z = z_2$. If the variation of y is sufficiently slow, we can neglect reflections and determine T_2 assuming the amplitude A_1 of the HE_{11} mode is nearly constant. The transverse field of the fundamental mode is then

$$A_1 \mathbf{e}_1 e^{-j\Phi_1(z)}, \tag{137}$$

where

$$\frac{d\Phi_1}{dz} = \beta_1. \tag{138}$$

The effect on the HE'_{11} mode of a small variation δy at $z = \xi$ is to produce at $z \gtrsim z_2$ a component

$$dT_2 = t_2(\xi)\delta y A_1 e^{-j\Phi_1(\xi)} e^{-j[\Phi_2(z_2)-\Phi_2(\xi)]}, \tag{139}$$

where

$$\frac{d\Phi_2}{dz} = \beta_2, \tag{140}$$

and from Eq. (86),

$$t_2(\xi) = \frac{-1}{a\beta_1 - a\beta_2} \sqrt{\frac{u_1 u_2}{a^2 \beta_1 \beta_2} \frac{1}{|(dy/du_1)(dy/du_2)|}}. \tag{141}$$

Note that both β_1 and β_2 are functions of z. From eq. (139), integrating from z_1 to z_2,

$$T_2(z_2) = \int_{z_1}^{z_2} dT_2 = A_1 e^{-j\Phi_2(z_2)} \times \int_{z_1}^{z_2} t_2(z) e^{+j[\phi_2(z)-\phi_1(z)]} \frac{dy}{dz} dz, \tag{142}$$

which assumes a very simple form when $ka \gg 1$, as discussed in the following section.

7.1 Conical horn with ka ≫ 1

Suppose the radius a varies linearly with z, as shown in Fig. 1 for $z > z_1$, that the flare angle α is very small, and

$$ka(z) \gg 1$$

for $z_1 < z < z_2$. It was shown in Section III that for $ka \gg 1$ the properties of a mode are entirely determined* by y/ka and, therefore, a mode will propagate without variation of its transverse field distribution if y/ka = constant. For this reason, it is reasonable to assume that mode conversion will be negligible if

$$\delta\left(\frac{y}{ka}\right) = 0. \tag{143}$$

Under this assumption, the effect on the amplitude of the HE'_{11} mode for $z = z_2$ of a small variation $\delta(y/ka)$ occurring at $z = \xi$ can be expressed in the form

$$dT_2 = \tau_2(\xi)\delta\left(\frac{y}{ka}\right) A_1 e^{-j\Phi_1(\xi)} e^{-j[\phi_2(z_2)-\phi_2(\xi)]}, \tag{144}$$

where, since eqs. (144) and (139) must agree in the particular case a = constant,

$$\tau_2(\xi) = kat_2(\xi). \tag{145}$$

From eq. (96),

$$\tau_2 = -\frac{u_1 u_2}{u_2^2 - u_i^2} \quad (ka \gg 1). \tag{146}$$

Therefore,

$$T_2(z_2) = -A_1 e^{-j\Phi_2(z_2)} \frac{u_1 u_2}{u_2^2 - u_1^2} \int_{z_1}^{z_2} e^{j[\Phi_1(\xi)-\Phi_2(\xi)]} d\left(\frac{y}{ka}\right). \tag{147}$$

Note that, since $ka \gg 1$,

$$a\beta_i \simeq ka - \frac{1}{2}\frac{u_i^2}{ka}. \tag{148}$$

Therefore, from eqs. (138) and (140)

$$\frac{d}{dz}[\Phi_1(z) - \Phi_2(z)] = \frac{k}{2}\frac{u_2^2 - u_1^2}{(ka)^2}. \tag{149}$$

Now, a varies linearly with z,

$$a = (z - z_0)\tan\alpha, \tag{150}$$

* Since now we are dealing with a conical waveguide, each mode is a spherical wave centered at the apex A of the waveguide, and the field distribution over a spherical wavefront is given to a first approximation (small α) by eqs. (35) and (37). To obtain the field distribution over a plane z = constant, we must therefore introdce in eqs. (35) and (37) a factor of exp $(-j\psi)$, $\psi = k(x^2 + y^2)/2R$, R being the distance of the plane from the apex.

where z_0 is the value of z at the apex of the waveguide. Therefore, from eqs. (149) and (150)

$$\Phi_1(z) - \Phi_2(z) = -\frac{1}{2}\frac{u_2^2 - u_1^2}{\mathrm{y}\tan\alpha}\frac{\mathrm{y}}{ka}. \tag{151}$$

Of particular interest is the case

$$\mathrm{y} = \text{constant}. \tag{152}$$

Then eq. (147) can be readily integrated. Taking into account eq. (151), we obtain

$$|T_2|^2 = |A_1|^2 \frac{4u_1^2u_2^2\mathrm{y}^2\tan\alpha^2}{(u_2^2 - u_1^2)^4} \times \left|1 - \exp\left[j\frac{1}{2}\frac{u_2^2 - u_1^2}{\mathrm{y}\tan\alpha}\frac{\mathrm{y}}{ka_1}\left(1 - \frac{a_1}{a_2}\right)\right]\right|^2, \tag{153}$$

where $a_i = a(z_i)$. Therefore,

$$|T_2|^2 \leqq |A_1|^2 \frac{16u_1^2u_2^2\mathrm{y}^2\tan\alpha^2}{(u_2^2 - \mathrm{u}_1^2)^4}, \tag{154}$$

where it is recalled that $u_2 = 5.1356$ and $u_1 = 2.4048$. For $\alpha = 4°$, which is the value chosen in the experiment of Ref. 17, this inequality gives for $\mathrm{y} = 1$

$$\frac{|T_2|^2}{|A_1|^2} \leqq 0.6636\ 10^{-4}, \quad (-41.8\ \mathrm{dB}), \tag{155}$$

which is a very small value for most aplications. For $\alpha = 16°$, on the other hand, we obtain -29.8 dB, which may no be negligible.

Note that $|T_2|^2$ and $|A_1|^2$ are, respectively, the powers carried by the HE_{11}' mode and the HE_{11} mode.

VIII. SUMMARY

In the feed of Fig. 1, when a TE_{11} mode is incident at the input, some of the incident power is in general reflected. Furthermore, some power may be converted to unwanted modes if the corrugated waveguide supports more than one mode at the input. Additional mode conversion may take place inside the feed if the variation of the radius and of the surface reactance is not gradual enough. As a result, a feed will have a nonzero input reflection and, at some frequencies, unwanted modes may illuminate the aperture of the feed. The consequences of these unwanted modes on the radiation characteristics—e.g., enhanced cross polarization—are pointed out in Section III and in Ref. 17, where the theory is compared with experiment.

These effects can be evaluated to good accuracy using the expressions derived in this article, as highlighted below. For large ka, eq. (35) ex-

presses the field shape for all copolarized hybrid modes of various radial harmonics with the same ϕ dependence, while eq. (37) corresponds to the cross-polarized modes. Equation (36) gives the mode amplitudes required to normalize the power carried by the modes.

A property of corrugated feeds is that the aperture field distribution does not remain constant with frequency, as in the case of a feed with smooth walls, but varies because of the frequency dependence of the surface reactance X_s. Thus, although the desired mode has no cross-polarized component at the resonant frequency of the corrugations, at other frequencies the *desired mode does radiate* some cross polarization. The ratio, C^2, between the maximum value of the cross-polarized power in the radiation pattern and the maximum value of the copolarized power (which occurs on axis) is given by eq. (39). From eq. (39), it follows that with large ka and thin disks one can maintain low cross-polarized power in the radiation pattern from the desired mode over an octave or more.

At a junction between waveguides of the same diameter but of different surface reactance, eq. (85) gives the general expression for the mode coupling coefficient to modes reflected from the transition, and Eq. (86) gives that for modes transmitted forward from the transition. Equations (97) and (96) are simplifications of eqs. (85) and (86), respectively, which apply for $ka \gg 1$. When the input waveguide is smooth, the mode coupling coefficient is given by eq. (100) for reflected TM modes, by eq. (102) for reflected TE modes, and by eq. (103) for hybrid modes transmitted forward from the transition. Since the transition from smooth to corrugated waveguide is a major source of unwanted modes in a corrugated horn, eq. (103) is very useful in determining mode purity. Another important formula is eq. (87), which determines the reflection coefficient of the dominant mode (return loss) for any transition in X_s and any ka.

Another source of generation of undesired modes is the mode conversion occurring along the conical taper of a corrugated horn. Equation (154) gives the mode-coupling coefficient for the transmitted undesired mode due to a conical taper.

In some cases a step in diameter may be used to match transitions between different surface impedances; eq. (134) determines the mode-coupling coefficients at a step in diameter.

APPENDIX A

Asymptotic Series for u and γ in Terms of 1/ka

We determine the asymptotic series for u and γ in terms of

$$\frac{1}{ka}, \tag{156}$$

under the assumption $y \neq 0$. It is convenient to introduce in eq. (20) the quantity[12]

$$F = \frac{uJ_1'(u)}{J_1(u)} = u\frac{J_0(u)}{J_1(u)} - 1. \tag{157}$$

Note that

$$\frac{dF}{du} = \frac{J_0(u)}{J_1(u)} - u - u\frac{J_1'(u)}{J_1(u)}\frac{J_0(u)}{J_1(u)}. \tag{158}$$

But from eq. (157)

$$\frac{J_0(u)}{J_1(u)} = \frac{F}{u} + \frac{1}{u}.$$

Therefore, eq. (158) gives

$$\frac{dF}{du} = -\frac{1}{u}F^2 + \left(\frac{1}{u} - u\right). \tag{159}$$

It follows that

$$u\frac{dF}{du} + F^2 - 1 + u^2 = 0, \tag{160}$$

$$u\frac{d^2F}{du^2} + \frac{dF}{du}(1 + 2F) + 2u = 0, \tag{161}$$

$$u\frac{d^3F}{du^3} + 2\frac{d^2F}{du^2}(1 + F) + 2\left(\frac{dF}{du}\right)^2 + 2 = 0, \tag{162}$$

etc.

In terms of F, eq. (20) can be rewritten

$$yu^2\frac{1}{ka}F - F^2 + 1 - u^2\left(\frac{1}{ka}\right)^2 = 0, \tag{163}$$

or, using eq. (159),

$$yu\frac{1}{ka}F + \frac{dF}{du} + u\left[1 - \left(\frac{1}{ka}\right)^2\right] = 0. \tag{164}$$

Now assume

$$u = u_{0m}\left[1 + \alpha_1\frac{1}{ka} + \alpha_2\left(\frac{1}{ka}\right)^2 + \cdots\right]. \tag{165}$$

Develop F in a Taylor series about the point $u = u_{0m}$,

$$F = F(u_{0m}) + \left(\frac{dF}{du}\right)_{u=u_{0m}}(u - u_{0m}) + \cdots. \tag{166}$$

From eqs. (160) to (162), with $u = u_{0m}$, taking into account that

$$F(u_{0m}) = -1, \tag{167}$$

we obtain for the derivatives appearing in eq. (166),

$$\left.\begin{aligned} \left(\frac{dF}{du}\right)_{u=u_{0m}} &= -u_{0m} \\ \left(\frac{d^2F}{du^2}\right)_{u=u_{0m}} &= -3. \\ \left(\frac{d^3F}{du^3}\right)_{u=u_{0m}} &= -2\frac{1+u_{0m}^2}{u_{0m}}, \text{ etc.} \end{aligned}\right\} \tag{168}$$

Substituting eqs. (165), (167), and (168) in eq. (166) one obtains F as a series of powers of $1/ka$; the coefficients of this series are algebraic expressions in u_{0m} and α_1, α_2, etc. Similarly, by developing dF/du in a Taylor series about the point $u = u_{0m}$, and then using eqs. (165), (167), and (168), we obtain dF/du as a series of powers of $1/ka$. Substituting eq. (165) and the above series expansions of F and dF/du in eq. (164), we can solve for the coefficients α_1, α_2, etc. We obtain eq. (31). Substituting eq. (31) in eq. (166) we obtain an expansion of F in powers of $1/ka$. Using these results, from

$$\gamma = -\frac{F}{\cos\theta_1} = -\frac{F,}{\sqrt{1-\left(\frac{u}{ka}\right)^2}}, \tag{169}$$

we get eq. (32).

Equations (31) and (32) have been obtained assuming eq. (165), which corresponds to the limiting case of eq. (28). If, instead of assuming eq. (165), we assume

$$u = u'_m = u_{2m}\left[1 + \beta_1\left(\frac{1}{ka}\right) + \beta_2\left(\frac{1}{ka}\right)^2 + \cdots\right], \tag{170}$$

we obtain eqs. (33) and (34).

APPENDIX B

Surface Wave Mode

In addition to the modes considered in Section III, there is a mode for which $\beta > k$. Thus, since for this mode $\cos\theta_1 > 1$, it is convenient to replace θ_1 with $j\theta_1$ in eqs. (16) and (18). Since

$$\left.\begin{aligned} \cos j\theta &= \cosh\theta_1 \\ \sin j\theta_1 &= j\sinh\theta_1 \\ J_1(jx) &= jI_1(x), \\ J'_1(jx) &= I'_1(x), \end{aligned}\right\}, \tag{171}$$

where $I_1(x)$ is the modified Bessel function of order 1, we obtain from eqs. (16) and (17)

$$\gamma = -\frac{I_1'(u)}{I_1(u)}\frac{u}{\cosh \theta_1}, \tag{172}$$

$$\frac{\mathrm{y}}{ka} = -\left[\frac{1}{u}\frac{I_1'(u)}{I_1(u)} + \frac{\cosh \theta_1}{u}\frac{1}{\gamma}\frac{1}{u}\right], \tag{173}$$

where $u = ka \sinh \theta_1$. We can verify from these two equations that $u \to \infty$ as $ka \to \infty$. Now, for large u,

$$I_1'(u) \approx I_1(u) \approx \frac{e^u}{\sqrt{2\pi u}}. \tag{174}$$

Therefore eq. (172) gives

$$\gamma \to \infty, \quad \text{as } ka \to \infty.$$

From eq. (173), for large u and ka,

$$\frac{\mathrm{y}}{ka} \approx -\frac{1}{u}. \tag{175}$$

Therefore, since $u = ka \sinh \theta_1$,

$$\sinh \theta_1 \approx -\frac{1}{\mathrm{y}}. \tag{176}$$

We now examine the behavior of the field components for $ka \to \infty$. Taking into account eq. (171), from eqs. (10) to (15), after replacing θ_1 with $j\theta_1$, we find for $\gamma = -\infty$ (i.e., for $B = 0$) that the only nonzero component of the magnetic field is H_ϕ and

$$-Z_0 H_\phi = \frac{1}{\sinh \theta_1} A I_1'\left(u\frac{r}{a}\right) \cos\phi \, e^{-j\beta z}.$$

Therefore, for $kr \gg \mathrm{y}$

$$-Z_0 H_\phi \approx A' \sqrt{\frac{a}{r}} \cos\phi \exp[k(r-a)\sinh \theta_1 - j\beta z],$$

where A' is a constant. This shows that the field is confined to the near vicinity of the wall, decaying approximately exponentially from the wall.

We can show that $ka > 1.81$, the surface wave mode in combination with the $\mathrm{HE}_{1\mathrm{m}}$ and $\mathrm{HE}_{1\mathrm{m}}'$ modes comprise the complete set of propagating modes whose E_z azimuthal dependence is $\cos\phi$.

APPENDIX C

Derivation of dy/du

From eq. (163),

$$\frac{y}{ka} = \frac{1}{u^2}F + \left(\frac{1}{(ka)^2} - \frac{1}{u^2}\right)\frac{1}{F}.$$

where

$$F = \frac{uJ_1'(u)}{J_1(u)}.$$

Furthermore, from eq. (164)

$$F' = -\frac{1}{u}F^2 + \left(\frac{1}{u} - u\right).$$

Therefore,

$$\frac{1}{ka}\frac{dy}{du} = -\frac{2}{u^3}\left(F - \frac{1}{F}\right) + \frac{1}{u}\left[\left(\frac{1}{ka}\right)^2 - 1\right] - \frac{1}{u^3}F^2 - \left(\frac{1}{u} - u\right)\left[\left(\frac{1}{ka}\right)^2 - \frac{1}{u^2}\right]\frac{1}{F^2}. \quad (178)$$

For the case $y \to \infty$, we now determine

$$\lim_{y\to\infty}\frac{dy}{du}.$$

For $y \eqsim \infty$ there are two types of modes: TE modes, in which case $\gamma \eqsim 0$, and TM modes, in which case $\gamma \eqsim \infty$. In the former case from eq. (169) we have $F \eqsim 0$ and, therefore, eq. (178) gives

$$\lim_{y\to\infty}\frac{dy}{du} = -\frac{1}{u}(u^2 - 1)\frac{(ka)^2 - u^2}{u^2 ka}\frac{1}{F^2} = -y^2(u^2 - 1)\frac{kau}{(ka)^2 - u^2}, \quad (\gamma \eqsim 0) \quad (179)$$

since from eq. (163) for $F \eqsim 0$

$$\lim_{y\to\infty} F = -\frac{1}{y}\frac{(ka)^2 - u^2}{kau^2}$$

In the latter case ($\gamma \eqsim \infty$) from eq. (169), we have $F \eqsim \infty$. More precisely, from eq. (163)

$$F \eqsim \frac{u^2}{ka}y,$$

and therefore from eq. (178)

$$\frac{dy}{du} \eqsim \frac{ka}{u^3}F^2 \eqsim -\frac{u}{ka}y^2 \quad (\gamma \eqsim \infty). \quad (180)$$

APPENDIX D

Cutoff Frequencies of the Modes of Equations (35) and (37)

From eqs. (16) and (17) we get

$$\gamma^2 + \gamma\omega - 1 = 0,$$

where

$$\omega = \frac{\mathrm{y}u^2}{ka\,\cos\theta_1}.$$

Therefore, either

$$\gamma = \frac{-\omega + \sqrt{\omega^2 + 4}}{2} \tag{181}$$

or

$$\gamma = \frac{-\omega - \sqrt{\omega^2 + 4}}{2}, \tag{182}$$

respectively, in the two cases of eqs. (35) and (37) (which correspond, respectively, to $\gamma \to 1$ and $\gamma \to -1$ for $ka \to \infty$). At cutoff $\beta \to 0$; i.e., $\cos\theta_1 \to 0$. For $\cos\theta_1 \simeq 0$, $\mathrm{y} \neq 0$, we have $|\omega| \to \infty$ and, therefore, from eq. (181)

$$\gamma \to \begin{cases} 0 & \text{if } \mathrm{y} > 0 \\ \infty & \text{if } \mathrm{y} < 0 \end{cases}, \tag{183}$$

(184)

whereas from eq. (182)

$$-\gamma \to \begin{cases} \infty, & \text{if } \mathrm{y} > 0 \\ 0, & \text{if } \mathrm{y} < 0 \end{cases}.$$

If $\gamma = 0$, the mode is of the TE type. Now, for a TE mode at cutoff, the only nonzero component of the magnetic field is H_z and therefore the surface reactance X_s has no effect on the cutoff frequency. This means that the cutoff frequency can be determined by replacing the corrugated wall with a smooth wall of radius a, and therefore the cutoff frequency is determined by the condition:

$$J_1'(ka) = 0.$$

If $\gamma = \infty$, on the other hand, the mode is of the TM type and the only nonzero component of the electric field at cutoff is E_z. It follows that if the disks are very thin ($t \simeq 0$), they can be removed without affecting the field. Thus, the cutoff frequency can in this case be determined by replacing the corrugated waveguide with a smooth waveguide of radius b. It is thus determined by the condition

$$J_1(kb) = 0.$$

When y = 0, $\gamma = \pm 1$, and from eq. (16), $J_1'(ka) = 0$ for $\cos\theta_1 = 0$. Thus, when y = 0 both types of mode have the same cutoff frequencies.

REFERENCES

1. A. F. Kay, "The Scalar Feed," U.S. Air Force, Cambridge Research Laboratory Report, 62-347, March 1964.
2. V. H. Rumsey, "Horn Antennas with Uniform Power Patterns Around Their Axes," IEEE Trans. Ant. Propag., *AP-14*, No. 5 (September 1966), pp. 656–658.
3. A. F. Kay, "A Wide Flare Horn—A Novel Feed for Low Noise Broadband with High Aperture Efficiency Antennas," U.S. Air Force, Cambridge Research Laboratory Report, 62-757, October 1962.
4. H. C. Minnett and B. MacA. Thomas, "A Method of Synthesising Radiation Patterns with Axial Symmetry," IEEE Trans. Ant. Propag., *AP-14*, No. 5 (September 1966), pp. 654–656.
5. T. S. Chu, unpublished work, 1967.
6. H. C. Minnett and B. MacA. Thomas, "Fields in the Image Space of Symmetrical Focusing Reflectors," Proc. IEE, *115*, No. 10 (October 1968), pp. 1419–1430.
7. G. H. Bryant, "Propagation in Corrugated Waveguides," Proc. IEE, *116*, No. 2 (February 1969), pp. 203–213.
8. B. MacA. Thomas, "Bandwidth Properties of Corrugated Conical Horns," Electron. Lett., *5*, No. 22 (October 30, 1969), pp. 561–563.
9. M. S. Narasimlion and B. V. Rao, "Hybrid Modes in Corrugated Conical Horns," Electron. Lett., *6*, No. 2 (January 22, 1970), pp. 32–34.
10. T. B. Vu and Q. H. Vu, "Optimum Feed for Large Radio Telescopes: Experimental Results," Electron. Lett., *6* (March 19, 1970), pp. 159–160.
11. B. MacA. Thomas, "Prime-Focus One- and Two-Hybrid-Mode Feeds," Electron. Lett., *6*, No. 15 (July 1970), pp. 460–461.
12. P. J. B. Clarricoats and P. K. Saha, "Propagation and Radiation Behaviour of Corrugated Feeds; Part 1—Corrugated Waveguide Feed," Proc. IEE, *118*, No. 9 (September 1971), pp. 1167–1176.
13. P. T. B. Clarricoats and P. K. Saha, "Propagation and Radiation Behaviour of Corrugated Feeds: Part 2—Corrugated-Conical-Horn Feed," Proc. IEE, *118*, No. 9 (September 1971), pp. 1177–1186.
14. B. MacA. Thomas, "Theoretical Performance of Prime-Focus Paraboloids Using Cylindrical Hybrid-Mode Feeds," Proc. IEE, *118*, No. 11 (November 1971), pp. 1539–1549.
15. T. B. Vu, "Low-Noise Dual-Hyhrid-Mode Horn—An Experimental Model," Int. J. Electron., *34*, No. 3 (1973), pp. 391–400.
16. M. Mizusawa, F. Takeda, and S. Betsudau, "Radiation Characteristics of a Corrugated Conical Horn," Electron. Commun. Jap., *56-B*, No. 1 (January 1973), pp. 42–47.
17. C. Dragone, "Characteristics of a Broadband Corrugated Feed: A Comparison Between Theory and Experiment," this issue, pp. 869–888.
18. J. Brown, "Electromagnetic Momentum Associated with Waveguide Modes," Proc. IEE, *113*, No. 1 (January 1966), pp. 27–34.
19. L. Lewin, *Theory of Waveguides*, New York: John Wiley, 1955.
20. R. E. Collin, *Field Theory of Guided Waves*, New York: McGraw-Hill, 1960.
21. D. Marcuse, *Light Transmission Optics*, New York: Van Nostrand, 1972.
22. E. R. Nagelberg and J. Shefer, "Mode Conversion in Circular Waveguides," B.S.T.J., *44*, No. 7 (September 1965), pp. 1321–1338.
23. K. K. Agarwal and E. R. Nagelberg, "Phase Characteristics of a Circularly Symmetric Dual-Mode Transducer," IEEE Trans. Microw. Theory Tech., *MTT-18*, No. 1 (January 1970), pp. 69–71.
24. P. H. Masterman and P. J. B. Clarricoats, "Computer Field-Matching Solution of Waveguide Transverse Discontinuities," Proc. IEE, *118*, No. 1 (January 1971), pp. 51–63.
25. B. MacA. Thomas, "Mode Conversion Using Circumferentially Corrugated Cylindrical Waveguide," Electron. Lett., *8*, No. 15 (July 27, 1972), pp. 394–396.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 6, July–August 1977
Printed in U.S.A.

Characteristics of a Broadband Microwave Corrugated Feed: A Comparison Between Theory and Experiment

By C. DRAGONE

(Manuscript received December 14, 1976)

A corrugated feed with nearly ideal radiation characteristics from 17 GHz to 29 GHz has been built using a novel fabrication technique. The bandwidth of single-mode operation was maximized by properly choosing the input parameters of the feed. As a result, only the fundamental mode can propagate at the input from 19 GHz to 28.8 GHz. In this frequency range, the far field is essentially polarized in one direction. At frequencies higher than 28.8 GHz, there is a cross-polarized component caused by an unwanted mode. An approximate calculation of the power in this mode is given. A simple formula for the input reflection coefficient is provided. Results are included that show how to compute mode conversion in a conical taper, cross polarization from a corrugated horn, including contributions from spurious modes, and the reflection coefficient from the smooth-guide to corrugated-guide transition. Comparison of theory and experiment shows good agreement.

I. INTRODUCTION

It has been shown by Thomas[1] that under certain conditions the field over the aperture of a corrugated feed is virtually constant over a very wide frequency range. For this behaviour to occur, the radius a of the aperture must be much larger than a wavelength, i.e.,

$$ka \gg 1 \left(k = \frac{2\pi}{\lambda}\right) \tag{1}$$

and furthermore, the aperture must be illuminated by a single mode. If both conditions are satisfied, we can show that the field distribution is essentially independent of the surface reactance of the corrugated horn wall, X_s, and, as a consequence, it is little affected by the variation of X_s with frequency. This result is very important, for it implies that it

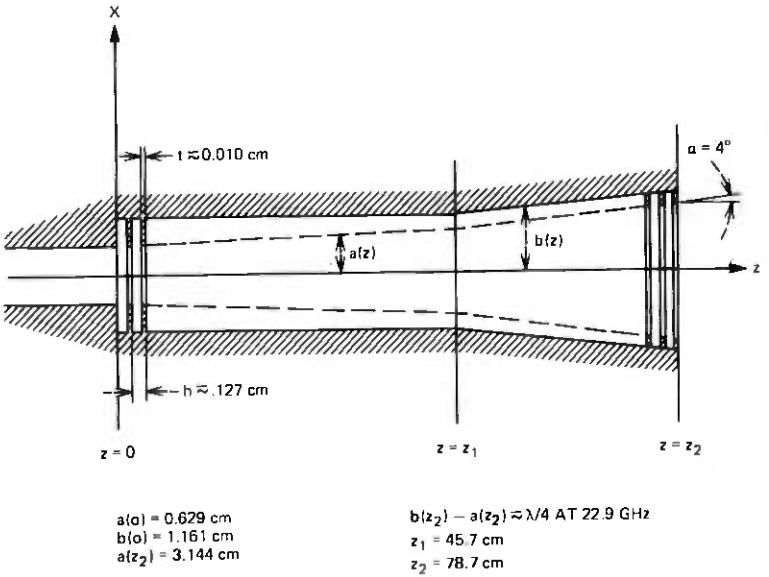

Fig. 1—Dimensions of the corrugated horn.

is possible to design a feed so that its radiation pattern is circularly symmetric and polarized in one direction over a very wide frequency range[1–3] (an octave or more). The most difficult condition to satisfy to obtain such large bandwidth is the requirement that a single mode be excited in the feed. This requirement is discussed in a separate article.[4] Here, after summarizing certain results of that article, we describe the results of an experiment. A long feed, with a flare angle of only 4°, was fabricated using a special technique described in Appendix D. At the input of the feed, which is shown in Fig. 1, the waveguide dimensions (the corrugated depth l and the radius a; see Fig. 1) were optimized so as to maximize the bandwidth of single mode propagation. As a consequence, unwanted modes were cut off (at the input) over the frequency range

$$\omega_1 < \omega < 1.6839\omega_1, \tag{2}$$

where $\omega_1 = 19$ GHz and $1.6839\omega_1 = 28.8$ GHz. The input reflection and the far field were then measured from 17 GHz to 35 GHz. The input reflection agrees very well with a simple formula given in eq. (42) and derived in Ref. 4. Over the frequency range (2), the far field is essentially polarized in one direction. At frequencies higher than 28.8 GHz, however, a strong cross-polarized component is caused by an unwanted HE_{11} mode, which is excited primarily at the input of the feed. A simple expression for the power converted into this mode is given in Ref. 4.

II. PRELIMINARY CONSIDERATIONS

In the experiment described in Section V, the feed is excited at the input by a linearly polarized TE_{11} mode. Therefore, consideration will be restricted to the modes arising for this particular excitation.[2–8]

Consider a disk-loaded waveguide centered around the z axis as in Fig. 1, and let r, ϕ, z be cylindrical coordinates defined by $x = r\cos\phi$ and $y = r\sin\phi$. Assume for the moment that the waveguide parameters h, a, b, and t are independent of z. The separation h of the disks, which occupy the region $a < r < b$, is assumed to be much smaller than a wavelength λ

$$kh \ll 1. \tag{3}$$

The region between two consecutive disks forms a radial line whose input reactance jX at $r = a$ is a function of the radial length $l = b - a$; for $ka \gg 1$,

$$jX \approx jZ_0 \tan kl, \tag{4}$$

where $Z_0 = \sqrt{\mu_0/\epsilon_0}$. Because of condition (3), the effect of the disks can be adequately accounted for by introducing an effective surface reactance[2]

$$jX_s = jX\left(1 - \frac{t}{h}\right), \tag{5}$$

where t is the thickness of the disks, and by requiring that the field for $r < a$ satisfy the boundary conditions

$$\left.\begin{aligned} E_\phi &\approx 0 \\ H_\phi &\approx -\frac{E_z}{jX_s} \end{aligned}\right\} \text{ for } r = a, \tag{6}$$

where E_ϕ, H_ϕ, E_z are the ϕ and z components of the electric and magnetic field.

2.1 The HE_{lm} and HE'_{lm} modes

The properties of the hybrid modes[2–8] of a corrugated waveguide are determined by the radius a and the surface reactance X_s of the waveguide. In general, there is no simple relation[2] between the propagation constant β of a mode and the two parameters a and X_s. If, however, condition (1) is satisfied and furthermore

$$\frac{\nu}{ka} \ll 1, \tag{7}$$

where

$$y = -\frac{Z_0}{X_s}, \tag{8}$$

then we can show[2,4] that for all the modes except one, described in Appendix B of Ref. 4, the propagation constant is independent of X_s and is simply given by

$$\beta a \approx \sqrt{(ka)^2 - u^2}, \tag{9}$$

where u is a constant that is either the mth root of $J_0(u) = 0$, or of $J_2(u) = 0$. Thus, either

$$u = u_m, J_0(u_m) = 0 \tag{10}$$

or

$$u = u'_m, J_2(u'_m) = 0. \tag{11}$$

The modes to which eq. (10) applies will be called HE_{lm} modes and those corresponding to eq. (11), HE'_{lm} modes. The HE_{lm} modes approach asymptotically, for large ka, the field distribution[2,4]

$$\mathbf{E} = J_0\left(\frac{r}{a}u\right) e^{-j\beta z}\,\mathbf{i}_x, \tag{12}$$

whereas the HE'_{lm} modes approach the distribution

$$\mathbf{E} = J_2\left(\frac{r}{a}u\right)(\cos 2\phi \mathbf{i}_x + \sin 2\phi \mathbf{i}_y)e^{-j\beta z}. \tag{13}$$

The absence of a z component in eqs. (12) and (13) is due to the fact that this component vanishes like $1/ka$, as $ka \to \infty$.

The HE_{lm} modes, which are given by eq. (12), have the important property that the electric field is polarized in one direction. Of special interest is the fundamental mode ($m = 1$) characterized by

$$u = u_1 = 2.4048. \tag{14}$$

Note that both eqs. (12) and (13) are frequency independent.

Thus, over the frequency range in which both conditions (1) and (7) are satisfied, the field distribution of an aperture illuminated by the HE_{11} mode is essentially frequency independent.[1,2] If, however, only condition (1) and not condition (7) is satisfied, then we must add[4] to the right-hand side of eq. (12) a component of the type (13), so that

$$E \approx J_0\left(\frac{r}{a}u\right)\mathbf{i}_x - \frac{\gamma - 1}{\gamma + 1}J_2\left(\frac{r}{a}u\right)(\cos 2\phi \mathbf{i}_x + \sin 2\phi \mathbf{i}_y), \tag{15}$$

where the factor $e^{-j\beta z}$ has been omitted. Both γ and u are functions of ka and X_s, and, therefore, they vary with frequency. If one develops γ,

u in power series of y and $1/ka$ one finds[4]

$$u = u_m \left\{1 - \frac{1}{2}\mathrm{y}\left(\frac{1}{ka}\right)\cdots\right\} \tag{16}$$

$$\gamma = 1 - u_m^2 \left\{\frac{\mathrm{y}}{2}\frac{1}{ka} - \frac{\mathrm{y}^2}{8}u_m^2(4 + u_m^2)\left(\frac{1}{ka}\right)^2\cdots\right\}. \tag{17}$$

2.2 Mode conversion in a conical horn[4]

Consider a conical horn of constant surface reactance and flare angle α, as shown in Fig. 1, from $z = z_1$ to $z = z_2$. If α is sufficiently small, and if the input of the horn is excited in the HE_{11} mode, then the resulting field inside the horn is very nearly a spherical wave[1,3] originating from the apex of the horn and given by

$$Ee^{-ikR}, \tag{18}$$

where E is given by eq. (15) and R is the distance from the apex of the horn. Since α is small,

$$R \approx z + \frac{r^2}{2z}, \tag{19}$$

r being the radial distance from the axis and z the axial distance measured from the horn apex. From eq. 17, the parameter γ which appears in eq. (15) is a function of both y and ka; assume $|\mathrm{y}| \neq \infty, 0$. Then, since ka increases with z, we have from eq. (17) that γ varies with z, and therefore the field distribution (15) does not remain constant with z. This variation is accompanied by generation of unwanted modes, an effect that will be negligible only if α is sufficiently small.

To determine how small α should be, consider the special case $ka(z_1) \gg 1$ treated in Ref. 4. Let P_c be the total power converted from the desired mode into the HE_{11}' mode and let P_0 be the power incident at the input. Then, we find that for $z = z_2$

$$P_c = P_0 \times 3.393\,(10^{-3})\mathrm{y}^2 \tan^2\alpha |1 - e^{j\psi}|^2, \tag{20}$$

where

$$\psi = \frac{10.295}{\mathrm{y}\tan\alpha}\frac{\mathrm{y}}{ka(z_1)}\left(1 - \frac{a(z_1)}{a(z_2)}\right). \tag{21}$$

Therefore,

$$P_c \leqq P_0 \times 1.357(10^{-2})\mathrm{y}^2 \tan^2\alpha, \tag{22}$$

where the equality sign is attained for $\psi = (2n + 1)\pi$. In the experiment, $\alpha = 4°$, in which case for $\mathrm{y} = 1$ we find $P_c \leqq 6.636 \times 10^{-5}$ (−41.8 dB), which is negligible for most practical purposes.

III. RADIATION CHARACTERISTICS[2,6,7]

Let the aperture of the horn be of sufficiently large diameter that the far field is simply proportional to the Fourier transform F of the aperture field. Also let α be so small that we can neglect the phase variation caused over the aperture plane by the variation of R in (18). Then, by taking the Fourier transform of eq. (15), we obtain for the field distribution at a great distance D from the aperture, for example, using the contour integral method of Ref. (4),

$$F = N_0(u,v)\mathbf{i}_x + \frac{\gamma - 1}{\gamma + 1} N_2(u,v)\,(\cos 2\phi \mathbf{i}_x + \sin 2\phi \mathbf{i}_y), \tag{23}$$

where

$$v = ka(z_2)\sin\theta, \tag{24}$$

$$N_k(u,v) = \frac{a^2}{v^2 - u^2}\,[vJ_k(u)J_1(v) - uJ_k(v)J_1(u)], \tag{25}$$

and ϕ, θ are spherical coordinates defined by $x = D\sin\theta\cos\phi, y = D\sin\theta\sin\phi$. Equation (23) gives, except for a factor independent of ϕ, θ, the far-field pattern. For small y/ka we have

$$u = u_1 = 2.4048.$$

In this case, from eqs. (23) to (25) the cross-polarization ratio C between the maximum value of $|E_y|$, which occurs for $v = 3.67$, and the maximum value of $|E_x|$, which occurs for $v = 0$, is given by

$$C^2 \equiv \frac{|E_y|^2_{\max}}{|E_x|^2_{\max}} = (0.26)^2 \left(\frac{\gamma - 1}{\gamma + 1}\right)^2. \tag{26}$$

From eq. 17, the asymptotic value of C^2 for large ka is

$$C^2 \to 0.14 \left(\frac{y}{ka}\right)^2. \tag{27}$$

If $ka \gg 1$, but y/ka is not small, C^2 has the behaviour of Fig. 2.

Equation (23) assumes that the aperture of the horn is illuminated by the HE_{11} mode. If, in addition to this mode, there is also some HE'_{11} mode of power P_c, then we have in addition to the component in (23), a component[4]

$$-e^{j\psi} \times \sqrt{\frac{P_c J_1^2(u)}{P_0 J_1^2(\bar{u})}}\, N_2(\bar{u},v)(\cos\phi\mathbf{i}_x + \sin\phi\mathbf{i}_y), \tag{28}$$

where P_c, P_0 are the powers carried by the two modes and ψ is their difference in phase for $z = z_2$. For $ka \gg 1$,

$$\bar{u} \simeq u'_1 = 5.1356. \tag{29}$$

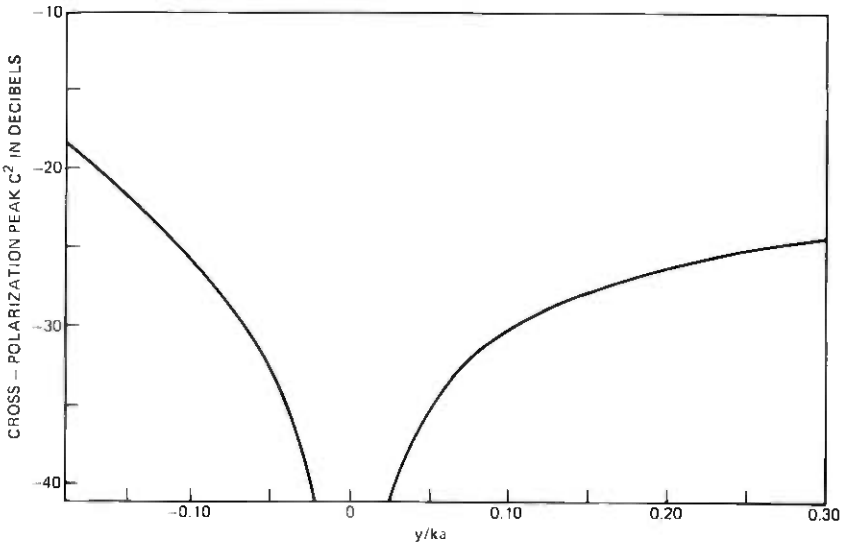

Fig. 2—Cross-polarization peak versus normalized surface susceptance.

Therefore, from eq. (25) with $J_2(\bar{u}) = 0$,

$$N_2^2(\bar{u},\upsilon) = \left(\frac{a^2\bar{u}}{\upsilon^2 - \bar{u}^2} J_2(\upsilon)J_1(\bar{u})\right)^2 \tag{30}$$

whose maximum value occurs for $\upsilon \approx 4.356$ and is

$$1.271 \times 10^{-3} \times a^4 J_1^2(\bar{u})\bar{u}^2. \tag{31}$$

Because of the component (28), we have for $\gamma = 1$ that the ratio C^2 between the maximum value of $|E_y|^2$, which now occurs for $\upsilon = 4.4$ and the maximum value of $|E_x|$ is

$$C^2 = 0.194 \frac{P_c}{P_0}. \tag{32}$$

Note from eq. (23) that the normalized radiation pattern of the HE_{11} mode for $ka \gg 1$ is simply given by

$$P(\theta) = \left[u^2 \frac{J_0(\upsilon)}{u^2 - \upsilon^2}\right]^2 \tag{33}$$

with $u = u_1 = 2.4048$ and $\upsilon = ka \sin\theta$. If θ_1 and θ_2 denote the values of θ for which $P(\theta) = 0.5$ and $P(\theta) = 0.1$, then from eq. (33)

$$ka \sin\theta_1 = 2.078 \text{ (3-dB point)} \tag{34}$$

and

$$ka \sin \theta_2 = 3.597 \text{ (10-dB point)}, \tag{35}$$

respectively.

IV. FEED DESIGN[2,4,9]

The feed of Fig. 1 is now described. Its measured characteristics are discussed in Section V. Let $\{\omega_2, \omega_1\}$ be the frequency range over which only the dominant mode, the HE_{11} mode, propagates at the input ($z \simeq 0$). Let ω_0 be the frequency at which the surface reactance is infinite at the aperture; then

$$b(z_2) - a(z_2) \simeq \frac{\lambda}{4}, \text{ at } \omega_0. \tag{36}$$

At the input, the ratio between the radii b and a is optimized for maximum ω_2/ω_1. This is shown in Appendix A to require

$$\frac{b(0)}{a(0)} = \frac{7.0155}{3.8317} = 1.8309. \tag{37}$$

Then*

$$\frac{\omega_2}{\omega_1} = \left(\frac{\omega_2}{\omega_1}\right)_{\text{MAX}} = 1.6839, \tag{38}$$

and the surface reactance can be shown to vanish for $\omega = \omega_2$,

$$y = \infty \text{ for } z = 0, \text{ at } \omega = \omega_2. \tag{39}$$

In the experiment, the frequency ω_1 was chosen equal to 19 GHz, which gives $\omega_2 = 28.8$ GHz, $a(0) = 0.556$ cm, and $b(0) = 1.161$ cm, as we obtain from eqs. (37) and (38) and the condition

$$kb = 7.0155 \text{ at } \omega = \omega_2, \tag{40}$$

shown by Fig. 9 of Appendix A.

From Fig. 1 the feed consists of two parts joined at $z = z_1$. From $z = 0$ to $z = z_1$, the outer radius is kept constant, so the cutoff frequency of the HE_{11}' mode remains constant, as shown in Appendix A, where the relations between a, b and the cutoff frequencies of the various modes of eqs. (12) and (13) are derived. Note that from $z = 0$ to $z = z_1$ the radius a increases, which implies that at any given frequency the surface reactance decreases with z. The importance of this requirement was first realized by Bryant[8]. From $z = z_1$ to the aperture, the surface reactance remains constant with z. The frequency ω_0 at which it is infinite was chosen in the experiment

* To put this ratio in perspective, the highest frequency of the 6-GHz (TH) common carrier band is 1.732 times the lowest frequency of the 4-GHz (TD-2) band; similarly, for the 18- and 30-GHz bands, the ratio is 1.695.

$$\omega_0 = 1.21\,\omega_1. \tag{41}$$

Over the frequency range $[\omega_1, \omega_2]$ only the HE_{11} mode is excited at the input, other modes being cut off. Therefore, since the variations of $a(z)$ and $b(z)$ from the input to the aperture in Fig. 1 are sufficiently gradual to ensure negligible generation of unwanted modes, we have that the aperture is essentially illuminated by a single mode, the HE_{11} mode, over the above frequency range. The radiation characteristics for $\omega_1 < \omega < \omega_2$ are, therefore, given accurately by eq. (23) with $a = a(z_2)$ and u given by eq. (16) for $m = 1$.

From eq. (39), the input surface reactance vanishes at ω_2. Therefore, since the corrugated waveguide is connected at the input to a smooth waveguide of the same diameter, the input reflection is essentially zero at $\omega = \omega_2$. For $\omega \neq \omega_2$, however, there is a reflection which, as we shall see in Section V, is given accurately by

$$|\rho|^2 = \left(\frac{\beta_1 - \beta_1'}{\beta_1 + \beta_1'}\right)^2, \tag{42}$$

where β_1 is the propagation constant of the TE_{11} mode and β_1' is that of the HE_{11} mode. Equation (42) is derived in Ref. 4.

At frequencies higher than ω_2, some of the incident power is converted into the HE_{11}' mode. If P_c denotes the converted power and P_0 denotes the incident power, then

$$\frac{P_c}{P_0} \approx \left[\frac{1}{a(\beta_1 - \beta_2')y}\right]^2 \frac{\beta_1}{\beta_2'(u^2 - 1)}, \tag{43}$$

where $u = 1.841$ and β_2' is the propagation constant of the HE_{11}' mode. Some of the incident power is also converted into the TM_{11} mode of the smooth waveguide. If P_c' denotes this power, which is reflected by the junction, we have[4]

$$\frac{P_c'}{P_0} \approx \frac{(\beta_2 - \beta_2')^2}{(\beta_1 + \beta_2')^2} \frac{\beta_2 \beta_1}{u^2 - 1} \frac{1}{k^2}, \tag{44}$$

where β_2 is the propagation constant of the TM_{11} mode. The above coupling equations are derived in Ref. 4 assuming $|y| \gg 1$.

In the experiment described in the following section, the corrugated waveguide is connected at the input to a smooth waveguide whose radius gradually increases from a relatively small value a' to the final value $a(0)$. The initial value a' is sufficiently small so that the TM_{11} mode is cut off and, as a consequence, the power P_c' is reflected back towards the junction where it is converted into the HE_{11}' mode of the corrugated waveguide. The total power converted into the HE_{11}'mode is thus in general different from P_c. It varies approximately between the two values

$$(\sqrt{P_c} - \sqrt{P_c'})^2 \text{ and } (\sqrt{P_c} + \sqrt{P_c'})^2, \tag{45}$$

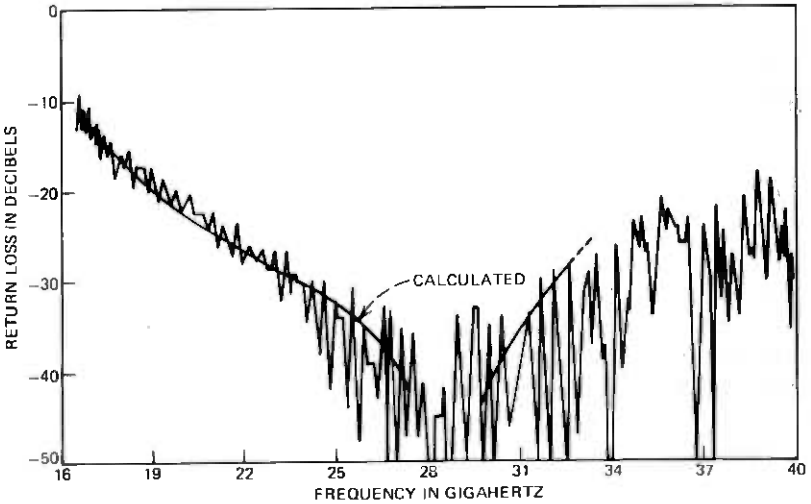

Fig. 3—Return loss of corrugated horn of Fig. 1 vs frequency.

depending on the phase angle of the TM_{11} mode incident on the junction.

V. EXPERIMENT

A corrugated feed of the dimensions shown in Fig. 2 was built using a technique described in Appendix C. Its measured reflection coefficient agrees* very well with eq. (42), as shown in Fig. 3. Its radiation characteristics are shown in Figs. 4 through 6. From 17.5 GHz to 32 to GHz, the radiation patterns are in good agreement with eq. (33), as shown in Fig. 4, which compares eq. (33) with two measured patterns of $|E_x|^2$ in the plane $\phi = 45°$. Furthermore, from 17.5 GHz to 32 GHz, there is little difference between the patterns of $|E_x|^2$ in the two principal planes ($\phi = 0$ and $\phi = 90°$) and the pattern for $\phi = 45°$. The difference is altogether negligible at $\omega_0 = 23$ GHz, which is the frequency for which $y = 0$ at the aperture. Figures 5 and 6 show a few examples of patterns measured in the two principal planes. The variations with frequency of the beamwidths $2\theta_1$ and $2\theta_2$ (respectively, the 3-dB and 10-dB beamwidths) agree very well with eqs. (34) and (35), as shown in Fig. 7, where the measured beamwidths in the plane $\phi = 45°$ are compared with the calculated values.

Finally, the cross-polarized component E_y is very small over the frequency range of eq. (2), as shown in Fig. 8. For $\omega > 28.8$ GHz, however,

* Note that the measured reflection coefficient includes a small reflection due to a transition from rectangular to circular waveguide, which was connected at the input of the feed during the measurement.

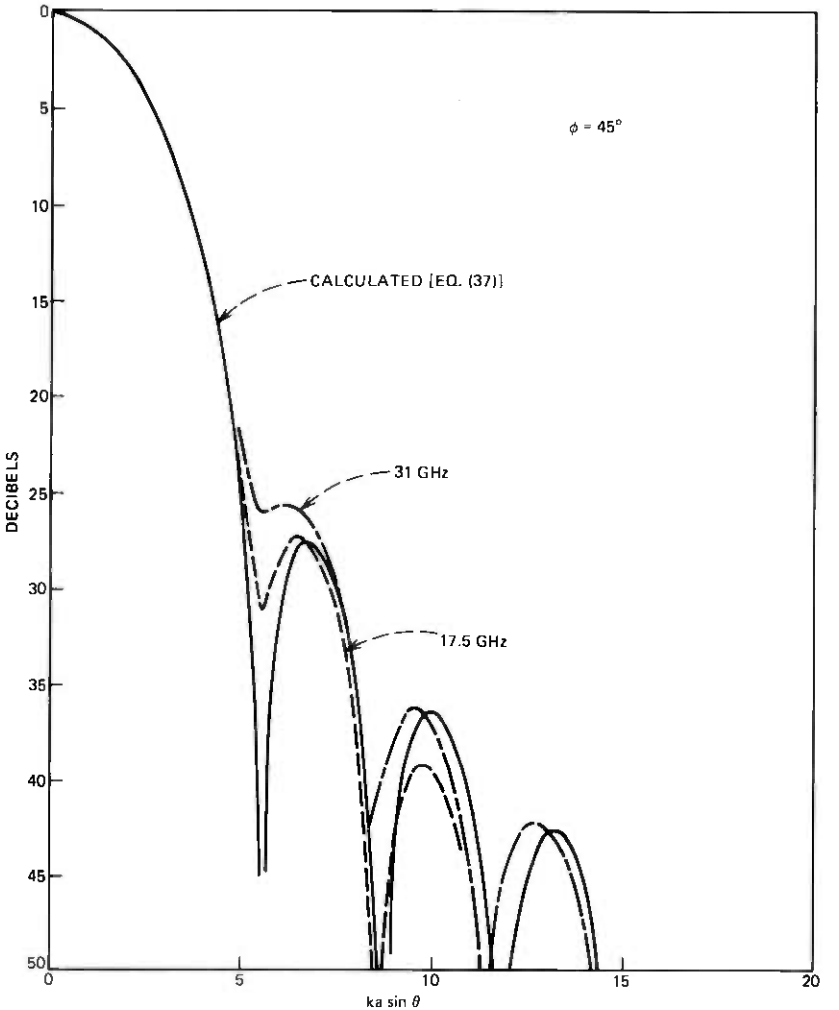

Fig. 4—Calculated and measured radiation patterns vs normalized angle at $\phi = 45°$.

the normalized peak C^2 of E_y increases rapidly with frequency as expected because of the HE'_{11} mode, which is excited at the input for $\omega >$ 28.8 GHz. The solid curves and the dashed curves in Fig. 8 correspond to eqs. (27) and (43), respectively. The agreement with the measurements is satisfactory, taking into account that eq. (43) is not expected to give exactly the total power converted into the HE'_{11} mode, for several reasons. In the first place, eq. (43) assumes a very large number of disks per wavelength,

$$h \ll \lambda,$$

whereas in the experiment, $h \simeq 0.137\lambda$ at 30 GHz; the effect of a finite number of teeth is briefly discussed in Appendix B. In the second place, eq. (43) assumes that only the TE_{11} mode is incident at the input, whereas in practice also the TM_{11} mode is incident, for the reason pointed out in Section IV. The total power carried by the TM_{11} mode is approximately given by eq. (44).

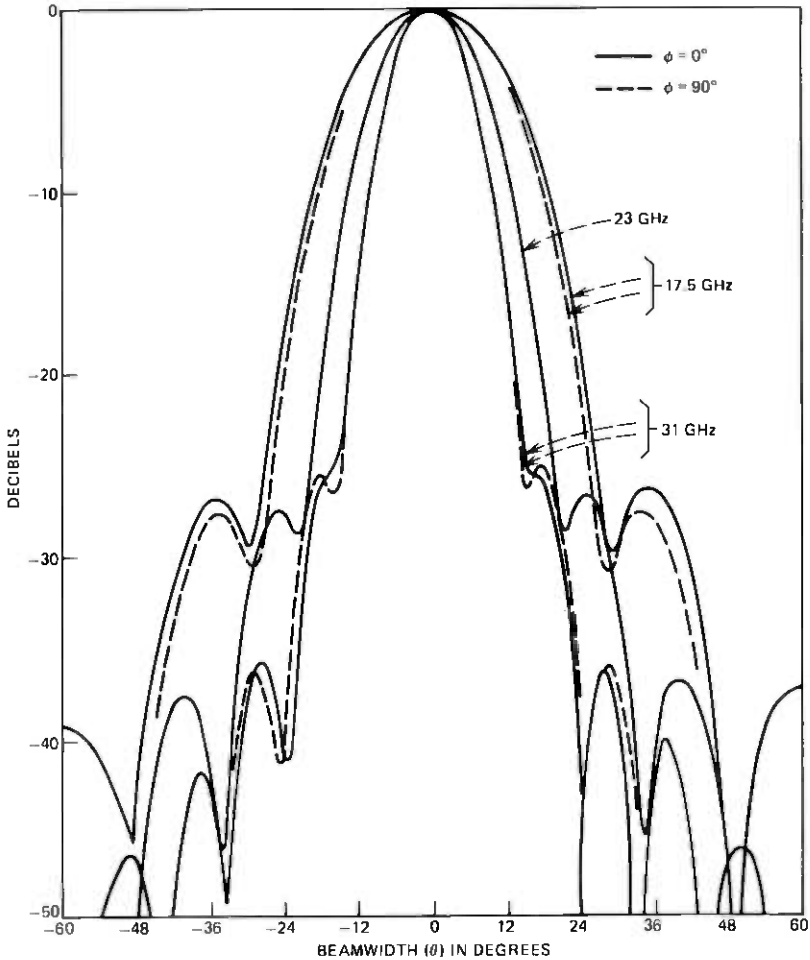

Fig. 5—Measured radiation patterns at $\phi = 0°$ and $\phi = 90°$ (principal planes).

Note, finally, that even if the total power converted into the HE'_{11} mode is calculated accurately, to determine the resulting cross-polarization peak C^2, the difference in phase between the HE_{11} and HE'_{11} modes must be determined.

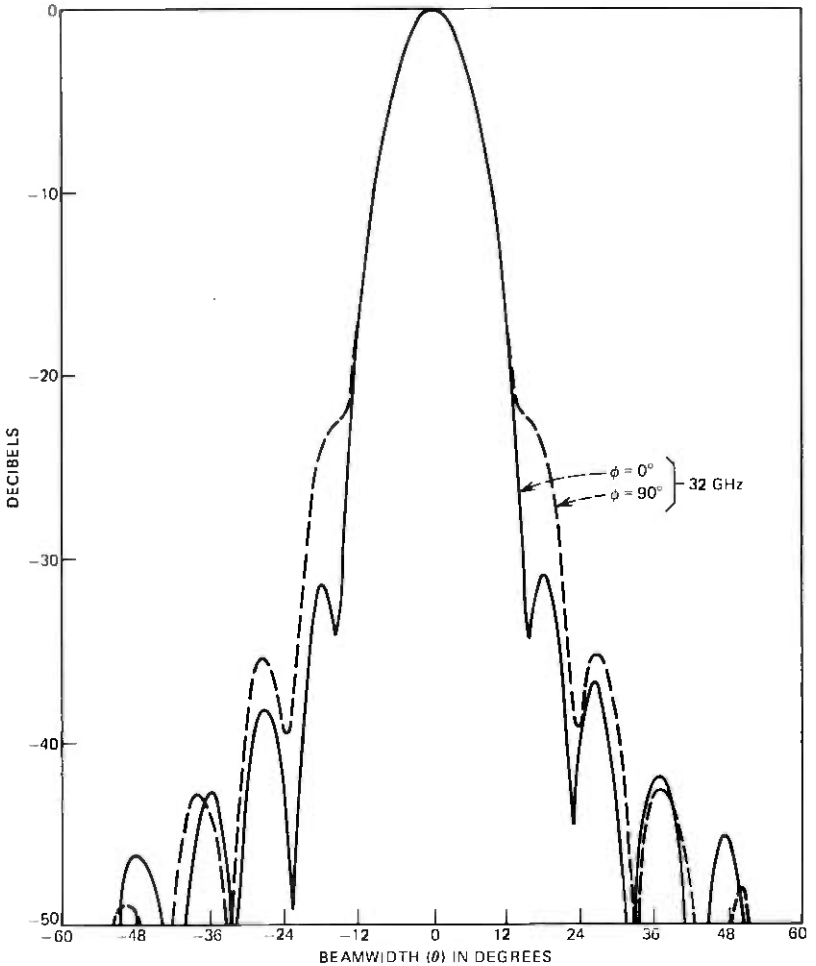

Fig. 6—Measured radiation pattern at $\phi = 0°$ and $\phi = 90°$ for 32 GHz (effect of HE_{11}' mode on pattern symmetry is evident at upper frequency limit of operation).

VI. CONCLUSIONS

Using a novel fabrication technique, described in Appendix C, which can be applied at very high frequencies, a corrugated feed of small flare angle was fabricated. Its input reflection, found to be given accurately by the simple formula

$$|\rho|^2 = \left(\frac{\beta_1 - \beta_1'}{\beta_1 + \beta_1'}\right)^2,$$

remained less than −30 dB from about 24 to 32 GHz. Over the frequency range $\omega_1 < \omega < \omega_2$, the far field was essentially polarized in one direction;

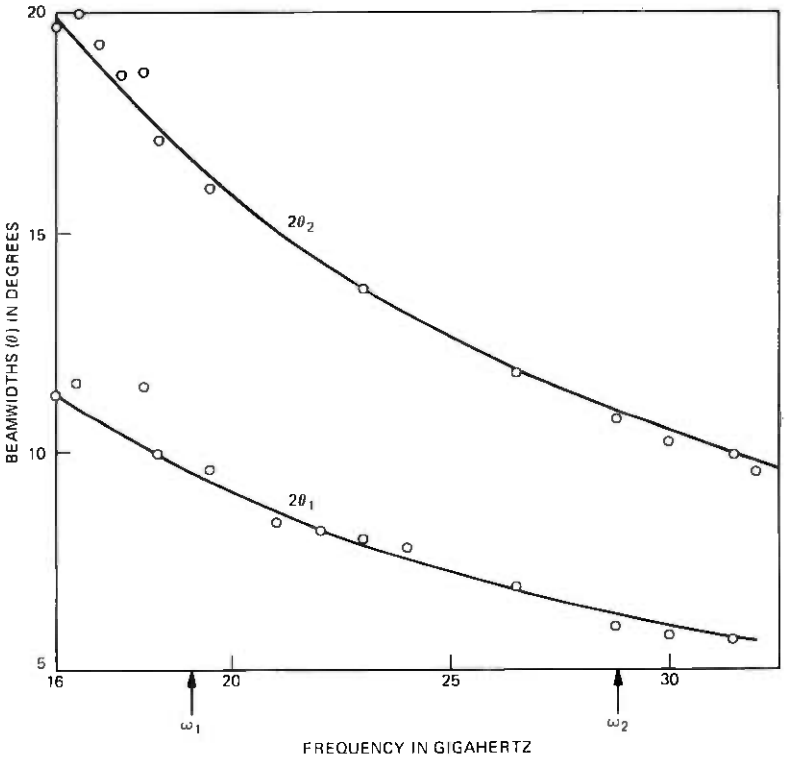

Fig. 7—Calculated and measured 3-dB and 10-dB beamwidths.

its pattern is simply given by

$$P(\theta) = \left(u^2 \frac{J_0(v)}{u^2 - v^2}\right)^2.$$

At frequencies higher than 28.8 GHz, the far field contained a cross-polarized component, as predicted by the theory of Ref. 4.

It was shown that a maximum bandwidth, expressed as the ratio ω_2/ω_1, of about 1.68 can be achieved for the type of corrugations considered here. By using the corrugations of Ref. 10, which, however, are difficult to realize at high frequency, greater values of ω_2/ω_1 may be achieved.

Both the input reflection and the cross-polarization ratio can be improved by increasing the thickness t of the disks at the input. Curves of $|\rho|^2$ and C^2 as a function of frequency for different values of t/h are given in Fig. 2 of Ref. 4. In the experiment, t/h was kept constant for reasons of simplicity, since our main concern was to verify the results of Sections II through IV and of Refs. 1 and 4, and to demonstrate the feasibility of the fabrication technique described in Appendix D.

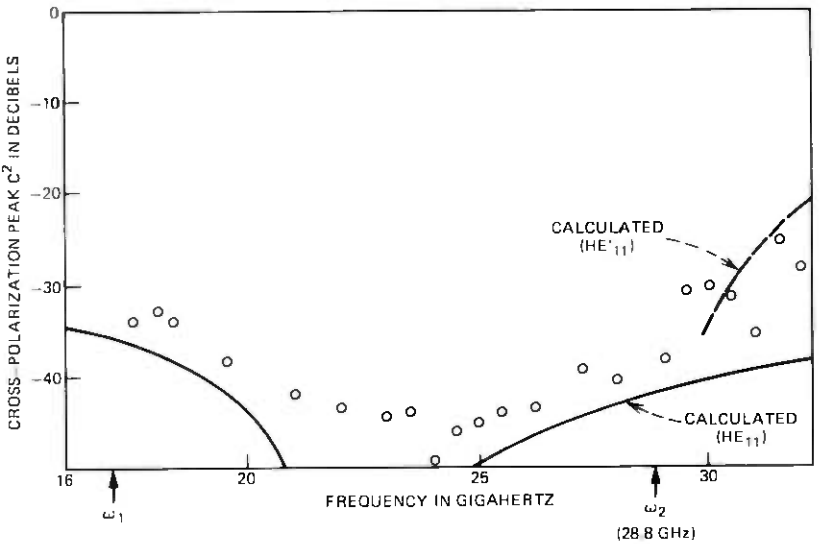

Fig. 8—Calculated and measured cross polarization; these are maximum values of cross polarization (see Section III, eq. 26).

ACKNOWLEDGMENT

The author wishes to acknowledge the assistance of W. E. Legg who carried out the measurements.

APPENDIX A

Cutoff Frequencies of the HE_{lm} and HE'_{lm} Modes

To determine eqs. (37) to (40) it is necessary to establish the relation between a, b and the cutoff frequencies of the various modes of eqs. (12) and (13). Assume the thickness of the disks is very small. Consider first the cutoff frequencies ω_{cm} of the HE_{lm} modes, which are characterized for large ka by the field distribution of eq. (12). In the vicinity of the cutoff frequency ω_{cm}, we have, for any one of the above modes,

$$\cos \theta_m = 0, \tag{46}$$

where $k \cos \theta_m$ is the propagation constant in the z direction,

$$\cos \theta_m = \sqrt{1 - \frac{u_m^2}{(ka)^2}}. \tag{47}$$

For $\omega \simeq \omega_{cm}$, we can show[2,4] that the mode degenerates into either a TE mode or a TM mode according to the following rule:

$$\left.\begin{array}{l}\text{TE mode, if } y > 0,\\ \text{TM mode, if } y < 0,\end{array}\right\}. \tag{48}$$

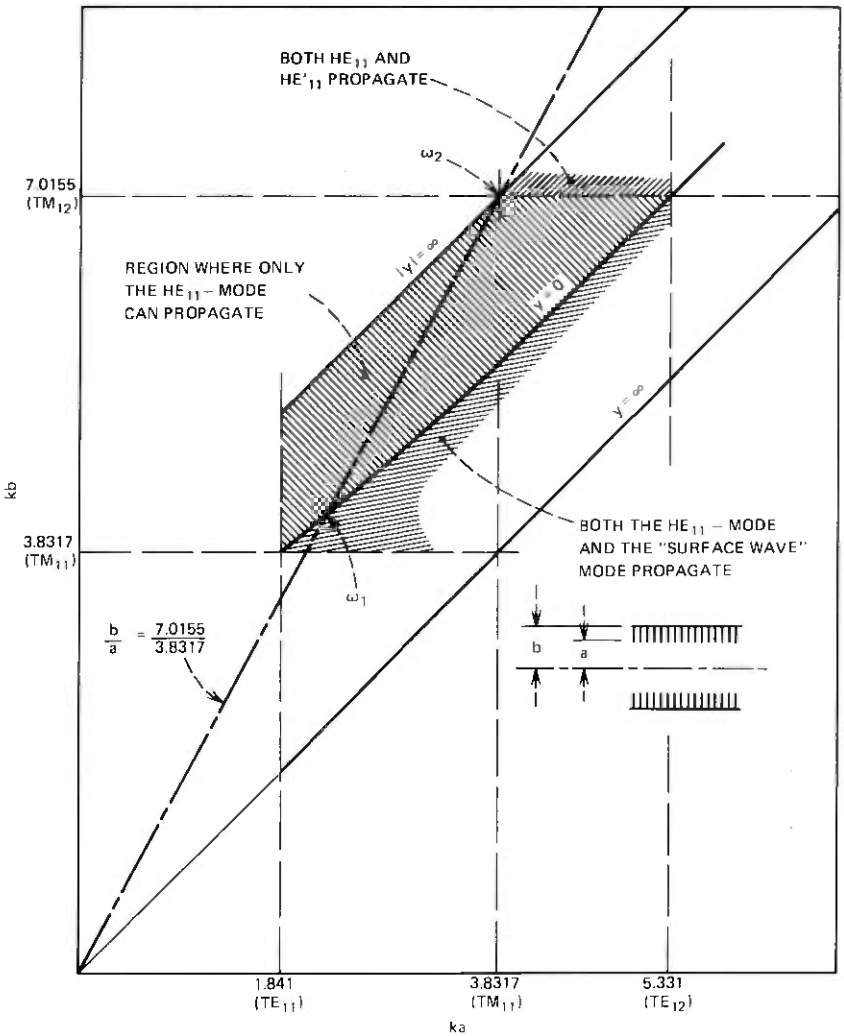

Fig. 9—Region where only the HE_{11} mode propagates. Note: As ω varies for a given a, b, a point ka, kb moves along a straight line through the origin.

In the former case, since the only nonzero component of the magnetic field of a TE mode near cutoff is H_z, we have $H_\phi = 0$, and therefore the second of the two boundary conditions (6) can be ignored, which implies that ω_{cm} is independent of y. The corrugated waveguide can be replaced, therefore, with a smooth waveguide of the same diameter whose cutoff frequencies for the TE modes are given by the roots of $J_1'(ka)$,

$$J_1'(ka) = 0 \text{ for } \omega = \omega_{cm} \ (y > 0). \tag{49}$$

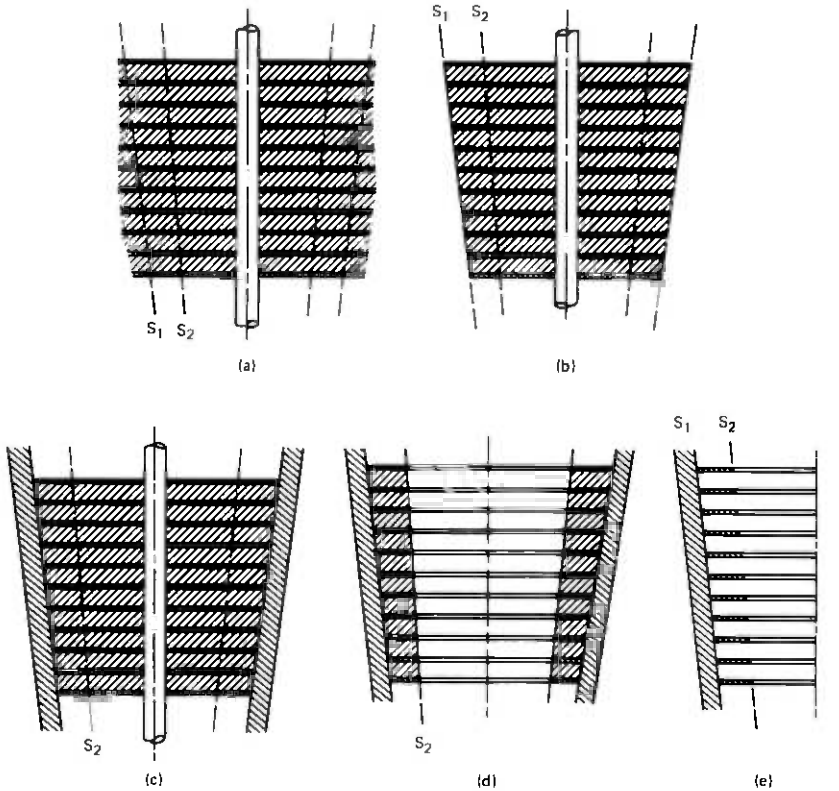

Fig. 10—Aluminum and brass disks assembled on mandrel. (a) Before machining. (b) After machining exterior surface. (c) With electroformed wall of copper. (d) After machining interior surface. (e) After etching away aluminum disks.

In the latter case, $y < 0$, the only component of the electric field for $\omega \simeq \omega_{cm}$ is E_z and, therefore, the very thin disks can be removed without affecting the field. The cutoff frequencies therefore coincide with those of the TM modes in a smooth waveguide of radius b, and are given by the roots of $J_1(kb)$,

$$J_1(kb) = 0 \text{ for } \omega = \omega_{cm}, \ (y < 0). \tag{50}$$

Analogous considerations[4] are valid for the HE'_{lm} modes, except that now, instead of the rules (48),

$$\left.\begin{array}{l} \text{TM mode, if } y > 0, \\ \text{TE mode, if } y < 0, \end{array}\right\}, \tag{51}$$

and therefore the cutoff frequencies ω'_{cm} are given by the conditions

$$\begin{aligned} J'_1(ka) &= 0 \text{ for } \omega = \omega'_{cm}, \text{ if } y < 0 \\ J_1(kb) &= 0 \text{ for } \omega = \omega'_{cm}, \text{ if } y > 0. \end{aligned} \tag{52}$$

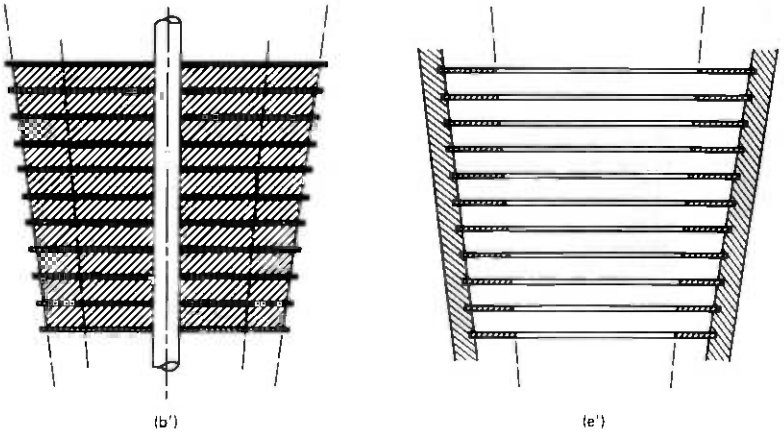

Fig. 11—Modification of assembly to obtain strong attachment to teeth of electroformed wall.

These results are illustrated by the diagram of Fig. 9 showing in the ka,kb plane the region* where only the HE_{11}-mode can propagate. We can see that ω_2/ω_1 is maximum when $b/a = 7.0155/3.8317 = 1.8309$, in which case $\omega_2/\omega_1 = 1.6839$, as pointed out in Section I before eq. 2.

APPENDIX B

Effect of a Finite Number of Teeth Per Wavelength

In Section IV, we assumed an infinite number of teeth per wavelength—i.e.,

$$h \ll \lambda,$$

a condition which seldom holds in practice. For instance, in the experiment

$$h > 0.13716\lambda$$

for $\omega > 30$ GHz. The effect of a finite number of teeth is not difficult to evaluate approximately if

$$ka \gg 1,$$

In this case, in fact, the field in the vicinity of the corrugations can be considered to be locally the field of a plane wave reflected by a corrugated plane tangent to the actual corrugated surface. The angle of incidence

* In Fig. 9 we have not indicated a region for $kb < 3.8317$ where mode propagation can occur, since that region is of no interest to us here. (The HE_{11} mode is cut off for $kb < 3.8317$).

is $90° - \theta_1$, where

$$\cos \theta_1 = \frac{\beta_1}{k}.$$

Thus, from Ref. 9, we find that a corrugated waveguide of radius $a \gg \lambda$ with a finite number of teeth is equivalent to one with $h \ll \lambda$, but with slightly larger inner radius a',

$$a' \simeq a + \frac{\ln 2}{\pi} h,$$

and the same outer radius b. Although this result is strictly valid only if $a \gg \lambda$, measurements[11] have shown that it is valid approximately even for $ka \simeq 4$.

APPENDIX C

Fabrication Technique

Even if the aperture of a feed is illuminated by a single mode, the far field contains a cross-polarized component with amplitude proportional to [eqs. (4), (5), and (27)]

$$\left|\frac{y}{ka}\right| = \left|\frac{1}{(1 - t/h)\, ka \tan kl}\right|. \tag{53}$$

To minimize this cross-polarized component, it is important that the thickness t of the disks be much smaller than their separation h,

$$t \ll h. \tag{54}$$

Since h is always appreciably smaller than $\lambda/4$, and since typically the depth l of the grooves is not much different from $\lambda/4$, condition (54) implies

$$t \ll \lambda/4 \tag{55}$$

$$t \ll l. \tag{56}$$

Corrugated feeds are difficult to fabricate. When a corrugated feed is electroformed, a mandrel of aluminum or other material is first prepared, and then the corrugated feed is electroformed around the mandril, which is then removed with a solvent. However, at high frequencies, say at frequencies higher than about 10 GHz, condition (55) demands that t be very small (much less than 0.318 cm). Then, taking into account (56), the above technique cannot be used.

Figures 10 and 11 illustrate a technique that can be used at very high frequencies, as high as 100 GHz, and which allows very small thicknesses t to be realized. First a set of disks of aluminum and brass is assembled, as shown in Fig. 10a, to form a single block whose outside surface is then

machined as shown in Fig. 10b. The central rod shown in Figs. 10a and b defines the axis of the corrugated waveguide. The surface S_1 corresponds to the bottom of the grooves, as shown in Fig. 10e. The thickness of the brass disks is t and, that of the aluminum disks, $h - t$. After the surface S_1 is machined, a wall of copper (or other metal) is electroformed, as shown in Fig. 10c. The central rod is then removed, and the inside surface S_2, which corresponds to the tops of the grooves (see Fig. 10e), is machined as shown in Fig. 10d. Finally, the aluminum is removed with a solvent and the final product is the corrugated waveguide of Fig. 10e. An important feature of this technique is that the two surfaces S_1 and S_2 can be machined very accurately.

If the brass disks are too thin, their mechanical adherence to the copper wall in Fig. 10e may not be satisfactory. In this case, we may modify the above procedure by adding, after step b, an extra step b′ in which the brass disks of Fig. 10b are replaced with disks of somewhat larger diameter, as shown in Fig. 11b′. The exact dimensions of the new disks are not important. The final result, after the aluminum is removed, is shown in Fig. 11e′. The disks are now embedded in the copper wall.

REFERENCES

1. B. MacA. Thomas, "Bandwidth Properties of Corrugated Conical Horns," Electron. Lett., *5*, No. 22 (October 30, 1969), pp. 561–563.
2. P. J. B. Clarricoats and P. K. Saha, "Propagation and Radiation Behaviour of Corrugated Feeds; Part I—Corrugated Waveguide Feed," Proc. IEE, *118*, No. 9 (September 1971), pp. 1167–1176.
3. P. T. B. Clarricoats and P. K. Saha, "Propagation and Radiation Behaviour of Corrugated Feeds; Part 2—Corrugated Conical-Horn Feed," Proc. IEE, *118*, No. 9 (September 1971), pp. 1177–1186.
4. C. Dragone, "Reflection, Transmission, and Mode Conversion in a Corrugated Feed," this issue, pp. 835–868.
5. V. H. Rumsey, "Horn Antennas With Uniform Power Patterns Around Their Axes," IEEE Trans. Ant. Propag., *AP-14*, No. 5 pp. 656–658.
6. H. C. Minnett and B. MacA. Thomas, "A Method of Synthesizing Radiation Patterns with Axial Symmetry," IEEE Trans. Ant. Propag., AP-14, No. 5 (September 1966), pp. 654–656.
7. M. S. Narasimlian and B. V. Rao, "Hybrid Modes in Corrugated Conical Horns," Electron. Lett., *6*, No. 2 (January 22, 1970), pp. 32–34.
8. G. H. Bryant, "Propagation in Corrugated Waveguides," Proc. IEE, *116*, No. 2 (February 1969), pp. 203–213.
9. R. E. Collin, *Field Theory of Guided Waves*, New York: McGraw-Hill, 1960.
10. Y. Takeichi, T. Hashimoto, and F. Takeda, "The Ring-Loaded Corrugated Waveguide," IEEE Trans. Microw. Theory Tech., *MTT-19*, No. 12 (December 1971), pp. 947–949.
11. M. J. Gans, private communication.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 6, July–August 1977
Printed in U.S.A.

Acoustic Properties of Longitudinal Displacement in Vocal Cord Vibration

By K. ISHIZAKA and J. L. FLANAGAN

(Manuscript received January 12, 1977)

We examine the acoustic significance of longitudinal displacement in the self-oscillatory behavior of the vocal cords, and inquire into the need for representing this detail in speech synthesis. We use computer techniques and a previously derived model of the vocal cords to study the contribution of longitudinal displacement to the total acoustic volume velocity generated at the vocal cords. This volume velocity is the effective sound source for production of voiced speech. From computational results, and from speech sounds synthesized by the programmed model, we find that the contribution of longitudinal displacement is not significant perceptually, and is not essential for modeling the dominant acoustic properties of voiced speech.

I. VOCAL-CORD MODEL

In earlier work[1,2] we derived an analytical model for the self-oscillatory motion of the human vocal cords. We consider the displacing tissue of each cord to be approximated by two stiffness-coupled masses (see Fig. 1). For normal (nonpathological) conditions of phonation, the oscillator is bilaterally symmetric, and the mechanical constants of the opposing cords are identical. The left-hand mass pair (denoted m_1, m_1') constitutes the bulk of the firm cord tissue, while the smaller right-hand mass pair (m_2, m_2') represents the more flaccid mucous membrane covering of the firmer tissue. Each mass has associated with it a restoring stiffness and a resistive loss. All the stiffnesses and resistances are substantially nonlinear,[1] and in the original work, these elements act to oppose lateral motion (x-direction) only. The restriction to lateral motion still permits, of course, phase differences in the motion of the coupled masses. Lateral displacement of each mass pair determines the cross-sectional area of opening at each position. If the length of the cords, or glottal opening, is taken as ℓ_g, then the cross-sectional glottal areas are taken as rectangular shapes whose areas are $A_{gi} = 2\ell_g x_i$, $i = 1, 2$, where the factor

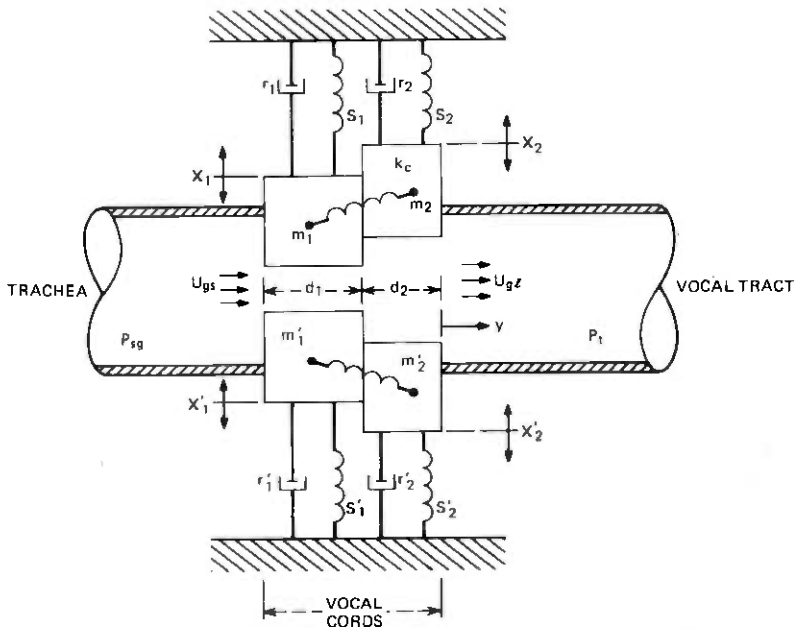

Fig. 1—Two-mass model of the vocal cords. Translational displacement is permitted in lateral (x) and longitudinal (y) directions.

2 arises from the bilaterally symmetric cord configuration. These cross-sectional areas determine the acoustic properties of the glottal volume current U_{gs}, which enters the cord orifice (from the subglottal system), and that which leaves it $U_{g\ell}$ (to pass into the larynx tube). The latter volume velocity is the effective sound source for all voiced speech sounds. The air pressure just to the left of (beneath) the vocal cords is the subglottal pressure P_{sg}, and the pressure just to the right of (above) the cords, at the entrance to the vocal tract, is P_t. The differential pressure $(P_{sg} - P_t)$ is the potential that creates the glottal volume currents.

The resulting volume currents depend upon serial acoustic impedances dictated by A_{g1} and A_{g2} and, hence, upon the cord motion, which, in turn, is conditioned by the intraglottal pressure distribution in the orifice and by the transglottal pressure $(P_{sg} - P_t)$. These serial acoustic impedances also are nonlinear (and flow dependent), and represent the mass (inertance) of air contained within the glottal orifice and the associated resistive flow losses.[1]

Additionally, there is another potential influence upon the glottal flow, namely, the volume of air displaced by the vibrating mass pairs. In general, this volume displacement can have components associated with lateral and longitudinal motion. In the original work, components of

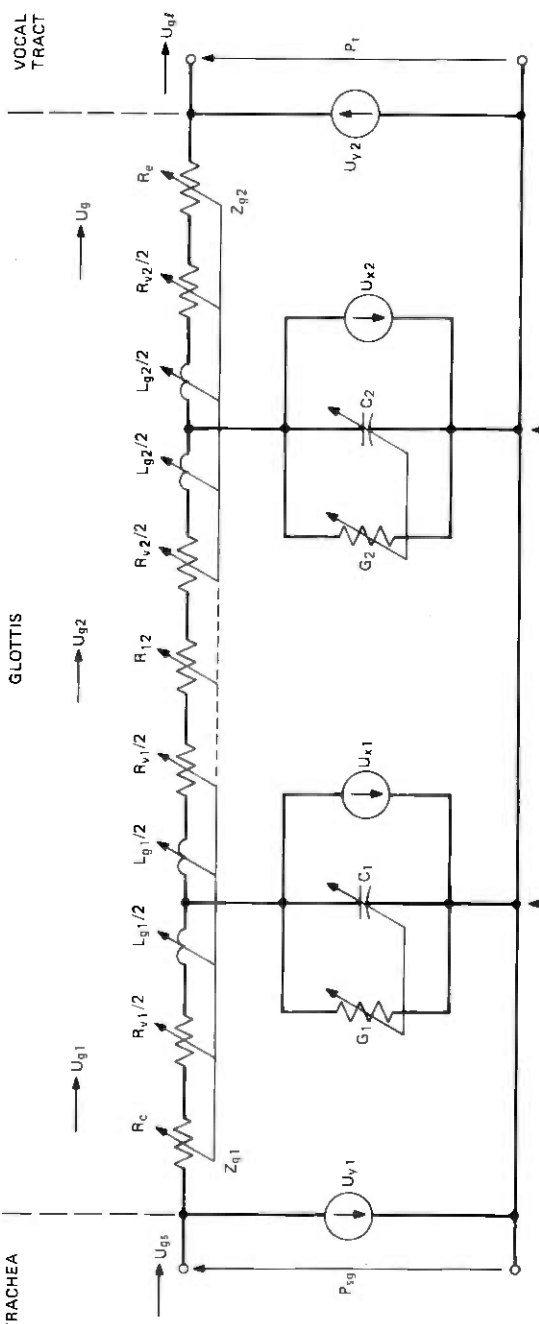

Fig. 2—Equivalent acoustic circuit including lateral and longitudinal displacement-volume velocities.

Table I — Values of impedance components of Fig. 2

Serial Impedances	$R_c = 1.37\frac{\rho}{2}\frac{\lvert U_{g1}\rvert}{A_{g1}^2}$,	$R_{12} = \frac{\rho}{2}\left(\frac{1}{A_{g2}^2} - \frac{1}{A_{g1}^2}\right)\lvert U_{g2}\rvert$
		$R_e = -\frac{\rho}{2}\frac{2}{A_{g2}A_1}\left(1 - \frac{A_{g2}}{A_1}\right)\lvert U_g\rvert$
	$R_{v1} = 12\mu\ell_g^2 d_1/A_{g1}^3$, $L_{g1} = \rho d_1/A_{g1}$,	$R_{v2} = 12\mu\ell_g^2 d_2/A_{g2}^3$ $L_{g2} = \rho d_2/A_{g2}$
Longitudinal Components	$U_{y1} = 2\ell_g(d_1 + d_2)\frac{dy}{dt}$,	$U_{y2} = 2\ell_g(d_1 + d_2)\frac{dy}{dt}$
Lateral Components	$U_{x1} = d_1\frac{d}{dt}(A_{g1})$,	$U_{x2} = d_2\frac{d}{dt}(A_{g2})$
	$= 2\ell_g d_1\, d_{x_1}/dt$,	$= 2\ell_g d_2\, d_{x2}/dt$
	$C_{g1} = A_{g1}d_1/\rho c^2$	$C_{g2} = A_{g2}d_2/\rho c^2$
	$G_1 = S_1\frac{\eta - 1}{\rho c^2}\sqrt{\frac{\lambda\omega_0}{2c_p\rho}}$,	$G_2 = S_2\frac{\eta - 1}{\rho c^2}\sqrt{\frac{\lambda\omega_0}{2c_p\rho}}$
	$S_1 = 2(\ell_g + 2x_1)d_1$	$S_2 = 2(\ell_g + 2x_2)d_2$
Constants (for vocal system, moist air at body temperature)*	$\rho = 1.14 \times 10^{-3}$gm/cm^3, air density	
	$\mu = 1.86 \times 10^{-4}$dyne-sec/cm^2, kinematic-coefficient of viscosity.	
	$c = 3.5 \times 10^4$cm/sec, sound velocity	
	$\eta = 1.4$, adiabatic constant	
	$\lambda = 0.055 \times 10^{-3}$cal/cm-sec-deg, coefficient of heat conduction	
	$\omega_0 = 2\pi(1000)$, mid audio range radian frequency	
	$c_p = 0.24$ cal/gm-degree, specific heat at constant pressure.	

* From J. L. Flanagan, *Speech Analysis, Synthesis and Perception*, second edition, New York: Springer Verlag, 1972.

glottal current corresponding to rate of volume displacement (both lateral and longitudinal) were neglected.

II. ACOUSTIC CIRCUIT

Recognizing that the cord dimensions are very small compared to sound wavelengths at the frequencies of interest, and that all mechanical velocities are small compared to the sound velocity, we derive a one-dimensional equivalent circuit for the acoustic quantities involved. Its complete form is shown in Fig. 2. The values of all impedance elements are given in Table I.

The serial elements (top branch in Fig. 2) are identical to those of our original work[1], and relate to time-variation of the acoustic impedance of the glottal opening.

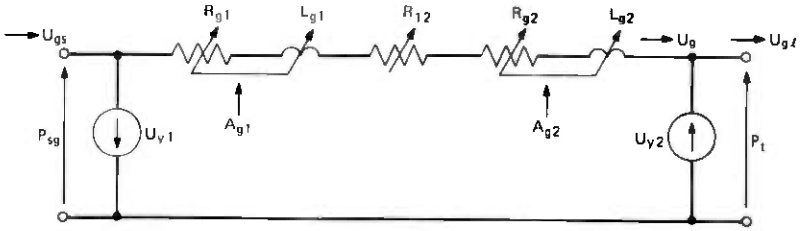

Fig.3—Simplified equivalent acoustic circuit, including longitudinal displacement currents.

All shunt elements relate to rate of displacement of air volume by the moving cord masses. The time variation of all shunt quantities is also determined by the motion of the cord masses.

Lateral motion of the cord masses (normal to the direction of glottal flow) displaces air volume at the rate of

$$U_{xi} = 2\ell_g d_i \frac{dx_i}{dt} \text{ cm}^3/\text{s}, i = 1, 2,$$

where x_i is the lateral displacement and d_i is the depth (thickness) of the cord element (mass). Again, the factor 2 arises from the two bilaterally opposing cords. The acoustic compliances, C_1 and C_2, represent the compressibility of the small air volumes contained between the opposing cords and the conductances G_1 and G_2 represent the heat-conduction loss at the tissue surfaces of the cords.

Longitudinal motion of the cord masses is assumed to occur cophasically and to be translational only. In this regard, consider the y-motion of the locked masses to be opposed by a nonlinear spring and loss similar to that of k_1 and r_1. The effective surface area exposed to the transglottal pressure difference is taken to be the product of cord length and total cord thickness, $\ell_g(d_1 + d_2)$. No cavity compliances or losses are associated with the longitudinal motion, and the longitudinal contribution to the total volume velocity is

$$U_{yi} = 2\ell_g(d_1 + d_2)\frac{dy}{dt}, i = 1, 2.$$

In other words, U_{y1} and U_{y2} are equal and oppositely poled.

Notice that in the earlier formulation,[1,2] the absence of the shunt elements imposes the constraint $U_{g\ell} = U_{gs} = U_g$. The presence of the shunt elements (all time-varying with displacements that are determined by the equations of motion for the mechanical system which, in turn, is forced by the intraglottal and transglottal pressures to close the feedback loop of the oscillator) makes the input flow U_{gs} and the output $U_{g\ell}$ typically different.

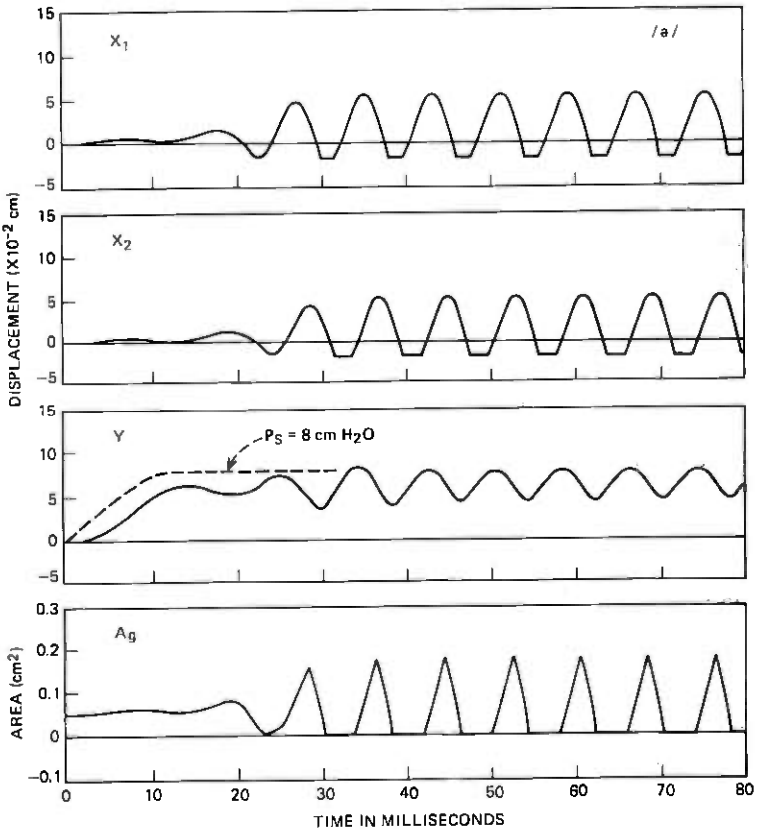

Fig. 4—Computed mechanical behavior of the vocal-cord/vocal-tract model. The vowel configuration is /ə/.

A recent related study[3] examined the influence of the U_{xi} upon $U_{g\ell}$. The present study considers separately the effects of the U_{yi}. For this purpose, the circuit of Fig. 3 is a simplification of Fig. 2.

III. PURPOSE OF PRESENT STUDY

In the original work,[1,2] we made estimates of the volume displacement currents, based upon long-wave assumptions and one-dimensional sound propagation, together with what we believed to be reasonable physiological estimates of cord velocities (compared with volume velocities responsive to transglottal pressure). We concluded that displacement currents are of second order, and in the original work we chose to neglect them in favor of elucidating dominant principles. The original formulation, therefore, treated only lateral displacement as it affects the serial glottal impedances. As a matter of completeness, we more recently have

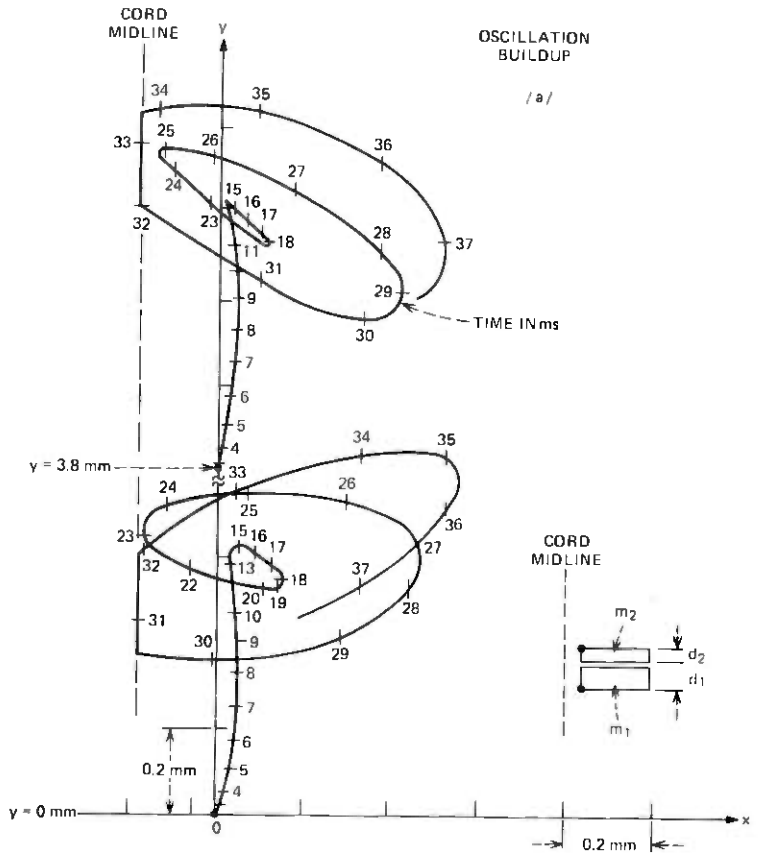

Fig. 5—X-Y trajectories for initiation of oscillation. Trajectories are for pellet positions shown in insert.

returned to a quantitative examination of these assumptions. A first study,[3] now completed, considered the importance of the shunt branches that represent the lateral components of volume velocity generated by the displacing masses—that is, from the volume current sources U_{x1} and U_{x2}. The results of the study support the original assumptions, and show the lateral components to be second order by comparison to the currents actuated by the pressure difference acting across the glottal opening.

The present study examines the contributions of the longitudinal displacement to the total glottal volume velocity (specifically, the contributions of U_{y1} and U_{y2}) and the importance of longitudinal displacement to the self-oscillatory dynamics of the cord model and to sound perception.

We take the longitudinal restoring stiffness k_y typically to be the same as the lateral restoring stiffness k_1, namely 80 kdynes/cm. The longitu-

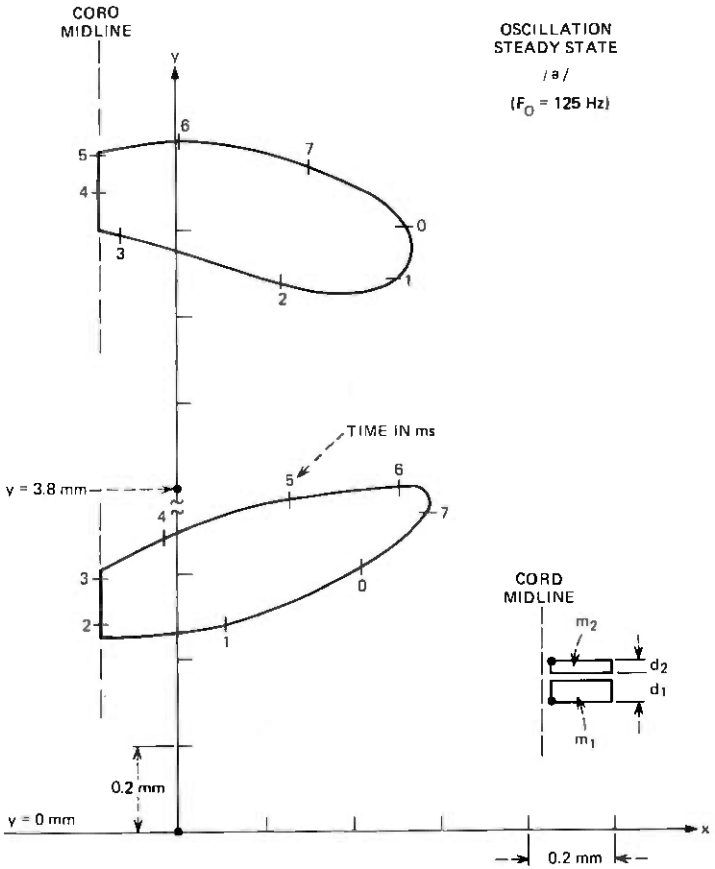

Fig. 6—X-Y trajectories for steady-state oscillation of the cord model.

dinal loss (or damping ratio) is also taken to be similar to the lateral one, namely $\zeta_y = \zeta_1 = 0.2$ These values are based upon clinical observations.[9] We examine these choices subsequently. Further, since the longitudinal and lateral motions of the cord masses are considered to be translational only, no rotational behavior is included. Still further, while the lateral translations of the coupled masses m_1 and m_2 can have large (and physiologically natural) phase differences, their longitudinal translations are considered to be cophasic, and the internal coupling stiffness is assumed to act only for lateral motion. The lateral and longitudinal motions are, therefore, coupled only through the acoustic variables that determine the oscillator forcing functions. In the course of our discussion, we will indicate comparisons to actual physiological data to assess the realism of these assumptions.

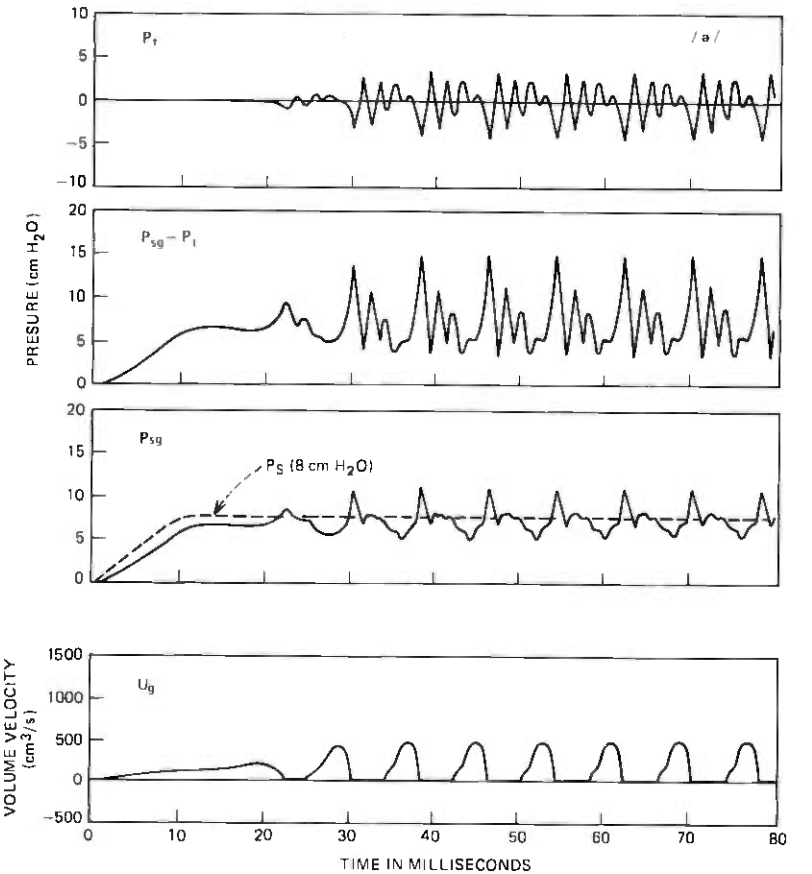

Fig. 7—Computed acoustic qualities for the vocal-cord/vocal-tract model. The vowel is /ə/.

IV. RESULTS OF COMPUTER SIMULATIONS

The vocal-cord model, as represented by Fig. 3, was combined with a transmission-line formulation of the vocal tract that we have used previously in speech synthesis studies.[4] The programmed vocal tract contains 20 sections which, in addition to the classical acoustic elements, represents the yielding soft walls of the tract and sound radiation from the yielding walls. This formulation is based upon measurements of tissue impedances that we reported earlier.[5] Also included for the present study is a transmission-line representation of the subglottal system. Six sections of line represent the trachea, bronchi and lungs, as previously described.[6] We implemented the entire system in terms of difference equations programmed on a laboratory computer by techniques we have described in detail previously.[1,2]

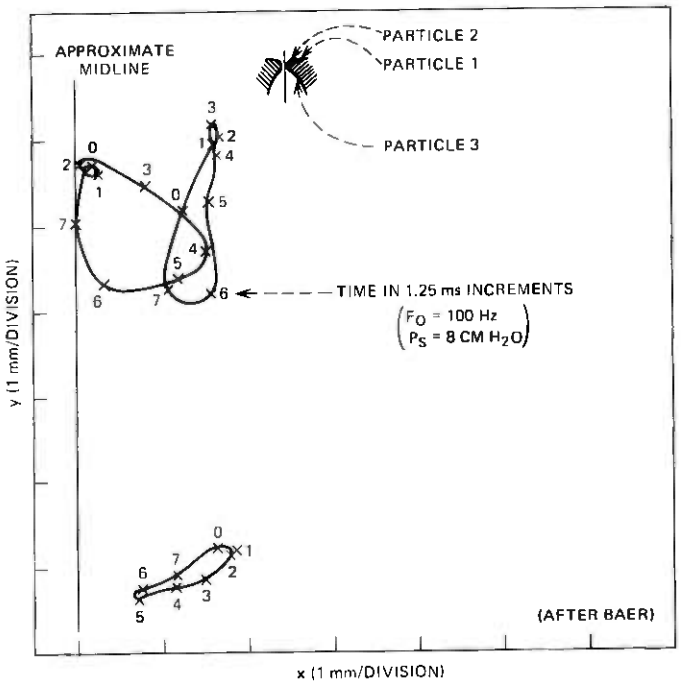

Fig. 8—X-Y trajectories observed from excised larynx of dog (after Baer).

Most of the data reported here are for the vocal tract configured in the shape for the neutral (schwa) vowel /ə/. Some data are also included for the vowels /i/ and /a/.

A first step is to ascertain if the cord oscillator, so arranged for longitudinal motion, performs realistically when compared with observations on the human larynx. A second step, then, is to determine the acoustic significance of the volume displacement current arising from longitudinal motion.

Throughout these calculations, the laryngeal parameters are set to the "standard" values used earlier for phonation by a man's voice in the chest register[1] (i.e., neutral glottal area $A_{go} = 0.05$ cm^2, cord tension parameter $Q = 0.78$, $d_1 = 0.25/Q$ cm, $d_2 = 0.05/Q$ cm). Recall that the Q parameter scales the values of mass and stiffness and, hence, also the values of the d_i. Phonation is initiated by raising the lung pressure P_s smoothly from zero to the standard value of 8 cm H_2O. The pressure is elevated in a 10-ms interval.

4.1 Mechanical behavior

As the lung pressure is elevated, the model commences a buildup of oscillation. After four or five transient swings, the oscillation settles into

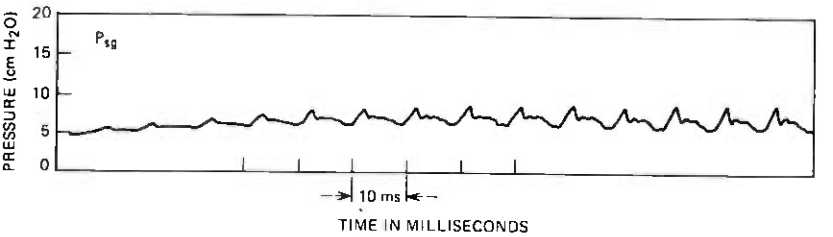

Fig. 9—Subglottal pressure variation measured on a human subject (after Sawashima).

a steady state behavior with a fundamental frequency (pitch) determined by the model parameters (the tension parameter Q has the dominant effect on pitch[1]). The initial 80 ms of this synthetic phonation is illustrated in Fig. 4 for the mechanical variables.

The top two curves show the displacements x_1 and x_2 of mass pair m_1 and mass pair m_2, respectively. The first collision of each mass pair is indicated by the first flat, negative-going portion of the displacement waveforms. For the $A_{g0} = 0.05$ cm^2 value, this occurs for $x_i = -0.0178$ cm. Note, too, that x_1 leads x_2 in phase by the order of 60°, which is

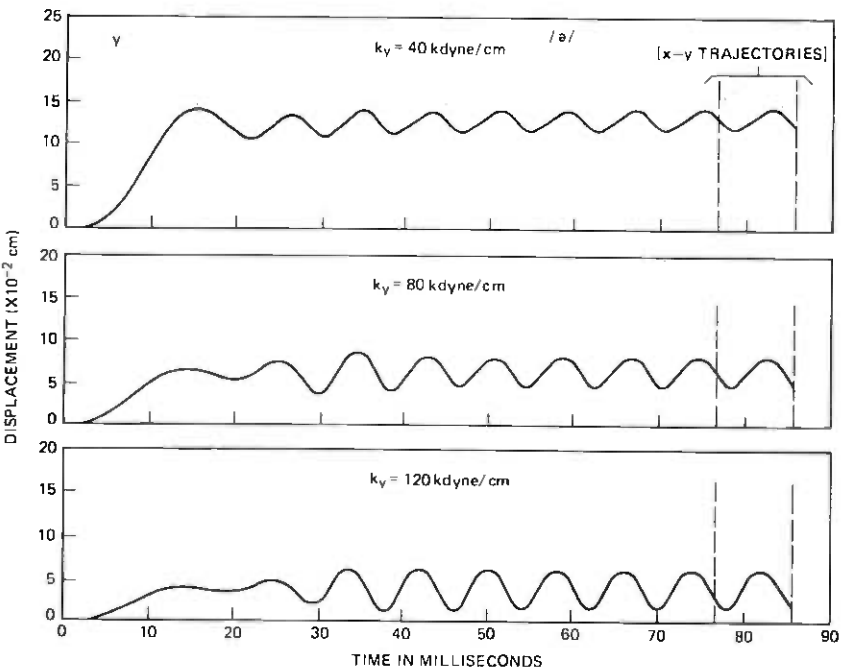

Fig. 10—Effect of longitudinal restoring stiffness k_y upon the longitudinal displacement, y. Data show oscillation buildup for a lung pressure P_s that is raised smoothly from zero to 8 cm H_2O.

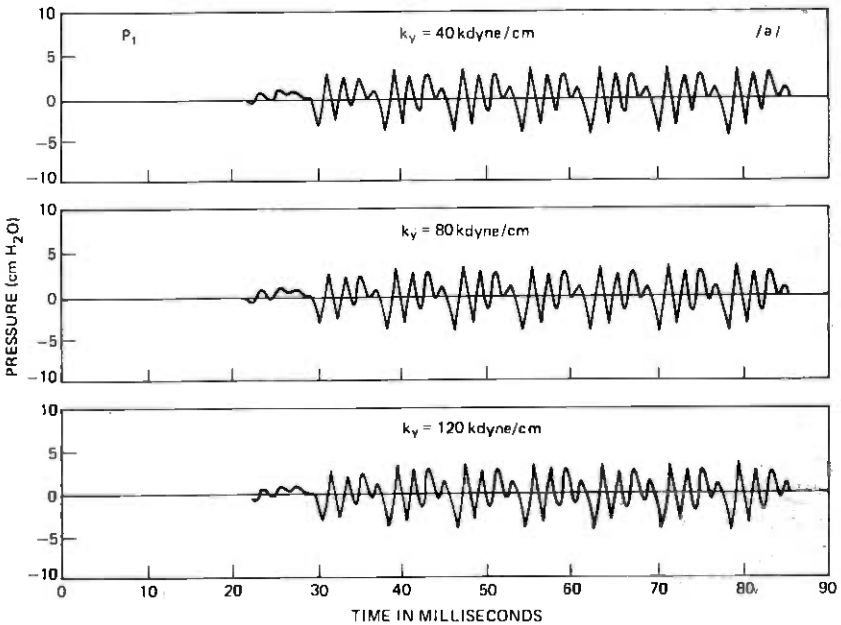

Fig. 11—Effect of k_y upon P_l.

consistent with observations from high-speed motion pictures of the human vocal cords. The third trace shows the longitudinal displacement, y, which bulges upward as P_s is raised. The y motion is roughly sinusoidal. The lower trace shows the net area of glottal opening A_g (namely the minimum of A_{g1} and A_{g2}). The y-displacement is seen to lead in phase the A_g wave, again consistent with the upward, rolling motion seen in high-speed photography of the real cords.

An x-y plot of the buildup transient portrays the behavior perhaps more graphically. Figure 5 shows the x_1 vs y and the x_2 vs y values with time as the parameter. Imagine pellets fixed to the lower and upper inner edges of one simulated cord, shown by the inserted anterior-posterior view of Fig. 5. The trajectories of the two pellets are plotted for the oscillation buildup. The y-axis is broken and re-originated at $y = (d_1 + d_2) = [(2.5 + 0.5\text{mm})/Q] = 3.8$ mm. The flat portion of the tracks, along the vertical midline, reflect collision with the opposing vocal-cord mass.

After several initial swings, the oscillator settles into a steady-state behavior. One cycle of this trajectory is shown in Fig. 6. The steady-state pitch frequency in this case is $Fo = 125$ Hz, or a period of $T = 8$ ms.

4.2 Acoustical behavior

The corresponding acoustical parameters, calculated for the same buildup period, are shown in Fig. 7. The acoustic pressure at the input

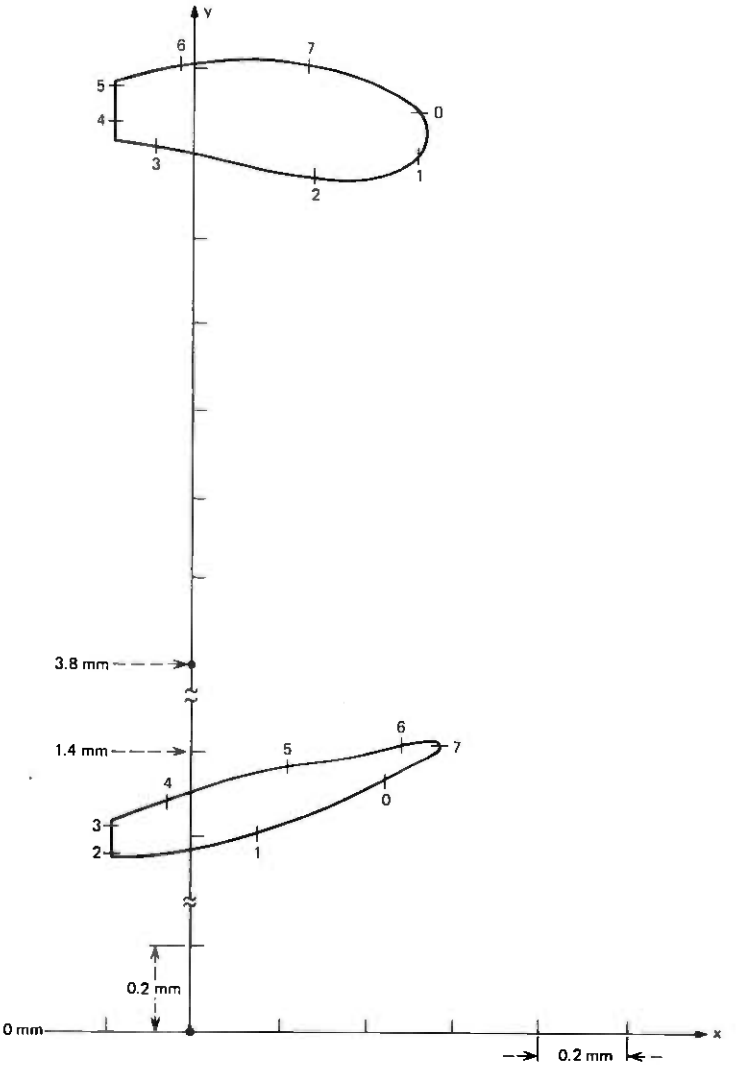

Fig. 12(a)—Steady-state X-Y trajectory for k_y = 40 kdyne/cm.

to the vocal tract P_t is shown in the top trace. It reflects strongly the eigenfrequency structure of the tract, in this case configured for /ə/ and having formant frequencies of approximately 500 Hz, 1500 Hz, 2500 Hz . . . The transglottal pressure ($P_{sg} - P_t$), which is the forcing function for the y-motion and the pressure potential for the volume flow through the glottal opening, exhibits a pronounced pitch-synchronous variation. Its peak values, in fact, approach twice the lung pressure value of P_s = 8 cm H_2O. Recall that P_s is the lung pressure input to the simulated

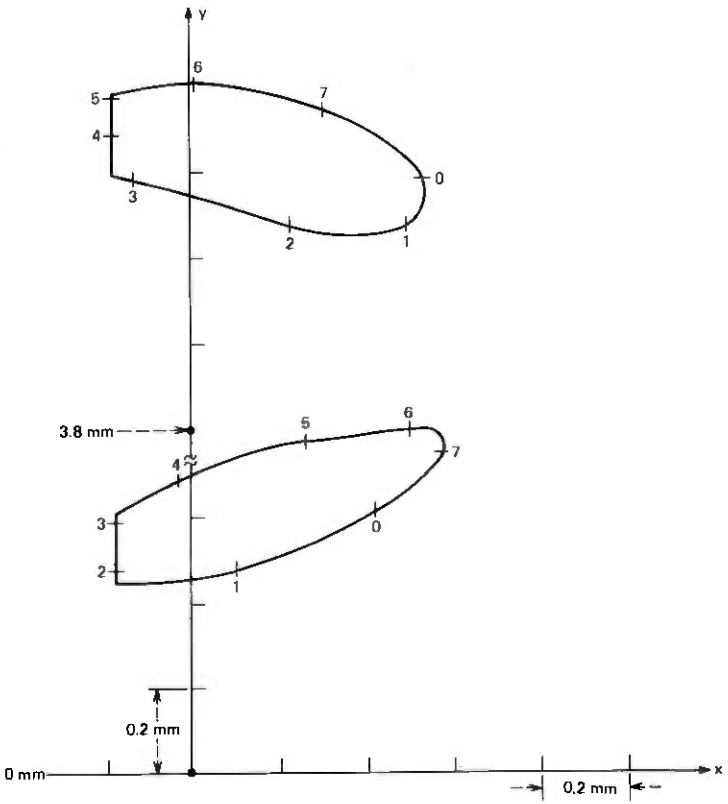

Fig. 12(b)—Steady-state X-Y trajectory for k_y = 80 kdyne/cm.

subglottal system, representing trachea and bronchi. But notice that the mechanical y-displacement (Fig. 4) does not respond with this detail. (Neither do the x_1 and x_2 displacements respond to high-frequency detail in their forcing functions—that is, the mechanical system, being mass-controlled, filters out this detail.)

The subglottal pressure P_{sg} (the pressure just beneath the vocal cords) also exhibits a pitch-synchronous fluctuation, but of somewhat less amplitude, namely about ±20 to ±30 percent of the mean subglottal pressure. Its positive peaks correspond to the closing epochs of the glottal port. The calculated volume velocity passing the glottal opening U_g (bottom trace) appears as a traditionally shaped, pulsive waveform. This wave is similar to that calculated in previous work (without longitudinal motion) but differs in that its values are modified by the effects that U_{y1} and U_{y2} couple into the pressure variables. That is, U_{y1} and U_{y2} can influence P_{sg} and P_t and, hence, $U_{g\ell}$. The latter three variables, in turn, close the oscillator feedback loop by constituting the forcing functions

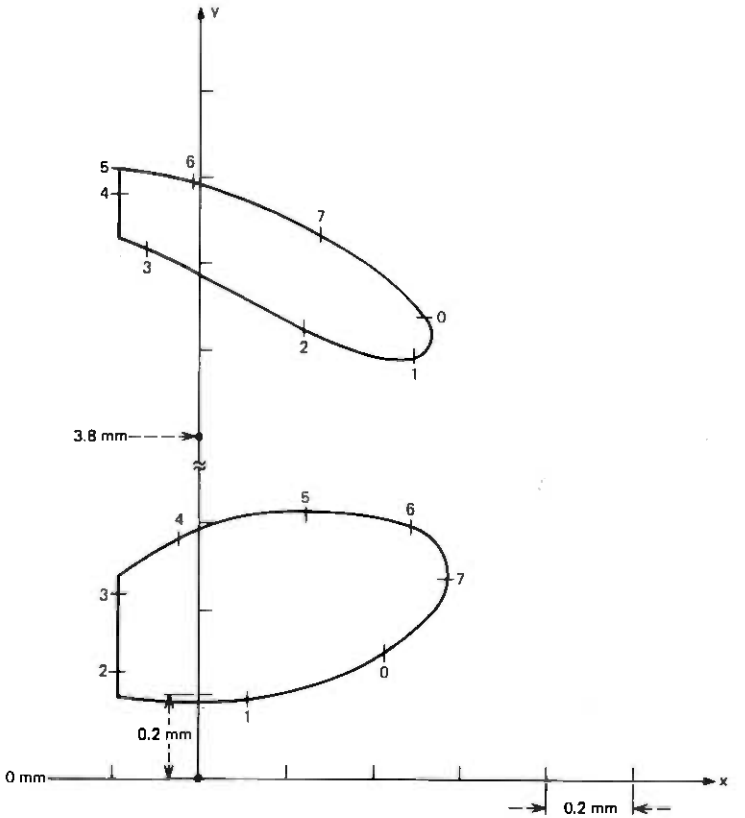

Fig. 12(c)—Steady-state X-Y trajectory for k_y = 120 kdyne/cm.

for the lateral displacement. As was the case in the mechanical variables, the U_g flow does not reflect a temporal fine structure comparable, say, to the $(P_{sg} - P_l)$ waveform. The resistive and inertive components of the glottal impedance (i.e., the serial components in Fig. 3) act effectively as a low-pass filter. It is not unusual, however, to see pronounced temporal structure that corresponds to the lowest eigenfrequency of the vocal tract, especially for low, back vowels (such as /a/), or for tightly articulated sounds.

A next question, then, is how do these mechanical and acoustical quantities, resulting from the model with longitudinal displacement, compare with physiological data.

4.3 Comparisons to physiological observations

One qualitative comparison can be made for the mechanical displacement behavior. Baer[7] performed studies on the excised larynx of

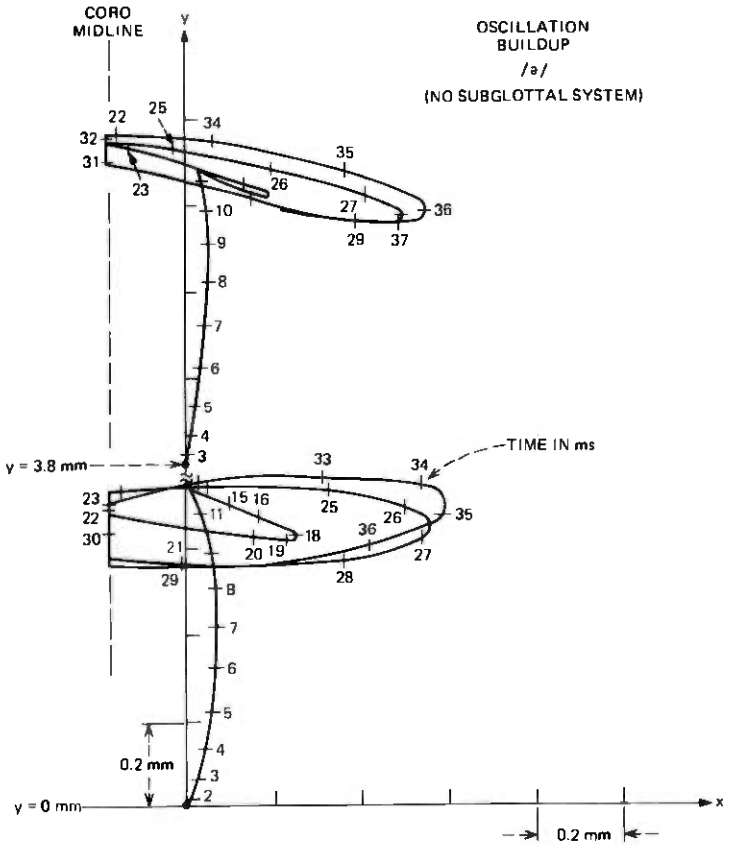

Fig. 13—Oscillation buildup without subglottal system.

a dog in which he fixed pellets to the displacing tissue and made optical observations under stroboscopic illumination. While his pellet positions do not correspond exactly to our mass-pair corners, we can roughly compare his observations with our data. Figure 8 shows x–y trajectories for one set of conditions for the dog larynx that approximates values used in human phonation (namely P_s = 8 cm H_2O, U_g = 275 cm^2/s, and Fo = 100 Hz). Particles (pellets) 2 and 3 are of interest. While the vibratory excursions of the excised dog larynx are larger than those we calculate with the model, the qualitative motions are gratifyingly similar. One question that arises is how much does the longitudinal (vertical) displacement depend upon the choice of longitudinal stiffness constant. We shall examine this question in more detail subsequently.

Another comparison can be made in the acoustic domain—namely, to the subglottal pressure variation P_{sg} shown previously in Fig. 7. Sawashima[8] has measured the subglottal pressure during phonation in a

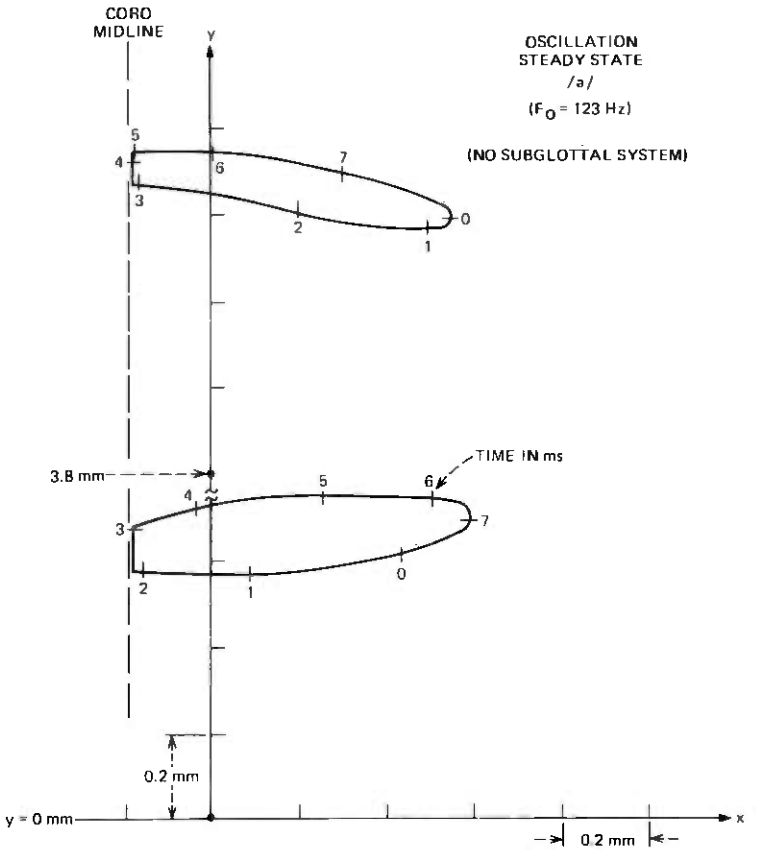

Fig. 14—Steady-state oscillation without subglottal system.

human subject. One of his results is shown in Fig. 9. The qualitative correspondence to the model calculation appears relatively good, and the acoustic interaction among the simulated cords, vocal tract, and subglottal system is realistic.

4.4 Effect of longitudinal stiffness constant

In view of uncertainties in the measurement of the stiffness constants in physiological preparations, it is important to examine how critical the value of k_y (the longitudinal restoring stiffness) is to the oscillatory behavior of the model.

For the bulk of our studies, we have taken k_y equal to our typical "standard" value of k_1, namely 80 kdynes/cm.[1] We have also used the standard value for the damping ratio, $\zeta_y = \zeta_1 = 0.2$. This choice is based upon the physiological measurements on cord tissue conducted by Ka-

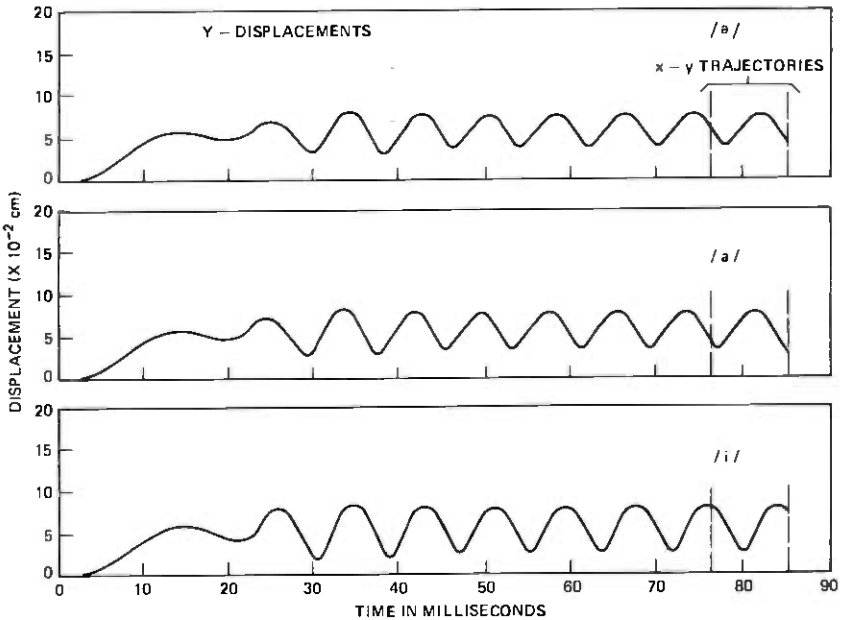

Fig. 15—Effect of vowel configuration upon longitudinal displacement.

neko,[9] who found the stiffness constants for lateral and for longitudinal displacement to be similar. Both stiffnesses are also taken to have the same nonlinearity, namely, a cubic nonlinearity in the restoring force (see Ref. 1).

To assess the model's sensitivity to large variations in k_y, we calculated the buildup of synthetic phonation for k_y = 40, 80, and 120 kdynes/cm. The resulting y-displacements for these values is shown in Fig. 10, and the sound pressure at the entrance to the vocal tract is shown in Fig. 11. Also, the x-y trajectories are shown in Fig. 12a, b, and c.

As would be expected, the greatest influence in this variation is reflected in the y-displacements. The "softer" k_y gives larger dc displacement and smaller peak-to-peak vibratory excursions. The P_t data indicate that the variations in acoustic behavior and glottal excitation are very small. The fundamental pitch is sensibly the same for all cases, namely 125 Hz. This factor is almost completely dominated by the lateral motion. In auditory assessment of the output synthetic sound, the differences are virtually imperceptible, suggesting that the longitudinal displacement current is insignificant for speech synthesis.

4.5 Effect of subglottal system

A side issue, of some interest in passing, is the effect of acoustic interaction between the subglottal system and the cord oscillator. If the

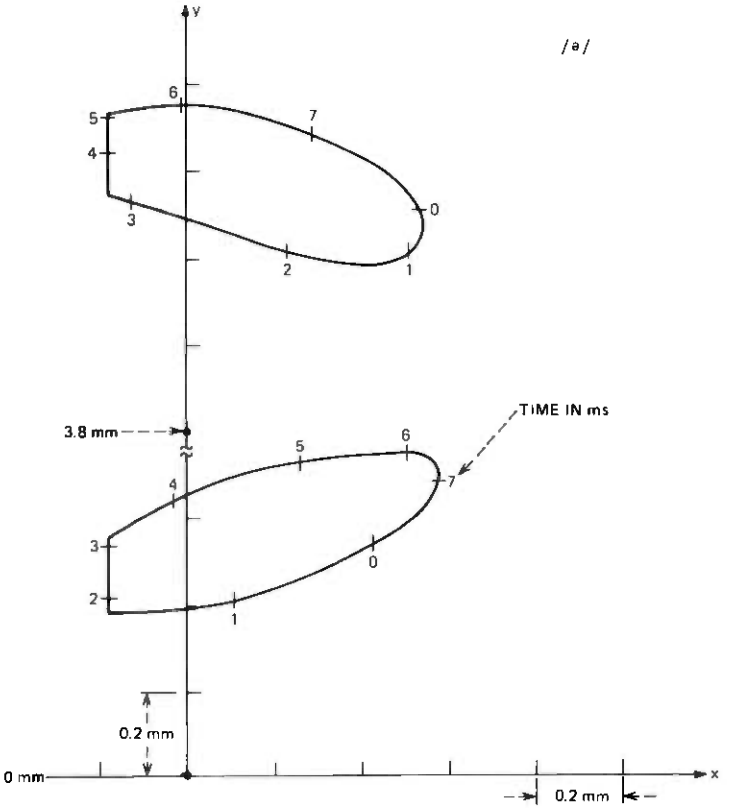

Fig. 16(a)—X-Y trajectory for the vowel /ə/.

trachea-bronchi-lung system is removed and the lung pressure applied directly to the cord input (as a zero-impedance source; i.e., P_{sg} becomes a pressure "battery" equal to P_s), then the temporal structure previously reflected in P_{sg} is eliminated, the transglottal pressure excursions simply equal P_t, and the longitudinal component of cord displacement is lessened. This is illustrated by the x-y trajectories for oscillation buildup and steady state shown in Figs. 13 and 14. For this case, k_y is reset to the typical value 80 kdynes/cm. Note the slight lowering of the fundamental frequency to 123 Hz.

4.6 Effect of vowel configuration

It also is instructive to consider the influence of vowel configuration upon the cord model, as presently formulated. Such studies were made in detail in the original work.[1] We therefore compare synthesized results for the vowel /ə, a and i/. (In this case the longitudinal stiffness constant

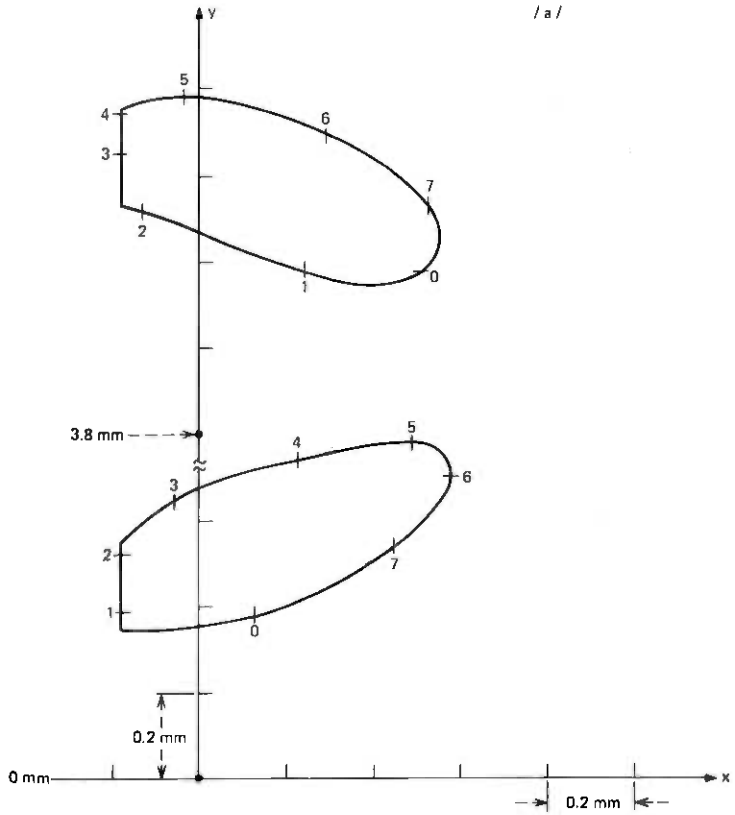

Fig. 16(b)—X-Y trajectory for the vowel /a/.

was set to k_y = 88 kdynes/cm through an inadvertent keypunch error. Because the differences are so small, it did not seem worthwhile to recompute the data for k_y = 80 kdynes/cm.)

The y-displacements are shown in Fig. 15, and the x-y trajectories for one period of steady-state oscillation are shown in Fig. 16a, b, and c. The longitudinal displacement is not greatly affected by vowel configuration, but the constricted articulations /a, i/ clearly lead to slightly greater longitudinal peak-to-peak excursions than does the open-pipe (neutral) vowel /ə/. This typically is owing to the greater acoustic interaction at the eigenfrequencies for the configurations with higher acoustic impedance levels, which in turn leads to greater transglottal pressure differences. This is well reflected in the acoustic variables resulting from this calculation. The corresponding acoustic quantities are shown in Figs. 17 thru 20. In these data, note especially how the tract eigenfrequencies are manifest, including in the synthetic output sound pressure from the

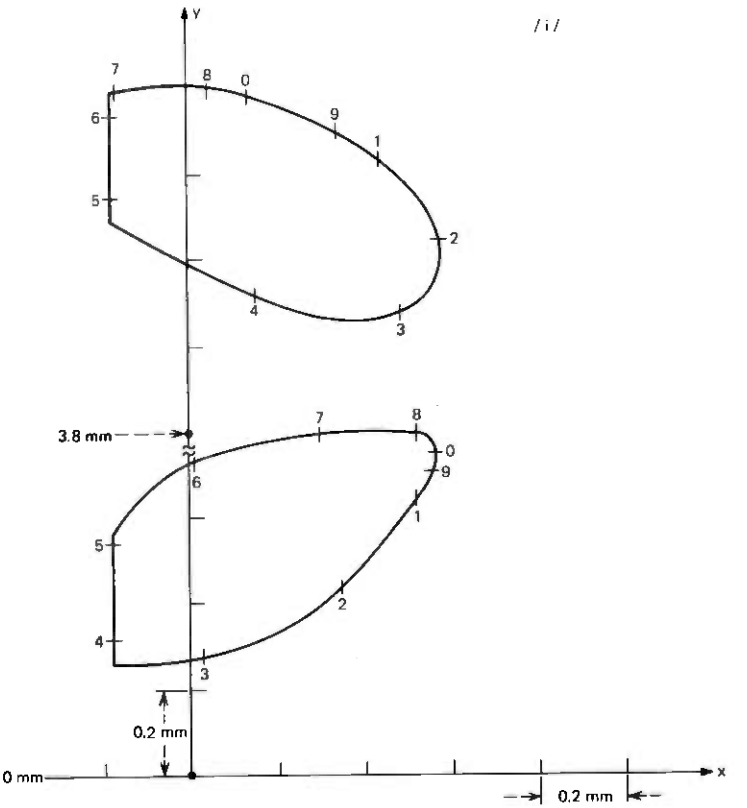

Fig. 16(c)—X-Y trajectory for the vowel /i/.

mouth, P_{out}. Note, too, that the resulting steady-state pitch frequencies are about 125 Hz for /ə/ and /a/ and about 120 Hz for /i/.

V. SIGNIFICANCE OF LONGITUDINAL DISPLACEMENT

Having established that the cord oscillator (with lateral and longitudinal degrees of freedom) appears to behave realistically, we consider the next questions:

(*i*) Is the longitudinal motion significant or necessary for proper self-oscillatory operation of the model?

(*ii*) Is the acoustic volume velocity contributed by the longitudinal motion physically or perceptually significant?

So far as the purposes of speech synthesis are concerned, we answer both of these questions in the negative.

What we wish to do, therefore, is compare the mechanical and acoustical behavior with and without longitudinal displacement. Because the longitudinal effects are coupled only through U_{y1} and U_{y2}, the lat-

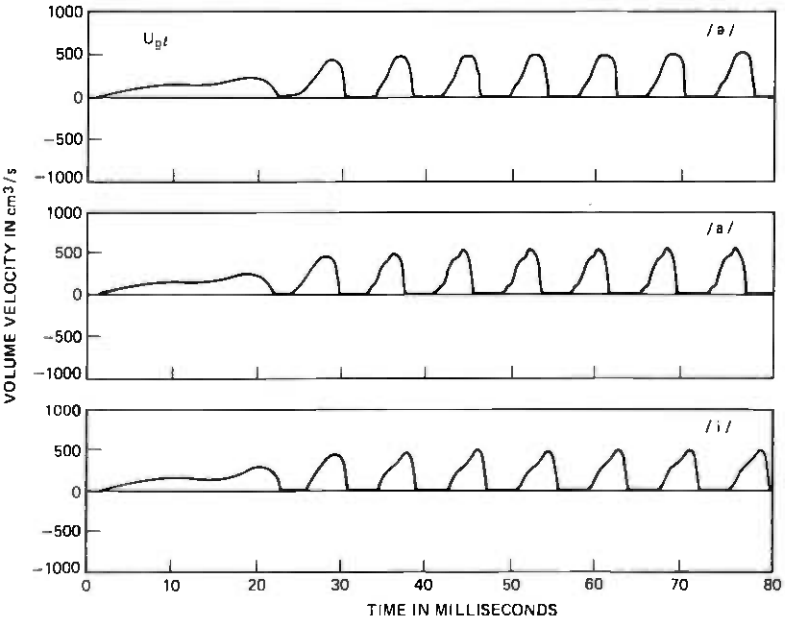

Fig. 17—$U_{g\ell}$ for the vowels /ə/, /a/, and /i/.

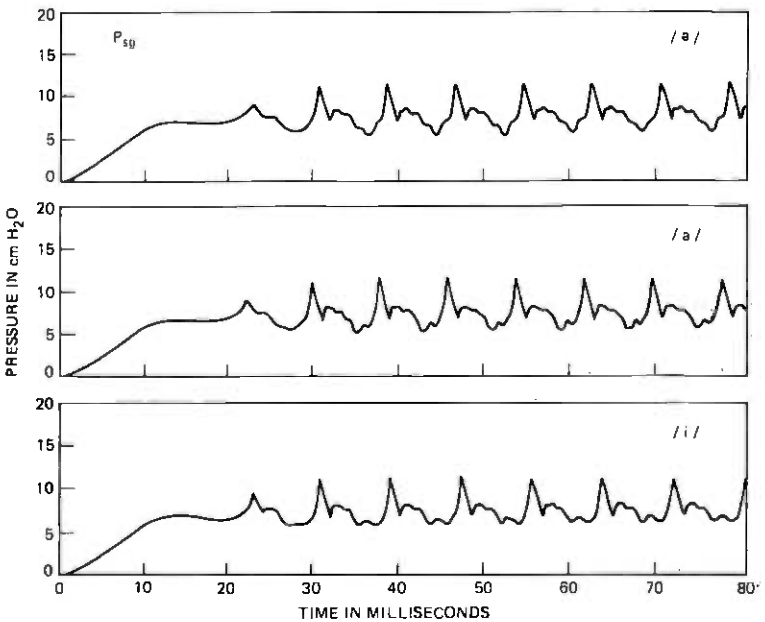

Fig. 18—P_{sg} for the vowels /ə/, /a/, and /i/.

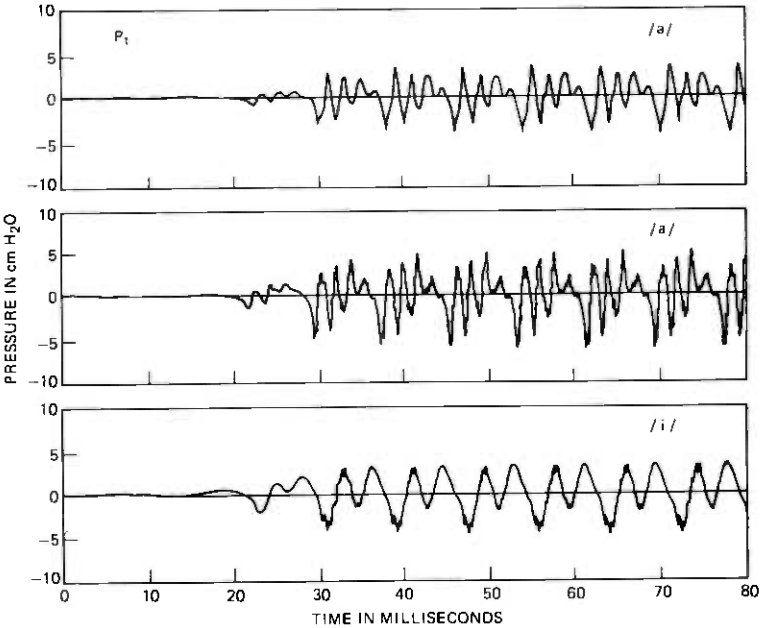

Fig. 19—P_t for the vowels /ə/, /a/, and /i/.

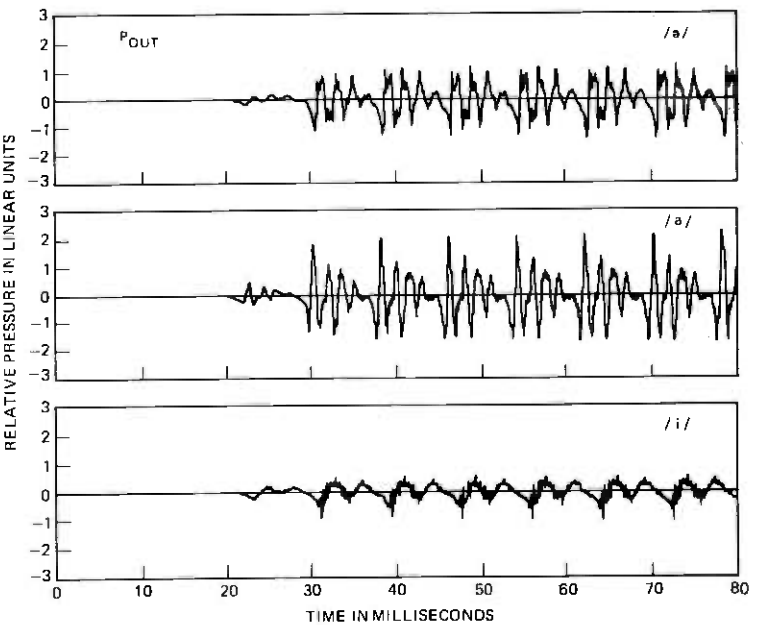

Fig. 20—P_{out} for the vowels /ə/, /a/, and /i/.

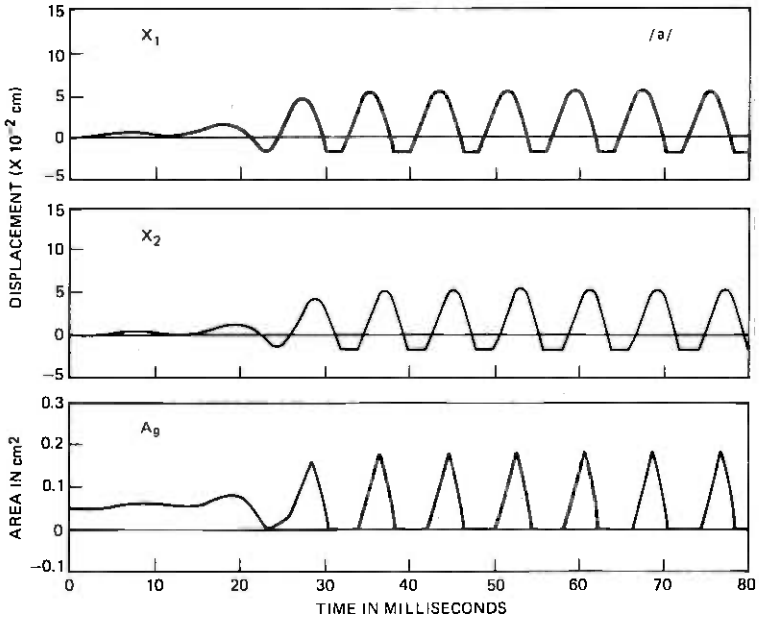

Fig. 21—Computed mechanical behavior of the cord/tract model without y-displacement.

eral-motion-only condition (that is, the original formulation of the model) is conveniently realized by setting $U_{y1} = U_{y2} = 0$. This condition, with no longitudinal displacement flow, yields the mechanical results shown in Fig. 21. These results are essentially identical (at least to three significant figures) to the corresponding quantities in Fig. 4. In particular, note that the crucial A_g waveforms are virtually identical.

We can now examine pertinent acoustic quantities. Calculation of the conditions with y-displacement yields the result shown in Fig. 22. The figure shows one cycle of the steady-state oscillation. Recall that $U_{g\ell}$ is the total volume velocity at the larynx tube entry to the vocal tract. U_g is the flow component through the actual opening of the glottis. Note that $U_{g\ell}$ is non-zero during the time the cords are actually closed, corresponding to an upward (vertical, longitudinal) displacement of air volume that adds positively to U_g. Similarly, later in the cycle, downward longitudinal displacement subtracts from U_g. The difference between these volume velocities is

$$(U_{g\ell} - U_g) = U_{y2} = U_{y1}$$

by virtue of the assumption of cophasic longitudinal motion. This difference is plotted on an X10 enlarged scale in Fig. 22. The peak value

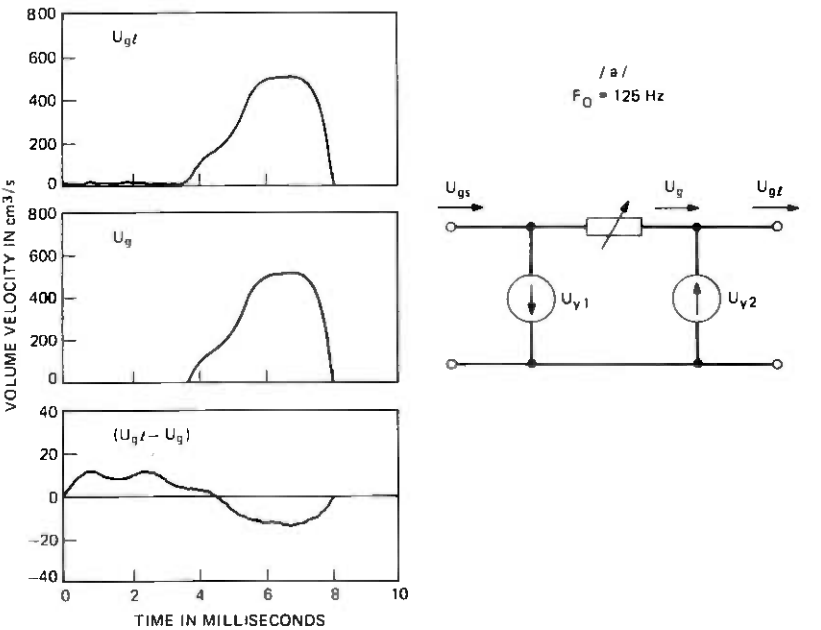

Fig. 22—Glottal volume velocities calculated with y-displacement (refer to Fig. 3).

of the difference for this condition is on the order of 10 cm^3/s, or about $1/50$ of the peak value of the total $U_{g\ell}$. This result is not just peculiar to this range of volume velocity, hut rather it scales comparably at louder and softer phonation. For example, if the lung pressure P_s is doubled say to 16 cm H_2O, the longitudinal displacement current increases because the transglottal pressure and the longitudinal displacement increase. But the U_g flow also increases and remains far and away the dominant quantity.

The amplitude spectra of these quantities provide convenient correlation with auditory percepts. The spectra for $U_{g\ell}$ and U_g are shown in Figs. 23a and b. A close comparison shows the differences to be less than 2 dB, an amount that is not significant perceptually. The more relevant comparison is obtained when the effect of y-motion is eliminated (by removing U_{y1} and U_{y2}). The corresponding glottal waveform for no y-displacement is illustrated in Fig. 24. It is denoted $U_{g\ell}^*$. Also reproduced is the $U_{g\ell}$ with y-displacement. Further, the difference between the longitudinal displacement and lateral-only conditions $(U_{g\ell} - U_{g\ell}^*)$ is shown on an X10 enlarged scale. During the glottis-closed time, this difference is identical to the $(U_{g\ell} - U_g)$ difference of Fig. 22, because $U_g = 0$. During the glottis-open time, the $(U_{g\ell} - U_{g\ell}^*)$ difference differs from the $(U_{g\ell} - U_g)$ difference. In other words, U_g differs from $U_{g\ell}^*$

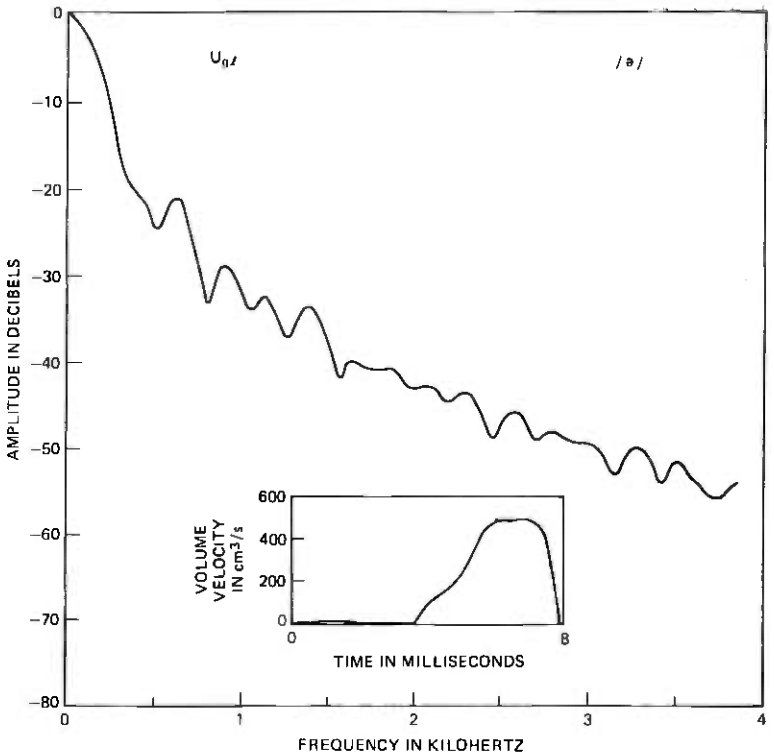

Fig. 23(a)—Amplitude spectrum of $U_{g\ell}$, which includes the effects of y-displacement.

essentially by the influence that U_{y1} and U_{y2} have upon the transglottal pressure difference $(P_{sg} - P_t)$.

Again, the more perceptually relevant comparison is to the amplitude spectrum. The spectrum of $U_{g\ell}^*$ is given in Fig. 25. A close comparison to the $U_{g\ell}$ spectrum of Fig. 23 shows the differences to be less than about 2 dB. Auditory assessment of the output synthetic vowels shows them to be indistinguishable even in close comparison.

VI. CONCLUSION

In view of these results, we conclude that realistic acoustic behavior (which is needed in speech synthesis) can be obtained in the cord model without the additional complexity of longitudinal displacement. Longitudinal displacement is not necessary for realistic self-oscillation of the model. The important vertical phase differences in the two-mass motion are adequately duplicated by lateral displacement only, as is the significant acoustic interaction between vocal tract and vocal cords. Further, the rate of volume displacement owing to longitudinal motion

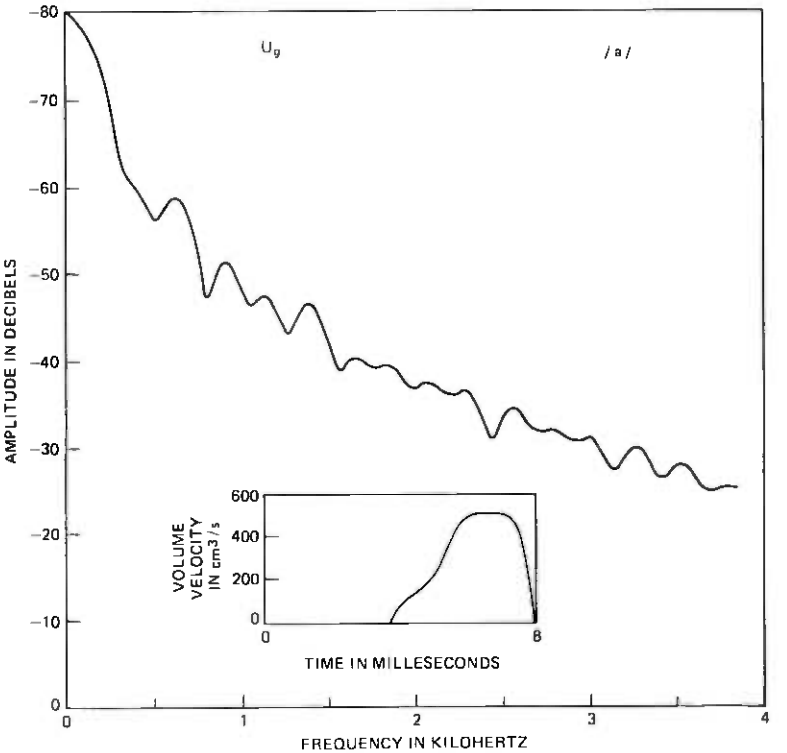

Fig. 23(b)—Amplitude spectrum of U_g calculated with y-displacement.

is clearly perceptually not significant and need not be represented with added detail.

These conclusions about the mechanical and acoustic behavior have a corollary in a companion study on the rate of displacement of air volume owing to lateral motion only.[3] This contribution was examined by making use of the shunt branches in Fig. 2 that include U_{x1}, U_{x2}. Calculations and computer simulations showed that the contribution to glottal volume velocity of the air extruded from the glottal port by lateral tissue displacement is barely discriminable in a differential auditory comparison. In fact, the perceptual effect for the lateral volume displacement is just slightly larger than for the longitudinal displacement. Both are quite second-order in importance.

We have found in the present study that proper acoustic and oscillatory behavior of the model does not depend significantly upon longitudinal displacement. The longitudinal motion is relatively insensitive to acoustic loading and to changes in longitudinal stiffness. The longitudinal motion influences fundamental frequency only slightly. What,

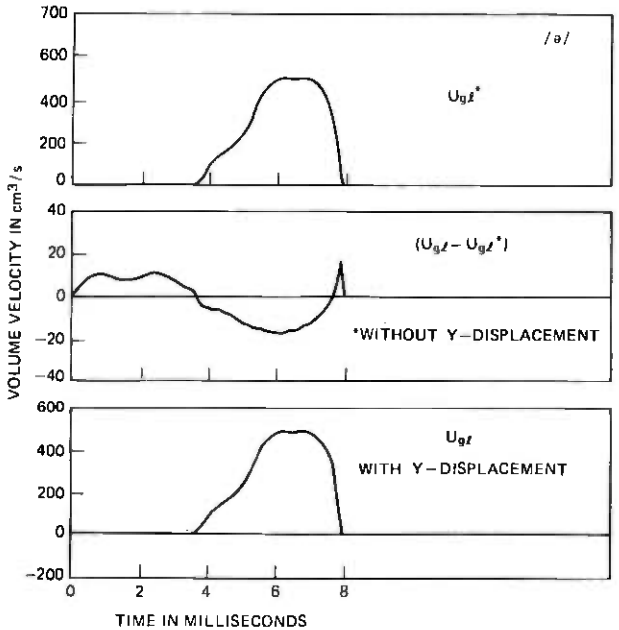

Fig. 24—Comparison of waveforms for $U_{g\ell}$, which includes y-displacement current, and $U_{g\ell}^*$, which is calculated without y-displacement.

then, are the critical and sensitive parameters of the cord model? In other words, what parameters are most influential upon the perceptual attributes of $U_{g\ell}$, since the end product—the output sound—depends directly upon $U_{g\ell}$? The results of our earlier work can be combined with the insights obtained here to consider this question.

The original study showed that the intra-glottal pressure distribution, and the fluid flow laws used to deduce it, are quite important to proper oscillatory behavior, to proper generation of the $U_{g\ell}$ flow, and to realistic acoustic interaction between the vocal tract and vocal cords. To a large extent this pressure distribution determines how the pitch frequency varies with subglottal pressure and with articulatory configuration. The mass-stiffness product (i.e., the natural frequency of the mechanical system) is quite dominant in determining pitch range. Subglottal pressure, assuming it to be above an initiation threshold of several cm H_2O, is primarily correlated with sound intensity, a relatively noncritical factor for voiced sounds. Mechanical parameters such as cord thickness, damping ratio, and nonlinearity are relatively noncritical except as they influence duty factor and "flow chopping" at collision (which yields a broad-spectrum $U_{g\ell}$ function). None of the mechanical variables, lateral or longitudinal, reflects the temporal fine structure of the acoustic variables, but both must and do reflect the open-close cycles of the vi-

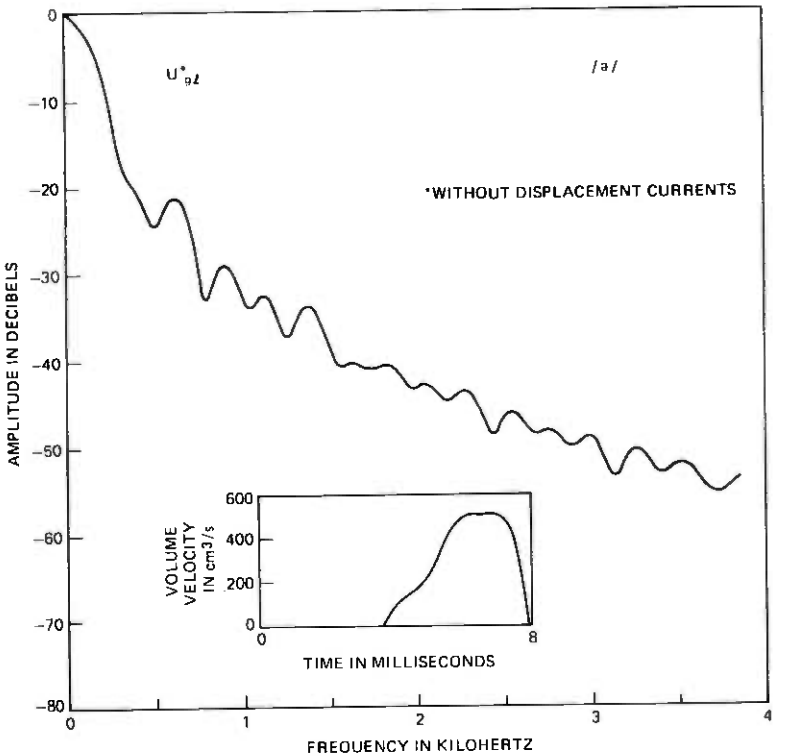

Fig. 25—Amplitude spectrum of U_g^*, the glottal volume velocity without y-displacement.

brating cords. The realistic phase differences in motion of the upper and lower edges of the cords (m_2 and m_1 in the model) allow phonation smoothly over a wide range of input impedances to the vocal tract (both inductive and capacitive), and this behavior can be obtained satisfactorily by permitting lateral displacement only of the stiffness-coupled masses. The computational complexity of anything more detailed does not seem necessary from the standpoint of duplicating realistic acoustic behavior, which is the objective in speech synthesis.

On the other hand, if the objective were a detailed study of tissue deformation (as might be the case in simulations for clinical diagnosis or for representing pathological conditions) then the computational complexity of longitudinal displacement might be considered. In such a case, the vocal-cord model should be treated as a more distributed system. For the representation and synthesis of normal speech, however, these details do not appear perceptually significant and are not needed to represent the dominant properties of vocal-cord vibration.

REFERENCES

1. K. Ishizaka and J. L. Flanagan, "Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords," B.S.T.J. *50*, (July–August 1972), pp. 1233–1268.
2. J. L. Flanagan and K. Ishizaka, "Automatic Generation of Voiceless Excitation in a Vocal-cord/Vocal-tract Speech Synthesizer," IEEE Trans. ASSP, *24* (April 1976), pp. 163–170.
3. K. Ishizaka and J. L. Flanagan, "Acoustic Characterization of the Air Volume Displaced by the Vibrating Vocal Cords," Proc. (British) Inst. Acoust. (Sept. 1976).
4. J. L. Flanagan, K. Ishizaka, and K. L. Shipley, "Synthesis of Speech From a Dynamic Model of the Vocal Cords and Vocal Tract," B.S.T.J., *54*, No. 3 (March 1975), pp. 485–506.
5. K. Ishizaka, J. C. French, and J. L. Flanagan, "Direct Determination of Vocal Tract Wall Impedance," IEEE Trans. ASSP, *23*, No. 4 (August 1975), pp. 370–373.
6. K. Ishizaka, M. Matsudaira, and T. Kaneko, "Input Acoustic-Impedance Measurement of the Subglottal System," J. Acoust. Soc. Am., *60*, No. 1 (July 1976), pp. 190–197.
7. T. Baer, "Investigation of Phonation Using Excised Larynxes," Ph.D. thesis, M.I.T., Cambridge, Mass., 1976.
8. M. Sawashima, H. Hirose, and Y. Kurauchi, "Subglottal Pressure during Phonation," Jpn. J. Logopedics and Phoniatrics, 5, No. 2 (August 1964), pp. 84–85 (in Japanese).
9. T. Kaneko, H. Asano, H. Miura, and K. Ishizaka, "Biomechanics of the Vocal Cords—On Stiffness," Practica Otologica (Kyoto), *64*, 1229–1235, 1971.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 6, July–August 1977
Printed in U.S.A.

Faulty-Trunk Detection Algorithms Using EADAS/ICUR Traffic Data

By J. S. KAUFMAN

(Manuscript received November 1, 1976)

A class of algorithms for detecting abnormally short-holding-time trunks has been developed that utilizes individual trunk data available in EADAS/ICUR (Engineering and Administrative Data Acquisition System/Individual Circuit Usage Recorder). This data consists of a two-dimensional statistic that compresses the raw trunk measurements—the state of the trunk (busy or idle) sampled every 100 or 200 seconds—into a manageable form. Because this data is essentially a sufficient statistic for the stochastic process used to model the (unobservable) trunk state measurements, one of the algorithms developed is Wald's sequential probability ratio test. Two of the algorithms developed have been implemented in ICAN (Individual Circuit Analysis Program), and are currently being used to test trunks associated with the No. 1 crossbar, No. 5 crossbar, crossbar tandem (1XB, 5XB, XBT), and step-by-step switching machines. The focus in this paper, however, is on the modeling and analysis aspects of the problem, and only slight attention is paid to the various trade-offs and real-world constraints encountered in implementing the algorithms.

I. INTRODUCTION

A message trunk, the basic connecting link in the switched telephone network, provides the communication path between switching machines as well as certain call setup capabilities, such as supervision, signaling, and ringing. For an important class of trunk faults that cause call failure, the trunk is released by the switching system upon customer abandonment and is again available to fail another call. As a result, a single undetected faulty trunk of this type can fail a significant fraction of the offered attempts to a group and will characteristically have an abnormally short holding time.

Because of their potential service impact, significant efforts have been made to understand and quantify the impact that such abnormally

short-holding-time trunks have on central office and network service.[1–4] It is now widely understood as a result of these studies that this generic trunk fault results in a fraction of service attempts "killed," which is out of all proportion to their number in the trunk population. Consequently, traffic data available from new and existing traffic data-acquisition systems has been viewed in the light of increasing interest in trunk-fault detection. In particular, with the advent of the Bell System EADAS/ICUR (Engineering and Administrative Data Acquistion System/Individual Circuit Usage Recorder) system,[5] it was natural to ask whether the new individual trunk data available could be used to detect such "killer" trunks.*

This paper discusses the theoretical aspects of a class of killer-trunk detection algorithms that utilize the individual trunk traffic data available in EADAS/ICUR. These algorithms were designed for, and practical versions of them are presently implemented in, the ICAN† portion of the EADAS/ICUR system. We focus here, however, on the problem formulation, modeling, and analysis aspects of the algorithms without bringing in many of the diverse factors and trade-offs encountered in the actual implementation.

Because the holding time of a trunk affects the statistical properties of the trunk data in EADAS/ICUR, it is natural to formulate the killer-trunk detection problem as a problem in the testing of statistical hypotheses. In this context our modeling effort is basically an attempt to precisely define the state of a trunk (normal or killer) and expose the relevant underlying distributions. Well-known aspects of the theory of hypothesis testing are then applicable and immediately suggest a number of different tests. Sequential tests are naturally considered since the EADAS/ICUR data evolve sequentially in time. Questions about the robustness of the models, and the structure and performance of statistical tests, are addressed using standard analytic tools.

The material in this paper has been organized into six major sections, whose content we briefly describe. After considering the data available in EADAS/ICUR (Section II), we proceed to model a trunk (Section 3.1), motivate an appropriate set of statistical hypotheses suitable to our problem (Section 3.2), and briefly review several classical tests for deciding between statistical hypotheses (Section 3.3). With these preliminaries out of the way, we develop individual trunk algorithms based solely on individual trunk data. Proceeding in a heuristic manner, we use the individual trunk data to "derive" an *ad hoc* killer-trunk-detection

* The term killer trunk has been widely adopted in referring to a faulty switching-machine-accessible trunk in a group whose average holding time is substantially smaller than the average group holding time.

† Individual circuit analysis program—a software program that analyzes much of the EADAS/ICUR traffic data and maintains the system data base.

algorithm (Section 4.1). Although the insight gained in proceeding in a heuristic fashion is significant, we shift our emphasis in Section 4.2 and rigorously derive an optimal test statistic. It is interesting to find that the ad hoc statistic is essentially one of two symmetric statistics which comprises the optimal test statistic. The relationship between these individual trunk statistics is further explored in Section 4.3.

In Section V we factor grouping information (which essentially identifies all trunks common to a trunk group) into the picture, and develop detection algorithms tailored to trunks associated with the No. 5 crossbar switching machine. This development necessitates modeling the 5XB trunk-group selection procedure, and several results due to Forys and Messerli[2] are utilized. In Section VI we shift our discussion to the performance of the 5XB group algorithms, deriving approximate expressions for the mean statistic update and mean detection time in Sections 6.1 and 6.2, respectively. The paper concludes in Section 6.3 with an approximate analysis of the false-alarm probability of the 5XB group algorithms.

II. EADAS/ICUR DATA

The structure of a killer-trunk detection algorithm is largely dependent on the type of individual trunk measurements available.* In EADAS/ICUR, the raw (unobservable) data consists of the state of each trunk (busy or idle) every 100 or 200 seconds. Fortunately, the data accumulations available essentially summarize all the relevent information in the raw data.

2.1 Switch-count and transition data

The data available from the EADAS/ICUR system, which can be used to distinguish between normal and killer trunks, is obtained by sampling individual trunks at 100 or 200 second intervals. This data consists of periodic accumulations (typically hourly, two-hourly, or three-hourly) of both the Busy states, and the State transitions. The busy state accumulation is usually referred to as the switch count. For the 200-second sampling option with a one-hour accumulation period, the switch-count is an integer between 0 and 18. A state transition occurs whenever the state of a trunk (busy or idle) is different at two successive scans. For the 200-second sampling option with a one-hour accumulation period, the number of state transitions is an integer between 0 and 17.

If we denote the ith scan during an accumulation period in which m scans occur by x_i, and let 0 and 1 correspond to trunk idle and trunk

* Until very recently, almost all trunk-traffic measurements were obtained on a group rather than on an individual trunk basis.

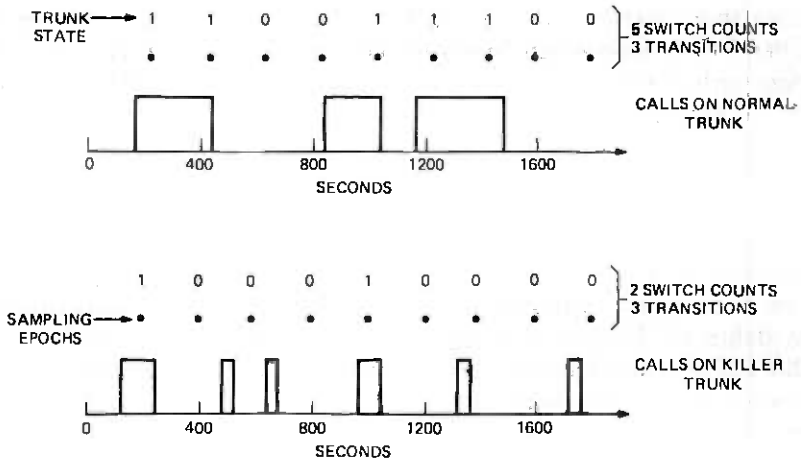

Fig. 1—ICUR data.

busy, respectively, the available data may be written

$$(i)\quad n(m) = \sum_{i=1}^{m} x_i \text{ (switch count)}$$

$$(ii)\quad t(m) = \sum_{i=2}^{m} |x_i - x_{i-1}| \text{ (state transitions).}$$

Thus, the raw (unavailable) data in the form of a binary sequence,

$$\mathbf{x}_m = (x_1, \cdots, x_m),$$

is compressed into the two statistics $n(m)$ and $t(m)$.

Because the holding time of a killer trunk is generally on the order of a few tens of seconds, it should have substantially more state transitions than a normal trunk, for a given switch count. Figure 1 illustrates the sampling process on both a normal and killer trunk.* For the purposes of this figure individual calls are represented by rectangles, call durations correspond to the width of the rectangles, and a half-hour accumulation period with the 200-second sampling option is used.

We note in passing that for the 200-second sampling option, very little information is lost by "compressing" the raw data $\mathbf{x}_m = (x_1, \cdots, x_m)$ into the two statistics $n(m)$ and $t(m)$. Thus, normal conversation lengths tend to be in the vicinity of 3 to 4 minutes and, hence, with the 200-second sampling option, we expect that only adjacent samples are significantly

* The realizations shown in Fig. 1 are more or less typical for a 5XB trunk group with a mean group holding time of approximately 4 minutes operating at about 40-percent occupancy, and having a killer trunk with a mean holding time of approximately 1 minute.

correlated. If $\{x_i\}_{i=1}^{m}$ is Markovian, then $[n(m),t(m)]$ is almost a sufficient statistic ($[n(m),t(m),x_1,x_m]$ is sufficient) for $\mathbf{x}_m$ (see Section 4.2). Note also that for a trunk in the killer state, succesive samples should be essentially independent (for both sampling options).

2.2 Grouping Information

In addition to the switch-count and transition data available from the EADAS/ICUR system, we are also able to utilize a system map to identify (*i*) all trunks common to a trunk group, and (*ii*) the trunk-selection procedure* associated with the trunk group. It turns out that using this grouping information,† in addition to the switch-count and transition data, enhances the detection potential considerably.

Thus, we divide the class of algorithms into two types according to whether or not grouping information is utilized. The first type, which uses only switch-count and transition data is referred to as an individual trunk algorithm. These individual trunk algorithms are applicable to all trunks—including two-way trunks—independent of the type of switching machine they are associated with. They do however assume knowledge of the trunks nominal holding time. The second type of algorithm uses the grouping information in addition to the switch-count and transition data and is referred to as a group algorithm. Group algorithms are "tailored" to a specific kind of trunk-selection procedure and, hence, apply to trunk groups associated with specific switching machines. For the purposes of this paper, the trunk-selection procedure considered is random selection of idle trunks. This procedure models the selection procedure of trunk groups associated with the 5XB switching machine. Group algorithms generally apply only to one-way trunk groups.

III. PRELIMINARY CONSIDERATIONS

In attempting to quantify the intuitive notion that a killer trunk will exhibit more transitions than a normal trunk (see Fig. 1), for a given switch count, it is natural to consider the transition probabilities:

$$P_{1,0} = P\{x_{t+\tau} = 0/x_t = 1\}$$

and

$$P_{0,1} = P\{x_{t+\tau} = 1/x_t = 0\},$$

* The map in EADAS/ICUR indicates the type of switching machine that the trunk group is associated with, and this allows us to model the trunk-selection procedures (see Ref. 2).

† We will consistently use "grouping information" to refer to both the identification of all trunks common to the group and the trunk-selection procedure associated with the group.

where x_t denotes the state of a trunk at epoch t and τ denotes the sampling interval. Of course, to evaluate these transition probabilities, we must be concrete about how we model a trunk.

Before tying ourselves down to any specific model, however, it is useful to view these conditional probabilities in a canonical form. Thus, suppose we begin by assuming only that the binary valued process x_t is stationary. We have then the following simple result:

Lemma 1: *Let x_t be a binary valued, stationary random process and let ρ and $R(\cdot)$ denote its mean and covariance function, respectively.* Then,

$$(i)\quad P_{1,0}(\rho,\tau) = (1-\rho)\left\{1 - \frac{R(\tau)}{R(0)}\right\} \tag{1}$$

$$(ii)\quad \rho P_{1,0}(\rho,\tau) = (1-\rho)P_{0,1}(\rho,\tau). \tag{2}$$

Proof: Part (i) is a consequence of the definition of $R(\cdot)$. That is,

$$\begin{aligned} R(\tau) &= E(x_t x_{t+\tau}) - \rho^2 \\ &= P(x_t = 1, x_{t+\tau} = 1) - \rho^2, \end{aligned}$$

where

$$\rho = E(x_t) = P(x_t = 1).$$

Part (ii) follows from the two identities:

$$\rho = \rho P_{1,1}(\rho,\tau) + (1-\rho)P_{0,1}(\rho,\tau)$$

and

$$1 = P_{1,1}(\rho,\tau) + P_{1,0}(\rho,\tau).$$

A consequence of this result is that uncorrelatedness and independence are equivalent:

Corollary 1: *For the process in lemma 1, $x_t, x_{t+\tau}$ are independent if and only if $R(\tau) = 0$.*

Note that the dependence of $R(\cdot)$ on ρ has been suppressed for convenience.

3.1 Modeling an individual trunk

A particularly simple way to model a trunk is as the server in a single server loss* system with a Poisson arrival process and an exponential service time distribution. This model is commonly denoted by M/M/1-loss.[6] Let x_t denote the state of the server:

* In a loss system, customers who are blocked depart without waiting.

$$x_t = \begin{cases} 1 & \text{if server is busy at epoch } t \\ 0 & \text{if server is idle at epoch } t \end{cases}$$

and $R(\cdot)$ the covariance function of x_t. It is easily shown[7] that for the M/M/1-loss system,

$$R(\tau) = R(0) \exp\{-(\lambda + \mu)\tau\},$$

where λ and μ are the mean arrival and service rates, respectively. Thus, $P_{1,0}(\rho,\tau)$ may be written as:

$$P_{1,0}(\rho,\tau) = (1 - p)\left\{1 - \exp\left(\frac{-\mu\tau}{1-\rho}\right)\right\}, \tag{3}$$

where the trunk occupancy ρ is equal to $\lambda/(\lambda + \mu)$. Throughout this paper we will be concerned with $\tau = 100$ or 200 seconds and a nominal holding time $1/\mu$ in the vicinity of 3 minutes. The mean holding time of a killer trunk $1/\mu^*$ will always be expressed as $1/r\mu$ with r typically in the range 5 to 15. Thus, if we denote $\mu\tau$ by S, we may write the transition probability $P\{x_{t+\tau} = 0/x_t = 1\}$ for a trunk with mean occupancy ρ as

$$P_{1,0}(\rho,r) = (1 - \rho)\left[1 - \exp\left(\frac{-rS}{1-\rho}\right)\right], \tag{4}$$

where $r = 1$ corresponds to a normal trunk. (Since $P_{1,0} = 1 - \rho$ implies that $x_\tau, x_{2\tau}, \cdots$ are independent, we will assume independence for r sufficiently large in subsequent sections.)

Figure 2 is a plot of $P_{1,0}$ vs ρ corresponding to $S = 10/9$ (200-second sampling and a 3-minute mean holding time) for several values of r. $P_{1,0}$ is essentially equal to $1 - \rho$ for $r \geq 5$. Figure 3 is a similar plot of $P_{1,0}$ vs ρ corresponding to 100-second sampling and a 3-minute mean holding time ($S = 5/9$). In this figure $P_{1,0}$ is essentially equal to $1 - \rho$ for $r \geq 7.5$.

Before putting too much emphasis on the transition probabilities in Figs. 2 and 3, it is prudent to consider the effect of factoring more realistic assumptions into the single server loss model. Thus, while the Poisson arrival process assumption is probably a reasonable assumption for a trunk in a 5XB trunk group (random selection of idle servers), it poorly models the overflow nature of the traffic offered to trunks in a 1XB/XBT trunk group.† In the latter case, it is more appropriate to model the input stream to a trunk as a peaked process.[6] Figure 4 is a plot of $P_{1,0}$ vs ρ parameterized by the peakedness (z) of the input stream. This figure is based on an expression for $P_{1,0}$ derived for a GI/M/1-loss model‡ with

† The trunk-selection procedure for 1XB and XBT trunk groups is essentially a two-sided ordered hunt.[2]

‡ GI/M/1-loss denotes a single server loss system with a renewal process input stream (GI) and an exponential (M) service time distribution.

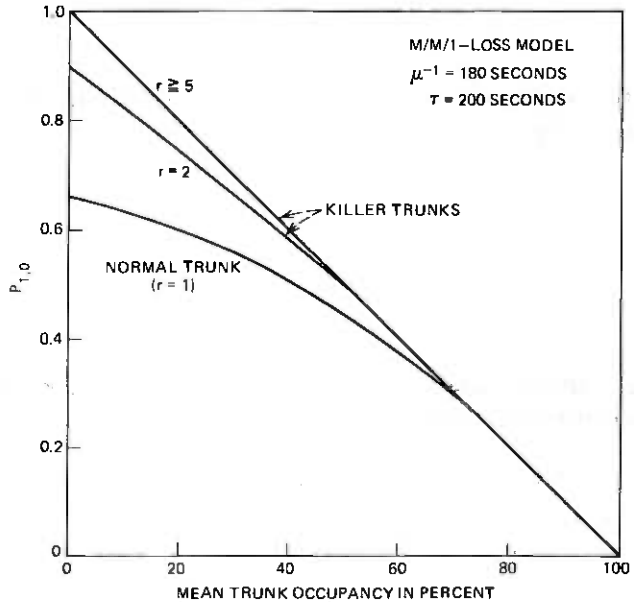

Fig. 2— 1 → 0 transition probability for the 200-second sampling option.

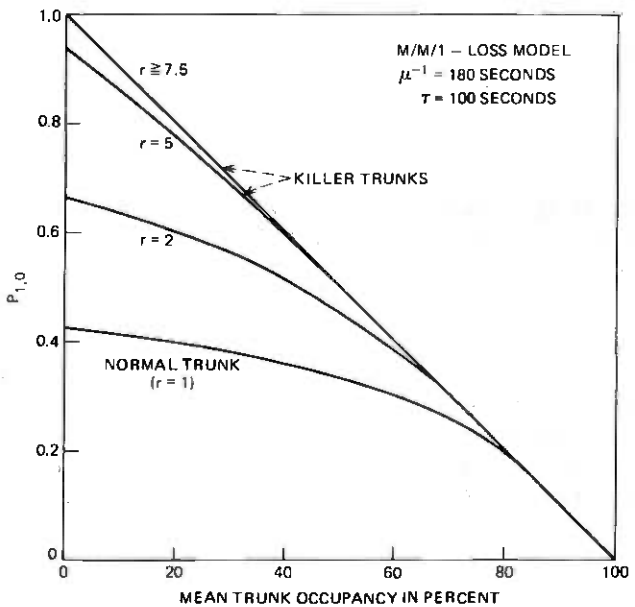

Fig. 3— 1 → 0 transition probability for the 100-second sampling option.

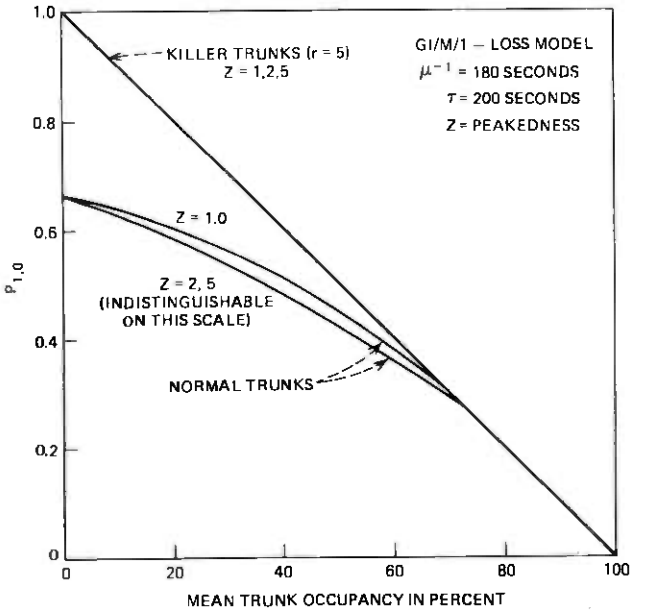

Fig. 4— $1 \rightarrow 0$ transition probability for the GI/M/1-loss model with a switched Poisson arrival process—200-second sampling option.

a switched Poisson input stream (commonly used to model overflow traffic).[8] Appendix A contains several details on the model and derivation. It is clear that the effect of peaked traffic on the transition probability is very small ($z = 1$ corresponds to a Poisson stream).

Recent data[9] indicates that the service time distribution of a normal trunk has a coefficient of variation significantly greater than 1 (the exponential case). Thus, in Appendix A we derive the covariance function of the server process x_t for an M/G/1-loss[‡] model with a mixed exponential type of service distribution. Figure 5 is a plot of $P_{1,0}$ vs ρ parameterized by the coefficient of variation of the mixed exponential service distribution. We see that increasing the coefficient of variation has a noticeable effect on the transition probabilities, but the effect is to increase the discrimination between the normal and killer-trunk transition probabilities.

Thus, it would appear that the transition probabilities based on the M/M/1-loss model are reasonably robust to perturbations in the trunk model. In addition, one suspects that using these transition probabilities in a detection scheme, which exploits the basic differences between killer and normal trunk transitions, might lead to a conservative design.

[‡] M/G/1-loss denotes a single server loss system with a Poisson (M) input stream and a general (G) service time distribution.

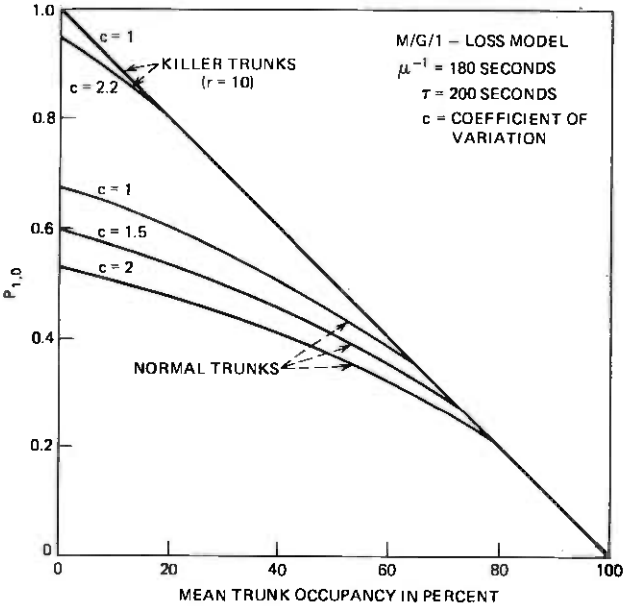

Fig. 5— $1 \rightarrow 0$ transition probability for the M/G/1-loss model with a mixed exponential service time distribution—200-second sampling option.

3.2 Testing statistical hypotheses—a basic idea

Suppose a trunk has constant mean occupancy ρ ($Ex_t = \rho$) and we observe it for h seconds during which n switch counts and t_{10} $1 \rightarrow 0$ state transitions accumulate. We may write

$$n = \sum_{k=1}^{m} x_{k\tau}$$

and

$$t_{10} = \sum_{k=2}^{m} [x_{k\tau} - x_{(k-1)\tau}]^{-},$$

where

$$m = \frac{h}{\tau} \text{ and } z^{-} = \begin{cases} 0 & z \geqq 0 \\ 1 & z < 0 \end{cases}.$$

Hence, we have

$$E(n) = \rho \left(\frac{h}{\tau}\right) \tag{5a}$$

and

$$E(t_{10}) = (m-1)E([x_{k\tau} - x_{(k-1)\tau}]^{-}) = \left(\frac{h}{\tau} - 1\right) \rho P_{1,0}. \tag{5b}$$

Thus, for example, a trunk at 20-percent occupancy (sampled at the 100-second rate) would accumulate 108 switch counts, on the average, in 15 hours. The mean number of $1 \rightarrow 0$ state transitions in this time interval corresponding to a normal trunk is 43, but the corresponding mean number of $1 \rightarrow 0$ state transitions for a killer trunk (during the same time interval) is 86. (Referring to Fig. 3, we see that $P_{1,0}(\rho = 0.2)$ is 0.40 and 0.80 for a normal and killer ($r \geqq 7.5$) trunk, respectively.)

With this example in mind, it is natural to consider a detection scheme of the following type:

(*i*) Wait until we accumulate n_0 switch counts on a trunk.

(*ii*) When $n = n_0$, compare the accumulated $1 \rightarrow 0$ state transitions t_{10} to some threshold T_t.

(*iii*) If $t_{10} > T_t$, decide that the trunk is a killer, otherwise decide that the trunk is normal. [t_{10} is not directly available (see Section 2.1), but it may be estimated by $t/2$.]

If the trunk occupancy were known and fixed, this scheme would appear to be very reasonable. The analogy to the usual scheme suggested for deciding between a fair and a biased coin is clear: n_0 is the number of (hopefully) independent experiments (analogous to the number of coin tosses), with each experiment having just two possible outcomes: the scan which follows the switch count is either 0 or 1. Thus, each switch count is associated with a $1 \rightarrow 0$ state transition ("heads") or a $1 \rightarrow 1$ state transition ("tails").

From the point of view of statistical hypothesis testing, we are thinking of two underlying states:

$$\text{Null hypothesis } H_0\text{: } P_{1,0}(\rho) = P \text{ trunk normal}$$

$$\text{Alternate hypothesis } H_1\text{: } P_{1,0}(\rho) = P^* \text{ trunk killer}$$

Thus, our intuition suggests that a threshold test of the type sketched above is natural for distinguishing between H_0 and H_1. We will see (Section 4.1) that a (nonoptimal) test of this form arises naturally from pursuing the coin tossing analogy further.

3.3 Problem formulation—sharpening the focus

To simplify matters, assume to begin with that

(*i*) The nominal mean holding time $1/\mu$ is known.

(*ii*) The trunk occupancy ρ is known and constant.

(*iii*) The switch-count and transition data accumulations $n(m)$ and $t(m)$ are continuously available (scan by scan).

With these assumptions, it is an easy matter to conceptually describe

the "optimum" scheme for deciding between the two simple hypotheses,

$$H_0\text{: Trunk normal (mean holding time } 1/\mu)$$

$$H_1\text{: Trunk killer (mean holding time } 1/r\mu),$$

with it understood that "trunk" refers to one of the specific models described in Section 3.1 (for concreteness assume the M/M/1-loss model).

Let $\mathbf{x}_m = (x_1, \cdots, x_m)$ be the sequence of trunk states up to and including the mth scan, and let the available data be (as before)

$$n(m) = \sum_{i=1}^{m} x_i,\ t(m) = \sum_{i=2}^{m} |x_i - x_{i-1}|.$$

Let

$$P_{im}(n,t) = P(n(m) = n, t(m) = t/H_i)\ i = 0 \text{ or } 1$$

and let

$$\ell_m(n,t) = \frac{P_{1m}(n,t)}{P_{0m}(n,t)}.$$

The joint probability distributions $P_{im}(n,t)$, $i = 0,1$, are well defined for any specific trunk model, but they may be nontrivial to derive. $\ell_m(.,.)$ viewed as a function of the vector random variable $[n(m),t(m)]$ is referred to as the likelihood ratio statistic and plays a central role in the theory of statistical hypothesis testing. More specifically, the optimum test (in a variety of senses) for deciding between two simple hypotheses involves suitably comparing ℓ_m to a threshold (or thresholds) in order to make a decision.

We briefly review two optimum tests, the Neyman-Pearson (fixed sample) test[10] and Wald's sequential probability ratio test (SPRT), using notation appropriate to our (discrete) problem.

3.3.1 The Neyman-Pearson test

Suppose α and β denote the type 1 and 2 errors* of the test,

$$\text{Choose } H_1 \text{ if } \ell_m \geqq T$$

$$\text{Choose } H_0 \text{ if } \ell_m < T,$$

and suppose α' and β' denote the type 1 and 2 errors of any other test (requiring m samples) for deciding between H_0 and H_1. Neyman and

* The type 1 and 2 errors, α and β, are often referred to as the probability of false alarm and the probability of miss, respectively (α = probability of choosing H_1 given H_0 is the true state, β = probability of choosing H_0, given H_1 is the true state.)

Pearson's classical result is: if $\alpha' \leqq \alpha$, then $\beta' \geqq \beta$. Thus, of all tests requiring m samples and having a false-alarm probability not exceeding α, the likelihood ratio test achieves the minimum probability of miss (maximum probability of detection). Since $\alpha = P(\ell_m > T/H_0)$, choosing a sample size m and threshold T to achieve $\alpha \leqq \alpha_0$ requires knowledge of the (conditional) distribution of ℓ_m. Similarly, having chosen m and T, calculating $\beta = P(\ell_m < T/H_1)$ requires the distribution of ℓ_m (conditioned on H_1). Note also that with such a fixed sample test, we decide in advance to accumulate exactly m samples before making a decision. In many contexts, data accumulates sequentially in time, and rigidly requiring m samples—independent of the particular realization that is unfolding—is not an optimal strategy.

3.3.2 Wald's sequential probability ratio test

Using Wald's SPRT,[10,11] we continue to update ℓ_k, $k = 1,2,\cdots$ and defer a decision as long as $\ell_k \epsilon (T_0, T_1)$. We make a decision the *first* time ℓ_k falls outside the interval (T_0, T_1). Thus,

$$\text{if } T_0 < \ell_k < T_1 \quad k = 1,2,\cdots,m-1$$

$$\text{and } \ell_m \notin (T_0, T_1),$$

$$\text{then choose } H_1 \text{ if } \ell_m \geqq T_1$$

$$\text{and choose } H_0 \text{ if } \ell_m \leqq T_0.$$

Clearly the stopping time m of the SPRT is a random variable, and the mean of m (given either hypothesis) is a measure of the time it takes to reach a decision. (Under a wide variety of circumstances, the SPRT terminates with probability 1.) Let $E_i(m)$ ($i = 0$ or 1) denote the mean stopping time, given that hypothesis i is in effect. Given a SPRT with type 1 and 2 errors α and β, and with mean stopping times $E_0(m)$ and $E_1(m)$, consider any other test (*sequential or not*) with type 1 and 2 errors α' and β', and with mean stopping times $E_0'(m)$ and $E_1'(m)$. The SPRT has the following optimal character[10]

$$\text{if } \alpha' \leqq \alpha \text{ and } \beta' \leqq \beta,$$

$$\text{then } E_0'(m) \geqq E_0(m) \text{ and } E_1'(m) \geqq E_1(m).$$

Thus a SPRT is superior to a fixed sample test, if both tests have the same type 1 and 2 errors, in the sense that on the average it reaches a decision more quickly (under either hypotheses).

In sharp distinction to the fixed sample test, the thresholds T_0 and T_1 required to approximately achieve specified type 1 and 2 errors are trivially determined.[11] On the other hand, even determining the mean and variance of the stopping time is often a difficult chore.

In Section 4.2, we explicitly calculate the SPRT* for the simple hypothesis testing problem described at the beginning of this section. Before looking at this optimum test, however, we describe an ad hoc algorithm which is very robust and consequently attractive from a practical point of view.

IV. INDIVIDUAL TRUNK ALGORITHMS

A basic underlying assumption in this section is that the normal mean holding time of a trunk is known. Thus, if the algorithms in this section are designed relative to a normal mean holding time of 3 minutes, they will not discriminate between normal trunks having a mean holding time in the vicinity of 40 seconds,† and an actual killer trunk with the same mean holding time—both of these trunks will be detected as killer trunks.

The rationale for studying this type of detection problem is two-fold: from the practical point of view the simplicity of implementation and general applicability‡ of these algorithms is attractive, and EADAS/ICUR can flag trunk groups which should not be studied by the killer trunk-detection algorithms (thus preventing false alarms on normal short-holding-time trunks). From the theoretical point of view, it was natural to consider this problem before factoring group information into the picture.

Another modeling assumption used in this section (as well as in subsequent ones) is that the arrival process is stationary within data accumulation intervals, but the mean arrival rate may change arbitrarily from one accumulation period to another. Since we use equilibrium analysis (e.g., in calculating $P_{1,0}$) we assume, in effect, that equilibrium is achieved instantaneously.

4.1 An ad hoc algorithm

The essential idea of the test suggested in Section 3.2, is to decide on the state of a trunk by comparing the number of $1 \rightarrow 0$ state transitions (t_{10}) to some threshold T_f, *conditional* on having accumulated a fixed number of switch-counts. We heuristically* proceed to derive such a test, using a standard likelihood ratio formulation, and explicitly take into account the time-variability of traffic.

Let $\mathbf{x}_m = (x_1, \cdots, x_m)$ correspond to the (unobservable) binary se-

* Based on the M/M/1-loss model for a trunk.

† Trunks in special-purpose trunk groups (credit checking, weather, etc.) will typically have mean holding times in the vicinity of 40 seconds.

‡ The individual trunk algorithms can be used to test *any* trunk—regardless of the type of switching machine the trunk is associated with.

* The distributional assumptions made in this section are intuitively motivated, but cannot be rigorously justified. We examine these assumptions carefully in Section 4.3.

quence of trunk states during an accumulation period in which m scans occur. Let $t_{10}(m)$ and $n(m)$ be the number of $1 \rightarrow 0$ state transitions and switch counts associated with $\mathbf{x}_m$. Denote the conditional probability

$$P(t_{10}(m) = t/n(m) = n)$$

for a normal and killer trunk by $P(t/n)$ and $P^*(t/n)$, respectively. These conditional distributions depend, of course, on the trunk's occupancy and on the particular trunk model we have in mind. [$P^*(t/n)$ also depends on the killer parameter r.] However, for the purposes of the heuristic development of this section, we do not precisely define which trunk model we have in mind.

Since each switch count is associated with either a $1 \rightarrow 0$ or a $1 \rightarrow 1$ state transition with probabilities $P_{1,0}$ and $P_{1,1} = 1 - P_{1,0}$, respectively for a normal trunk, and since we expect successive transition events on a trunk to be essentially independent,* it seems reasonable to assume that $P[t_{10}(m) = t/n(m) = n]$ for a normal trunk is binomially distributed with parameters n and $P_{1,0}$. This same argument applies to a killer trunk. Denote the binomial distribution with parameters n and p by $b(k;n,p)$ $k = 0, \cdots, n$, where

$$b(k;n,p) = \binom{n}{k} p^k(1-p)^{n-k}.$$

Thus, we may think of a trunk with occupancy ρ during an accumulation period as having a conditional distribution

$$P[t_{10}(m) = t/n(m) = n] = b[t;n,P_{1,0}(\rho,r)], \tag{6}$$

with $r = 1$ and $r = r_0$ corresponding to the normal and killer states of the trunk. (Recall that $P_{1,0}(\rho,r)$ is essentially independent of r for $r \geq r_0$ with $r_0 = 7.5$ and 5.0 for 100- and 200-second sampling, respectively.) With these assumptions, we may think of testing the two simple hypotheses:

$$H_0\colon P(t/n) = b(t;n,P_{1,0}) \quad P_{1,0} = P_{1,0}(\rho,1)$$

$$H_1\colon P(t/n) = b(t;n,P^*_{1,0}) \quad P^*_{1,0} = P_{1,0}(\rho,r_0).$$

If the $1 \rightarrow 0$ transition and switch-count accumulations for two successive and contiguous accumulation periods are (t_1,n_1) and (t_2,n_2) respectively, we assume that

$$P(t_1,t_2/n_1,n_2) = P(t_1/n_1)P(t_2/n_2).$$

The idea here is that the only dependence between the two successive

* The idea is that if significant correlation extends only one or two scans back, then successive transition events (events "triggered" by switch counts) should be essentially independent.

bit streams $\boldsymbol{x}_m^1 = (x_1,\cdots,x_m)$ and $\boldsymbol{x}_m^2 = (x_{m+1},\cdots,x_{2m})$ is essentially due to the dependence between x_m and x_{m+1}.

Thus, if we denote the transition and switch-court accumulations for the ith accumulation period (in which m_i scans occur) by $[t_{10}(m_i),n(m_i)]$ during which the trunk has occupancy ρ_i, we have

$$P[t_{10}(m_1) = t_1,\cdots,t_{10}(m_k) = t_k \,|\, n(m_1) = n_1,\cdots,n(m_k) = n_k] = \prod_{i=1}^{k} b[t_i;n_i,P_{1,0}(\rho_i,r)], \quad (7)$$

where ρ_i $i = 1,\cdots,k$ are the occupancies for the k accumulation periods. If $\boldsymbol{t}_k = (t_1,\cdots,t_k)$ and $\boldsymbol{n}_k = (n_1,\cdots,n_k)$ consider the likelihood ratio:

$$\ell(\boldsymbol{t}_k/\boldsymbol{n}_k) = \prod_{i=1}^{k} \frac{b[t_i;n_i,P_{1,0}(\rho_i,r_0)]}{b[t_i;n_i,P_{1,0}(\rho_i,1)]}. \quad (8)$$

Denote the log likelihood ratio† log $\ell(\boldsymbol{t}_k/\boldsymbol{n}_k)$ by $\hat{\ell}(\boldsymbol{t}_k/\boldsymbol{n}_k)$ and note that

$$\hat{\ell}(\boldsymbol{t}_k/\boldsymbol{n}_k) = \sum_{i=1}^{k} \hat{\ell}(t_i/n_i),$$

where

$$\hat{\ell}(t_i/n_i) = \log \frac{b[t_i;n_i,P_{1,0}(\rho_i,r_0)]}{b[t_i;n_i,P_{1,0}(\rho_i,1)]}.$$

The expression $\hat{\ell}(t_i/n_i)$ can be written as

$$\hat{\ell}(t_i/n_i) = \alpha(\rho_i)t_i - \alpha(\rho_i)n_i$$

with

$$\alpha(\rho) = \log \frac{1 - P_{1,0}(\rho,1)}{1 - P_{1,0}(\rho,r_0)} \quad (9a)$$

and

$$\alpha(\rho) = \alpha(p) + \log \frac{P_{1,0}(\rho,r_0)}{P_{1,0}(\rho,1)}. \quad (9b)$$

Thus, we have

$$\hat{\ell}(\boldsymbol{t}_k/\boldsymbol{n}_k) = \sum_{i=1}^{k} [\alpha(\rho_i)t_i - \alpha(\rho_i)n_i]. \quad (9c)$$

Unfortunately, the occupancy in the ith accumulation period (ρ_i) is unknown and hence equation (9c) cannot be used as a test statistic. One obvious "fix" is to estimate ρ_i by $\hat{\rho}_i = n_i/m_i$, where n_i and m_i are the

† $l > T$ iff $g(\ell) > g(T)$ if g is monotone increasing, so the tests $\ell > T$ and $g(\ell) > g(T)$ are equivalent.

switch count and the number of scans, respectively, during the ith accumulation period. In stationary traffic, the estimate

$$\tilde{\rho}_i = \sum_{j=1}^{i} n_j \Big/ \sum_{j=1}^{i} m_j$$

would be used [$\sigma(\tilde{\rho}_i) \doteq 1/\sqrt{i}\ \sigma(\hat{\rho}_i)$ if $m_j = m$ for all j].

Corresponding to the sequence of accumulations, (t_1,n_i,m_i) $i = 1,2,\cdots$ we define r_i and R_i, $i = 1,2,\cdots$ by

$$r_i = \alpha(\hat{\rho}_i)t_i - \alpha(\hat{\rho}_i)n_i,\ \hat{\rho}_i = n_i/m_i \tag{10a}$$

and

$$R_i = R_{i-1} + r_i \text{ with } R_0 = 0. \tag{10b}$$

Thus, we arrive at the sequential test:

(i) Compute R_i, $i = 1,2,\cdots$ and defer making a decision as long as $T_0 < R_i < T_1$.

(ii) If $i = k$ corresponds to the first accumulation period for which $R_i \notin (T_0,T_1)$, then

$$R_k \leqq T_0 \Rightarrow \text{Trunk normal}$$

$$R_k \geqq T_1 \Rightarrow \text{Trunk killer.}$$

If we ignore the fact that we are estimating ρ_i by $\hat{\rho}_i$, and by assuming that the various assumptions made are valid (see Section 4.3), we identify the above test as Wald's SPRT and as such T_0 and T_1 can be calculated as follows:[11] to approximately achieve type 1 and 2 errors, α and β, respectively, $\alpha + \beta < 1$, choose

$$T_0 = \log\left(\frac{\beta}{1-\alpha}\right) \tag{10c}$$

and

$$T_1 = \log\left(\frac{1-\beta}{\alpha}\right). \tag{10d}$$

Throughout this section, we have assumed that the $1 \to 0$ transitions (t_{10}) are available when, in fact, only the total transitions (t) are available. It should he clear that t_{10} can differ from $t/2$ by at most $\pm\frac{1}{2}$. To be precise, let $t_{10}(m)$, $t_{01}(m)$ be the number of $1 \to 0$ and $0 \to 1$ state transitions corresponding to a bit stream $\mathbf{x}_m = (x_1, \cdots, x_m)$. If $n(m)$ is the switch count corresponding to $\mathbf{x}_m$, then we have

$$n(m) = t_{10}(m) + t_{11}(m) + x_m \tag{11a}$$

and

$$n(m) = t_{01}(m) + t_{11}(m) + x_1, \tag{11b}$$

where $t_{11}(m)$ is the number of $1 \rightarrow 1$ state transitions. Therefore

$$t_{10}(m) + x_m = t_{01}(m) + x_1,$$

which together with $t(m) = t_{01}(m) + t_{10}(m)$ yields

$$t_{10}(m) = \frac{1}{2}t(m) + \left(\frac{x_1 - x_m}{2}\right) \tag{12a}$$

and

$$t_{01}(m) = \frac{1}{2}t(m) - \left(\frac{x_1 - x_m}{2}\right). \tag{12b}$$

Thus, we can write the statistic update (eq. 10a) as

$$\alpha\frac{t}{2} - \text{an} + \left(\frac{x_1 - x_m}{2}\right)\alpha.$$

[It is easy to show that $E[(x_1 - x_m)\alpha(\hat{\rho})] = 0$.]

We conclude this section with an interpretation of the statistic update. Rewriting the statistic update as

$$r = (\alpha - a)t_{10} - a(n - t_{10})$$

and using eq. (11a), we obtain

$$r = (\alpha - a)t_{10} - at_{11} - ax_m. \tag{13}$$

Now, from eqs. (9a) and (9b), it is clear that $\alpha > a > 0$. Thus, each $1 \rightarrow 0$ transition is weighted positively (evidence of a killer) while each $1 \rightarrow 1$ transition is weighted negatively (evidence of a normal trunk). This is an intuitive explanation of the fact that the random walk (eq. (10b))

$$R_k = R_{k-1} + r_k$$

has a positive drift if the trunk is a killer and a negative drift if the trunk is normal.

The fact that the update assigns a negative weight $(-a)$ whenever the last bit (x_m) is 1 uncovers a modeling deficiency. Recall that in eq. (6) we assumed

$$P(t_{10}(m) = t/n(m) = n) = b(t;n,P_{10}),$$

even though $x_m = 1$ can not contribute to an observable $1 \rightarrow 0$ transition. In this way we effectively modeled in a bias towards making "trunk normal" decisions. We can easily correct eq. (6) by conditioning on whether $x_m = 0$ or 1, obtaining:

$$P(t_{10}(m) = t/n(m) = n) = (1 - \rho)b(t;n,P_{10}) + \rho b(t;n - 1,P_{10}).$$

Now, proceeding as before in formulating the log likelihood ratio yields

a statistic $\tilde{R}_k$, where

$$\tilde{R}_k = \tilde{R}_{k-1} + \tilde{r}_k$$

and

$$\tilde{r}_k = r_k + q_k,$$

where r_k is defined by eq. (10a),

$$q_k = \log \frac{1 + \left(\frac{\hat{\rho}_k}{1-\hat{\rho}_k}\right)\left(1 - \frac{t_k}{n_k}\right)\left(\frac{1}{1 - P_{1,0}(\hat{\rho}_k, r_0)}\right)}{1 + \left(\frac{\hat{\rho}_k}{1-\hat{\rho}_k}\right)\left(1 - \frac{t_k}{n_k}\right)\left(\frac{1}{1 - P_{1,0}(\hat{\rho}_k, 1)}\right)} \tag{14}$$

and $\hat{\rho}_k$, t_k, and n_k are the trunks occupancy estimate, $1 \rightarrow 0$ state transitions, and switch count, respectively, during the kth accumulation period.

Thus, we obtain our original test statistic with the correction term q_k added on. Note that $q_k \geqq 0$, $q_k \rightarrow 0$ as $\hat{\rho}_k \rightarrow 0$, and $q_k \rightarrow$ a as $\hat{\rho}_k \rightarrow 1$, which is just the type of behavior expected, to offset the bias term in r_k.

Having heuristically developed an ad hoc sequential algorithm that is intuitively appealing and easily implementable, it is natural to ask: how does it compare to the *optimum* sequential algorithm? In the following section, we rigorously develop an optimum sequential test.

4.2 An optimal algorithm

Consider the two simple hypotheses:
H_0: Trunk normal (mean occupancy ρ, mean holding time $1/\mu$)
H_1: Trunk killer (mean occupancy ρ, mean holding time $1/r_0\mu$).
The optimum test for deciding between the two hypotheses—in the sense of minimizing the mean decision time—for given type 1 and 2 errors, is Wald's SPRT (see Section 3.3), and it is based on the likelihood ratio statistic $\ell_m(t,n)$ given by

$$\ell_m(t,n) = \frac{P^*(t(m) = t, n(m) = n)}{P(t(m) = t, n(m) = n)}. \tag{15}$$

Thus, it is clear that the ad hoc test described in Section 4.1 is not optimal, based as it is on an assumed conditional distribution,

$$P(t_{10}(m) = t/n(m) = n).$$

Before proceeding to study eq. (15), we must define the trunk model precisely. In the developments that follow, we model a trunk as the server in an M/M/1-loss system (see Section 3.1). The model implies that the sequence of trunk states x_t, $t = k\tau$, $k = 1,2,\cdots$ is Markovian. Note that

although this appears to be a reasonable model for a normal trunk with 200-second sampling, it ignores the conditional dependence "2 samples back," which is more important for 100-second sampling—e.g., x_t given $x_{t-\tau}$ is independent of $x_{t-2\tau}$ for the M/M/1-loss model. Taking this dependence into account in a trunk model would not be useful however, since the data needed to implement dependence "two scans back," is not available.

Since we are modeling the sequence of trunk states as a binary valued Markov process $x_{k\tau}$, k = 1,2,···,in equilibrium,it is clear that this process is characterized by $\theta = (P_{1,0},P_{0,1})$, where $P_{1,0}$ and $P_{0,1}$ are the transition probabilities

$$P(x_{t+\tau} = 0/x_t = 1) \text{ and } P(x_{t+\tau} = 1/x_t = 0),$$

respectively. (In general, a binary valued Markov process $x_{k\tau}$, $k = 1,2,\cdots$ in equilibrium, can be characterized by any two of the three quantities ρ, $P_{1,0},P_{0,1}$. For our special Markov process (based on the M/M/1-loss model), both $P_{1,0}$ and $P_{0,1}$ and hence the process itself is determined by ρ alone.) Now having observed any m-tuple of the samples, which we denote by $\mathbf{x}_m = (x_1,\cdots,x_m)$, it is trivial to show that the statistic

$$\boldsymbol{T}(\mathbf{x}_m) = (t(m),n(m),x_1,x_m)$$

is a sufficient statistic for θ. Thus, except for the initial and terminal states (x_1 and x_m), the transition and switch-count accumulations summarize all the "relevant information" in $\mathbf{x}_m$.

Our hypothesis-testing problem can now be formulated as follows: $x_1,x_2,\cdots$ is a binary-valued Markov chain in equilibrium with parameter $\theta = (P_{0,1},P_{1,0})$ or $\theta^* = (P^*_{0,1},P^*_{1,0})$. That is, our two states are

$$H_0\text{: } \{x_i\} \text{ Markovian, characterized by } \theta = (P_{0,1},P_{1,0})$$

$$H_1\text{: } \{x_i\} \text{ Markovian, characterized by } \theta^* = (P^*_{0,1},P^*_{1,0}).$$

Now, because $(t(m),n(m),x_1,x_m)$ is a sufficient statistic for θ, we know that the likelihood-test statistic based on the raw (unobservable) data $\mathbf{x}_m = (x_1,\cdots,x_m)$ will be expressible in terms of $t(m)$, $n(m)$, x_1 and x_m only. Thus, instead of studying eq. (15), we proceed (for simplicity) to study the likelihood-ratio statistic:

$$\hat{\ell}(\mathbf{x}_m) = \log \frac{P^*(\mathbf{x}_m)}{P(\mathbf{x}_m)}. \tag{16}$$

In Appendix B we study $\hat{\ell}_m(t,n)$ and find that it differs from $\hat{\ell}(\mathbf{x}_m)$ only in an end-effect term. In $\hat{\ell}(\mathbf{x}_m)$ this term depends on x_1 and x_m, whereas in $\hat{\ell}_m(t,n)$ the corresponding term is a function of t and n.

Since

$$P(\mathbf{x}_m) = P(x_1) \prod_{i=2}^{m} P(x_i/x_{i-1}),$$

we may write

$$P(\mathbf{x}_m) = P(x_1)\, P_{10}^{t10}\, P_{11}^{t11}\, P_{01}^{t01}\, P_{00}^{t00}$$

so

$$\hat{\ell}(\mathbf{x}_m) = \log \frac{P^*(x_1)}{P(x_1)} + t_{10} \log \frac{P_{10}^*}{P_{10}} + t_{11} \log \left(\frac{1 - P_{10}^*}{1 - P_{10}}\right) + t_{01} \log \frac{P_{01}^*}{P_{01}} + t_{00} \log \left(\frac{1 - P_{01}^*}{1 - P_{01}}\right). \quad (17)$$

Note that a trunk with mean occupancy ρ is busy and idle with probability ρ and $1 - \rho$ respectively, independently of the state it is in (normal or killer). Thus,

$$\log \frac{P^*(x_1)}{P(x_1)} = 0 \text{ and eq. (17) can be written}$$

$$\hat{\ell}(\mathbf{x}_m) = [(\alpha - a)t_{10} - at_{11}] + [(\beta - b)t_{01} - bt_{00}], \quad (18)$$

where the parameters b and β are defined by

$$b = \log \left(\frac{1 - P_{01}}{1 - P_{01}^*}\right) \quad (19a)$$

$$\beta = b + \log \left(\frac{P_{01}^*}{P_{01}}\right), \quad (19b)$$

and the parameters a and α are defined as in Section 4.1 (eqs. (9a) and (9b)). $P_{0,1}^*$ and $P_{1,0}^*$ correspond to $P_{0,1}(\rho,r)$ and $P_{1,0}(\rho,r)$ with $r = r_0$.

Before discussing the symmetric structure of the optimum statistic [eq. (18)], we examine the $P_{0,1}$ characteristics for the M/M/1-loss model. Using eqs. (2) and (3), we can obtain $P_{0,1}$ vs mean-trunk occupancy ρ for a normal ($r = 1$) and killer ($r = r_0$) trunk. Figures 6 and 7 are plots for the 200- and 100-second sampling option, respectively, with a mean holding time of 180 seconds. It is clear from Fig. 6 that a $0 \rightarrow 1$ transition is just marginally more likely to occur on a killer trunk than on a normal trunk with a 200-second sampling rate. Although, the difference in the $0 \rightarrow 1$ transition probabilities between a normal and killer trunk increases substantially with the 100-second sampling rate, it is clear that these differences are still quite small—compared to the spread between the $P_{1,0}$ and $P_{1,0}^*$ plots (see Figs. 2 and 3). Note that eqs. (2) and (4) show that

$$\frac{P_{0,1}^*}{P_{0,1}} = \frac{1 - \exp \dfrac{-r_0 s}{1 - \rho}}{1 - \exp \dfrac{-s}{1 - \rho}} > 1 \text{ for } r_0 > 1 \quad (19c)$$

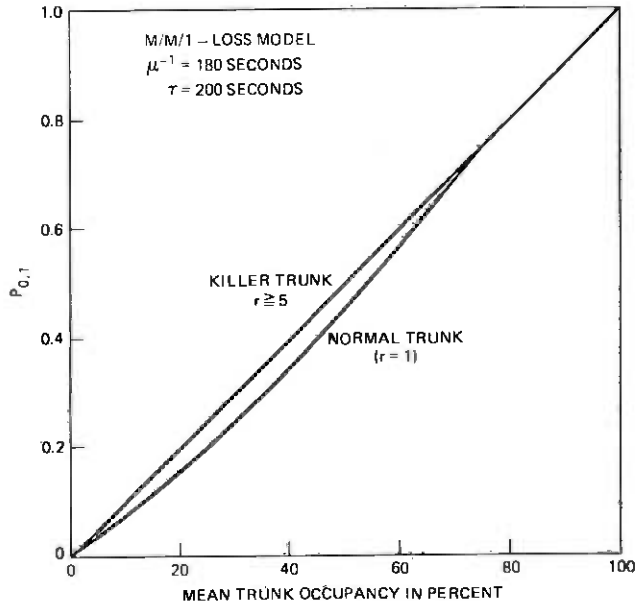

Fig. 6— 0 → 1 transition probability for the 200-second sampling option.

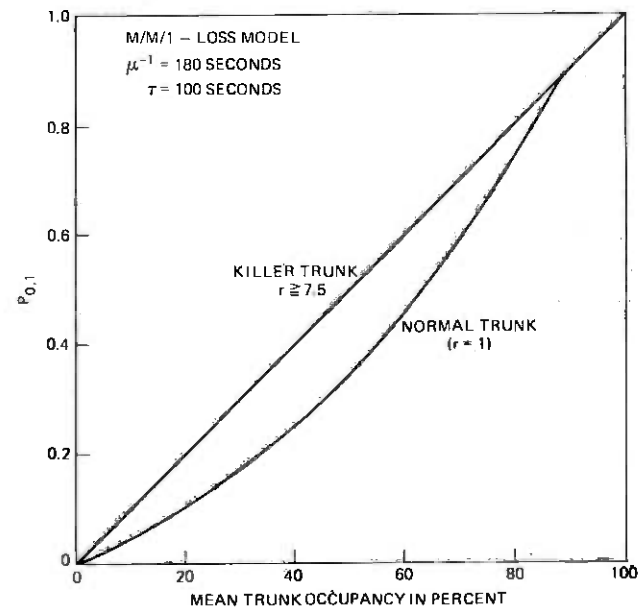

Fig. 7— 0 → 1 transition probability for the 100-second sampling option.

and, hence, we have $\beta > b > 0$. Using eqs. (4) and (19c), we see that $P^*_{0,1}/P_{0,1} = P^*_{1,0}/P_{1,0}$ and, therefore,

$$\beta - b = \alpha - a \tag{19d}$$

Equation (18) shows that the optimum statistic is the sum of two *symmetric* statistics:

(*i*) The statistic $[(\alpha - a)t_{10} - at_{11}]$, which is essentially the *ad hoc* statistic (see eq. (13) and related discussion).

(*ii*) An *additional* statistic $[(\beta - b)t_{01} - bt_{00}]$, which weights $0 \rightarrow 1$ transitions positively (evidence of a killer) and $0 \rightarrow 0$ transitions negatively (evidence of a normal trunk).

Note that by interchanging the role of 0 and 1 in either of these two statistics, we obtain the other—b is obtained from a and β is obtained from α by replacing $P_{1,0}$ with $P_{0,1}$.

By using eq. (11a) and the analogous equation

$$m - n(m) = t_{00}(m) + t_{01}(m) + x^c_m, \; (x^c_m = 1 - x_m) \tag{20}$$

in eq. (18), the optimum statistic can be written

$$\hat{\ell}(\mathbf{x}_m) = [\alpha t_{10}(m) - an(m)] + [\beta t_{01}(m) - b(m - n)] + e_1(x_m), \tag{21}$$

where the end-effect term $e_1(x_m)$ is given by

$$e_1(x_m) = ax_m + bx^c_m.$$

To implement $\hat{\ell}(\mathbf{x}_m)$ with only $t(m)$ and $n(m)$ available, necessitates estimating both $t_{10}(m)$ and $t_{01}(m)$ by $t(m)/2$. That is, using eqs. (12a) and (12b) in eq. (21) yields.

$$\hat{\ell}(\mathbf{x}_m) = \left[\alpha \frac{t(m)}{2} - an(m)\right] + \left\{\beta \frac{t(m)}{2} - b[m - n(m)]\right\} + e(x_1, x_m), \tag{22a}$$

where

$$e(x_1, x_m) = (\alpha - \beta)\left(\frac{x_1 - x_m}{2}\right) + e_1(x_m)$$

$$= \left(\frac{a - b}{2}\right)(x_1 + x_m) + b^{\dagger} \tag{22b}$$

or

$$\hat{\ell}(\mathbf{x}_m) = \left(\frac{\alpha + \beta}{2}\right) t(m) - (a - b)n(m) - bm + e(x_1, x_m). \tag{23}$$

† Recall that $\alpha - \beta = a - b$ [eq. (19d)].

In the development of the ad hoc algorithm, we assumed that the statistics corresponding to successive accumulation periods are *independent*. We conclude this section by examining the independence assumption and with some remarks on implementation.

Thus, turning to multiple accumulation periods, suppose $\mathbf{x}_m^i$, $i = 1,2,\cdots,k$ are the (unobservable) bit streams for k successive (and contiguous) accumulation periods, where $\mathbf{x}_m^i = (x_{(i-1)m+1},\cdots,x_{im})$. Assuming stationary traffic, and noting that $\{x_i\}_{i=1}^{km}$ is Markovian, we can write

$$P(\mathbf{x}_m^1,\cdots,\mathbf{x}_m^k) = \prod_{i=1}^{k} P(\mathbf{x}_m^i) \times \prod_{i=1}^{k-1} \frac{P(x_{im+1}/x_{im})}{P(x_{im+1})} \tag{24}$$

and, therefore,

$$\hat{\ell}(\mathbf{x}_m^1,\cdots,\mathbf{x}_m^k) = \sum_{i=1}^{k} \hat{\ell}(\mathbf{x}_m^i) + \sum_{i=1}^{k-1} \log\left\{\frac{P^*(x_{im+1}/x_{im})/P^*(x_{im+1})}{P(x_{im+1}/x_{im})/P(x_{im+1})}\right\}, \tag{25}$$

where $P(\cdot)$ and $P^*(\cdot)$ denote the distribution under H_0 (trunk normal) and H_1 (trunk killer), respectively. But, as we have seen,

$$\frac{P^*(x_{im+1}/x_{im})}{P^*(x_{im+1})} = 1 - \exp\left\{\frac{-r_0 s^i}{1-\rho}\right\} \tag{26}$$

and hence x_{im+1} and x_{im} are essentially independent for r_0 sufficiently large. Therefore,

$$P^*(x_{im+1}/x_{im}) = P^*(x_{im+1})$$

and, hence, eq. (25) may be written as

$$\hat{\ell}(\mathbf{x}_m,\cdots,\mathbf{x}_m^k) = \sum_{i=1}^{k} \hat{\ell}(\mathbf{x}_m^i) - \sum_{i=1}^{k-1} I(x_{im};x_{im+1}),$$

where

$$I(x_{im};x_{im+1}) = \log\frac{P(x_{im+1}/x_{im})}{P(x_{im+1})} \tag{27}$$

is recognized as the mutual information random variable, which plays a central role in information theory.[12] It is well known[12] that (under H_0) $E\{I(x_{im};x_{im+1})\}$ is non-negative, and hence to ignore the end-effect term

$$\sum_{i=1}^{k-1} I(x_{im};x_{im+1})$$

by implementing the statistic

$$\sum_{i=1}^{k} \hat{\ell}(\mathbf{x}_m^i)$$

would tend to make a normal trunk look more like a killer trunk on the average. In Appendix C, however, we show that the mean end effect $E\{I(x_{jm};x_{im+1})\}$ is negligible compared to the mean statistic update $E\{\hat{\ell}(\mathbf{x}_m^i\}$.

The "optimal" sequential algorithm is implemented in the same manner as the ad hoc algorithm (see eqs. (10a) and (10b)) except that now, corresponding to the sequence of accumulations $(t_i,n_i,m_i)i = 1,2,\cdots$, we define r_i by

$$r_i = \left(\frac{\alpha+\beta}{2}\right) t_i - (a-b)n_i - bm_i.$$

The term t_i denotes the total number of transitions in the ith accumulation interval. As is the case for the ad hoc algorithm (which corresponds to $b = \beta = 0$), the weights are functions of the trunk occupancy estimate $\hat{\rho}_i = n_i/m_i$. In a practical nonstationary environment, no claims of optimality are made or implied. The term "optimal" is applicable only in the context of the equilibrium (e.g., stationary) model with known trunk occupancy.

4.3 The ad hoc algorithm reviewed

The assumption that $t_{10}(m)$ conditioned on the switch count $n(m)$ is binomially distributed, is the basic assumption in the development of the ad hoc statistic. Although this assumption is incorrect (as we will soon see), the ad hoc statistic is essentially (except for an end-effect term) one of two symmetric statistics whose sum is the optimum statistic. Our purpose in this section is to examine the binomial assumption and to explain the relationship found between the ad hoc and optimal statistics.

Since the optimal statistic was developed for a trunk modeled as a server in an M/M/1-loss system, it is natural to examine the binomial assumption (eq. (6)):

$$P[t_{10}(m) = t/n(m) = n] = b[t;n,P_{1,0}(\rho,r)], \tag{28}$$

where $P_{1,0}(\rho,r)$ is given by eq. (4) in this context. Consider a killer trunk with r sufficiently large and suppose $\mathbf{x}_m = (x_1,\cdots,x_m)$ is the bit stream for a killer trunk during some accumulation period. Then, for all practical purposes [see eq. (26)], the trunk states x_i $i = 1,2,\cdots,m$ are independent and identically distributed Bernoulli random variables:

$$P(x_i = x) = \begin{cases} \rho & \text{if } x = 1 \\ 1-\rho & \text{if } x = 0. \end{cases}$$

Thus, it is clear that the switch-count distribution on a killer trunk is the binomial:

$$P(n(m) = n) = b(n;m,\rho). \tag{29}$$

Now suppose $A_m(t,n)$ denotes the number of binary m-tuples having exactly t $1 \rightarrow 0$ transitions and n ones. If $\mathbf{x}_m = (x_1,\cdots,x_m)$ is a sequence of trunk states for a killer trunk with $n(m) = n$, it is clear that each such sequence has probability

$$P(\mathbf{x}_m) = \rho^n(1-\rho)^{m-n}$$

and therefore

$$P(t_{10}(m) = t, n(m) = n) = A_m(t,n)\rho^n(1-\rho)^{m-n} \tag{30}$$

for a killer trunk (with r sufficiently large). Equations (29) and (30) show that for a killer trunk,

$$P(t_{10}(m) = t/n(m) = n) = \frac{\binom{m-n}{t}\binom{n}{t}}{\binom{m}{n}}, \tag{31}$$

where we have used the fact that

$$A_m(t,n) = \binom{m-n}{t}\binom{n}{t}. \tag{32}$$

It is interesting to note that while our assumed distribution for a killer trunk (28) differs from the correct distribution (31)—note that (31) is independent of ρ—there are some interesting similarities. For example, the assumed distribution peaks in the vicinity of $(n+1)(1-\rho)$ and has mean equal to $n(1-\rho)$† whereas the true distribution peaks in the vicinity of $(n+1)[1-(n/m)]$ and has mean equal to $n[1-n/m]$. Note that for "typical" realizations $(x_1,\cdots,x_m)$, we have

$$\frac{n}{m} \doteq \rho$$

and, hence, the two distributions have the same general location and scale. [In fact, expression (31) is a hypergeometric distribution, which converges to (28) as $m \rightarrow \infty$ if $n = \rho m$ (Ref. 13).] Thus, although incorrect, the binomial distribution approximates the true distribution of the killer trunk.

The following result helps to put the relationship between the ad hoc and the optimal statistics in perspective.

Lemma 2: If $\{x_i\}$ is a binary state stationary Markov chain with transition probabilities $P_{0,1}$ and $P_{1,0}$, and if $\mathbf{x}_m = (x_1,\cdots,x_m)$, then we

† $P_{1,0}(\rho,r) \rightarrow 1-\rho$ as $r \rightarrow \infty$.

have

$$P(\mathbf{x}_m) = b(t_{10};n,P_{1,0}) \times b(t_{0,1};m-n,P_{0,1}) \times q, \tag{33a}$$

where

$$q = \frac{P(x_1)(1-P_{1,0})^{-x_m}(1-P_{0,1})^{-x_m{}^c}}{\binom{n}{t_{10}}\binom{m-n}{t_{01}}}. \tag{33b}$$

Proof:

$$P(\mathbf{x}_m) = P(x_1) \prod_{i=2}^{m} P(x_i/x_{i-1})$$

$$= P(x_1)\, P_{1,0}^{t_{10}}\, P_{1,1}^{t_{11}}\, P_{0,1}^{t_{01}}\, P_{0,0}^{t_{00}}$$

$$= P_{1,0}^{t_{10}}(1-P_{1,0})^{t_{11}}\, P_{0,1}^{t_{01}}(1-P_{0,1})^{t_{00}} \times P(x_1).$$

Using eqs. (11a) and (18) to express t_{11} and t_{00} in terms of t_{10} and t_{01}, respectively, yields the result.

Thus, given a binary state stationary Markov chain $\{x_i\}$, it is clear from the above lemma that the log likelihood ratio

$$\hat{\ell}(\mathbf{x}_m) = \log \frac{P^*(\mathbf{x}_m)}{P(\mathbf{x}_m)}$$

formulated for the two hypotheses

$$H_0\text{: } \{x_i\} \text{ Markovian, characterized by } (P_{0,1},P_{1,0})$$

$$H_1\text{: } \{x_i\} \text{ Markovian, characterized by } (P^*_{0,1},P^*_{1,0})$$

is the sum of three terms:

$$\hat{\ell}(\mathbf{x}_m) = \log \frac{b(t_{10};n,P^*_{1,0})}{b(t_{10};n,P_{1,0})} + \log \frac{b(t_{01};m-n,P^*_{0,1})}{b(t_{01};m-n,P_{0,1})} + \log\left(\frac{q^*}{q}\right).$$

The first term is the ad hoc statistic $(\alpha t_{10} - an)$, the second term is the additional statistic $[\beta t_{01} - b(m-n)]$, and the third term is an end-effect term $(ax_m + bx^c_m)$.

$$\log\left(\frac{q^*}{q}\right) = ax_m + b\,x^c_m + \log \frac{P^*(x_1)}{P(x_1)} = ax_m + bx^c_m,$$

since, $P^*(x_1) = P(x_1) = \rho$.

The ad hoc algorithm, although based on the approximate binomial distribution, is very attractive for a number of practical reasons:

(*i*) For the 200-second sampling option, it is essentially optimum in a *practical* sense, since the $P_{0,1}$ characteristics for a normal and killer trunk are not far enough apart to exploit (see Fig. 6).

(*ii*) When we exploit the grouping information for the 5XB trunk

group in Section V, it will become obvious that the additional part of the optimum statistic is quite sensitive to the trunk occupancy and, hence, to the trunk-selection procedure modeled. (The sensitivity of the "additional" statistic to trunk occupancy stems from the fact that $P_{0,1}$ is "almost" proportional to ρ.) On the other hand, we will see that the ad hoc 5XB algorithm is relatively robust to minor perturbations in the trunk occupancy (and therefore to the trunk-selection procedure) and hence might be expected to perform well in a real 5XB environment.

V. THE 5XB TRUNK-GROUP ALGORITHMS

In addition to utilizing individual trunk switch-count and transition accumulations the 5XB group algorithms exploit the following:

(*i*) The identity of all trunks common to a group.
(*ii*) The trunk-selection procedure.

The resulting 5XB group algorithms typically are faster* than their individual trunk counterparts and are also less sensitive to the groups nominal holding time.

5.1 The 5XB trunk-group model

For the purposes of this paper, we model a 5XB trunk group (with all trunks normal) as an M/M/N-loss model with *random selection of idle trunks.*[2] The same assumptions apply if the group contains one or more killer trunks, but in this case we assume that killer trunks have a mean holding time equal to $1/r$ that of the normal mean holding time. In addition to being convenient theoretically, this idealized model has also been very useful in developing the 5XB group algorithms presently implemented in ICAN.

If all N trunks are normal, the random selection rule implies that all trunks have the same mean occupancy. In Ref. 2, the birth and death equations for the above model with a single killer trunk were solved in closed form, and in Ref. 14 this was generalized to an arbitrary number of killer trunks. These analytic results turn out to be quite useful, and in what follows we will need the following results derived in Ref. 2.

Theorem 1: For the above 5XB trunk group model having a single killer trunk with parameter r and an offered load of a erlangs, the blocking probability $\hat{B}(N,a,r)$ *and the mean occupancy* $\rho_r^*(N,a,r)$ *of the killer trunk are given by:*

* For given type 1 and 2 errors, the group algorithms typically have a considerably smaller mean decision time than their individual trunk algorithm counterparts.

$$(i)\quad \hat{B}(N,a,r) = \frac{NB(N,a)}{Nr - (r-1)a[1 - B(N,a)]} \tag{34a}$$

$$(ii)\quad \rho_r^*(N,a,r) = \frac{1}{1 + \dfrac{rN}{a} - r[1 - B(N-1,a)]}, \tag{34b}$$

where $B(N,a)$ is the usual erlang B blocking associated with an M/M/N-loss system with all trunks normal and an offered load of a erlangs.

It is easy to see that the occupancy ρ_r of each of the $N-1$ normal trunks must satisfy the conservation equation:

$$r\rho_r^* + (N-1)\rho_r = a[1 - \hat{B}], \tag{34c}$$

and the trunk-group occupancy ϕ_r^* is defined by:

$$\phi_r^* = \frac{\rho_r^* + (N-1)\rho_r}{N}. \tag{34d}$$

(For a 5XB trunk group having a killer trunk with parameter r: ρ_r^* and ρ_r denote the mean occupancy for a killer and normal trunk and ϕ_r^* denotes the mean group occupancy.)

Although eqs. (34a) through (34d) define an implicit relationship between $\rho_r^*(N,a,r)$ and $\phi_r^*(N,a,r)$, it will be very useful to have a simple explicit relationship. If the blocking term in eq. (34b) is ignored and if we "associate" ϕ_r^* with a/N, an approximation suggested is:

$$\rho_r^* \doteq \frac{\phi_r^*}{r - (r-1)\phi_r^*}. \tag{35a}$$

This approximation, although quite good for large N, is rendered obsolete by the following exact result:

Theorem 2: Consider a 5XB trunk-group model with all trunks normal and mean-group occupancy ϕ. (We will let ϕ denote the (mean) group occupancy for a 5XB trunk group with all trunks normal.) *If one of the trunks is replaced by a killer with parameter r, then*

$$\rho_r^* = \rho(\phi,r),$$

where,

$$\rho(\phi,r) = \frac{\phi}{r - (r-1)\phi}. \tag{35b}$$

Of course

$$\phi = \frac{a[1 - B(N,a)]}{N}$$

where $B(\cdot\,,\cdot)$ is the usual erlang B blocking expression.

This surprising result, which follows easily from eq. (34b), is proved in Appendix D. As a consequence of this theorem, "5XB group occupancy" will be used to denote the occupancy of a 5XB group with all trunks normal.

The mean occupancy of the normal trunks in a 5XB trunk group model having a single killer trunk no longer is given exactly by ϕ. But the following result, derived in Appendix D, shows that ϕ is a good approximation.

Theorem 3: Consider a 5XB trunk group model with N trunks having a single killer trunk with parameter $r \geqq 1$. The mean occupancy (ρ_r) of the $N - 1$ normal trunks satisfy

$$\phi \geqq \rho_r \geqq \left\{ \frac{r - \left(\frac{N}{N-1}\right)(r-1)\phi}{r - (r-1)\phi} \right\} \times \phi, \tag{36}$$

where ϕ is the mean-group occupancy with all trunks normal.

Theorems 2 and 3 are proved in Appendix D, where an exact expression for ρ_r is also derived. These results are special cases of general results obtained for the random selection model.[14]

5.2 Exploiting the 5XB Grouping Information

To simplify matters, we assume that a trunk in a 5XB group* with mean-group occupancy ϕ has mean occupancy $\rho(\phi,r)$ given by

$$\rho(\phi,r) = \frac{\phi}{r - (r-1)\phi}, \tag{37a}$$

where $r = 1$ corresponds to a normal trunk. Thus, if a group has no killer trunks, all normal trunks satisfy $\rho = \phi$ and eq. (37a) with $r = 1$ yields the correct occupancy. If, however, the group has a killer trunk, then all normal trunks satisfy inequality (36) so eq. (37a) with $r = 1$ is an approximation that increases in accuracy with the size of the group. Of course eq. (37a) is exact for a (single) killer trunk in a 5XB trunk group.

It is clear from eq. (37a) that

$$\rho(\phi,r) = \frac{\rho(\phi,1)}{r - (r-1)\rho(\phi,1)} \tag{37b}$$

* We use "5XB group" and our idealized model of a 5XB trunk group interchangeably.

and hence $\rho(\phi,r)$ is typically much smaller than $\rho(\phi,1)$. (For $\phi = 0.50$ and $r = 10$, $\rho(\phi,r) = 2/11\ \rho(\phi,1)$. Thus, it would appear that considering the $1 \rightarrow 0$ transition probability as a function of the 5XB group occupancy would effectively "spread" the $P_{1,0}$ characteristics in Figs. 2 and 3 further apart. That is, for a given ϕ, we propose comparing

$$P_{1,0}[\rho(\phi,1)] \text{ and } P^*_{1,0}[\rho(\phi,r)]$$

[rather than $P_{1,0}(\rho)$ and $P^*_{1,0}(\rho)$, as in Section IV].

Denoting the composition $P_{1,0}[\rho(\phi,r)]$ by $P_{1,0}(\phi,r)$, we have

$$P_{1,0}(\phi,r) = (1 - \rho(\phi,r)) \left(1 - \exp\left\{\frac{-rS}{1 - \rho(\phi,r)}\right\}\right), \tag{38}$$

which is plotted in Figs. 8 and 9 for the 200- and 100-second sampling options, respectively. The normal holding time used in these figures is 180 seconds, and the killer-trunk characteristics are drawn for $r = 5$, 10, and 15.

The increased "spread" between normal and killer $P_{1,0}$ characteristics obtained in this way is simply a consequence of exploiting the distinctly different occupancies of a normal and killer trunk in a 5XB trunk group. Figure 10 is a three-dimensional sketch of the composition of $P_{1,0}$ and ρ. Because all normal trunks in a 5XB group have the same mean occupancy, we see that a *single* $P_{1,0}$ vs ϕ characteristic suffices to describe

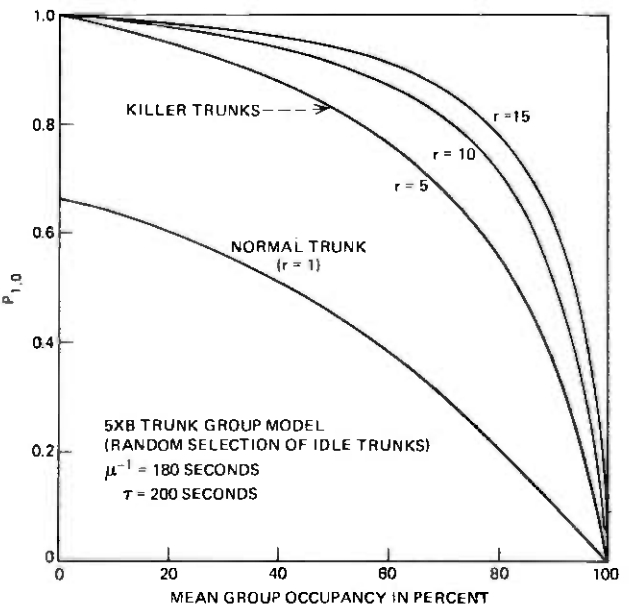

Fig. 8— $1 \rightarrow 0$ transition probability for the 200-second sampling option.

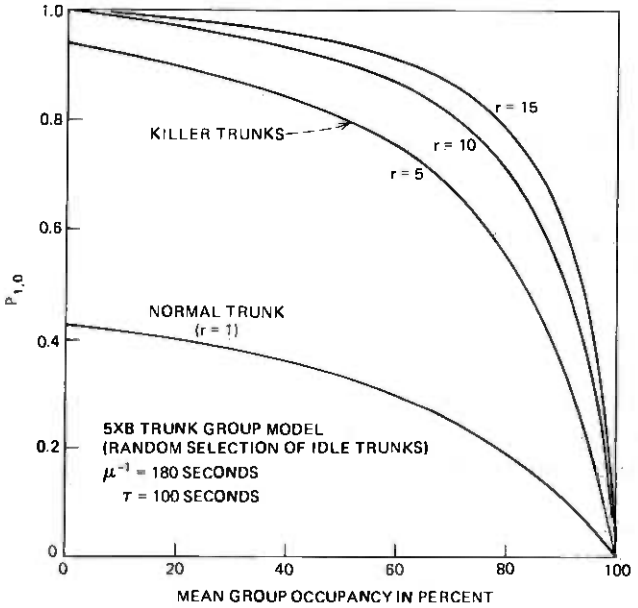

Fig. 9— 1 → 0 transition probability for the 100-second sampling option.

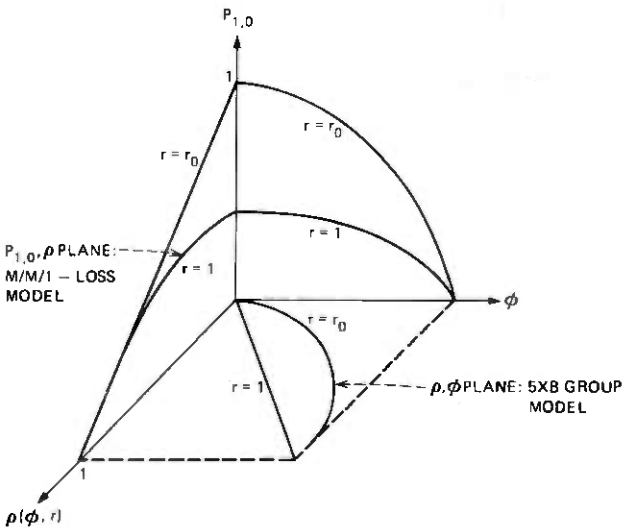

Fig. 10—Sketch of the composition of $P_{1,0}(\rho)$ with $\rho(\phi,r)$.

all normal trunks. This fact allows us to translate the individual trunk algorithm's development to this 5XB context with essentially only notational changes.

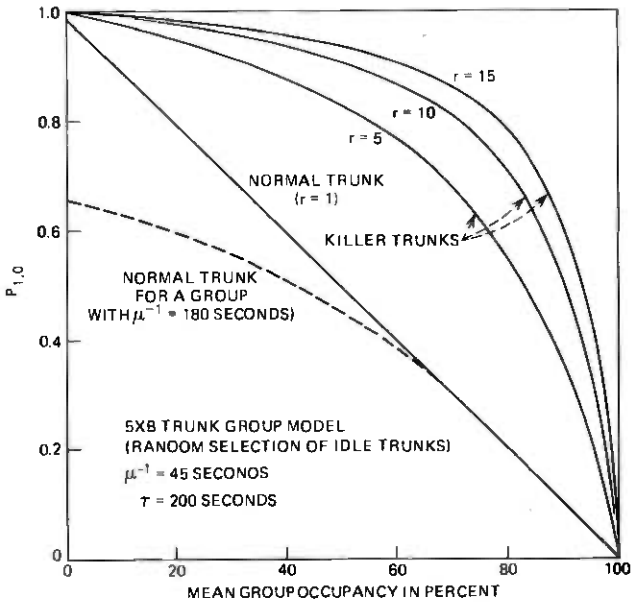

Fig. 11— $1 \rightarrow 0$ transition probability for the 200-second sampling option (mean group holding time = 45 seconds).

Figures 11 and 12 are plots of eq. (38) drawn for a normal group mean holding time of 45 seconds.* The normal trunk characteristic corresponding to 180 seconds is shown in dashed lines. We see that with 5XB grouping information factored into the picture, considerable discrimination exists between both normal trunk characteristics as well as between the normal trunk having a holding time of 45 seconds and the killer trunks. The discrimination that exists between the normal trunks permits us to make the 5XB group algorithm adaptive to the group mean holding time. [Although we will not pursue this topic, the basic idea is that $\Sigma_j t_{10}(j)/\Sigma_j n(j)$ (sums are over all trunks in the group) is an estimate of $p_{1,0}$ and can be used to decide which (of several) normal $p_{1,0}$ characteristics constitutes H_0.]

5.3 The ad hoc and the optimal 5XB group algorithms

We assume that the mean group holding time is known and consider formulating a hypothesis-testing problem similar to that in Section 4.1. Thus, we denote $P(t_{10}(m) = t/n(m) = n)$ by $P(t/n)$ and consider the two hypotheses:

* For normal holding times in the vicinity of 45 seconds, a killer parameter r in the range 3 to 5 probably is typical. An r of 10 or 15 in this context is unrealistic.

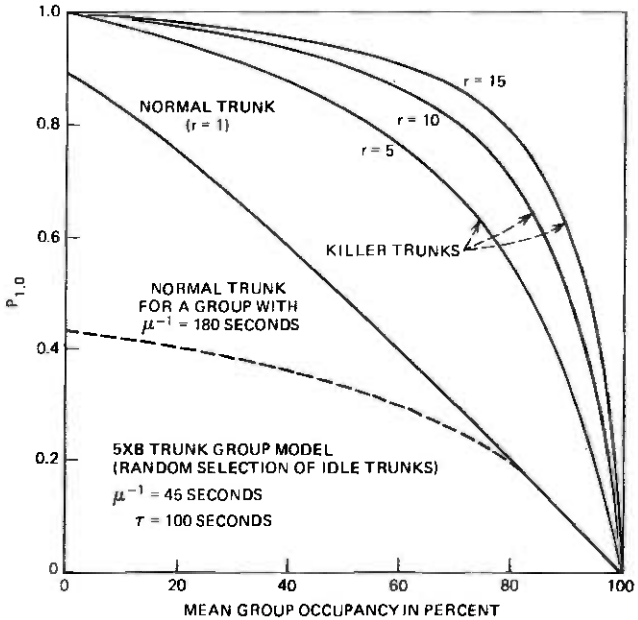

Fig. 12— 1 → 0 transition probability for the 100-second sampling option (mean grouping holding time = 45 seconds).

$$H_0\colon P(t/n) = b[t;n,P_{1,0}(\phi,1)]$$
$$H_1\colon P(t/n) = b[t;n,P_{1,0}(\phi,r)], \quad r \geq r_0. \tag{39}$$

There are two differences between this formulation and the one in Section 4.1:

(*i*) The trunk occupancy ρ in Section 4.1 is replaced by the group occupancy ϕ.

(*ii*) The alternate hypothesis H_1 is composite since $P_{1,0}(\phi,r)$ for $r \geq r_0$ are distinct.

The approach taken in dealing with (*ii*) is a natural one often adopted;[11] since $P_{1,0}(\phi,r)$ is monotone increasing in r {this follows from eq. (38) upon noting that $r/[1 - \rho(\phi,r)] = r + \phi/(1 - \phi)$}, then testing between H_0 and the *simple* alternate hypothesis

$$H_1\colon P(t/n) = b[t;n,P_{1,0}(\phi,r_0)],$$

say with type 1 and 2 errors α and β, respectively, implies that if the true state of nature is H_1 with $r = r_0' > r_0$ the resulting type 2 error will not exceed β. With this approach, we can simply translate the ad hoc algorithm results developed in Section 4.1 to this 5XB group context by making the appropriate changes in notation.

Thus, the ad hoc 5XB group algorithm can be described as follows: corresponding to the sequence of data accumulations (t_i^j, n_i^j) $i = 1,2,\cdots$ for the jth trunk in a group of N trunks, where t_i^j and n_i^j are the $1 \to 0$ transition and switch-count accumulations, respectively, during the ith accumulation period in which m_i scans are made, define s_i^j and S_i^j $i = 1,2,\cdots; j = 1,\cdots; N$ by

$$s_i^j = \alpha(\phi_i)t_i^j - a(\phi_i)n_i^j \tag{40a}$$

and

$$S_i^j = S_{i-1}^j + s_i^j \text{ with } S_0^j = 0, \tag{40b}$$

where ϕ_i is the group's occupancy during the ith accumulation period. The sequential test for the jth trunk in the group, $j = 1,\cdots,N$ is defined by

(*i*) Compute S_i^j, $i = 1,2,\cdots$ and defer making a decision as long as $T_0 < S_i^j < T_1$.

(*ii*) If $i = k$ corresponds to the first accumulation period for which

$$S_i^j \notin (T_0, T_1),$$

then

$$S_k^j \leqq T_0 \Rightarrow \text{trunk } j \text{ is normal}$$

$$S_k^j \geqq T_1 \Rightarrow \text{trunk } j \text{ is a killer.}$$

The weights $a(\phi)$ and $\alpha(\phi)$ are defined by

$$a(\phi) = \log \frac{1 - P_{1,0}(\phi,1)}{1 - P_{1,0}(\phi,r_0)} \tag{41a}$$

and

$$\alpha(\phi) = a(\phi) + \log \frac{P_{1,0}(\phi,r_0)}{P_{1,0}(\phi,1)}, \tag{41b}$$

where $P_{1,0}(\phi,r)$ is defined by eq. (38).

Just as in the individual trunk algorithm, the actual occupancy required to choose the weights a and α is unknown and must be estimated. Thus, the group occupancy ϕ_i during the ith accumulation period is estimated by

$$\hat{\phi}_i = \frac{1}{N} \sum_{j=1}^{N} n_i^j / m_i \tag{42}$$

Note that $\hat{\phi}_i$ is a "better" estimator than $\hat{\rho}_i = n_i/m_i$ (the estimator used in the individual trunk algorithm) in the following sense: given a 5XB group with all trunks normal and mean-group occupancy ϕ (in equilibrium), we have

(*i*) $E(\hat{\rho}) = \rho = \phi$
(*ii*) $E(\hat{\phi}) = \phi$
(*iii*) $\text{var}(\hat{\phi}) < \text{var}(\hat{\rho})$.

In addition to the better occupancy estimate available on a group basis, the fact that the group $P_{1,0}$ characteristics are "flatter and broader" then the individual trunk $P_{1,0}$ characteristic implies that the group algorithm more faithfully tracks the required weights, than does the individual trunk algorithm.

The ad hoc 5XB group algorithm has the same pleasant intuitive interpretation that the ad hoc individual trunk algorithm had (see eq. (13) and related discussion). It is also easy to show how the optimal individual trunk algorithm development of Section 4.2 carries over to the 5XB group context.

Thus, consider the two states of a trunk to be described by:

H_0: $\{x_i\}$ Markovian, characterized by $\theta = (P_{0,1}, P_{1,0})$

H_1: $\{x_i\}$ Markovian, characterized by $\theta^* = (P^*_{0,1}, P^*_{1,0})$,

where $P_{1,0} = P_{1,0}(\phi, 1)$, $P^*_{1,0} = P_{1,0}(\phi, r_0)$, and [see eq. (2)]

$$P_{0,1}(\phi,r) = \frac{\rho(\phi,r)}{1 - \rho(\phi,r)} P_{1,0}(\phi,r) \tag{43}$$

with $\rho(\phi,r)$ defined by eq. (37a). The assumptions that lead to a consideration of these two statistical hypotheses as a model of the normal and killer states of a trunk can be found in Section 4.2.

Proceeding as in Section 4.2 leads us to the optimum statistic $\hat{\ell}(\mathbf{x}_m)$ for distinguishing between the two simple hypotheses under consideration:

$$\hat{\ell}(\mathbf{x}_m) = [(\alpha - a)t_{10} - at_{11}] + [\tilde{b}t_{00} - (\tilde{\beta} - \tilde{b})t_{01}] + \log \frac{P^*(x_1)}{P(x_1)}, \tag{44}$$

where the parameters a and α are defined by eqs. (41a) and (41b) and the parameters $\tilde{b}$ and $\tilde{\beta}$ are defined by

$$\tilde{b} = \log \frac{1 - P_{0,1}(\phi,r_0)}{1 - P_{0,1}(\phi,1)} \tag{45a}$$

and

$$\tilde{\beta} = \tilde{b} + \log \frac{P_{0,1}(\phi,1)}{P_{0,1}(\phi,r_0)}. \tag{45b}$$

As in eq. (39), the alternate hypothesis is really composite since $[P_{0,1}(\phi,r), P_{1,0}(\phi,r)]$ for $r \geq r_0$ are distinct. Since $P_{0,1}(\phi,r)$ is monotone decreasing in r, the approach discussed earlier of treating H_1 as a simple

hypothesis with $r = r_0$ is followed. The parameters $\hat{b}$ and $\hat{\beta}$ are defined a bit differently than were the parameters b and β for the optimal individual trunk algorithm (see eq. (20a) and (20b)) in order to obtain nonnegative weights. Thus, in the individual trunk algorithm context, we had $P^*_{0,1} > P_{0,1}$ (eq. (19c)), but in the present 5XB group context we have $P_{0,1} > P^*_{0,1}$. The reason for this "flip-flop" is easy to see: for a *given group occupancy* ϕ, we are now contrasting the $0 \rightarrow 1$ transition probability for two trunks which differ, not only in their hang-up rates but also in their occupancies as well. Thus, the difference in the two occupancies dominates the effect that the hang-up rates alone have. Roughly speaking, the $0 \rightarrow 1$ transition probability of a trunk is approximately equal to its occupancy (conditioning on the last scan has little effect) and hence since $\rho(\phi,r_0) \ll \rho(\phi,1)$ it is clear that we should have $P_{0,1}(\phi,r) < P_{0,1}(\phi,1)$. Figures 13 and 14 are plots of $P_{0,1}(\phi,r)$ for the 200- and 100-second sampling options, respectively. In both figures, the killer trunk characteristics have been plotted for $r = 5$, 10, and 15 and a normal mean holding time of 180 seconds is assumed. These figures are very insensitive to the assumed normal mean holding time, since they essentially reflect eq. (37a), which is independent of the mean group holding time.

The "additional" statistic which appears in eq. (44),

$$\hat{b}t_{00} - (\hat{\beta} - \hat{b})t_{01}$$

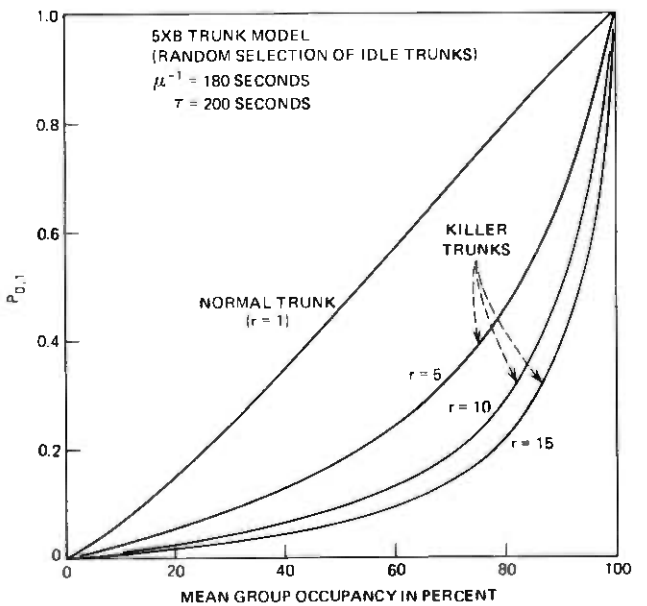

Fig. 13— $0 \rightarrow 1$ transition probability for the 200-second sampling option.

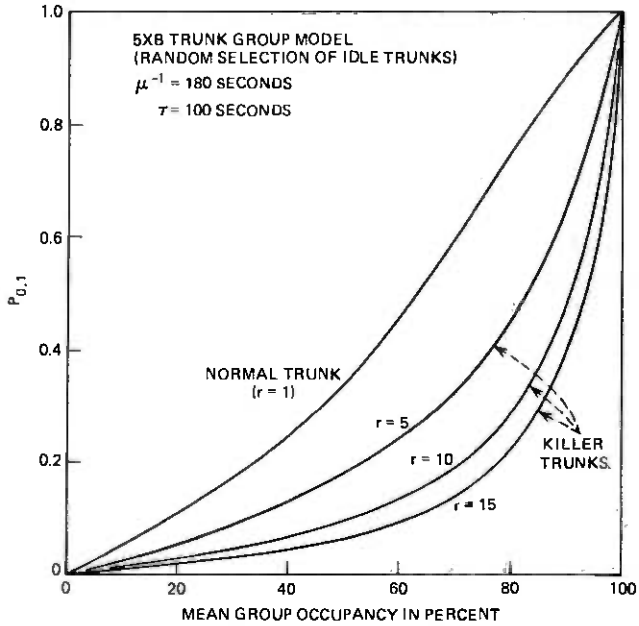

Fig. 14—0 → 1 transition probability for the 100-second sampling option.

shows that 0 → 0 transitions are weighted positively (evidence of a killer trunk*) and 0 → 1 transitions are weighted negatively (evidence of a normal trunk). This additional statistic is strongly influenced by the occupancy of a trunk, and only slightly by its hang-up rate.

Note also that the term $\log P^*(x_1)/P(x_1)$ is nonzero in the 5XB context since

$$\frac{P^*(x_1)}{P(x_1)} = \begin{cases} \dfrac{1}{r-(r-1)\phi} & \text{if} \quad x_1 = 1 \\ \dfrac{r}{r-(r-1)\phi} & \text{if} \quad x_1 = 0. \end{cases}$$

VI. PERFORMANCE OF THE 5XB GROUP ALGORITHMS

In common with all sequential detection algorithms, the time required by the killer-trunk detection algorithms to reach a decision (trunk normal or killer) is a random variable. In this section, we obtain an approximate formula for the *mean time* required by the 5XB group algorithms to reach a decision. This result is used to contrast the performance

* Killer trunks in the 5XB group model have very low occupancy, and hence 0 → 0 transitions are likely.

of the ad hoc and optimal algorithms as well as to point out the considerable effect that the sampling rate has on each. In addition, we also obtain an approximate expression for the false-alarm probability of the 5XB group algorithms. The analysis for the individual trunk algorithms, although differing in several respects from the group algorithms, involves the same sort of considerations and is omitted.

The analysis in this section assumes a server system in equilibrium and, therefore, the mean trunk-group occupancy ϕ is assumed constant. In addition, to simplify the analysis, we assume that ϕ is known; an approximate analysis which does not require this assumption is sketched in Section 6.1. A consequence of this assumption is that the algorithm weights are treated as constants rather than random variables. This assumption is not unreasonable because for multiple-hour accumulation periods $\text{var}(\hat{\phi})$ is quite small ($\hat{\phi}$ is the switch-count estimate of ϕ). ($\text{Var}(\hat{\phi})$ has been derived for an M/M/N-loss system.[15])

6.1 Mean statistic update

Corresponding to a sequence of trunk states $x_1, x_2, \cdots, x_m$ in an accumulation period with m scans, define a sequence of transition updates $z_2, \cdots, z_m$ by

$$z_n = \begin{cases} b & \text{if} \quad (x_{n-1}, x_n) = (0,0) \\ -(\beta - b) & \text{if} \quad (x_{n-1}, x_n) = (0,1) \\ (\alpha - a) & \text{if} \quad (x_{n-1}, x_n) = (1,0) \\ -a & \text{if} \quad (x_{n-1}, x_n) = (1,1) \end{cases} \tag{46}$$

The optimum 5XB statistic (eq. (44)) may therefore be written:

$$\hat{\ell}(\mathbf{x}_m) = \sum_{i=2}^{m} z_i + \log \frac{P^*(x_m)}{P(x_m)} \tag{47}$$

In practice the end-effect term cannot be implemented and all the transitions in eq. (44) must be estimated in terms of $t(m)$, $n(m)$, and m. Thus, if we denote the implementable version of eq. (44) by $S_m(\phi, r_0)$,† use eqs. (11), (12), and (20) in eq. (44), and drop all end-effect terms we obtain

$$S_m(\phi, r_0) = \left[\alpha(\phi) \frac{t(m)}{2} - a(\phi) n(m) \right] + \left[\tilde{b}(\phi)[m - n(m)] - \tilde{\beta}(\phi) \frac{t(m)}{2} \right]. \tag{48}$$

† r_0 is the value of the killer parameter used in defining the alternate hypothesis and hence the algorithm weights (see eqs. (41) and (45)).

Now using eqs. (11), (12), and (20) once again, we see that $S_m(\phi,r_0)$ may be written

$$S_m(\phi,r_0) = \sum_{i=2}^{m} z_i + (\tilde{b}x_m^c - ax_m) + (\alpha + \tilde{\beta})\left(\frac{x_1 - x_m}{2}\right). \tag{49}$$

Since the 5XB group is in equilibrium, $z_2,\cdots,z_m$ are identically distributed and it is easily verified that their common mean is given by

$$E(z) = (\alpha P_{1,0} - a)\rho + (\tilde{b} - P_{0,1}\tilde{\beta})(1 - \rho). \tag{50}$$

We recall that ρ (eq. (37a)) as well as the transition probabilities (eqs. (38) and (43)) are functions of the group occupancy ϕ and the state r of the trunk. The weights a, α, $\tilde{b}$ and $\tilde{\beta}$ are only functions of ϕ for a specified choice of the parameter (r_0), which characterizes the alternate hypothesis. Thus, the mean statistic update for the optimum 5XB statistic and its "implementable version" is given by:

$$E\{\hat{\ell}(\mathbf{x}_m)\} = (m-1)E(z) + E\left\{\log\frac{P^*(x_1)}{P(x_1)}\right\} \tag{51a}$$

and

$$E\{S_m(\phi,r_0)\} = (m-1)E(z) + \tilde{b}(1-\rho) - a\rho. \tag{51b}$$

Note that it is easily shown that

$$E\left\{\log\frac{P^*(x_1)}{P(x_1)}\right\} = -\log\left[r_0 - (r_0-1)\phi\right] + \frac{r(1-\phi)}{r-(r-1)\phi}\log r_0$$

which is negative for $r = 1$ and positive for $r \geqq r_0$.

Although the increments (transition updates) defined in eq. (46) are identically distributed, they are not independent. In fact, since the state sequence $\{x_i\}$ has been modeled as a Markov chain (see Section 4.2), it is easy to see that the sequence $\{z_i\}$ defined by (46) is also a Markov chain. Relabeling the four natural states,

$$\tilde{b},\ -(\tilde{\beta} - \tilde{b}),\ (\alpha - a),\ -a,$$

of the chain $\{z_i\}$ by

$$0, 1, 2, 3,$$

respectively, it is easily seen that the one-step transition matrix π for this chain is given by

$$\pi = \begin{pmatrix} 1-P_{0,1} & P_{0,1} & 0 & 0 \\ 0 & 0 & P_{1,0} & 1-P_{1,0} \\ 1-P_{0,1} & P_{0,1} & 0 & 0 \\ 0 & 0 & P_{1,0} & 1-P_{1,0} \end{pmatrix},$$

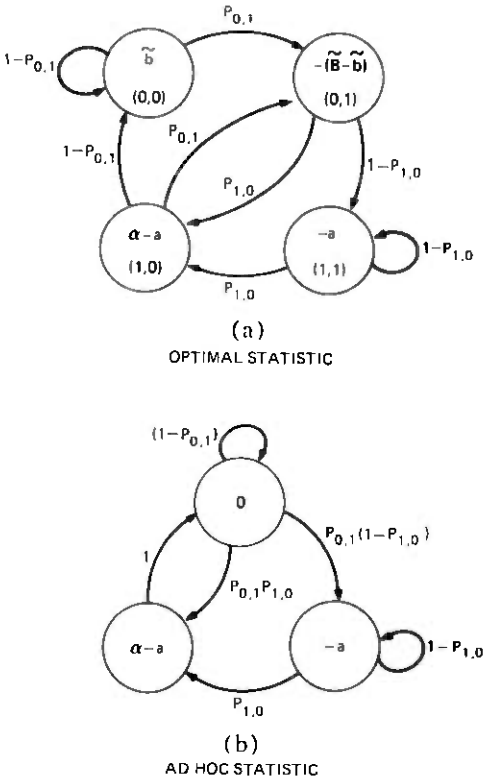

Fig. 15—State diagram for the ad hoc and optimal 5XB group statistics. (a) Optimal statistic. (b) Ad hoc statistic.

where $\pi_{i,j} = P(z_n = j/z_{n-1} = i)$. Of course, the stationary distribution $\boldsymbol{P}$ satisfying $\boldsymbol{P}\pi = \boldsymbol{P}$ is

$$\boldsymbol{P} = [(1-\rho)(1-P_{0,1}), (1-\rho)P_{0,1}, \rho P_{1,0}, \rho(1-P_{1,0})].$$

The state diagram corresponding to the Markov chain $\{z_i\}$ is shown in Fig. 15a. If $\tilde{b} = \tilde{\beta} = 0$, we obtain the ad hoc algorithm, for which the sequence $\{z_i\}$ is a three-state Markov chain, with natural states 0, $\alpha - a$ and $-a$. The state diagram for this chain is shown in Fig. 15b.

Figures 16 and 17 are plots of the mean statistic update vs group occupancy for the implementable version of the 5XB group algorithms. (Log base 10 is used in this paper.) These figures are drawn for a killer trunk with parameter $r = 10$. Also shown is the corresponding plot for a normal trunk ($r = 1$).

We close this section by indicating an approximate analysis of $E(z)$ which doesn't assume that ϕ is known. Thus, in practice, ϕ is unknown

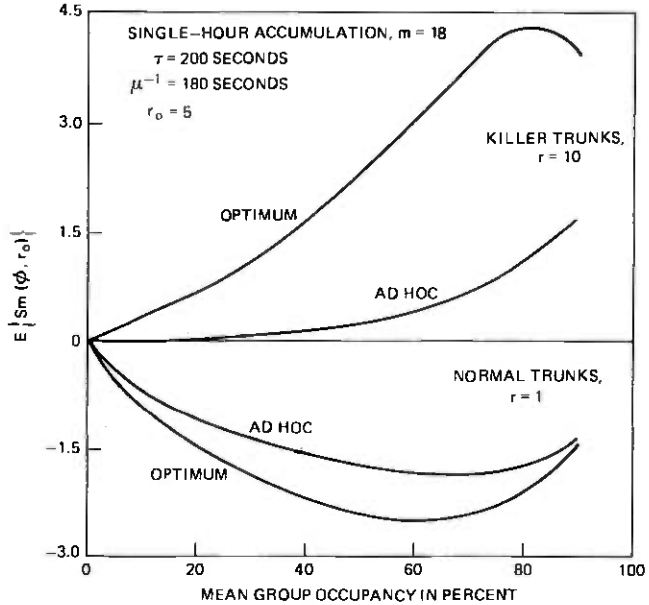

Fig. 16—Mean statistic update for the 5XB group algorithms—200-second sampling option.

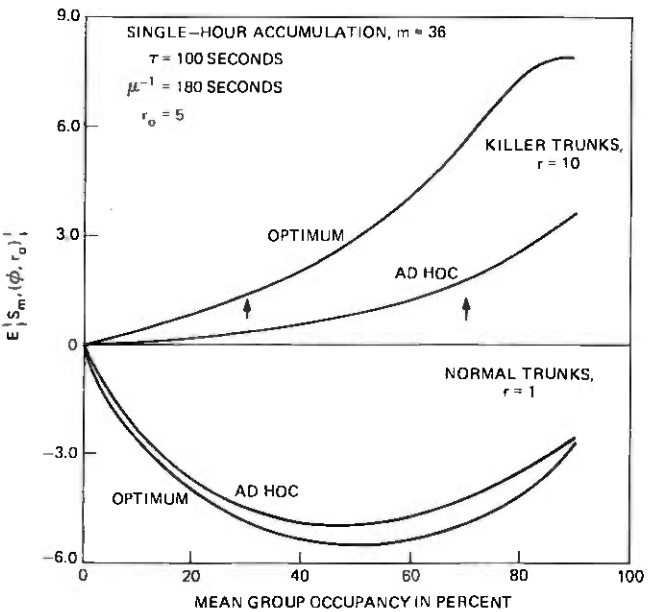

Fig. 17—Mean statistic update for the 5XB group algorithms—100-second sampling option.

and is estimated by $\hat{\phi}$. Hence eq. (46) should read:

$$z_n = \begin{cases} \tilde{b}(\phi_0) & \text{if} \quad (x_{n-1},x_n) = (0,0) \quad \text{and} \quad \hat{\phi} = \phi_0. \\ \vdots \end{cases}$$

Conditioning on $\hat{\phi}$, we have

$$E(z) = E\{E(z/\hat{\phi})\},$$

where

$$E(z/\hat{\phi}) = \tilde{b}(\hat{\phi})P(x_{n-1} = 0,x_n = 0/\hat{\phi}) + \cdots.$$

Now by assuming that (x_{n-1},x_n) and $\hat{\phi}$ are independent,* we get eq. (50) with the constant weights $\alpha(\phi),\cdots$ replaced by the mean values $E\{\alpha(\hat{\phi})\},\cdots$. The mean values can be approximated in either one of two ways:

(*i*) $E\{\alpha(\hat{\phi})\} \doteq \alpha[E(\hat{\phi})] = \alpha(\phi)$, which obviously amounts to assuming ϕ is known.

(*ii*) $$E\{\alpha(\hat{\phi})\} \doteq \alpha(\phi) + \frac{\text{var}(\hat{\phi})}{2!} \frac{d^2\alpha^\dagger}{d\hat{\phi}^2}\bigg|_{\hat{\phi}=\phi,}$$

which factors the available variance of $\hat{\phi}$ into the picture.

6.2 Mean time to detection

The basic structure of all the detection algorithms in this paper are the same: a statistic s_i is evaluated at the end of the ith accumulation period, $i = 1,2,\cdots$ and a decision is made the first time that the sum $s_1 + s_2 + \cdots$ falls outside an interval (T_0,T_1). Presumably, the random walk type statistic S_i has a negative drift under H_0 (trunk normal) and a positive drift under H_1 (trunk killer). Wald's SPRT always has the appropriate drift: if H_0 and H_1 correspond to the probability distributions $P_0(w)$ and $P_1(w)$, respectively, and if $\hat{\ell}(w)$ is defined by

$$\hat{\ell}(w) = \log \frac{P_1(w)}{P_0(w)},$$

then $E\{\hat{\ell}(w)\} < 0$ under H_0 and $E\{\hat{\ell}(w)\} > 0$ under H_1. The proof is immediate by using the inequality[12]

$$-\sum_i p_i \log p_i < -\sum_i p_i \log q_i,$$

* For reasonable-size trunk groups, we expect very little dependence between the sampled state process of an individual trunk $(x_1,\cdots)$ and the group process ϕ.

† This is a Taylor expansion to second order.

where $\{p_i\}$ and $\{q_i\}$ are distinct probability distributions. In Appendix E, we show that both the ad hoc ($\tilde{b} = \tilde{\beta} = 0$) and the additional ($a = \alpha = 0$) parts of $E(z)$ (eq. 50) have the appropriate drift. This in turn shows that both the ad hoc and additional parts of the 5XB group statistic (eq. (51a)) have the appropriate drift.

Suppose $Y_1, Y_2, \cdots$ are i.i.d. random variables with common mean μ and consider a random walk

$$S_n = \sum_{i=1}^{n} Y_i \quad n = 1,2,\cdots$$

with absorbing barriers at T_0 and T_1. If μ is small compared to T_0 and T_1, then the mean stopping time (mean number of steps to absorption) $E(n)$ is approximately given by

$$\frac{P_0 T_0 + P_1 T_1}{\mu},$$

where P_0 and P_1 are the probabilities of absorption at T_0 and T_1, respectively. This follows from Wald's identity[11]

$$E(S_n) = \mu E(n)$$

if we approximate the mean value of the random walk at absorption by $P_0 T_0 + P_1 T_1$.

In our detection theory context, P_0 and P_1 correspond to β and $1 - \beta$, respectively, if the trunk is a killer (β = probability of miss) and $1 - \alpha$ and α, respectively, if the trunk is normal (α = probability of false alarm). If we denote the mean number of accumulation periods needed to reach a decision under H_0 and H_1 by $E(T_N)$ and $E(T_k)$, respectively, and assume (*i*) successive statistic updates are independent and (*ii*) the mean statistic update is small compared to T_0 and T_1, we obtain

$$E(T_N) \doteq \frac{(1 - \alpha)T_0 + \alpha T_1}{E\{|S_m(\phi, r_0)|\}} \tag{52a}$$

and

$$E(T_k) \doteq \frac{\beta T_0 + (1 - \beta)T_1}{E\{|S_m(\phi, r_0)|\}}, \tag{52b}$$

where the mean statistic update $E\{|S_m(\phi, r_0)|\}$ is evaluated for $r \doteq 1$ in (52a) and for $r \geqq r_0$ in (52b).

The assumption that successive statistic updates are independent is not strictly true if the successive statistic updates are contiguous. However, one expects that the slight (end-effect) dependence will not give rise to very much error.

Let T_{k_0} be the time required to decide (incorrectly) that a killer trunk is normal. Similarly, let T_{k_1} be the time required to decide (correctly)

that a killer trunk is a killer. Then we have

$$E(T_k) = \beta E(T_{k_0}) + (1 - \beta)E(T_{k_1}),$$

which suggests that

$$E(T_{k_1}) \doteq E(T_k) \quad \text{if} \quad \beta \ll 1. \tag{53}$$

The moments of the conditional stopping times T_{k_0} and T_{k_1} can be obtained by using a well known technique of Wald's,* and one finds that approximation (53) is reasonable if T_1 and $(-T_0)$ are sufficiently large.

In our faulty-trunk detection context, a type 1 error ("false alarm") may result in the misuse of craft resources (e.g., testing a perfectly good trunk). A type 2 error ("miss") on the other hand, will result in an increased time to detection. Assuming type 1 and 2 errors of 10^{-6} and 10^{-2}, respectively (realistic implementation values), implies approximate thresholds $T_0 = -2$ and $T_1 = 6$ (formulae 4.5c and d, log base 10). With these parameter values, the mean detection time $E(T_{k_1})$ can be approximated by

$$E(T_{k_1}) \doteq \frac{T_1}{E\{S_m(\phi,r_0)\}}. \tag{54}$$

For a normal trunk, the mean statistic update is comparable to T_0 and hence expression (52a) isn't applicable, nor is it needed since the 5XB group algorithm reaches a decision on a normal trunk after one or two updates.

Figures 18 and 19 are plots of the mean-detection time vs group occupancy for the implementable versions of the 5XB group algorithms. [The dashed line portion of Figs. 18 and 19 indicates where approximation (54) involves considerable error (e.g., the region in which $E\{S_m(\phi,r_0)\}$ is a significant fraction of T_1).] It is apparent from these figures that

(*i*) The mean detection time for both the ad hoc and optimal algorithms is enhanced by using the 100-second sampling option. This enhancement is far more pronounced for the ad hoc algorithm.

(*ii*) The optimal algorithm is "faster" than the ad hoc algorithm. This contrast is greater for the 200-second sampling option.

6.3 False alarm probability of the 5XB group algorithms

If $B(\alpha,\beta)$ and $A(\alpha,\beta)$ are the test thresholds that result in type 1 and 2 errors α and β, respectively, for a SPRT, Wald showed[11] that using

* See eq. 158 and 159 in Appendix A.5.2 of Ref. 11. The method ignores (as usual) the "excess over the boundaries" and hence yields approximate results.

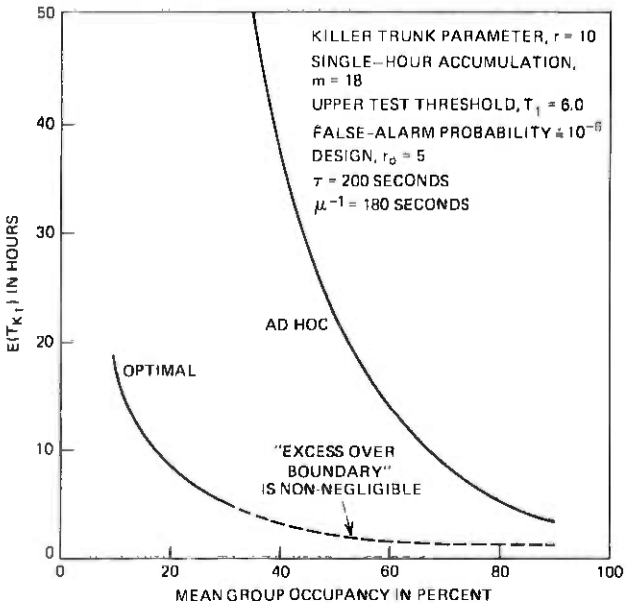

Fig. 18—Mean time to detection for the 5XB group algorithms—200-second sampling option.

thresholds T_0 and T_1 defined by*

$$T_0 = \log\left(\frac{\beta}{1-\alpha}\right) \quad \text{and} \quad T_1 = \log\left(\frac{1-\beta}{\alpha}\right) \text{ (assuming } \alpha + \beta < 1)$$

yield type 1 and 2 errors α' and β' which satisfy

$$\beta' \leqq \frac{\beta}{1-\alpha} \quad \text{and} \quad \alpha' \leqq \frac{\alpha}{1-\beta}.$$

The proof of this result is trivial and depends only on the assumption that the SPRT terminates with probability 1. This assumption is satisfied by a wide class of SPRTs,[16] including the case of interest to us, where the underlying distribution is "Markovian". Thus, the probability of false alarm (α') for a SPRT (an example of which is the "optimal" 5XB group algorithm) satisfies $\alpha' \leqq 10^{-T_1}$.

Because the ad hoc algorithm is not a SPRT (the true underlying distribution is not binomial, see Section 4.3), we may wish to study the consequences of using the above thresholds appropriate for a SPRT. To do this, we can proceed in (at least) two distinct ways:

(i) consider the statistic $S_m(\phi, r_0)$ to be the basic update in the algorithm.

* This assumes that the log likelihood ratio is used in defining the SPRT.

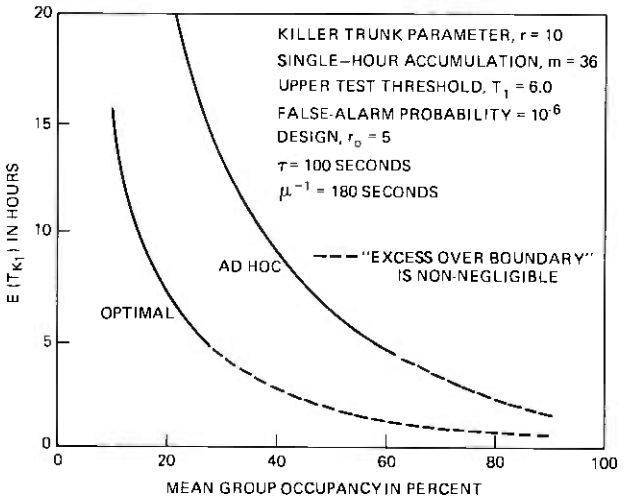

Fig. 19—Mean time to detection for the 5XB group algorithms—100-second sampling option.

(*ii*) Consider the transition update $Z_i(x_{i-1},x_i)$ to be the basic update in the algorithm.

In either case, we study a process of the form $S_j = \Sigma_{i=1}^{j} x_i$, $j = 1,2,\cdots$ with absorbing barriers $T_0 < 0 < T_1$. The advantage of proceeding as in (*i*) above is that the increments x_i may be assumed to be i.i.d. We briefly sketch this approach.

Consider a random walk S_j, $j = 1,2,\cdots$ with i.i.d. increments $\{x_i\}$. If n is a stopping time associated with S_j, $j = 1,2,\cdots$, Wald's fundamental identity[11,17] is given by

$$E\{e^{S_n t}\chi(t)^{-n}\} = 1 \text{ (all } t \text{ satisfying } |\chi(t)| \geq 1), \tag{55}$$

where

$$\chi(t) = E\{e^{xt}\}$$

is the moment generating function corresponding to the common distribution of the increments. If P_{T_0} and P_{T_1} denote the probabilities of absorption at T_0 and T_1, respectively, then rewriting (55) in a standard way yields

$$P_{T_0}E\{e^{S_n t}\chi(t)^{-n}/S_n \leqq T_0\} + P_{T_1}E\{e^{S_n t}\chi(t)^{-n}/S_n \geqq T_1\} = 1. \tag{56}$$

Now if the x_i take on both positive as well as negative values with nonzero probability and have a nonzero mean, then the equation

$$\chi(t) = 1$$

has only one nonzero root h_0 with the property that h_0 and $E(x)$ have opposite sign.* That is, $E(x) > 0 \Rightarrow h_0 < 0$ and $E(x) < 0 \Rightarrow h_0 > 0$. Assuming that the excess of S_n over the boundaries is small, eq. (56) yields the standard approximation[11,17]

$$P_{T_0} \doteq \frac{e^{h_0 T_1} - 1}{e^{h_0 T_1} - e^{h_0 T_0}} \tag{57a}$$

and

$$P_{T_1} \doteq \frac{1 - e^{h_0 T_0}}{e^{h_0 T_1} - e^{h_0 T_0}} \tag{57b}$$

Note that the probability of false alarm (type 1 error) corresponds to P_{T_1} if the random walk increment is that of a *normal* trunk $[S_m(\phi,r_0), r = 1]$. Similarly, the probability of miss (type 2 error) corresponds to P_{T_0} if the random walk increment is that of a killer trunk $[S_m(\phi,r_0), r \geqq r_0]$.

To use these approximations, we must compute the moment generating function

$$\chi(\mu) = E\{e^{[\alpha(\phi)t/2 - a(\phi)n]\mu}\} \tag{58}$$

by using the joint distribution $p(t,n)$ derived in Appendix B. [Note that our discussion applies equally well to the additional and optimal statistics. In general, we need $E\{\exp[S_m(\phi,r_0)\mu]\}$, where $S_m(\phi,r_0)$ is given by eq. (48).] Choosing the test thresholds T_0 and T_1 for the ad hoc algorithm according to the Wald SPRT formulae [eqs. (10c) and (10d)] typically results in $P_{T_0} < \beta$ and $P_{T_1} < \alpha$.

VII. SUMMARY

A class of killer-trunk detection algorithms has been developed that use the individual trunk usage and transition accumulations available in EADAS/ICUR. Because this data is essentially a sufficient statistic for the Markov chain used to model the (unobservable) sampled data, one of the algorithms developed is Wald's celebrated SPRT.

The detection algorithms developed can be partitioned in two natural ways:

(*i*) By sampling rate (100 or 200 seconds).
(*ii*) According to whether grouping information is used.

The algorithms which do not use grouping information are applicable to all trunks (one way or two way) independent of the type of switching machine used. A version of one of these individual trunk algorithms is currently in use in ICAN, testing trunks on 1XB, XBT and step-by-step

* See Appendix A.2.1 of Wald's original treatise (Ref. 11).

switching machines. The algorithms that exploit grouping information detect killers more quickly but are tailored to a specific switching machine. A "group" algorithm of this type is currently being used to test trunks associated with 5XB switching machines.

In the course of this study several problems of independent interest were studied. These include:

(*i*) The server covariance in a M/G/1-loss and GI/M/1-loss system.

(*ii*) The structure of the likelihood statistic that arises in testing simple hypotheses characterized by a binary valued Markov chain.

(*iii*) The occupancy of nonidentical trunks in a random-selection (Markovian) loss system.

The major conclusion in this study is, of course, that accumulated switch-count and state-transition data on *individual* trunks (based on sampling intervals on the order of a normal holding time) can be used to *reliably* detect abnormally short holding time trunks. Moreover the (near) optimal sequential detection algorithms using this accumulated data are easily exhibited, simple in structure, and intuitively appealing.

VIII. ACKNOWLEDGMENTS

Various stages of this study benefited from discussions with A. E. Eckberg, R. L. Franks, S. Horing, E. J. Messerli, and T. E. Rutt.

APPENDIX A

Sensitivity of the Transition Probabilities to Modeling Assumptions

To get an idea about the sensitivity of the algorithms to some of the modeling assumptions, the transition probability $P_{1,0} = P(x_{t+\tau} = 0/x_t = 1)$ was studied for the following two cases:

(*i*) M/G/1-loss, where the service distribution function $F(\cdot)$ is the mixed exponential given by

$$F(t) = 1 - d_1 e^{-x_1 t} - d_2 e^{-x_2 t} \quad t \geq 0$$

(*ii*) GI/M/1-loss, where the arrival process is the switched-Poisson process[8] commonly used to model overflow traffic.

Because the methods used to obtain $P_{1,0}$ for these two models differ, we discuss these models separately.

A.1 The M/G/1-loss model

An observer viewing the server in an M/G/1-loss system sees an alternating sequence of busy and idle intervals. The busy intervals are distributed according to some distribution $F(\cdot)$ and are independent.

The idle intervals are exponentially distributed with mean λ^{-1} (λ = mean arrival rate) and are independent. Thus, the sequence of alternating busy and idle intervals constitutes an alternating renewal process. In this context, the conditional probability

$$P_{1,0} = P(x_{t+\tau} = 0/x_t = 1), \tag{59}$$

where

$$x_t = \begin{cases} 1 & \text{if server is } \textit{busy} \text{ at epoch } t \\ 0 & \text{if server is } \textit{idle} \text{ at epoch } t \end{cases}$$

has already been studied,[18] and we have the following result:

Theorem 4: Consider an M/G/1-loss system in equilibrium with a service time distribution $F(t)$ having Laplace transform $f^*(s)$. If $P^*_{1,0}(s)$ denotes the Laplace transform of $P_{1,0}(\tau)$, then we have

$$P^*_{1,0}(s) = \frac{\mu[1 - f^*(s)]}{s\{s + \lambda[1 - f^*(s)]\}} \tag{60}$$

where λ and μ are the mean arrival and service rates, respectively.

Proof: See Section 7.4 of Ref. 18 [$f^*_2(s) = \lambda/s + \lambda$ and $f^*_1(s) = f^*(s)$].

For the mixed exponential service time distribution mentioned above let

$$x_1 = 2\mu d$$

and

$$x_2 = 2\mu(1 - d), \quad 0 \leqq d \leqq 1.$$

If T is distributed according to this two-parameter (μ,d) family of distributions, then

(*i*) $E(T) = \mu^{-1}$.

(*ii*) $\text{var}(T) = \mu^{-2}\left(\frac{2 - \delta}{\delta}\right)$, where $\delta = 4\,d(1 - d)$.

(*iii*) $c(T) = \frac{\sigma(T)}{E(T)} = \left(\frac{2 - \delta}{\delta}\right)^{1/2}$.

Thus, the mean is fixed at μ^{-1} and the coefficient of variation satisfies $c \geqq 1$, with equality occurring when $\delta = 1$, which corresponds to the M/M/1-loss system.

Equation (60) is easily inverted for this family of mixed exponential distributions, obtaining the following result:

$$P_{1,0}(\rho,\tau) = (1 - \rho) + \frac{r_1 + \delta}{r_1(r_1 - r_2)} e^{r_1\mu\tau} + \frac{r_2 + \delta}{r_2(r_2 - r_1)} e^{r_2\mu\tau}, \tag{61}$$

where

$$r_1 = -\left(\frac{a+2}{2}\right) + \left[(1-\delta)(1+a) + \left(\frac{a}{2}\right)^2\right]^{1/2}$$

$$r_2 = -\left(\frac{a+2}{2}\right) - \left[(1-\delta)(1+a) + \left(\frac{a}{2}\right)^2\right]^{1/2}$$

$$\delta = 4d(1-d), \quad a = \frac{\rho}{1-\rho}$$

and $\rho = \lambda/(\lambda + \mu)$ is the mean occupancy of the server.

In Fig. 5, $P_{1,0}(\rho,\tau)$ [eq. (61)] is plotted vs ρ for several different values of c (c = 1, 1.5, and 2) assuming

(*i*) μ^{-1} = 180 seconds
(*ii*) τ = 200 seconds.

Also shown is a plot of $P_{1,0}$ vs ρ for a killer trunk with r = 10 [μ in (61) is replaced by $r\mu$]. The normal trunk $P_{1,0}$ characteristic with c = 1 corresponds to the M/M/1-loss system.

A.2 The GI/M/1-loss model

The covariance function $R(\cdot)$ for the GI/M/1-loss model with a switched Poisson arrival process has the form:

$$R(\tau) = c_1 e^{w_1\tau} + c_1 e^{w_2\tau} + c_3 e^{-(\omega+\gamma)\tau}, \tag{62}$$

where the coefficients c_i and the exponents w_i are messy expressions involving the three switch parameters ω, γ, and λ (Ref. 8) and the mean service rate μ. The derivation of this covariance function is straightforward but tedious and is therefore omitted. (For the switched Poisson arrival prccess, the Markovian state equations can be solved for $P(x_t,x_{t+\tau})$, where x_t = state of server at epoch t.) Our purpose here is to explain how eq. (62) was used in generating Fig. 4.

If ρ is the mean occupancy of the server ($\rho = E(x_t)$) and a is the offered load in a GI/M/1-loss system, then it is easy to show that the peakedness[6] z satisfies

$$z = a\left(\frac{1-\rho}{\rho}\right). \tag{63}$$

For GI/M/1-loss system the call congestion is $\phi(\mu)$ and is related to the time congestion (ρ) by $a(1 - \phi(\mu)) = \rho$. So using $z(\mu) = 1/[1 - \phi(\mu)] - a$ yields the result.[2] ($\phi(\cdot)$ is the L.S. transform of the interarrival time distribution.) Therefore, specifying ρ and z uniquely determines a. Hence, with a and z known, we obtain the equivalent random parameters and use the three-moment match to obtain the switch parameters (see

Ref. 8). Using this procedure, we obtain $R(\tau)$ vs ρ parameterized by z. Equation (1) then yields $P_{1,0}$.

APPENDIX B

The Likelihood Statistic Based on the Observable Data

For ease of derivation, the likelihood statistic derived in Section 4.2 was based on the raw data $\mathbf{x}_m = (x_1,\cdots,x_m)$ rather than on the observable data $[t(m),n(m)]$; $t(m)$ and $n(m)$ are defined in Section 2.1. We will now study $\hat{\ell}_m(t,n)$ and verify that the two statistics differ only in their end-effect terms. We will also examine the end-effect term based on $t(m)$ and $n(m)$ and show how it "tracks" the end-effect term based on x_1 and x_m.

We begin by expressing the probability of $\mathbf{x}_m = (x_1,\cdots,x_m)$ in terms of $t(m)$, $n(m)$, x_1, and x_m:

Lemma 3: If $\{x_i\}$ is a binary state stationary Markov chain with transition probabilities $P_{0,1}$ and $P_{1,0}$ and if $\mathbf{x}_m = (x_1,\cdots,x_m)$, then

$$P(\mathbf{x}_m) = P(x_1) \times Q(x_1,x_m) \times P_{1,0}^{t/2}(1-P_{1,0})^{n-(t/2)} \times P_{0,1}^{t/2}(1-P_{0,1})^{m-n-(t/2)}, \quad (64)$$

where

$$Q(x_1,x_m) = P_{1,0}^{(x_1-x_m)/2}\, P_{1,1}^{-(x_1+x_m)/2}\, P_{0,1}^{-(x_1-x_m)/2}\, P_{0,0}^{(x_1+x_m)/2+1} \quad (65)$$

and

$$P(x_1) = \begin{cases} \rho & \text{if} \quad x_1 = 1 \\ 1-\rho & \text{if} \quad x_1 = 0 \end{cases}.$$

Proof: This result is obtained from lemma 2 using eqs. (12a) and (12b).

For convenience, we introduce the following notation

(*i*) $f\left(\frac{t}{2};n,P_{1,0}\right) = P_{1,0}^{t/2}(1-P_{1,0})^{n-(t/2)}$. If t is even then of course

$$f\left(\frac{t}{2};n,P_{1,0}\right) = \frac{b\left(\frac{t}{2};n,P_{1,0}\right)}{\binom{n}{t/2}}.$$

(*ii*) $g\left(\frac{t}{m};m-n,P_{0,1}\right) = P_{0,1}^{t/2}(1-P_{0,1})^{m-n-t/2}$.

(*iii*) $P_{x,y}(t,n) = P(t(m) = t, n(m) = n, x_1 = x, x_m = y)$.

(*iv*) $S_{x,y}(t,n)$ = number of binary m-tuples satisfying $t(m) = t$, $n(m) = n$, $x_1 = x$, and $x_m = y$.

Thus, we can write

$$P(t(m) = t, n(m) = n) = \sum_{x=0}^{1} \sum_{y=0}^{1} P_{x,y}(t,n) \tag{66}$$

and

$$P_{x,y}(t,n) = f\left(\frac{t}{2}; n, P_{1,0}\right) g\left(\frac{t}{2}; m-n, P_{0,1}\right) P(x) S_{x,y}(t,n) Q(x,y). \tag{67}$$

Equations (66) and (67) imply that

$$P(t, n(m) = t, n(m) = n)$$
$$= f\left(\frac{t}{2}; n, P_{1,0}\right) g\left(\frac{t}{2}; m-n, P_{0,1}\right) \left\{ \sum_{x=0}^{1} \sum_{y=0}^{1} P(x) S_{x,y}(t,n) Q(x,y) \right\} \tag{68}$$

and therefore it is easily seen that:

$$\hat{\ell}_m(t,n) = \log \frac{P^*(t(m) = t, n(m) = n)}{P(t(m) = t, n(m) = n)},$$
$$= \left[\alpha \frac{t}{2} - an \right] + \left[\beta \frac{t}{2} - b(n-m) \right] + E(t,n), \tag{69}$$

where

$$E(t,n) = \log \left\{ \frac{\sum_{x=0}^{1} \sum_{y=0}^{1} P^*(x) S_{x,y}(t,n) Q^*(x,y)}{\sum_{x=0}^{1} \sum_{y=0}^{1} P(x) S_{x,y}(t,n) Q(x,Y)} \right\}. \tag{70}$$

Comparing eqs. (70) and (22), we see that $\hat{\ell}_m(t,n)$ and $\hat{\ell}(\mathbf{x}_m)$ differ only in their respective end-effect terms. The following result is perhaps a bit surprising:

Lemma 4: $\hat{\ell}(t,n) = \hat{\ell}(\mathbf{x}_m)$ *if* t *is odd.*

Proof: t odd $\Rightarrow x_1 \neq x_m \Rightarrow e(0,1) = e(1,0) = a + b/2$ (see eq. (22b)). t odd $\Rightarrow S_{0,0}(t,n) = S_{1,1}(t,n) = 0$ so $E(t,n)$ for t odd may be written:

$$E(t,n) = \log \frac{(1-\rho) S_{0,1} Q^*(0,1) + \rho S_{1,0} Q^*(1,0)}{(1-\rho) S_{0,1} Q(0,1) + \rho S_{1,0} Q(1,0)}.$$

Now using (4) $E(t,n)$ for t odd can be manipulated into the following form

$$E(t,n) = \log \frac{(1-\rho) S_{0,1} Q^*(0,1) + \rho S_{1,0} Q^*(1,0)}{(1-\rho) S_{0,1} Q(0,1) + \rho S_{1,0} Q(1,0)}.$$

but $P_{0,1}/P_{1,0} = P^*_{0,1}/P^*_{1,0} = \rho/1-\rho$ which completes the proof.

If t is even then $x_1 = x_m$,

$$e(0,0) = b$$

and

$$e(1,1) = a.$$

For even t, eq. (70) can be written as

$$E(t,n) = \log \left\{ \frac{S_{0,0} + S_{1,1}}{\left(\frac{1-\rho}{1-P_{01}}\right) S_{0,0} + \left(\frac{\rho}{1-P_{10}}\right) S_{1,1}} \right\}.$$

Thus, for even t, $E(t,n)$ is a complicated function* of t and n. Note however that $E(t,n) \leqq 0$ and

$$E(t,n) = \begin{cases} a & \text{if } S_{0,0}(t,n) = 0 \\ b & \text{if } S_{1,1}(t,n) = 0 \end{cases}$$

It can be shown that

$$\frac{S_{1,1}(t,n)}{S_{0,0}(t,n)} = \frac{n - \frac{t}{2}}{m - n - \frac{t}{2}}$$

if (t,n) is such that $S_{1,1}$ and $S_{0,0}$ are nonzero.

APPENDIX C

The End Effect $E\{I(x_{lm};x_{lm+1})\}$

The end effect

$$E\{I(x_{im};x_{im+1})\} = H(x_{im+1}) - H(x_{im+1}/x_{im}),$$

where

$$H(x_{im+1}/x_{im}) = (1-\rho)H(x_{im+1}/0) + \rho H(x_{im+1}/1)$$

and

$$H(x_{im+1}),\ H(x_{im+1}/0) \text{ and } H(x_{im+1}/1)$$

are the binary entropy function $\mathcal{H}(x)$ evaluated at ρ, $P_{0,1}$ and $P_{1,0}$, respectively ($\mathcal{H}(x) = -x \log x - (1-x) \log (1-x)$, $0 \leqq x \leqq 1$).

Table I exhibits $E\{I(x_{im};x_{im+1})\}$ and $E\{\hat{\ell}(\mathbf{x}_m)\}$ as a function of the trunks occupancy ρ—for a normal trunk, with a single-hour accumula-

Table I — End effect as a function of occupancy

	(τ = 100 seconds)		(τ = 200 seconds)	
ρ	$E\{I(x_{im+1};x_{im})\}$	$E\{\hat{\ell}(x_m)\}$	$E\{I(x_{im+1};x_{im})\}$	$E\{\hat{\ell}(x_m)\}$
0.10	0.04	−1.36	0.01	−0.22
0.20	0.05	−1.60	0.01	−0.20
0.30	0.04	−1.50	0.01	−0.15
0.40	0.03	−1.21	0.01	−0.09
0.50	0.02	−0.84	0.00	−0.04
0.60	0.01	−0.47	0.00	−0.01

tion period (m = 36 and 18 for the 100- and 200-second sampling options, respectively). The M/M/1-loss model is used with $1/\mu$ = 180 seconds. It is clear that the mean end effect is negligible compared to the mean statistic update.

APPENDIX D

Occupancy Formulae for a Random-Selection Loss System

Theorem: *Consider an N server Markovian loss system with random selection of idle servers and*

(*i*) $N - 1$ *servers with mean service rate* μ.
(*ii*) *1 server with mean service rate* $r\mu$ $(r > 0)$.
(*iii*) *Mean arrival rate* λ.

Let ρ_r *and* ρ_r^* *denote the mean occupancy of the servers with mean service rates* μ *and* $r\mu$, *respectively, and let* $\hat{B}$ *denote the blocking (call congestion). Also let* ϕ *and* ϕ^- *denote the carried load per server in an N server and* $N - 1$ *server Markovian loss system, respectively, (assuming all servers have rate* μ), *given an offered lead* $a = \lambda/\mu$. *That is,* $\phi = a[1 - B(N,a)]/N$ *and* $\phi^- = a[1 - B(N-1,a)]/(N-1)$, *where* $B(\cdot,\cdot)$ *is the Erlang blocking formula. Then*

$$(i)\quad \rho_r^* = \frac{\phi}{r - (r-1)\phi}. \tag{71}$$

$$(ii)\quad \hat{B}(N,a,r) = \frac{B(N,a)}{r - (r-1)\phi} \quad [B(\cdot,\cdot) \text{ is the Erlang} \tag{72}$$

blocking formula].

$$(iii)\quad \phi \geqq \rho_r \geqq \left(\frac{r - (r-1)[N/(N-1)]\phi}{r - (r-1)\phi}\right)\phi. \tag{73}$$

Proof:

$$(iv)\quad \rho_r = \left(\frac{r - (r-1)\phi^-}{r - (r-1)\phi}\right)\phi. \tag{74}$$

(*i*) This part of the theorem follows immediately from eq. (34b)

$$\rho_r^* = \frac{1}{1 + \dfrac{rN}{a} - r[1 - B(N-1,a)]}$$

by using the well known Erlang B recursion[6]

$$\frac{B(N,a)}{1 - B(N,a)} = \frac{a}{N} B(N-1,a)$$

and the expression for the mean group occupancy

$$\frac{a[1 - B(N,a)]}{N} = \phi.$$

(*ii*) This part of the theorem follows immediately from eq. (34a) by using the above expression for mean-group occupancy.

(*iii*) The lower bound part of eq. (73) follows from eq. (34c) by noting that $\hat{B}(N,a,r) \leqq B(N,a)$† and using eq. (71) of this theorem. Thus, we obtain

$$\frac{r\phi}{r - (r-1)\phi} + (N-1)\rho_r \geqq a[1 - B(N,a)] = N\phi$$

which can be arranged to yield the lower bound in eq. (73). The upper bound in (73) is an immediate consequence of (74) since $\phi^- \geqq \phi$. Thus, it remains to prove (74).

(*iv*) We prove this part as follows: Using eqs. (71) and (72) in the conservation eq. (34c),

$$r\rho_r^* + (N-1)\rho_r = a(1 - \hat{B}) \tag{75}$$

yields

$$\rho_r = \frac{a(r-1)(1-\phi) + (N-r)\phi}{(N-1)[r - (r-1)\phi]}. \tag{76}$$

Therefore, eq. (74) holds if and only if

$$[r - (r-1)\phi^-] \times \phi = \frac{a(r-1)(1-\phi) + (N-r)\phi}{(N-1)}. \tag{77}$$

The right-hand side of eq. (77) can be rewritten as:

$$\text{rhs} = \frac{a}{N(N-1)}\{r(N-1-a) + a - B[N - (a+1)r + a]\} \tag{78}$$

† This follows eq. (34a) by noting that $a[1 - B(N,a)] \leqq N$.

using

$$\phi = \frac{a[1 - B(N,a)]}{N}.$$

Similarly, using $\phi^- = a[1 - B(N-1,a)]/(N-1)$, the left-hand side of eq. (77) can be written as:

$$\text{lhs} = \{r - (r-1)(a/N - 1)[1 - B(N-1,a)]\}(a/N)[1 - B(N,a)]. \quad (79)$$

Now using the Erlang B recursion formula in eq. (79) and rearranging terms yields eq. (78). Thus, eq. (77) and consequently eq. (74), of the theorem is proved.

APPENDIX E

The Sign of the Mean Statistic Update

The mean statistic update of the optimal 5XB group algorithm is given by eq. (51a):

$$E\{\hat{\ell}(\mathbf{x}_m)\} = (m-1)\{(\alpha P_{1,0} - a)\rho + (\tilde{b} - \tilde{\beta} P_{0,1})(1-\rho)\} + \log \frac{P^*(x_1)}{P(x_1)}.$$

The mean update of the ad hoc and the additional statistics correspond to $\tilde{b} = \tilde{\beta} = 0$ and $a = \alpha = 0$, respectively. The following lemma is needed to study the sign of $E\{\hat{\ell}(\mathbf{x}_m)\}$:

Lemma 5: If $0 < q < p < 1$ *define*

$$a = \log\left(\frac{1-q}{1-p}\right) \text{ and } \alpha = a + \log\frac{p}{q},$$

then

$$q < \frac{a}{\alpha} < p.$$

Proof: Consider the log likelihood ratio statistic $\hat{\ell}(m)$ for this simple hypothesis testing context: the observed process is $n(m) = \sum_{i=1}^{m} x_i$, where x_i are i.i.d. Bernoulli random variables with $P(x_i = 1)$ given by q and p under H_0 and H_1, respectively. Therefore,

$$\hat{\ell}(m) = \log \frac{b(n;m,p)}{b(n;m,q)} = \alpha n - am$$

and hence the mean of $\hat{\ell}(m)$ is $m(\alpha q - a)$ under H_0 and $m(\alpha p - a)$ under H_1. But as we noted in Section 6.2, $E\{\hat{\ell}\}$ is negative under H_0 and positive under H_1 in a general discrete setting.* Thus, we must have

$$\alpha q - a < 0$$

and

$$\alpha p - a > 0,$$

which completes the proof.

An immediate consequence of this lemma is that

$$(i)\quad P_{1,0}(\phi,1) < \frac{a(\phi,r_0)}{\alpha(\phi,r_0)} < P_{1,0}(\phi,r_0)$$

and

$$(ii)\quad P_{0,1}(\phi,r_0) < \frac{\tilde{b}(\phi,r_0)}{\tilde{\beta}(\phi,r_0)} < P_{0,1}(\phi,1),$$

where a and α are defined by eqs. (41a) and (41b) and $\tilde{b}$ and $\tilde{B}$ are defined by eqs. (45a) and (45b). Since $r \geqq r_0$ implies that $P_{1,0}(\phi,r) \geqq P_{1,0}(\phi,r_0)$ and $P_{0,1}(\phi,r) \leqq P_{0,1}(\phi,r_0)$, we see that

$$\operatorname{sgn}(\alpha(\phi,r_0)P_{1,0}(\phi,r) - a(\phi,r_0)) = \operatorname{sgn}(\tilde{b}(\phi,r_0) - P_{0,1}(\phi,r)\tilde{B}(\phi,r_0))$$

$$= \begin{cases} \text{positive} & \text{if } r \geqq r_0 \\ \text{negative} & \text{if } r = 1. \end{cases}$$

REFERENCES

1. A. Klimontowicz, "Grade of Service for Full Available Trunk Groups with Faulty Trunks," Proc. Sixth International Teletraffic Conference, Munich, 1970, pp. 212/1–212/6.
2. L. J. Forys and E. J. Messerli, "Analysis of Trunk Groups Containing Short-Holding Time Trunks," B.S.T.J., Vol. 54, No. 6, July–August, 1975, pp. 1127–1154.
3. E. J. Messerli, "Faulty Trunks: A New Look at an Old Problem," Bell Laboratories Record, June 1975, pp. 278–284.
4. W. S. Gifford and J. Shapiro, "Effect of the Change in Mean Holding Time Associated with an Equipment Irregularity on Network Trouble Detection and Customer Service," IEEE Trans. Commun., *COM-21*, No. 1 (January 1973), pp. 1–3.
5. R. E. Machol, Acquiring Data for Network Planning and Control," Bell Laboratories Record, October 1974, pp. 279–285.
6. R. B. Cooper, *Introduction to Queueing Theory*, New York: MacMillan, 1972.
7. E. Parzen, *Stochastic Processes*, Holden-Day, 1962.
8. A. Kuczura, "The Interrupted Poisson Process," B.S.T.J., *52*, No. 3 (March 1973), pp. 437–448.
9. V. B. Iverson, "Analysis of Traffic Process Based on Data Obtained by the Scanning Method, Proc. Seventh International Teletraffic Conference, Stockholm, 1973, pages 224/1–224/10.
10. E. L. Lehman, *Testing Statistical Hypotheses*. New York: John Wiley, 1959.
11. A. Wald, *Sequential Analysis*, New York: John Wiley, 1947.
12. R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.
13. W. Feller, *An Introduction to Probability Theory and Its Applications*, New York: John Wiley, 1957.
14. J. S. Kaufman, "Distrihution of Busy Trunks in a Random Hunt Group Containing Killer Trunks," unpublished work.
15. V. E. Benes, *Mathematical Theory of Connecting Networks and Telephone Traffic*, New York: Academic, 1965.
16. Z. Govindarajulu, *Sequential Statistical Procedures*, New York: Academic, 1975.
17. D. R. Cox and H. D. Miller, *Theory of Stochastic Processes*, Wiley: Methuen, 1965.
18. D. R. Cox, Renewal Theory. Methuen; Wiley, 1962.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 6, July–August 1977
Printed in U.S.A.

Cancellation of Polarization Rotation in an Offset Paraboloid by a Polarization Grid

By T. S. CHU

(Manuscript received July 20, 1976)

The polarization rotation properties of the field radiated from a polarization grid have been found to be similar to those in the aperture of an offset parabolic antenna. This observation suggests broadband cancellation of the polarization rotation in a large offset reflector by the opposite prerotation of the incident feed radiation via a polarization grid. Reduction of cross polarization from −24 dB without the grid to −39 dB with the grid wires is predicted in a numerical example. A previously unexplained polarization rotation measured using an offset parabolic grid is shown to be in good agreement with calculation.

I. INTRODUCTION

Current interest in frequency reuse through orthogonal polarizations creates a strong incentive for improving polarization properties of antennas. Cross-polarized radiation from an offset reflector[1] is often regarded as a blemish on an otherwise excellent antenna, which offers both low sidelohe level and good impedance matching.[2] Although the cross polarization can be minimized using a large effective f/D ratio, the corresponding requirements of small offset angle and large feed aperture are not always convenient in applications. Recently a trimode feed horn[3] has been discussed as a means of reducing the cross-polarized radiation of offset reflectors; however, a multimode arrangement is inherently narrow in bandwidth. Our purpose here is to propose a broadband method for reducing the cross-polarized radiation of the offset reflector antenna. This scheme is based upon cancellation of polarization rotation due to reflector curvature[1] by opposite rotation generated by a polarization grid.[4]

In Section II we examine the similarity between cross polarization in the aperture of an offset reflector and that of a polarization grid. The condition of cancelling the first-order cross-polarization terms is then

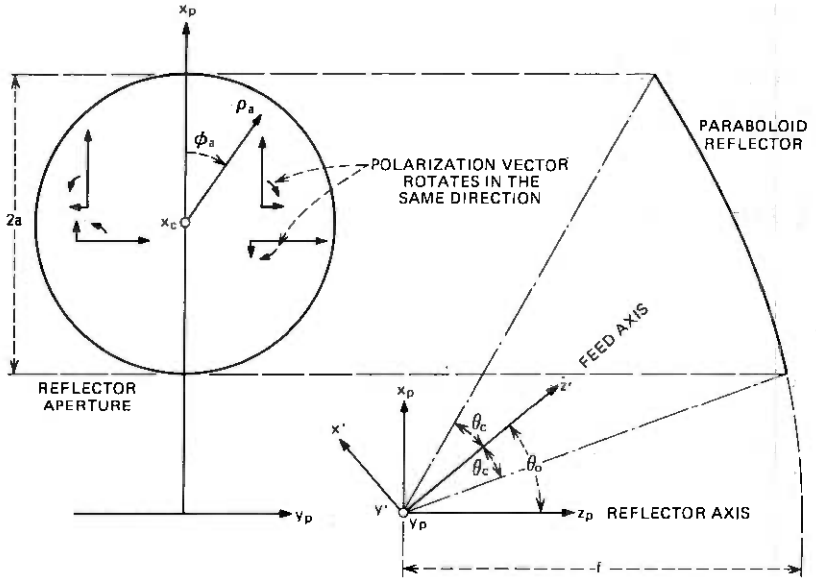

Fig. 1—Geometry of offset reflector.

deduced. In Section III numerical calculations of practical examples are given.

In the appendix the previously unexplained measured polarization rotation* of a parabolic grid[5] is calculated. This example is given here as experimental evidence of agreement with the predicted rotation of radiation from a polarization grid.

II. CANCELLATION OF POLARIZATION ROTATION

Let us first briefly review the salient properties of the cross-polarized field in the aperture of an offset paraboloid as shown in Fig. 1. For a balanced feed radiation,

$$\mathbf{E}_f = F(\theta', \phi') \begin{bmatrix} \cos\phi' & \sin\phi' \\ \hat{\theta}' \mp & \hat{\phi}' \\ \sin\phi' & \cos\phi' \end{bmatrix} \frac{\exp(-j\rho)}{\rho}, \tag{1}$$

where (ρ, θ', ϕ') are spherical coordinates with respect to the z' axis and $(\hat{\theta}', \hat{\phi}')$ are the corresponding unit vectors. Since the reflected field from the paraboloid is

$$\mathbf{E}_r = -\mathbf{E}_f + 2\hat{n}(\mathbf{E}_f \cdot \hat{n}),$$

* The design was initially proposed by Comsat for the Comstar satellite.

where $\hat{n}$ is a unit vector normal to the reflector surface. The principal and cross-polarized field components in the reflector aperture can be written respectively:[1]

$$M = \mathbf{E}_r \cdot \frac{\hat{x}_p}{\hat{y}_p} = \frac{F(\theta', \phi')}{t\rho} [\sin\theta' \sin\theta_0 \cos\phi' - \sin^2\phi' \cdot (\cos\theta_0 + \cos\theta') - \cos^2\phi'(1 + \cos\theta_0 \cos\theta')] \quad (2)$$

$$N = \mathbf{E}_r \cdot \frac{\hat{y}_p}{\hat{x}_p} = \mp \frac{F(\theta',\phi')}{t\rho} [\sin\theta' \sin\theta_0 \sin\phi' - \sin\phi' \cos\phi'(1 - \cos\theta')(1 - \cos\theta_0)], \quad (3)$$

where $t = 1 + \cos\theta' \cos\theta_0 - \sin\theta' \sin\theta_0 \cos\phi'$, $M^2 + N^2 = F^2/\rho^2$, and N vanishes when $\theta_0 = 0$. The offset angle θ_0 is between the feed axis and the reflector axis. The sign combination in eqs. (2) and (3) indicates that the rotation of the polarization vector due to offset in a paraboloidal aperture has the same magnitude and is in the same direction as illustrated in Fig. 1 for any orientation of the incident linear polarization. The projection of the intersection of a circular cone (with vertex at the focus) and the offset paraboloid onto the $x_p y_p$ plane is a circular aperture with center

$$x_c = \frac{2f \sin\theta_0}{\cos\theta_0 + \cos\theta_c} \quad (4)$$

and radius

$$a = \frac{2f \sin\theta_c}{\cos\theta_0 + \cos\theta_c}, \quad (5)$$

where θ_c is the half angle of the cone. Equations (4) and (5) will be used later to obtain the relations in eqs. (12) and (13).

Radiation from transmitting and reflecting wire grids can be obtained by magnetic and electric current sheet models, respectively. The principal and cross-polarized components are[4]

$$P = -C[1 - \cos^2\phi' (1 - \cos\theta') - \sin\theta' \cos\phi' \tan\delta] \quad (6)$$

$$X = \pm C[\sin\phi' \cos\phi'(1 - \cos\theta') + \sin\theta' \sin\phi' \tan\delta], \quad (7)$$

where C is a proportionality constant, θ' and ϕ' are the spherical coordinates of the feed (z') axis, and δ is the angle between the conducting wires and the $x'y'$ plane as shown in Fig. 2; the expressions inside the brackets are identical to those of eqs. (9) and (10) in Ref. 4, provided one makes the following substitutions: $\phi' = \phi - 90°$ and $\delta = 90° - \gamma$. The changes of notation are made for the purpose of comparison with eqs. (2) and (3). The upper and lower signs in eq. (7) correspond to the transmitting and reflecting cases. The orientations for the transmitting

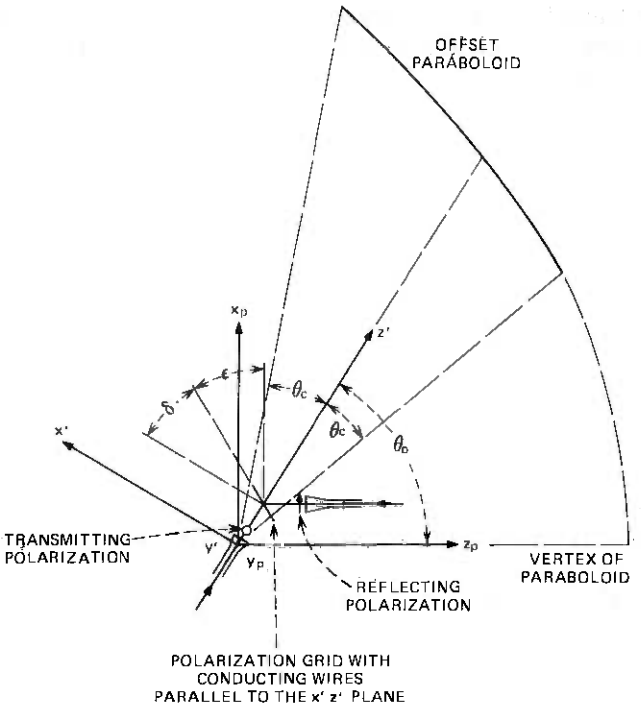

Fig. 2—Configuration for cancellation between polarization rotations of an offset paraboloid and a polarization grid.

and reflecting polarizations together with a given grid geometry are shown in Fig. 2 where the conducting wires are parallel to the plane of the figure.

One notes that the leading terms, which contain first power of θ', in eqs. (3) and (7) have the same sinusoidal dependence on θ' and ϕ'. Furthermore, the sign combination in eqs. (6) and (7) indicates that the transmitting and reflecting orthogonal polarizations rotate in the same direction, opposite to the rotation in the aperture of an offset paraboloid.

Let us take the first-order approximation—i.e., $\cos\theta' \approx 1$—in eqs. (2), (3), (6) and (7), but $\sin\theta'$ is retained. Then the cross polarization in eq. (3) normalized with respect to eq. (2) cancels that in eq. (7) normalized with respect to eq. (6):

$$\frac{X}{P} + \frac{N}{M} = 0 \text{ if } \delta = \frac{\theta_0}{2}. \tag{8}$$

Polarization rotation of the radiation from a wire grid, as predicted by eqs. (6) and (7), also explains data measured using a cylindrical re-

flector made of a curved wire grid. The report on this experiment[5] misinterpreted the polarization rotation as a consequence of diffraction and the offset geometry. A comparison between the calculated and measured polarization rotations of this case is given in the Appendix.

III. NUMERICAL EXAMPLES

Since the cancellation of polarization rotation discussed in the preceding section only eliminates the leading terms, it is of interest to determine the residual cross polarization. Assuming that the offset reflector is located in the far zone of the radiation from a wire grid, as shown in Fig. 2, the principal and cross-polarized components in the reflector aperture can be written

$$P = F(\theta')[1 - \cos^2 \phi_p\,(1 - \cos \theta_p) + \sin \theta_p \cos \phi_p \tan \epsilon] \tag{9}$$

$$X = \mp F(\theta')\,[\sin \phi_p \cos \phi_p\,(1 - \cos \theta_p) - \sin \theta_p \sin \phi_p \tan \epsilon], \tag{10}$$

where $F(\theta')$ is the feed-radiation pattern and

$$\theta' = \cos^{-1}\,[\cos \theta_p \cos \theta_0 + \sin \theta_p \sin \theta_0 \cos \phi_p]. \tag{11}$$

The above equations are simply a decomposition of the grid radiation into the two orthogonal components of a balanced feed whose axis coincides with the paraboloidal axis. The expressions inside the brackets of eqs. (9) and (10) are of the same form as those of eqs. (6) and (7); but θ_p and ϕ_p are the spherical coordinates with respect to the paraboloidal (z_p) axis instead of the feed axis, and $\epsilon = (\theta_0 - \delta)$ is the angle between the conducting wires and the $x_p y_p$ plane, as shown in Fig. 2.

To relate eqs. (9) and (10) to the normalized aperture coordinates $r = (\rho_a/a)$ and ϕ_a, the following expressions can be obtained with the aid of eqs. (4) and (5):

$$\theta_p = 2 \tan^{-1}\left[\frac{\sqrt{(r \sin \theta_c \sin \phi_a)^2 + (\sin \theta_0 + r \sin \theta_c \cos \phi_a)^2}}{\cos \theta_c + \cos \theta_0}\right] \tag{12}$$

$$\phi_p = \tan^{-1}\left[\frac{r \sin \theta_c \sin \phi_a}{\sin \theta_0 + r \sin \theta_c \cos \phi_a}\right]. \tag{13}$$

Numerical examples of several combinations of parameters (θ_0, θ_c and ϵ) have been calculated for the principal and residual cross-polarization components from eqs. (9) and (10). The feed pattern has a gaussian shape with 10-dB taper at the edge of the reflector. The principal polarization is close to unity (0 dB) around the center of the reflector aperture. The maximum, calculated, residual cross polarization is given in Table I for a number of examples. Fig. 3 shows a plot of both principal and cross polarizations for the case $\theta_0 = 50°$, $\theta_c = 20°$, and $\epsilon = 25°$. Only half of the

Table I — Cross Polarization in the aperture of an offset reflector

θ_0 (Deg)	θ_c (Deg)	ϵ (Deg)	Max. Residual Cross Pol. With Grid (dB)	Max. Cross Pol.[6] Balanced Feed Without Grid (dB)
50	20	25	−38.6	−24.0
50	20	23	−36.4	−24.0
50	20	27	−36.1	−24.0
60	20	30	−38.0	−22.5
60	30	30	−30.9	−18.0
90	20	45	−34.3	−17.5
90	14	45	−40.5	−20.0

aperture needs to be shown for each polarization because of the symmetry. The maximum residual cross polarization in this case is −38.6 dB, reduced from −24 dB for the same reflector aperture illuminated by a balanced feed without a polarizer. Keeping the same set of parameters, $\theta_0 = 50°$ and $\theta_c = 20°$, the calculated residual cross polarization

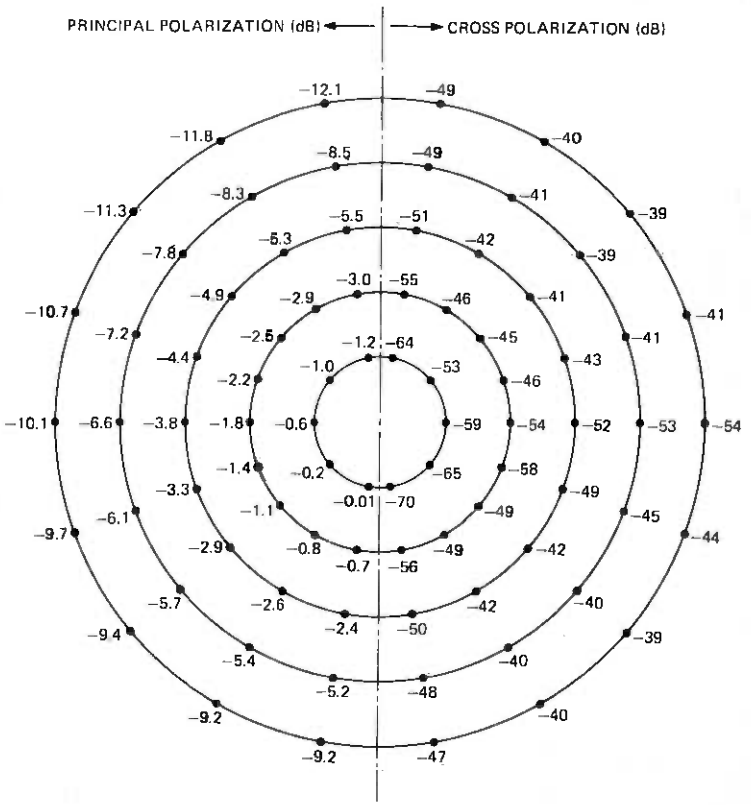

Fig. 3—Relative amplitude levels in the aperture of an offset paraboloid. $\theta_0 = 50°$, $\theta_c = 20°$, $\epsilon = 25°$.

becomes −36.4 dB for $\epsilon = 23°$ and −36.1 dB for $\epsilon = 27°$. These results indicate that the residual cross polarization is not overly sensitive to a slight departure from the optimum orientation of $\epsilon = \theta_0/2$.

The examples for $\theta_0 = 60°$ and $\epsilon = 30°$ show residual cross polarizations of −38.0 dB and −30.9 dB for $\theta_c = 20°$ and 30°, respectively. The second-order terms are not quite negligible at $\theta_c = 30°$; however, the cancellation of cross polarization appears to be significant.

IV. DISCUSSION

In view of the residual second-order $(1 - \cos \theta')$ terms, the half-cone angle of the reflector subtended at the focus should not exceed about 20° in order to take full advantage of the cancellation. If the reflectors, such as in an offset cassegrain configuration, do not cause significant cross polarization, the conducting direction of a polarization diplexing grid should be oriented to avoid introducing any polarization rotation, as discussed in Ref. 4.

When $\theta_0 - \theta_c$ is less than about 30° and $\epsilon = \delta = \theta_0/2$, the feed horn associated with the polarization reflected from the grid will introduce some blockage, as shown in Fig. 2. The blocking problem can be eased by using a smaller value of ϵ. This practical difficulty may prevent optimum orientation of the grid wires for reflectors of small offset angle, and hence reduce the effectiveness of the cancellation. The practical application of this scheme should indeed lie in reflectors with large offset angle.

The expression $\epsilon = \theta_0/2$ implies that the grid wires are approximately parallel to the tangent plane at the center of the offset reflector. One notes the similarity between this case and a symmetrical small-cone-angle paraboloid illuminated by a grid-covered feed.

ACKNOWLEDGMENT

The author is indebted to M. J. Gans and D. C. Hogg for valuable discussions, and to Mrs. Diane Vitello for assistance with the computation.

Appendix

Polarization Rotation of an Offset Parabolic Grid

An offset cylindrical-reflector system fed by two line sources of pillbox type (as shown in Fig. 4) was proposed[5] as a dual-polarized antenna with an elliptically shaped coverage pattern. The reflecting system consists of a vertically polarized grid attached to the surface of a parabolic cylinder, the front surface of the grid having the same curvature as the cylindrical reflector. The measured data[5] showed excellent orthogonality (within 1°) between vertical and horizontal polarizations over the whole

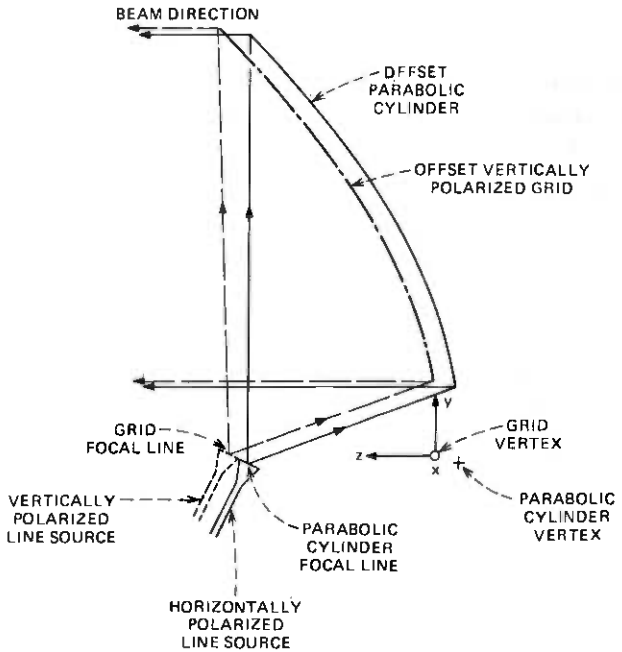

Fig. 4—Configuration of a dual-polarized cylindrical reflector antenna (Ref. 5).

6.8° × 3.4° (3 dB) elliptical beam. However, significant polarization rotations, in the same direction for both polarizations, were observed. Maximum rotations of about 2° occur on the major axis (xz plane) of the half-power ellipse of the beam. No adequate explanation was given for this measured polarization rotation. Here we explain the rotation using electric and magnetic current sheet models for the grid.

The first-order approximation of polarization rotation by a wire grid is simply $\sin\theta \cos\phi \cot\gamma$ (eq. 7 in Section II or eq. (10) in Ref. 4), where θ is the angle off the beam axis, $\cos\phi$ is unity in the xz plane of maximum rotation, and γ the angle between the conducting wires and the beam axis. Since the wires of the curved grid have a variable direction, the rotation Δ in the plane of maximum rotation can be obtained by averaging $\cot\gamma$ over the parabolic curve, $y^2 = 4fz$.

$$\Delta = \frac{\sin\theta}{\int E\,ds} \int E \cot\gamma\, ds = \frac{\sin\theta \int E \dfrac{dz}{dy} \sqrt{(dy)^2 + (dz)^2}}{\int E \sqrt{(dy)^2 + (dz)^2}}, \quad (14)$$

where E is the aperture field distribution. Let $u = y/2f$, eq. (14) be transformed into

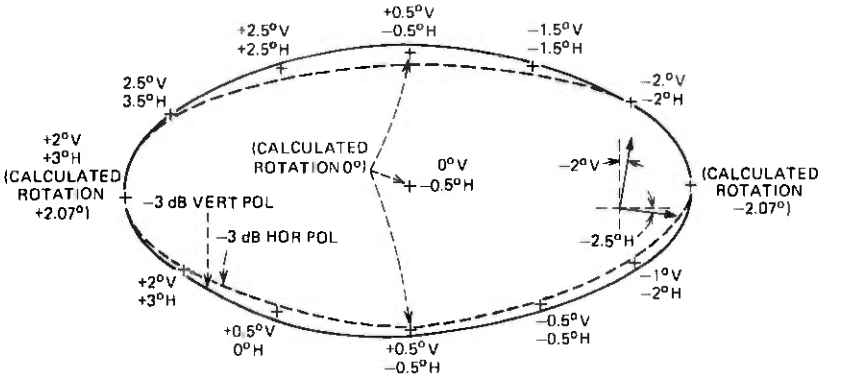

Fig. 5—Measured angles of nominal vertical and horizontal polarizations around −3-dB pattern contours of 6.8° × 3.4° for a dual-polarized cylindrical antenna with a parabolic grid; + indicates locations of measurement.

$$\Delta = \frac{\sin\theta \int_{0.1763}^{1} E\sqrt{1+u^2}\,u\,du}{\int_{0.1763}^{1} E\sqrt{1+u^2}\,du}, \tag{15}$$

where the upper and lower limits of integration are obtained from $y = 2f \tan \psi/2$ with $\psi = 90°$ and 20°, respectively. Δ is not sensitive to the aperture distribution. Numerical calculation gives $\Delta = 0.61 \sin\theta$ for both cases when E is assumed to be uniform and when E has a 10-dB edge taper—i.e., when

$$E = -0.396 + 4.746\left(\frac{y}{2f}\right) - 4.034\left(\frac{y}{2f}\right)^2. \tag{16}$$

The quadratic form has been chosen to perform the integration in closed form. When $\theta = 3.4°$, $\Delta = 0.0362$ rad = 2.07°; agreement between this calculated value and Wilkinson's measured rotation shown in Fig. 5 is indeed very good.

Furthermore, eqs. (6) and (7) in Section II indicate that the transmitted and reflected orthogonally polarized fields from the grid rotate in the same direction, just as in the measured rotations. The planes of measured maximum and null rotation also agree with the predictions. Since the two polarizations have the same sense of rotation, orthogonality is preserved. However, if this rotation occurs in the feed radiation, nonorthogonal elliptically polarized radiations from the reflector will result.

REFERENCES

1. T. S. Chu and R. H. Turrin, "Depolarization Properties of Offset Reflector Antennas," IEEE Trans. Ant. Propag., *AP-21* (May 1973), pp. 339–345.

2. C. Dragone and D. C. Hogg, "The Radiation Pattern and Impedance of Offset and Symmetrical Near-Field Cassegrainian and Gregorian Antennas," IEEE Trans. Ant. Propag. *AP-22* (May 1974), pp. 472–475.
3. A. W. Rudge and N. A. Adatia, "New Class of Primary-Feed Antennas for Use with Offset Parabolic-Reflector Antennas," Electron. Lett. *11* (November 1975), pp. 597–599.
4. T. S. Chu, M. J. Gans, and W. E. Legg, "Quasi-Optical Polarization Diplexing of Microwaves," B.S.T.J., *54* (December 1975), pp. 1665–1680.
5. E. J. Wilkinson, "A Dual-Polarized Cylindrical Reflector Antenna for Satellite Communication," Microw. J., *16* (December 1973), pp. 27–30, 62.
6. M. J. Gans and R. A. Semplak, "Some Far-Field Studies of an Offset Launcher," B.S.T.J., *54* (September 1975), pp. 1319–1340.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 6, July–August 1977
Printed in U.S.A.

Ten Years of Power Aging of the Same Group of Submarine Cable Semiconductor Devices

By A. J. WAHL

(Manuscript received January 18, 1977)

The active devices in the first Bell System transistorized submarine cable system (SF) are unpassivated, diffused-base, germanium transistors and oxide-passivated, silicon diodes. At the date of this writing these devices have accumulated over 550 million device hours of powered operation in service and on the aging rack without failure or impaired device performance for a demonstrated failure rate of less than 0.00045 percent per thousand hours (4.5 FITs) with 90-percent confidence. This paper reports details of the behavior of 500 of these devices that have reached ten years of controlled power aging. The overall results indicate that initial reliability objectives are being achieved and that the semiconductor devices should not be expected to limit the desired life span of SF submarine cables.

I. INTRODUCTION

In the initial stages of the project, the decision to use germanium transistors in the first Bell System transistorized submarine cable was a rather bold gamble to take advantage of their then superior frequency capability while also attempting to mitigate any reliability inferiority to silicon transistors. When the decision was made, in the time period of 1960–61, silicon transistors were clearly the wave of the future and, although still inferior to germanium transistors in frequency capability, had already indicated their reliability superiority. Nevertheless, the results reported here, together with the results of regular aging and actual service experience thus far, offer rather convincing evidence that the desired reliability objectives are being achieved. The first submarine cable using these transistors, a relatively short cable with 136 repeaters, has been in service over eight years. A transatlantic cable with 363 repeaters has been in service nearly seven years. No semiconductor device failure nor degradation toward failure has yet been observed in any of the systems that have no redundancy and in which the failure of even

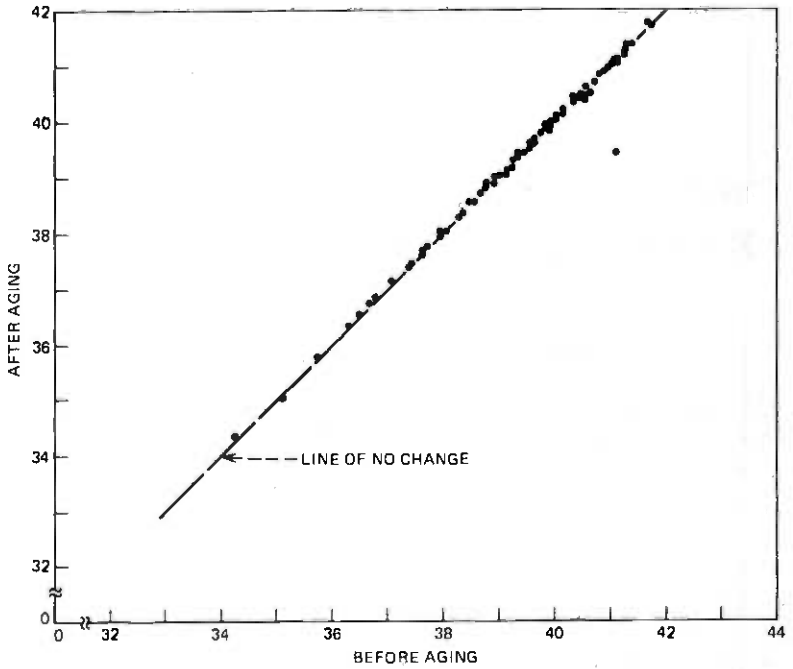

Fig. 1—Collector-base breakdown voltage in volts at 100 μA for L2287 transistors—before and after 10 years of power aging at 75 mW.

one device could disable the system. At the time of this writing the device-hours of service in these cables totals more than 250 million for both transistors and diodes. On the aging rack these devices have accumulated over 300 million device-hours of power aging—also without failure. On the basis of these results the demonstrated failure rate, with 90 percent confidence, is less than 0.00045 percent per thousand hours (4.5 FITs).

The method of providing semiconductor devices for the SF submarine cable system has been described previously.[1] As a further part of this program, a hundred devices of each of the five basic device codes of transistors and pn diodes have been retained on extended power aging. A milestone was observed in this program when the aging time on these devices reached ten years. The primary purpose of this paper is to show in a condensed way the behavior of these devices across ten years of power aging.

The transistors are of diffused-base germanium with no passivation or junction protection on the active chip. The diodes are of diffused silicon with grown oxide passivation over the junctions. All devices are encapsulated in hermetic packages.

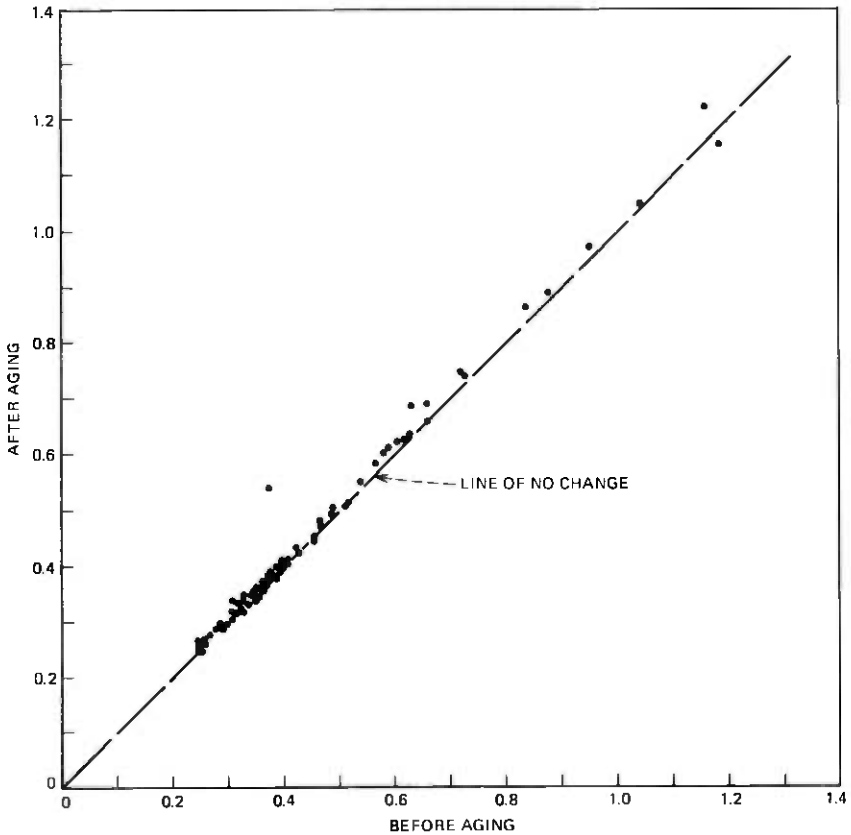

Fig. 2—Collector-base reverse leakage current in μA at 15 V for L2287 transistors—before and after 10 years of power aging at 75 mW.

II. DATA COLLECTION

All transistor parameters are measured before the devices are placed on long-term power aging, which is a period of about six months for candidate devices. From previous work with devices of this kind, we found that device current gain was the only parameter expected to show any consistent aging trend, and it is also the most critical parameter for the repeater amplifier circuit. For practical testing reasons during long-term aging, the dc current gain h_{FB} was measured. From the standpoint of the repeater circuit, however, the critical parameter is the low-frequency ac current gain h_{fb} rather than the dc current gain h_{FB}. The two are not the same, but they bear a relationship to each other, as described in any standard text on transistor theory. If one shifts, the other shifts in the same direction.

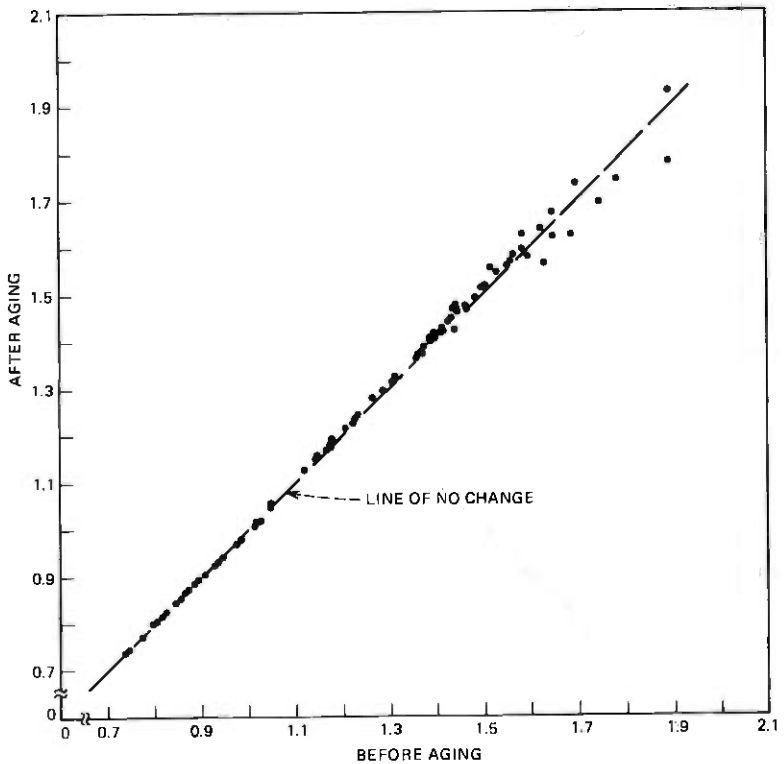

Fig. 3—Emitter-base breakdown voltage in volts at 100 μA for L2287 transistors—before and after 10 years of power aging at 75 mW.

Previous work with transistors of this kind had also indicated that any interruption of the bias on the device would cause the current gain to suffer a disturbance that would tend to obscure the real aging trend. For this reason, during long-term power aging, only the dc current gain was measured, and the measurement was made very precisely without disturbing the bias on the device. From the standpoint of device parameters directly relevant to circuit performance, therefore, the complete picture of the aged transistor cannot be known until the device is removed from long-term aging and all device parameters have again been measured. For such transistors, the aging time interval is ten years rather than six months.

Since the aging trends of diode parameters were not disturbed by change of electrical bias for testing purposes, most of the sensitive diode parameters were tested during aging. The diodes, therefore, were not removed from the aging rack at the end of the ten-year aging time. Interim aging information on these diodes has already been published.[2]

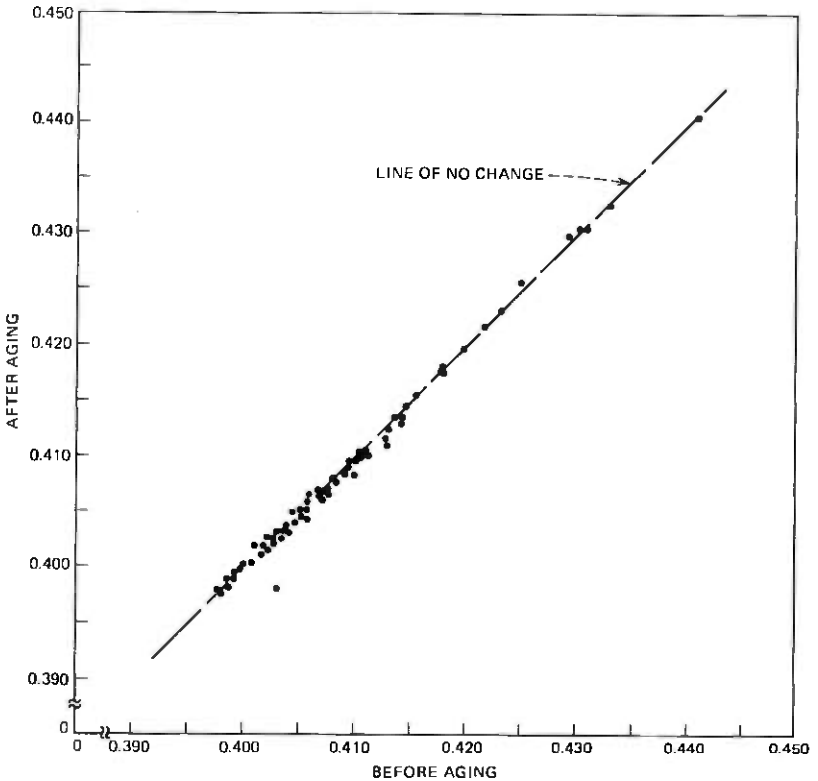

Fig. 4—Emitter-base forward voltage in volts at 15 mA for L2287 transistors—before and after 10 years of power aging at 75 mW.

III. PRESENTATION OF THE AGING RESULTS

A rather simple and clear picture of device aging behavior can be shown in the scatter plot in which the initial value of a device parameter is plotted vs the final value across the aging period for each device. For the transistors in this case, the information obtained at times between the initial and final values is not available for reasons discussed previously. For the diodes the information is available, but not significant since general aging trends, where they exist at all, are monotonic. For simplicity and consistency, therefore, all the ten-year aging results are shown here in the form of scatter plots.

IV. COMMENTS ON THE RESULTS

4.1 General

In most cases where any detectable parameter shifts have occurred across ten years of aging, the magnitudes of the shifts have been small

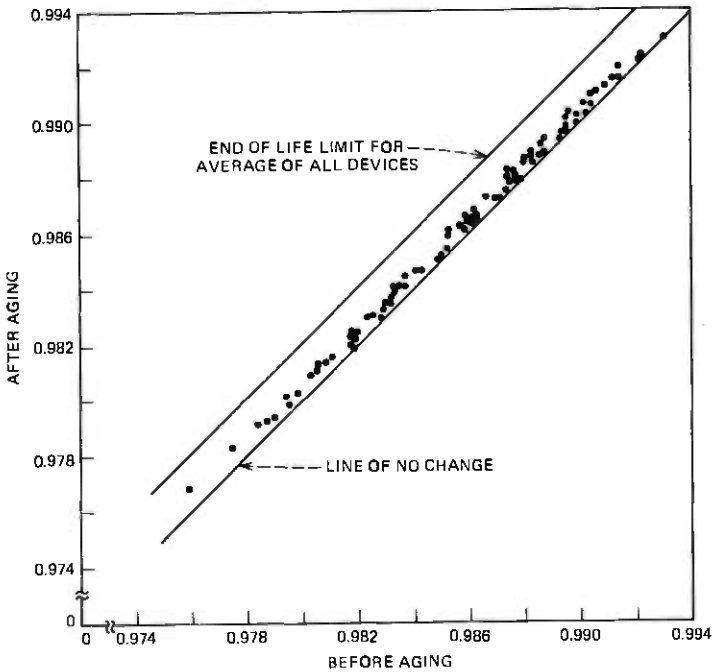

Fig. 5—Low-frequency common-base current gain for L2287 transistors—before and after 10 years of power aging at 75 mW.

and very uniform among all the devices, almost as if the devices had shifted in unison. Such behavior immediately raises the suspicion that the apparent shift may actually be an offset in testing rather than a real

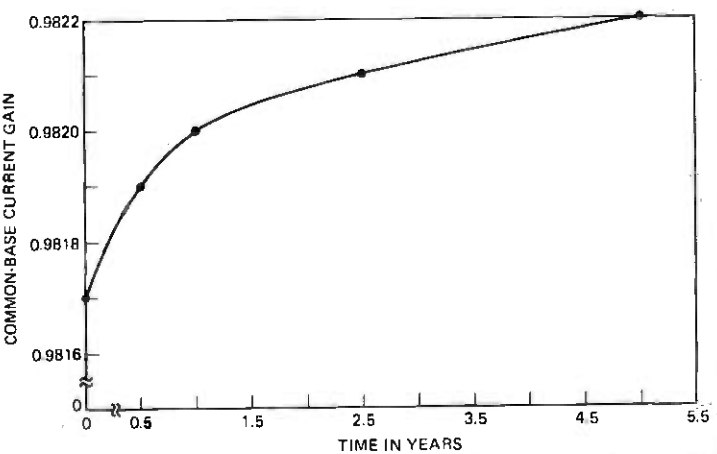

Fig. 6—Typical behavior of dc common-base current gain for diffused-base germanium transistors on long-term aging.

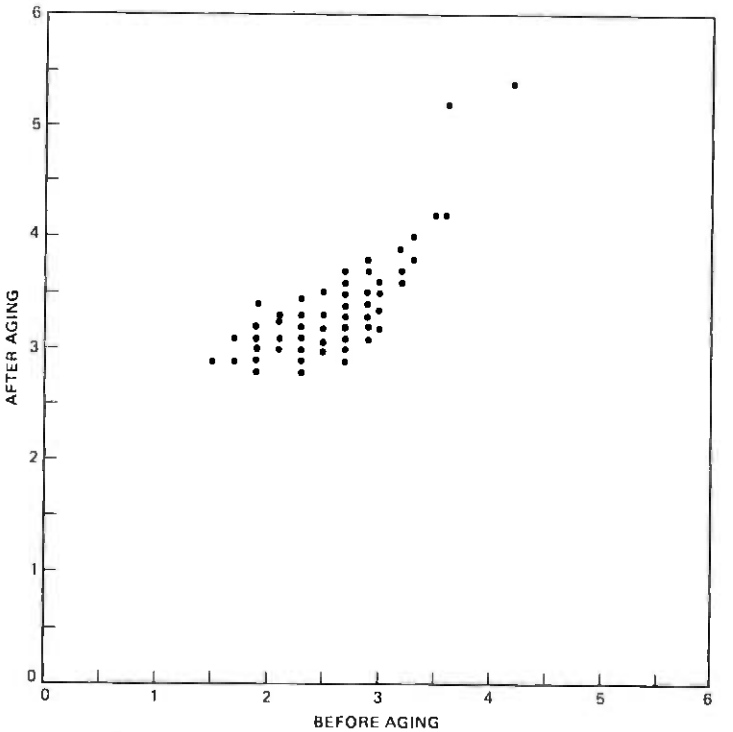

Fig. 7—Noise figure in dB at 5.5 MHz for L2287 transistors—before and after power aging at 75 mW.

shift. Unless the testing offset can be definitely confirmed, however, the shift must be accepted at face value. In the latter case, the uniformity of the shift makes the extrapolation very simple for more devices and longer time. The extrapolated result is then quite obvious and comfortable. Possibly the only deviation from this situation is the behavior of the leakage current of the L2320 diodes. Further discussion of this case follows under the heading of that particular diode.

4.2 L2287 transistors

L2287 transistors are split into 38A, 38B, and 38D codes for use in the first two stages of the amplifier and in the oscillator of the SF repeater. The two amplifier stages operate at power levels of 25 and 75 mW, respectively. The oscillator operates at a lower power level. During aging, all of these devices are held at a power level of 75 mW, composed of 5 V and 15 mA. The extended aging program has included one hundred of these transistors.

As shown in Fig. 1, the collector-base breakdown voltage of these transistors showed almost no discernible change after ten years of power

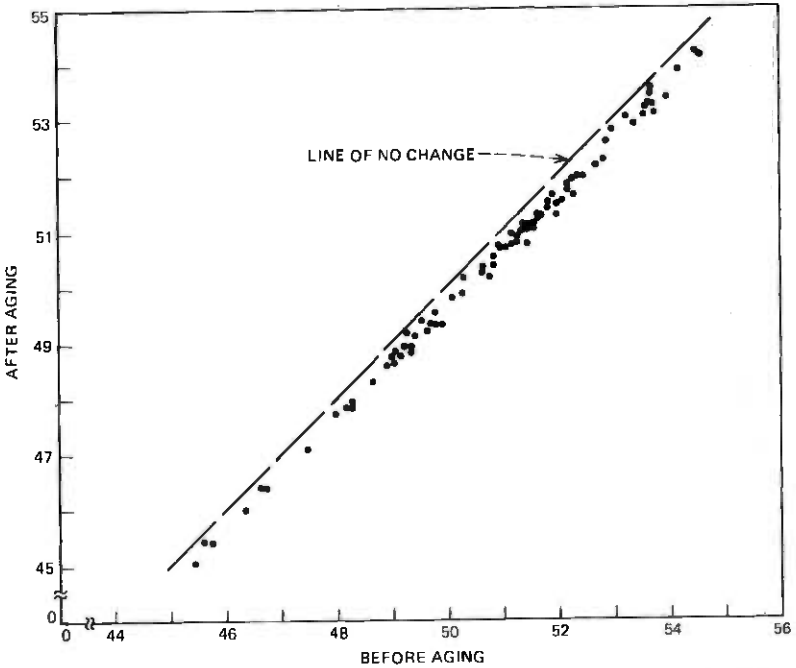

Fig. 8—Collector-base breakdown voltage in volts at 100 μA for L2288 transistors—before and after 10 years of power aging at 0.99 W.

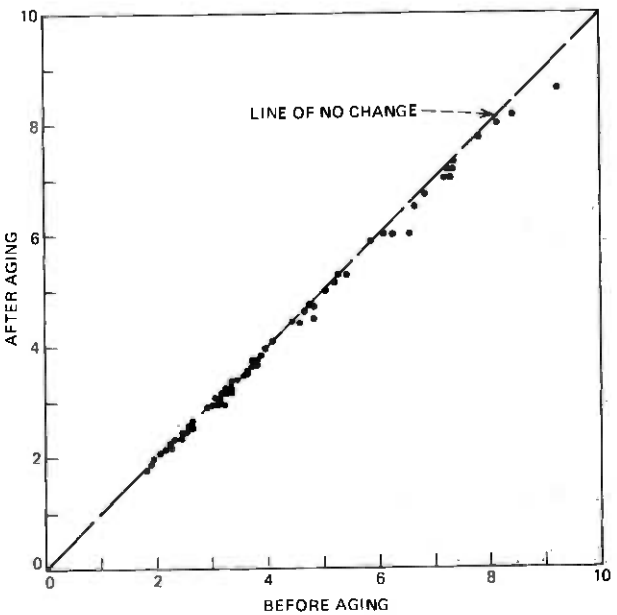

Fig. 9—Collector-base reverse-leakage current in μA at 15 V for L2288 transistors—before and after power aging at 0.99 W.

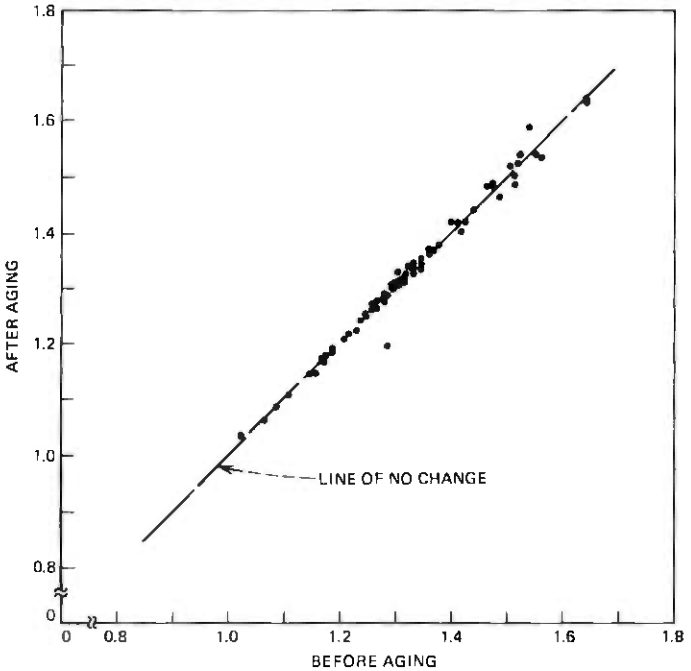

Fig. 10—Emitter-base breakdown voltage in volts at 1 mA for L2288 transistors—before and after 10 years of power aging at 0.99 W.

aging, with the exception of one device which showed a decrease of about 1.5 V. This device showed an increase in collector-base leakage current of about 0.15 μA, as seen in the lower part of Fig. 2, and a decrease in the emitter base forward voltage of about 5 mV, as seen in the lower part of Fig. 4. These changes, however, would have no effect whatever on the operational reliability of the device. No attempt has been made to determine the cause of the changes since all parameters of the device are still well within initial specification limits.

Figure 2 shows that the lower values of collector-base reverse-leakage current show no discernible change while those above about 0.5 μA show a very slight upward shift. Such a small shift, however, is of no practical significance.

The emitter-base breakdown voltages of most of the devices show remarkable stability, as depicted in Fig. 3. Only those above about 1.6 V show a small amount of scatter in the behavior. With the one exception, previously noted, the emitter-base forward voltages showed essentially no discernible change after ten years of power aging, as seen in Fig. 4.

Perhaps the most gratifying result of this extended aging program is the behavior of the low-frequency common-base current gain h_{fb}, as

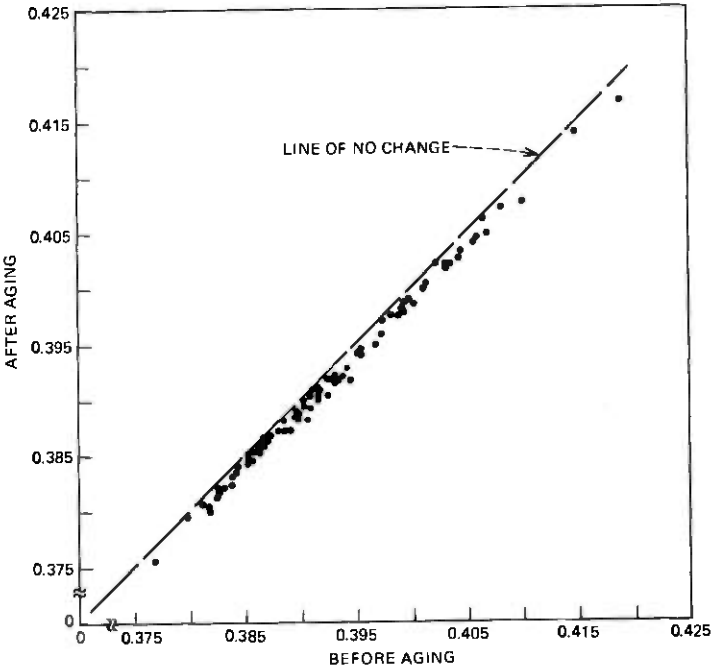

Fig. 11—Emitter-base forward voltage in volts at 90 mA for L-2288 transistors—before and after ten years of power aging at 0.99 W.

shown in Fig. 5. From the beginning, this parameter was expected to show a significant aging trend. Even under the conservative assumption that the aging of this parameter is linear with time, Fig. 5 indicates that in ten years of aging the average of the parameter has shifted only about a quarter of the way to end-of-life for the group. Actually, the aging of this parameter is known to occur at a decreasing rate with increasing time, as shown in Fig. 6 for a typical case of the aging of the dc common-base current gain. With similar behavior of the ac common-base current gain, the end-of-life for the group of devices would appear to be at some indeterminate time far beyond forty years.

Only one-third of the L2287 transistors (those for the first stage of the amplifier) are required to meet the initial noise figure specification limit of 4 dB at 5.5 MHz. Figure 7 shows that in fact nearly all of these devices still meet this limit after ten years of power aging. In the evolution and refinement of techniques for measuring the noise figure over the ten-year period of extended device aging, it was more important to provide the best and most realistic noise figure data than to maintain reference to the past. Other experience in aging indicates that if the other device parameters remain well-behaved, the noise figure does not change.

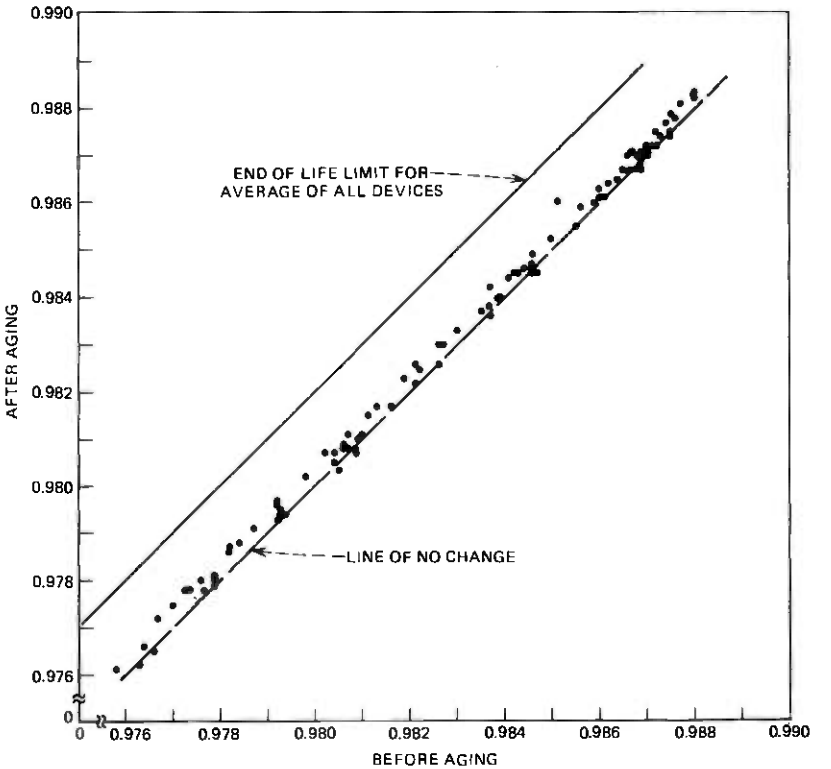

Fig. 12—Low-frequency common-base current gain for L2288 transistors—before and after power aging at 0.99 W.

4.3 L2288 transistors

The L2288 transistors, coded as 38C and operating at a power level of 0.99 W, are used in the output stage of the amplifier in the SF repeater. During aging, all of these devices are held at this same level of power, composed of 11 V and 90 mA. The extended aging program has included one hundred of these devices.

Figure 8 shows that the collector-base breakdown voltage of these transistors has very uniformly shifted downward by about a quarter volt after ten years of power aging. This is the kind of result which arouses suspicion that the apparent shift may actually be due to an offset between tests conducted over ten years apart. However, since the fact of testing offset could not be definitely established, the shift must be accepted at face value. Even if it is real, such a shift is of no practical significance. Under a simple linear extrapolation the voltage would be decreasing at the rate of 1 V in forty years, and a decrease of at least 30 V would be required before even the first device would begin to impair system performance.

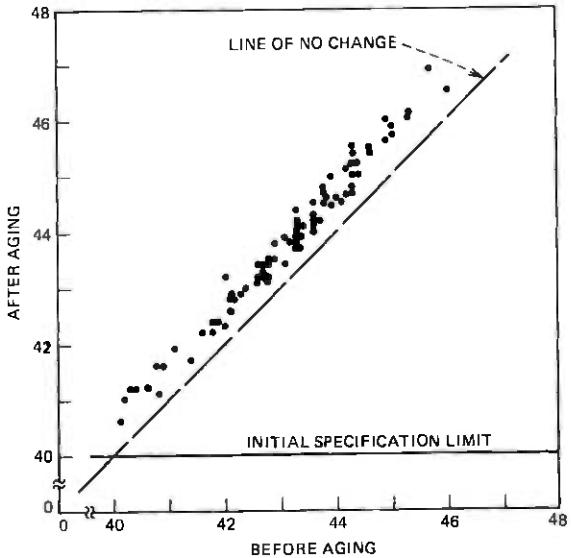

Fig. 13—Second-harmonic distortion in dB below fundamental for L2288 transistors—before and after power aging at 0.99 W.

In Fig. 9 the lower values of collector-base leakage current (below about 4 μA) show no discernible change after ten years of power aging. Those above about 4 μA show, if anything, a very slight decrease.

The emitter-base breakdown voltages shown in Fig. 10 indicate very consistent behavior, with the possible exception of one device whose voltage decreased by about 0.1 V and another whose voltage increased by about 0.05 V, again with no significant effect on circuit or system reliability since the emitters are always forward biased.

The emitter-base forward voltages shown in Fig. 11 indicate a small but consistent decrease of about 1 mV after ten years of power aging. Such a decrease is not significant enough to argue about whether it is real or an offset in testing.

The behavior of the low-frequency common-base current gain shown in Fig. 12 is again very gratifying. An increase in this parameter was expected, but the magnitude of increase has turned out to be smaller than expected. If the first ten years of aging serve as any indication, the end of life limit as an average for all devices will probably never be reached for all practical purposes.

Harmonic distortion for these devices was determined by measuring the level of the second and third harmonic at the output with a sine wave input at a frequency of 1.0 MHz at a power level of 5.0 dBm. Measurement precision is not in the same class with that of the dc and low-frequency parameters, especially for the initial data. During the ten-year

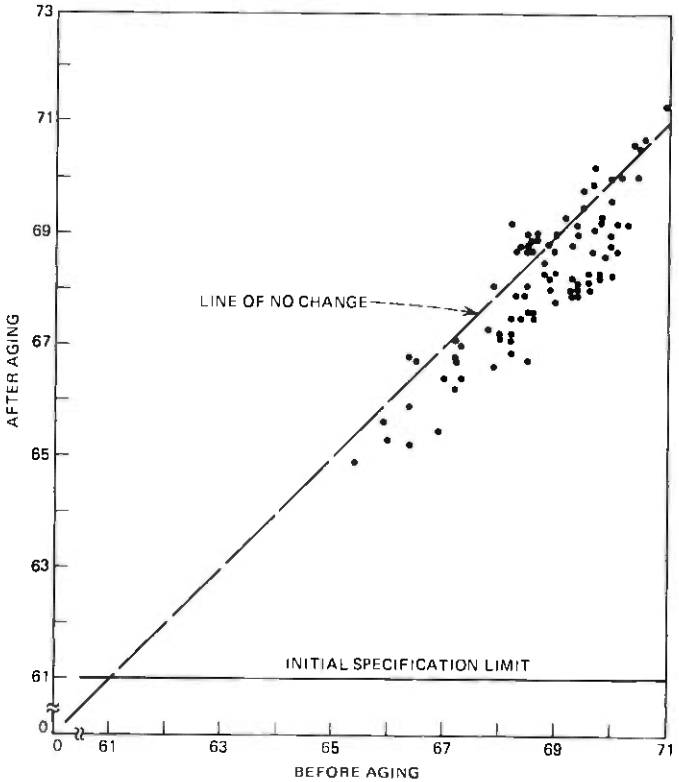

Fig. 14—Third-harmonic distortion in dB below fundamental for L2288 transistors—before and after power aging at 0.99 W.

period of extended aging, the precision of measurement was improved considerably. Nevertheless, the aging results shown in Figs. 13 and 14 should be considered merely indicative that the distortion parameters are quite well behaved and that after ten years of power aging, they still meet all of the device initial requirements.

4.4 L2317 diodes

The L2317 diodes, coded as 467A, are large area pn junction protective devices used in the power path of the SF repeater. During normal system operation, each device merely draws reverse-leakage current. In the event of a cable fault, however, it would conduct a large transient of reverse current to protect the repeater circuit from excess voltage. These devices are aged with 13-V reverse bias, which is a few volts below breakdown. Although the ten-year aging results are again presented in the form of initial and final values for each parameter, a large number of interme-

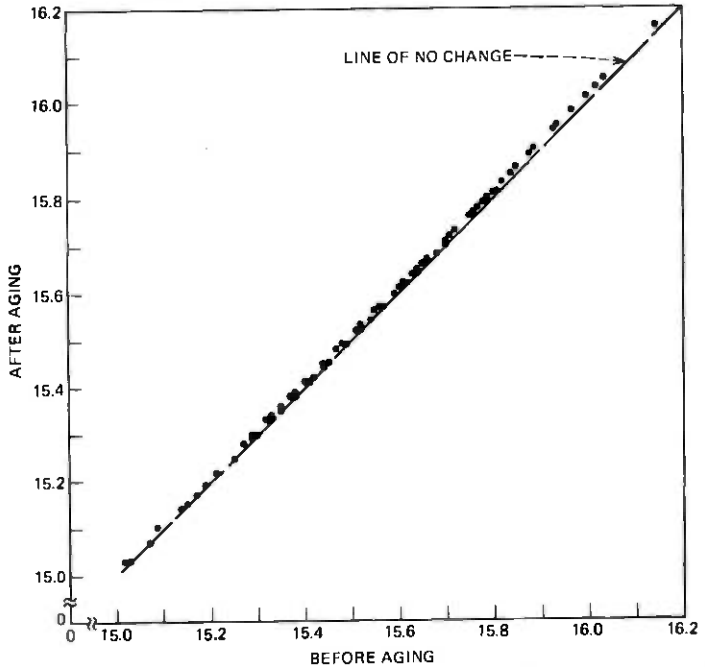

Fig. 15—Breakdown voltage in volts at 1.0 mA for L2317 diodes—before and after 10 years of power aging at 13 V reverse bias.

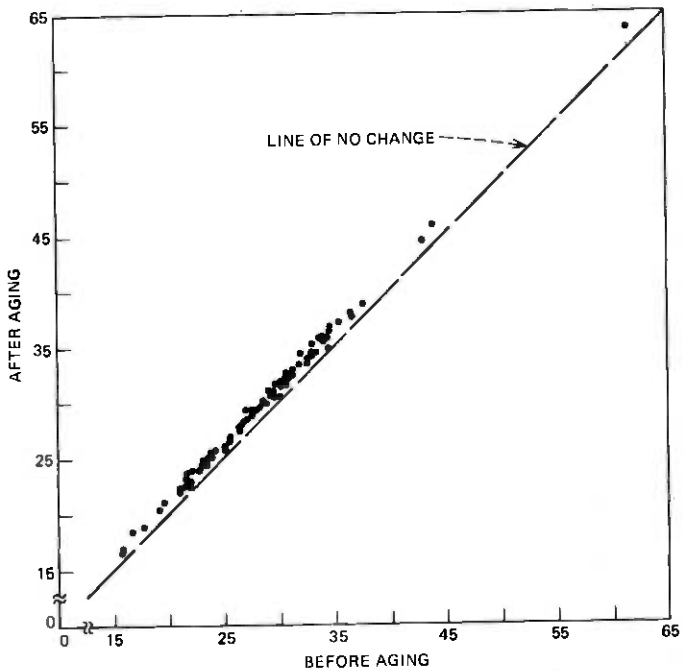

Fig. 16—Reverse-leakage current in nA at 13 V for L2317 diodes—before and after 10 years of aging at 13 V reverse bias.

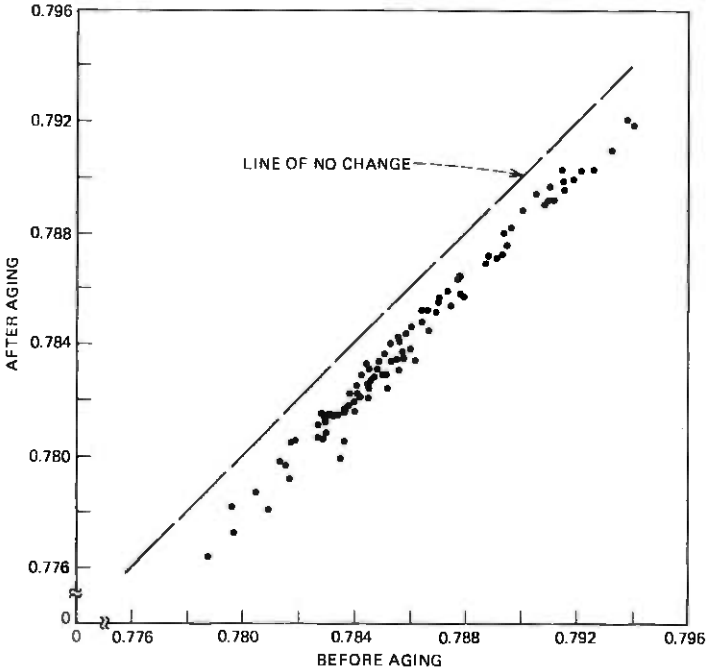

Fig. 17—Forward voltage in volts at 500 mA for L2317 diodes—before and after 10 years of power aging at 13 V reverse bias.

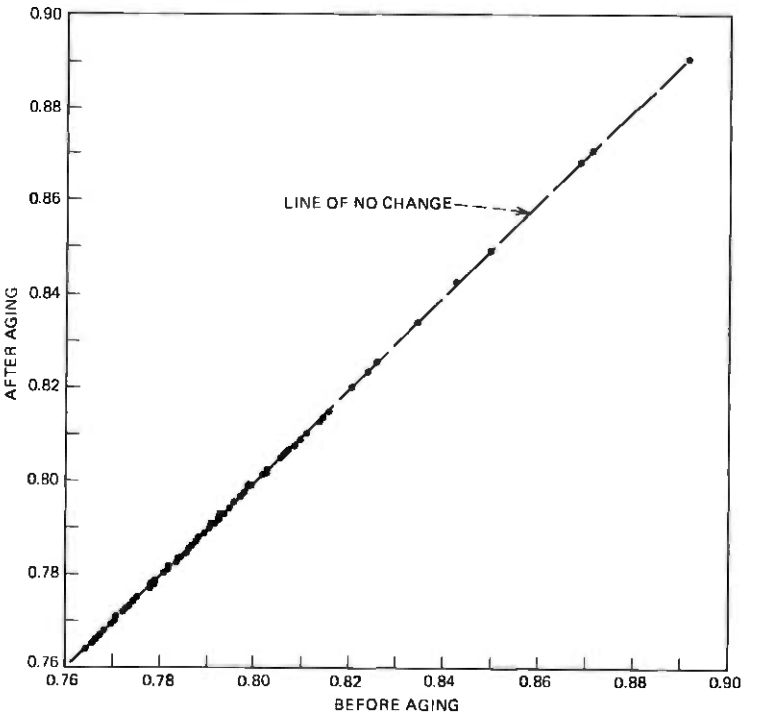

Fig. 18—Forward voltage in volts at 200 mA for L2318 diodes—before and after 10 years of aging at 10 mA rms and 60 Hz.

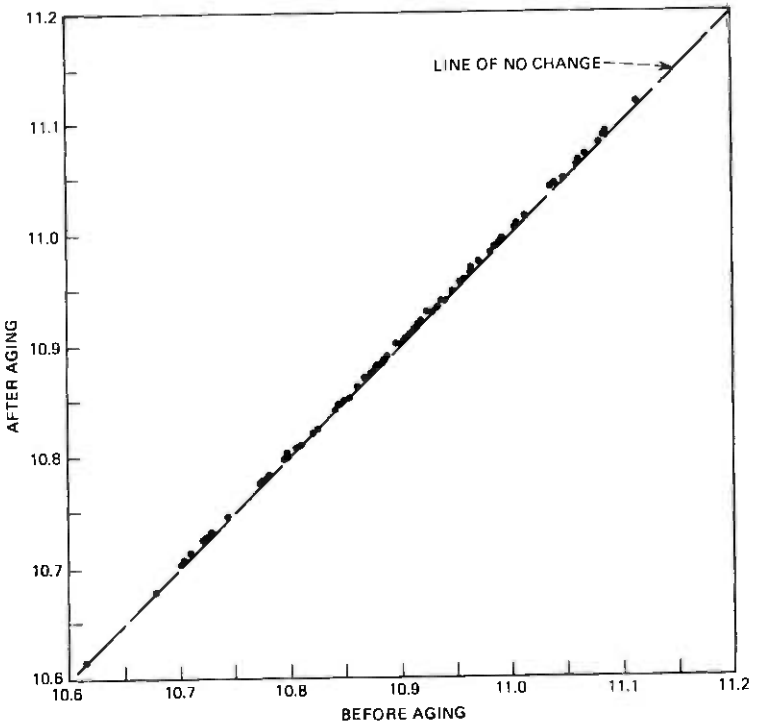

Fig. 19—Breakdown voltage in volts at 2 mA for L2320 diodes—before and after 10 years of aging at 2 mA in breakdown region.

diate data points taken during the ten years show no significant excursions away from the end points shown.

Figure 15 shows the behavior of the breakdown voltage of one hundred L2317 diodes across ten years of power aging. A very consistent upward shift of about 0.01 V is readily apparent but is too small to be of any practical significance. The behavior of the reverse leakage current shown in Fig. 16 also exhibits a small but very consistent upward shift of about 1 nA, while Fig. 17 shows a consistent downward shift of about 2 mV in the forward voltage.

4.5 L2318 diodes

The L2318 diodes, coded as 468A and 468B, are used for surge protection in the oscillator circuit of the SF repeater. Each device consists of two silicon pn junction diode chips mounted on a common metallic heat sink and oppositely poled so that only a forward characteristic is observed with an applied voltage of either polarity. During long-term power aging, a 60-Hz voltage is applied to make the diode conduct 10 mA rms in each direction. The forward voltage drop is measured at two

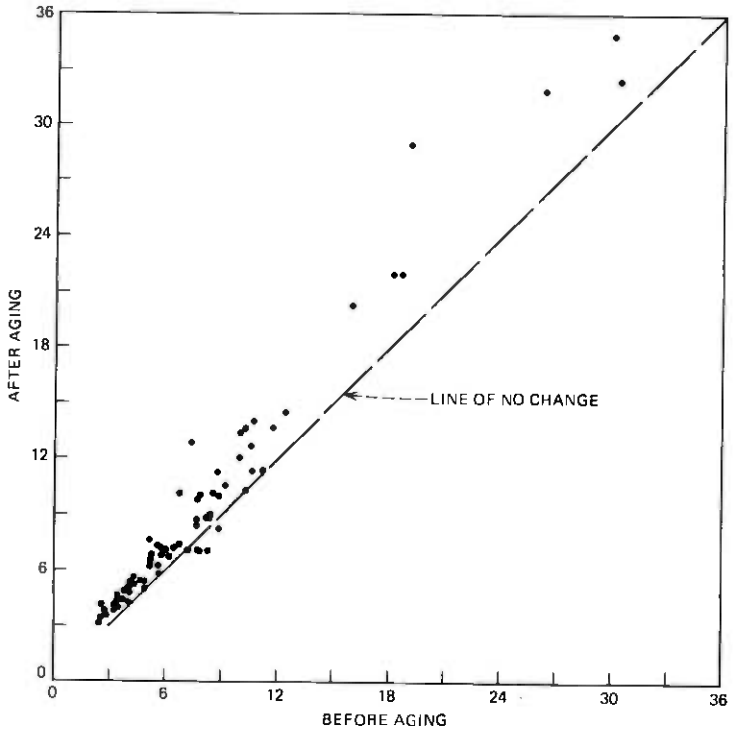

Fig. 20—Reverse leakage current in nA at 8 V for L2320 diodes—before and after 10 years of aging at 2 mA in breakdown region.

current levels in each direction during long-term power aging.

Figure 18 shows the behavior of the forward voltage of one hundred L2318 diodes in one direction and at one current level across ten years of power aging. No discernible aging can be seen. The behavior of the other forward voltages is not shown because they look no different.

4.6 L2320 diodes

The L2320 diodes, coded as 467B, are silicon pn junction limiter devices used in the third stage of the amplifier of the SF repeater. Operation is at 2.0 mA in the breakdown region, and the devices are aged under these power conditions.

Figure 19 shows the behavior of the breakdown voltage of one hundred L2320 diodes across ten years of power aging. No discernible shift in this parameter is evident.

In Fig. 20, a definite upward shift in the reverse-leakage currents can be seen. This is the one case where a significant amount of scatter is encountered in the magnitudes of the parameter shifts among the hundred devices. The data suggests that percent change rather than

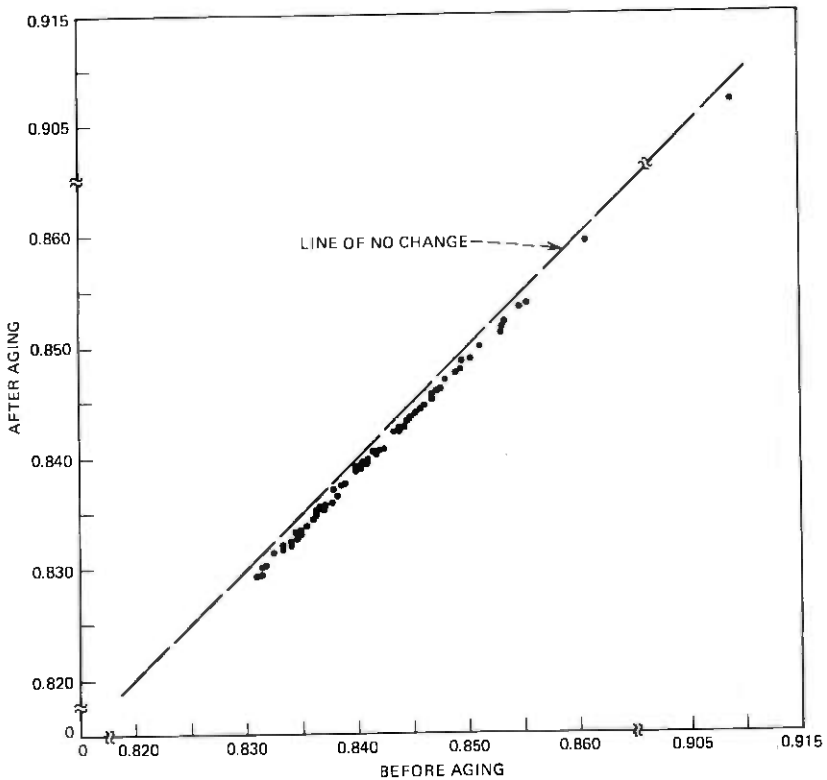

Fig. 21—Forward voltage in volts at 200 mA for L2320 diodes—before and after 10 years of aging at 2 mA in breakdown region.

incremental (delta) current change is probably the appropriate way to regard the behavior of individual devices. A leakage-current magnitude of at least 100 μA would be required in any one device to produce even the start of any impairment in circuit operation.

For this particular situation, even the aging of 100 devices for ten years still leaves some degree of uncertainty in how best to interpret the behavior in the tails of the distribution where the device would be found with potentially the greatest change in leakage current. In consideration of this situation and in consideration of the fact that a long cable might typically have 400 of these devices compared to the 100 shown here, it has been conservatively estimated[3] that with high (90 percent) confidence, a 100-μA leakage current would not be reached in any one device in less than 60 years.[3]

According to Fig. 21, the forward voltages of these diodes have shown a small but very consistent downward shift of about 1 mV in ten years of power aging. Such a shift is of no significance to system reliability.

V. GENERAL COMMENTS

The regular procedure for producing submarine cable semiconductor devices can be divided into three general stages: manufacturing, screening, and aging. A selection process is employed in each of these stages. The devices in this extended aging program progressed through the first two stages in the usual way, but did not have the benefit of having completed the standard aging stage of about six months with ensuing selection before beginning the extended aging period. Nevertheless, any of these devices, even after ten years of power aging, could still be considered candidates for cable use.

For the next generation of submarine cable (SG), the first of which was laid across the Atlantic Ocean in the early part of 1976, silicon transistors have been used throughout. Probably no new systems will ever be designed with germanium transistors. Nevertheless, the reliability record thus far of the germanium transistors used in the SF system is rather impressive in its own right.

VI. ACKNOWLEDGMENT

Over many years, a number of people contributed to this project. To list all of them would not be possible and to list only part of them would not be fair. Overall, however, the project stands as a shining example of team effort between Bell Laboratories and Western Electric.

REFERENCES

1. A. J. Wahl, W. McMahon, N. G. Lesh, and W. J. Thompson, "SF System: Transistors, Diodes, and Components," B.S.T.J., *49*, No. 5 (May–June 1970), pp. 683–698.
2. L. E. Miller, "Reliability of Semiconductor Devices for Submarine Cable Systems," Proc. IEEE, *62*, No. 2 (February 1974), pp. 230–244.
3. J. A. Tischendorf, private communication.

Contributors to This Issue

Ta-Shing Chu, B.S., 1955, The National Taiwan University; M.S., 1957, and Ph.D., 1960, Ohio State University; Bell Laboratories, 1963—. Mr. Chu has been engaged in research on microwave antennas and tropospheric wave propagation for satellite communication and terrestrial microwave network. Senior Member, IEEE; member, Commissions B and F of URSI, Sigma Xi, Pi-Mu-Epsilon.

Corrado Dragone, Laurea in E.E., 1961, Padua University (Italy); Libera Docenza, 1968, Ministero della Pubblica Istruzione (Italy); Bell Laboratories, 1961—. Mr. Dragone has been engaged in experimental and theoretical work on microwave antennas and solid-state power sources. He is currently concerned with problems involving electromagnetic wave propagation and microwave antennas.

James Flanagan, Sc.D. Electrical Engineering, 1955, Massachusetts Institute of Technology; Bell Laboratories, 1957—. Mr. Flanagan has worked in voice communications, acoustics, and digital techniques for signal coding and transmission. He is Head of the Acoustics Research Department. Fellow, IEEE; fellow and currently vice president, Acoustical Society of America; member Sigma Xi and Tau Beta Pi.

Kenzo Ishizaka, B.S. (E.E.), 1953, M.S. (E.E.), 1955, and Ph.D. (E.E.), 1972, Tohoku University, Japan; Toyo Communication Equipment Co., 1955–62; Professor of Electrical Engineering, University of Electro-Communications, Tokyo, Japan, 1962—. Professor Ishizaka was on leave to Bell Laboratories, 1970–71 and 1972–76. He worked in the Acoustics Research Department on computer simulation of speech production and on applications to speech synthesis. Member, Acoustical Society of America, Acoustical Society of Japan, Institute of Electronics and Communication Engineers of Japan.

Joseph S. Kaufman, B.E.E., 1965, Pratt Institute; M.S.E.E., 1966 and Ph.D., 1970 (computer, information, and control engineering), University of Michigan; Assistant Professor of Electrical Engineering and Computer Science, Columbia University 1971–1973; Bell Laboratories, 1973—. Since Joining Bell Laboratories, Mr. Kaufman has been concerned with the modeling and analysis of telephone traffic systems. Member FAS, IEEE.

A. J. Wahl, B.S., 1942, University of Kansas; Ph.D., 1950, Princeton University; United States Air Force, 1951–1953; Bell Laboratories, 1950–1951, 1953—. Since 1953, Mr. Wahl has been engaged in various phases of transistor development. His early work was in the area of surface effects relating to device behavior, particularly in regard to device reliability. Later he was concerned with providing the semiconductor devices for the Telstar satellites and has since supervised the provision of transistors and diodes for submarine cable systems. He is currently supervisor of a group concerned with the development of microwave transistors and submarine cable devices. Member, IEEE, Sigma Xi.

Abstracts of Papers by Bell System Authors Published in Other Journals

COMPUTING

A "DOGLEG" Channel Router. D. N. Deutsch, Des. Automation Conf. Proceed., No. 13 (June 1976), pp. 425–433. A new algorithm for interconnecting two sets of terminals across an intervening channel is presented. The routing is done on two distinct layers, one for horizontal paths, the other vertical. Typical channels (300 terminals, 100 nets) are routed within one track of the mathematical bound (density) in a few seconds.

Impact of On-Line Systems on a Literature Searching Service. D. T. Hawkins, Spec. Libr., 67 (December 1976), pp. 559–567. Some experiences with on-line literature searching systems in the Bell Laboratories Libraries and Information Systems Center during 1975 are described. A total of 604 sessions, averaging 15 minutes each, occurred. Average cost per session was approximately $22, including connect time, communications, and off-line printing. The actual cost of many sessions, however, was $10 or less. On the average, on-line searching costs $1/minute and 10¢/off-line print. A search on a specific subject took an average of two sessions. Details on types of searches performed, the method of charging costs to the user, and some effects on library services are given, as well as some of the advantages of using an on-line system for information retrieval.

LTX—A System for the Directed Automatic Design of LSI Circuits. G. Persky, D. N. Deutsch, and D. G. Schweikert, Des. Automation Conf. Proceed., No. 13 (June 1976), pp. 399–407. LTX is a minicomputer-based design system for large-scale integrated circuit chip layout which offers a flexible set of interactive and automatic procedures for translating a circuit connectivity description into a finished mask design. The system encompasses algorithms for two-dimensional placement, string placement, and channel routing.

Pro—An Automatic String Placement Program for Polycell Layout. G. Persky, Des. Automation Conf. Proceed., No. 13 (June 1976), pp. 417–424. PRO is a minicomputer-based string placement program that reorders cells in polycell rows such that channel routing in the adjacent channels needs fewer routing tracks, and is not blocked by cyclic constraints. PRO utilizes cell reflections and pairwise neighbor exchanges, simultaneously monitoring both channels adjacent to the row being reordered.

ELECTRICAL AND ELECTRONIC ENGINEERING

Birefringent Coupler for Integrated Optics: Comment 2. A. Albanese and J. A. Arnaud, Appl. Opt., *15* (September 1976), p. 2029. A proof is given of the impossibility to realize lossless and one-directional couplers.

Control of Limit Cycles in Recursive Digital Filters by Randomized Quantization. R. B. Kieburtz, V. B. Lawrence, and K. V. Mina, Proc. 1976 IEEE Int. Symp. Circuits Systems, April 1976, pp. 624–627. The different types of limit cycles that can occur in single second-order sections respond differently to efforts to reduce them. Length 1 (dc) limit cycles play an important complicating role, especially in cascaded filters. We describe a simple method, involving random requantization of multiplier outputs, which can reduce or eliminate limit cycles in digital filters. We compare limit cycle and roundoff noise in three typical low-pass filters with and without the method.

Impurity and Substrate Diffusion in a Thin Contact Layer. R. P. Goel and F. E. Bader, J. Electrochem. Soc., *123*, No. 8 (August 1976), pp. 1242–1245. Based on simplifying assumptions, diffusion phenomena of substrate and impurity through a thin contact layer have been compared. Results are presented in terms of nondimensionalized parameters k (a measure of diffusion rate, time, and contact thickness) and l/l (film thickness/contact thickness). A single master curve is shown to represent each diffusion process.

Low-Noise GaAs MESFETS. B. S. Hewitt, et al., Electron. Lett., *12*, No. 12 (June 10, 1976), pp. 309–310. GaAs MESFETs with optimum noise figures of 1.6 dB at 6 GHz have been fabricated by projection photolithography. An equation has been developed for the calculation of optimum noise figure which gives good agreement between calculated and measured values.

Measurement of Land to Plated-Through-Hole Interface Resistance in Multilayer Boards. J. N. Hines, 14th Annual Proceed., 1976 Reliability Phys. Symp., April 20, 1976, pp. 132–134. Thermally induced cracks in the bond between an innerlayer land and barrel will occur during the lifetime of an MLB if processing is faulty. A technique to detect incipient failures that features four-terminal resistance measurements with a minimum of bulk resistance to overcome the insensitivity of series circuit measurements is presented.

Mechanical-Chemical Technique for Removal of Epitaxial Spikes. L. E. Katz and W. C. Erdman, J. Electrochem. Soc., *123*, No. 8 (August 1976), pp. 1249–1251. A new mechanical-chemical spike removal technique has been devised which effectively removes epitaxial spikes on silicon substrates. A mask damage technique was used to evaluate masks subjected to spike removal. Results show effective spike removal as well at no subsequent mask damage.

Millimeter Wave Integrated Circuit Technologies. R. H. Knerr and L. F. Moose, Microwave J., *19*, No. 11 (November 1976), pp. 23–24. Introduction to a two part series of short articles authored by the participants in a panel discussion on "Millimeter Wave Integrated Circuit Technologies" at the 1976 IEEE-MTT-S International Microwave Symposium.

A New Digital Echo Canceler for Two-Wire Full-Duplex Data Transmission. K. H. Mueller, IEEE Trans. Commun., *COM-24*, No. 9 (September 1976), pp. 956–962. A new approach to echo canceling for two-wire full-duplex data transmission is proposed. The canceling signal is directly synthesized from the binary data, using a transversal filter approach, and the usual multiplications are replaced by additions and subtractions, thus allowing efficient operation of a large number of taps as required for the canceling of distant echoes. As a specific application, a system processing one sample per baud is discussed where timing signals at both communicating stations are assumed to be synchronized.

Optimization of Digital Postdetection Filters for PSK Differential Detectors. R. D. Gitlin and K. H. Mueller, IEEE Trans. Commun., *COM-24*, No. 9 (September 1976), pp. 963–970. A digital transversal postdetection filter for use in differential PSK detectors is proposed, and a method for optimizing the weighting coefficients is presented. It is shown that even though the detector output signal is quadratic, the optimization problem can be formulated in a way similar to the well-known MMSE equalizer for linear signals.

MATERIALS SCIENCE

Application of Weibull-type Analysis to the Strength of Optical Fibers. B. K. Tariyal and David Kalish, Mater. Sci. Eng., *27* (January 1977), pp. 69–71. The strength distribution for 400 measurements on 0.61-m-length fused silica optical fiber specimens exhibits a bimodal behavior. The cumulative failure probability is presented for Weibull-type distributions and is analyzed by assuming that the strength distribution is comprised of two distinct populations having finite lower and upper strength limits. This procedure provides the best prediction of the strength in long lengths.

Many-Body Effects in Transition Metals: Role of the Density of States. G. K. Wertheim and L. R. Walker, J. Phys. F Metal Phys., *6*, No. 12 (December 1976), pp. 2297–2306. The effects of structure in the density of electron-hole pair excitations is examined by numerical integration of the many-body screening formalism. The results are applied to Ir and Pt.

Oxidation State of Tungsten in the Na_xWO_3 Bronzes. G. K. Wertheim, M. Campagna, J.-N. Chazalviel,* H. R. Shanks†, Chem. Phys. Lett., *44*, No. 1 (November 15, 1976) pp. 50–52. X-ray photoemission spectra of W 4f electrons in vacuum-cleaved cubic Na_xWO_3 covering the entire composition range give no evidence for the existence of multiple W valence states in the bulk. Strong oxidation effects are observed in air exposed

surfaces. * Laboratoire PMC, Ecole Polytechnique, France. † Ames Laboratory, Iowa State Univ.

Thin Film Interaction in Aluminum and Platinum. S. P. Murarka, I. A. Blech, and H. J. Levinstein, J. Appl. Phys., *47* (December 1976), pp. 5175–5181. Aluminum and Platinum interacted very rapidly with the formation of several intermetallics (namely Pt_3Al_2, Pt_5Al_3, $PtAl_2$, $PtAl_3$ and $PtAl_4$) which resulted in large volume changes and loss of adhesion. The Al-Pt interaction rate was dependent on the annealing ambient being higher in forming gas, argon or helium than in vacuum or air.

MECHANICAL AND CIVIL ENGINEERING

Free Vibrations of a Beam-Mass System With Elastically Restrained Ends. R. P. Goel, J. Sound Vib., *47*, No. 1 (July 1976), pp. 9–14. The vibration problem of a beam with an arbitrarily placed concentrated mass and elastically restrained against rotation at either end is solved. The effects on eigenfrequencies of the system produced by varying the mass, stiffness, and position ratios are presented.

Tranverse Vibrations of Tapered Beams. R. P. Goel, J. Sound Vib., *47*, No. 1 (July 1976), pp. 1–7. Transverse vibrations of linearly tapered beams, elastically restrained against rotation at either end, have been investigated. Results for the first three eigenfrequencies, with different values of stiffness ratios and taper ratios are presented. Cases of a tapered cantilever beam with a concentrated mass at the free end and spring hinged at the other end have also been presented.

PHYSICS

Absence of the Low-Temperature Specific-Heat Anomaly in bcc ^{3}He. D. S. Greywall, *37*, No. 2 (July 12, 1976), pp. 105–107. The specific heat of bcc ^{3}He has been measured along three isochores for temperatures T between 50 mK and the melting curve. The data for $T < 0.3$ K can be described well by the function $C_v = \alpha T^{-2} + \gamma T^3$ and are inconsistent with the existence of the low-temperature specific-heat anomaly reported previously by others.

Digital Normalization of Iodine Filter Structure in Quasielastic Light Scattering. K. B. Lyons and P. A. Fleury, Phys. Rev. Lett., *47* (November 1976), pp. 4898–4900. A molecular iodine filter may be used to reduce the elastic component in a light scattering experiment but also introduces extraneous structure into the spectrum. A technique is reported here to compensate quantitatively for this structure by a digital normalization scheme.

Enhanced Optical Fluorescence by Resonant Mössbauer Excitation. C. P. Lichtenwalner, H. J. Guggenheim, and L. Pfeiffer, Phys. Lett., *56A*, No. 2 (March 8, 1976), pp. 117–118. Enhanced optical Eu^{2+} fluorescence in Eu-doped CaF_2 crystals is observed following resonant absorption of Mössbauer γ-rays by Eu^{3+} nuclei. The effect is due to scintillation fluorescence from Eu^{2+} ions excited by conversion electrons from decaying 21.6 keV ^{151}Eu.

Particle Brownian Motion Near a Hydrodynamic Instability. J. B. Lastovka and J.-P. Boon,* Phys. Rev., *14*, No. 4 (October 1976), pp. 1583–1586. We consider the alteration of the normal Brownian motion of seed particle brought about by the enhanced velocity fluctuations that exist near the threshold of a hydrodynamic instability. We utilize our results to quantitatively predict the fluctuation induced effects for the threshold region of the Bénard convective instability. * Université Libre de Bruxelles.

Pressure-Induced Valence Change in Cerium Phosphide. A. Jayaraman, W. Lowe, L. D. Longinotti, and E. Bucher. The pressure-volume relationship for cerium monophosphide (CeP) has been determined to 200 kbars pressure. From the data, we conclude that there is an electronic transition involving a change in the valence state of Ce from 3^+ towards the 4^+ state near 100 kbar pressure. Qualitative observations of the reflectivity of CeP under pressure support the valence transition.

The Use of Our Specially Designed Nonachromatic Microscope Objective to Examine Tracks in Nuclear Emulsions. J. S. Courtney-Pratt, Opt. Eng., *15*, No. 4 (July–August 1976), pp. 379–383. We have shown that our specially designed nonachromatic objective allows us to produce sharp images of particle tracks in nuclear emulsions, for visual study and/or photographic recording, with a lateral resolution of about one micron, and a depth of field many times greater than can be achieved with conventional objectives.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 6
Printed in U.S.A.

B.S.T.J. Briefs

Adhesive Sandwich Optical Fiber Ribbons

By M. J. SAUNDERS and W. L. PARHAM

(Manuscript received April 15, 1977)

In 1974 a proposal was made to put optical fibers together into easily handled units for optical communication purposes.[1] This proposal suggested "the use of fiber ribbons consisting of linear arrays of fibers embedded in a thin, flexible supporting medium as components of a cable for fiber transmission systems." This note is a brief description of fiber ribbons made by sandwiching fibers between two layers of polyester-backed adhesive (adhesive sandwich ribbons).

The machine for making these ribbons has evolved around two

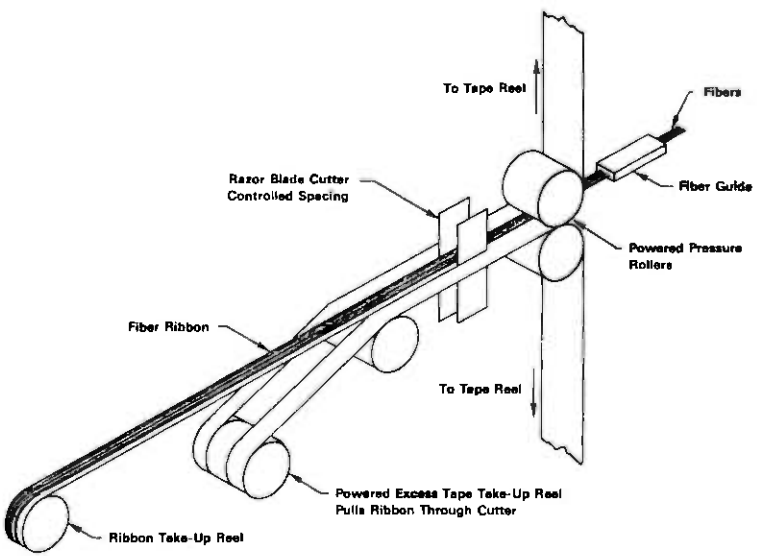

Fig. 1—Schematic diagram of adhesive sandwich ribbon machine.

motor-driven pressure rollers. Figure 1 is a schematic diagram of the machine. Twelve fibers from payout reels are directed through the fiber guide where they are made contiguous. The fibers then pass between the two pressure rollers and are sandwiched between two layers of polyester backed adhesive tape supplied from large reels. After emerging from the pressure rollers, the sandwich is cut by a blade cutter to a width somewhat larger than the width of the 12 fibers. The excess tape that has been cut away from the fiber-carrying ribbon is attached to a reel that supplies the power to pull the tape through the system. The ribbon with the fibers is either wound on a reel or is permitted to fall, under zero tension, into a container so that loss measurements can be made. Currently, we are making kilometer lengths of ribbon at speeds of about 0.2 m/s. However, shorter lengths of ribbon have been made at a speed of 0.8 m/s. We have made about 100 adhesive sandwich ribbons, including ribbons for the Atlanta Fiberguide System Experiment,[2] varying in length from 0.2 to 2 km.

The added loss of fibers in these ribbons, caused by microbending,[3] is quite small. The average added loss of the fibers in the ribbons used to make the two cables, each 1 km in length, for the Atlanta Fiberguide System Experiment was 0.95 ± 0.07 dB/km for 137 fibers and 0.55 ± 0.05 dB/km for 132 fibers.[4] Some fiber breakage occurred, caused both by defects in the machine and by the fibers being wound improperly on the reels.

We thank M. J. Buckler for making the loss measurements, E. D. Knab for much of the design work, and R. Sabia for originally suggesting the concept of the adhesive sandwich ribbon.

REFERENCES

1. R. D. Standley, "Fiber Ribbon Optical Transmission Lines," B.S.T.J., *53*, No. 6 (July–August 1974), pp. 1183–1185.
2. J. S. Cook, J. H. Mullins, and M. I. Schwartz, "An Experimental Fiber Optics Communications System," Conf. on Laser and Electro-Optical Systems, San Diego, California, May 25–27, 1976. Sponsored by IEEE and OSA.
3. D. Gloge, "Optical-Fiber Packaging and Its Influence on Fiber Straightness and Loss," B.S.T.J., *54*, No. 2 (February 1975), pp. 245–262.
4. M. J. Buckler, unpublished work.