THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 52 April 1973 Number 4

Normal Modes in Overmoded Dielectric-Lined Circular Waveguide

By J. W. CARLIN and P. D'AGOSTINO

(Manuscript received October 30, 1972)

The propagation constants (loss and phase) are determined for the normal modes in dielectric-lined overmoded (guide radius ≫ wavelength) circular guide by numerical solutions of the appropriate characteristic equations. It is shown that the heat loss of the TM_{pn} modes is quite low and decreases with increasing frequency in contrast to the increase in heat loss predicted by the perturbation theory for thin linings. The primary effect of the low-loss TM_{pn} modes on a transmission system using the TE_{01} mode is a reduction in the route bend losses.

I. INTRODUCTION

Communications systems are currently under development by the Bell System and abroad which utilize as the transmission medium dielectric-lined circular waveguide (DLG) excited in the low-loss circular electric (TE_{01}) mode. The presence of a thin dielectric liner, bonded to the wall of the waveguide, eliminates the undesirable degeneracy which exists between the phase constants of the TE_{01} and TM_{11} modes in unlined metallic guide. It also significantly modifies the propagation constant (phase and attenuation) and field distribution of many other modes compared to their counterparts in unlined waveguide.

In order to determine the TE_{01} transmission characteristics of lined waveguide, it is necessary to determine the characteristics of the normal modes of the structure. The general problem of determination of the modes in a hollow, perfectly conducting metallic cylinder with a lossless dielectric lining has been of much interest since the initial investigations of Bucholz[1] and Wachowski and Beam.[2]

Unger[3,4] discusses the basic propagation characteristics of a thinly lined DLG suitable for use as a wide-band carrier in the millimeter-wave region. He determines the phase velocities of the normal modes exactly, but uses a first-order perturbation theory, with lining thickness as the perturbation parameter, to obtain the heat loss of the normal modes of DLG. It can be shown that the actual heat loss characteristics of the normal modes of DLG differ significantly from those predicted by the simple perturbation theory for a thin lining, in that modes not of the circular electric class may have very low heat loss at high frequencies. This was discussed briefly in an earlier paper[5] which treated the loss characteristics of circular symmetric modes (TE_{on} or TM_{on}) in DLG. The present paper is a continuation of that work with a more complete discussion of the characteristics of the normal modes in overmoded DLG. Representative values for the purely real normal-mode propagation constants based on a numerical solution of the eigenvalue equation resulting from the classical boundary value problem formulation for a lossless DLG are given. The losses were determined using two different methods. The methods ("induced current" and "wall impedance") will be explained in a later section of this paper and we will see that the results are consistent. The normal-mode loss behavior in DLG was found to have a simple physical explanation in terms of the plane wave reflection coefficient of an equivalent infinite impedance plane. Whenever possible, the algebraic details are omitted or deferred to a suitable appendix. Some plots of the electric field lines of the normal modes and the energy density distribution over the guide cross section are also given.

II. BOUNDARY VALUE PROBLEM FORMULATION

An idealization of the waveguide structure under investigation is shown in Fig. 1. It is a perfectly conducting cylinder of radius b to which a lossless dielectric of relative permittivity ϵ_r and thickness t is bonded. The interior region ($r < a$) is to be filled with an inert gas and its electrical properties are those of free-space over the frequency range of interest.

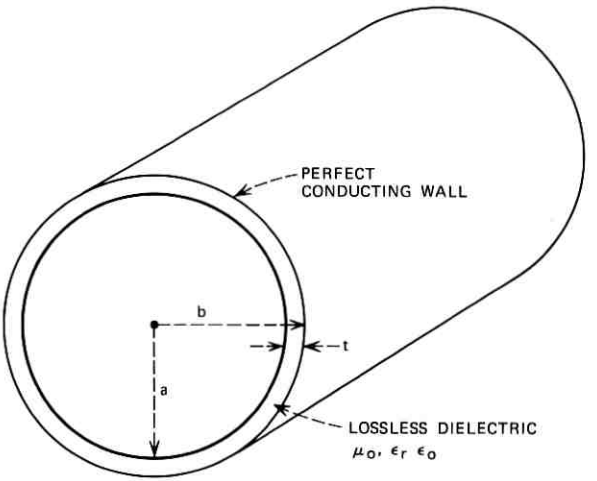

Fig. 1—The idealized lossless model for dielectric-lined guide.

The fields of a normal mode (TE_{pn} or TM_{pn}) in the structure may be expanded in the usual manner in terms of two scalar functions over the two regions of the guide cross section. In the empty region of the guide ($r < a$), they are of the form

$$\begin{aligned} T_n &= N_n J_p(\chi_n r) \sin p\phi \\ T'_n &= N_n J_p(\chi_n r) \cos p\phi \end{aligned} \qquad r < a \tag{1}$$

where N_n is a suitably chosen normalization constant. On applying the usual boundary conditions at the free-space dielectric (continuity of the tangential electric and magnetic fields) and metallic wall (zero tangential electric field) boundaries, we obtain a characteristic equation which must be solved for the eigenvalue k_n. Symbolically, we have

$$E(k_n, f, b, \epsilon_r, p, t) = 0 \tag{2}$$

where k_n (the only unknown) is defined as

$$k_n = \chi_n a, \tag{3}$$

p is the circumferential order of the mode, and f is the frequency of interest. The fields vary along the z axis as $e^{-jh_n z}$ for an assumed $e^{j\omega t}$ time dependence. The propagation constant h_n is related to k_n by

$$h_n^2 = k^2 - \chi_n^2 \tag{4}$$

where k is the free-space propagation constant.

The development of the characteristic equation is discussed by Unger[3,4] and is outlined in Appendix A for reasons of continuity. The notation used in this paper is consistent with that of Unger.[4] It should be noted that the normal mode fields corresponding to the solution of (2) are not strictly transverse electric (TE) or magnetic (TM) as in hollow metallic waveguide. (The circularly symmetric modes are pure TE or TM.) However, we carry the usual TE or TM nomenclature over with the understanding that this is what the field structure tends to in the limit of zero lining thickness.

For a lossless structure, the eigenvalue k_n as given by the solution of eq. (2) is either pure real or pure imaginary as is the characteristic eq. (2). Because of this, it is always possible to obtain a solution of (2) by the well-known "bisection" method on a digital computer.

On solving (2) for k_n, we can then determine the propagation constant h_n for the TE_{pn} or TM_{pn} mode of interest. The differential propagation constant ($\Delta\beta$) between the $TE_{pn}(TM_{pn})$ mode and the TE_{01} mode is an important parameter in the estimation of TE_{01} loss characteristics. We define $\Delta\beta$ as

$$\Delta\beta = h_n - h_{01}(\text{rad/m}). \tag{5}$$

Although the model used is an idealization of the actual lossy structure, $\Delta\beta$ as given by (5) is still an accurate representation of the difference in propagation constants. We can estimate the attenuation constant α_n(Np/m) for a $TE_{pn}(TM_{pn})$ mode by using a simple physical approximation.

The fields in the lossless structure are approximately the same as in the actual lossy guide. Hence they may be used as a very accurate first-order approximation to obtain the conduction currents in the lossy metallic walls and the displacement currents in the dielectric lining. The total heat loss (wall losses and dielectric losses) readily follows. We shall refer to this as the "induced current" method in the following discussion. The details are given in Appendix A. We now define the differential attenuation constant ($\Delta\alpha$), which is also needed to determine the TE_{01} transmission characteristics in DLG, as

$$\Delta\alpha = \alpha_n - \alpha_{01}. \tag{6}$$

In the preceding section we have indicated one well-known method of determining the attenuation and propagation constants of the normal modes in DLG. There is another approach, first introduced by Unger,[4] which aids greatly in understanding in an intuitive fashion the normal mode characteristics of DLG. This is in essence a quasi-optics ap-

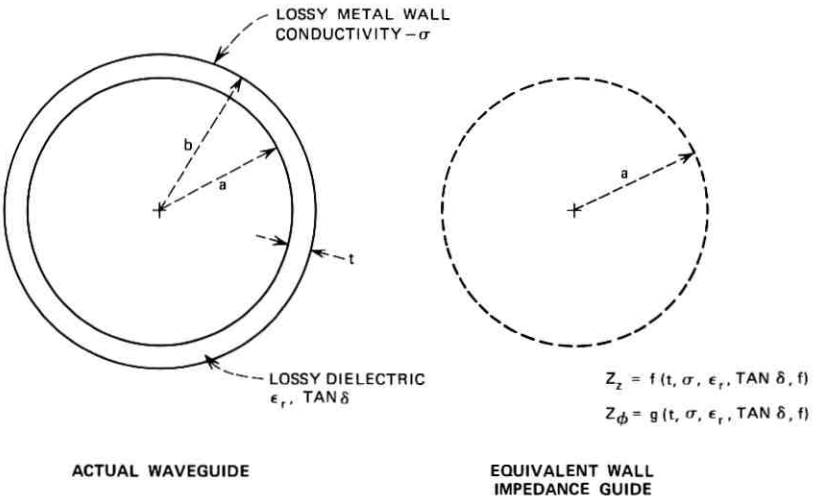

Fig. 2—The wall-impedance model for dielectric-lined guide.

proach, as we replace that portion of the DLG in the region $r > a$ of Fig. 1 by an impedance wall at $r = a$ as indicated in Fig. 2. The values of the impedances Z_z and Z_ϕ, which approximate the actual boundary conditions relating the tangential fields at the free-space dielectric boundary,

$$Z_z = -\frac{E_z}{H_\phi}\bigg| r = a,$$
$$Z_\phi = \frac{E_\phi}{H_z}\bigg| r = a, \tag{7}$$

are given by a simple plane wave analysis as outlined at the end of Section III. We can then obtain a characteristic equation for the equivalent wall-impedance structure of Fig. 2. Again we defer the details to Appendix B. The resultant equation has the form

$$E(k_n, f, Z_z, Z_\phi, a) = 0. \tag{8}$$

Here Z_z and Z_ϕ are *complex* quantities for a lossy metal wall with a lossy liner, as is the resulting eigenvalue k_n. We use the well-known Newton-Raphson method to solve (8) for k_n numerically on a digital computer. In this case, the propagation constant γ_n is complex (the fields vary along the guide axis as $e^{-j\gamma_n z}$ for $e^{j\omega t}$ time dependence) and given by

$$\gamma_n^2 = k^2 - \chi_n^2 \tag{9}$$

where

$$k_n = \chi_n a,$$

and k is the free-space propagation constant. We define the differential attenuation constant ($\Delta\alpha$) and propagation constant ($\Delta\beta$) as in (6) and (7), where it is understood that

$$\gamma_n = h_n + i\alpha_n.$$

III. RESULTS

In this section we give results characteristic of those expected of the DLG for use in typical waveguide transmission systems. The results are for a polyethylene liner bonded to copper waveguide walls. The general behavior exhibited, however, is representative of that to be expected of any low-loss dielectric liner on a good conductor.

Figure 3 is a typical plot of the eigenvalue (k_n) for the first four circular symmetric modes for the *lossless* guide of Fig. 1 based on numerical solution of the exact characteristic equation. The eigenvalues exhibit a cyclic or periodic behavior as the lining thickness increases. The eigenvalue for a $TE_{on}(TM_{on})$ mode in DLG is equal to that of a

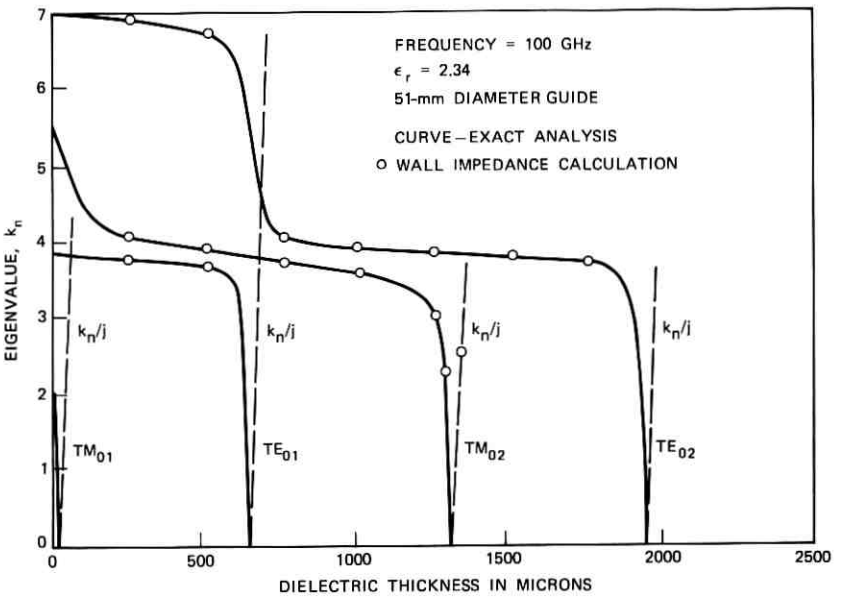

Fig. 3—Eigenvalues vs lining thickness for the TE_{on} and TM_{on} modes, lossless guide; wall-impedance model and the exact solution.

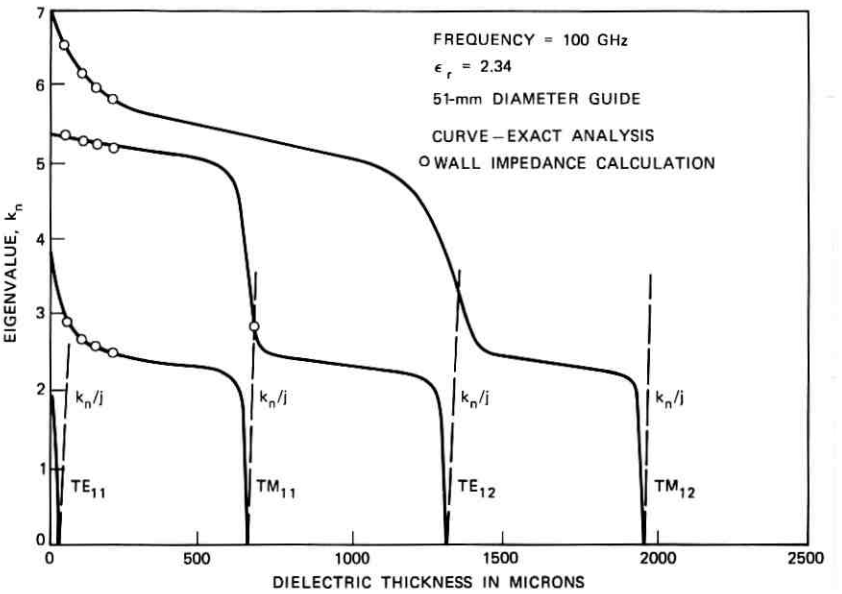

Fig. 4—Eigenvalues vs lining thickness for the TE_{1n} and TM_{1n} modes, lossless guide; wall-impedance model and the exact solution.

$TE_{o,n-1}(TM_{o,n-1})$ in unlined guide when the lining is "equivalent" to a half wavelength in thickness. The "equivalent wavelength" in the dielectric is given by $\lambda/\sqrt{\epsilon_r - 1}$, where λ is the free-space wavelength. The interested reader is referred to the discussion preceding (11) for details. Similarly, the eigenvalue for a $TE_{on}(TM_{on})$ mode in DLG is equal to that for a $TM_{o,n}(TE_{o,n-1})$ in unlined guide when the lining is a quarter-wavelength thick.

It can be seen that the eigenvalues of the TM_{on} modes change rapidly with an initial increase in lining thickness for thin linings, while those of the TE_{on} mode do not. This is due to a strong electric field at the wall for TM_{on} modes and a vanishing electric field for the TE_{on} modes.

The TM_{01} mode eigenvalue behaves in an interesting fashion for very thin linings. It decreases rapidly to zero as the lining thickness increases and then becomes imaginary. This indicates that the TM_{01} mode propagates as a slow wave or as a surface wave closely bound to the lining region in DLG. The other circular symmetric modes exhibit the same surface wave behavior for thicker linings, i.e., a quarter-wave lining, half-wave lining, etc.

The hybrid mode eigenvalues exhibit the same sort of repetitive behavior as seen in Fig. 4. The eigenvalues repeat as the lining thickness

increases by a half wavelength. The TM_{1n} eigenvalues change rapidly as the lining thickness increases for thin linings and then undergo little further change until the lining is a quarter-wave thick. The TE_{11} eigenvalue drops to zero and then becomes imaginary for very thin linings, indicative of a surface-wave-type mode. The TE_{1n} mode eigenvalues ($n > 1$) change very little for thin linings but undergo rapid

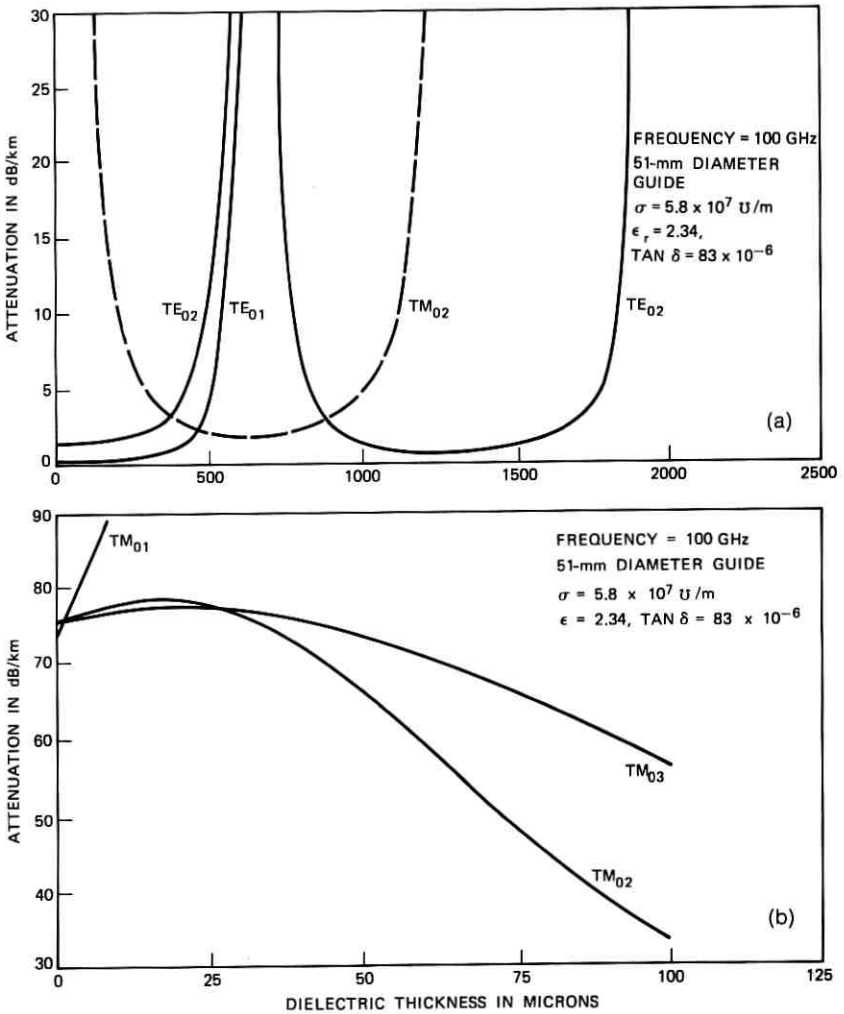

Fig. 5—Heat loss vs lining thickness for (a) the TE_{on} and TM_{on} modes; induced current method and (b) the TM_{on} modes; induced current method.

changes in the quarter-wavelength lining-thickness region. Figures 3 and 4 also furnish a good comparison of the exact solution (the curves shown in Figs. 3 and 4) for the eigenvalues and the "wall impedance" solution (the ○'s shown in the figures at selected points). The agreement is quite good with observed errors of approximately 2 percent or smaller for the cases shown.

In Figs. 5 and 6, the behavior of the total heat loss for some typical normal modes of DLG versus lining thickness is shown. We see that the TE_{on} modes are low loss for o, $\lambda/2$, λ, $\cdots$ thick linings while the TM_{on} modes exhibit low loss for $\lambda/4$, $3\lambda/4$, $\cdots$ thick linings.[5] The normal-mode heat loss exhibits the same periodic behavior vs lining thickness as the eigenvalues. The TE_{01} loss increases slowly with lining thickness for a thin lining; as the lining thickness approaches a quarter wavelength, the TE_{01} loss increases rapidly as the mode propagates as a surface wave.

The TE_{02} loss for a $\lambda/2$ lining is less than in unlined guide but still greater than that for the TE_{01} mode in unlined guide, as it is equivalent to the TE_{01} mode in a slightly smaller unlined guide.

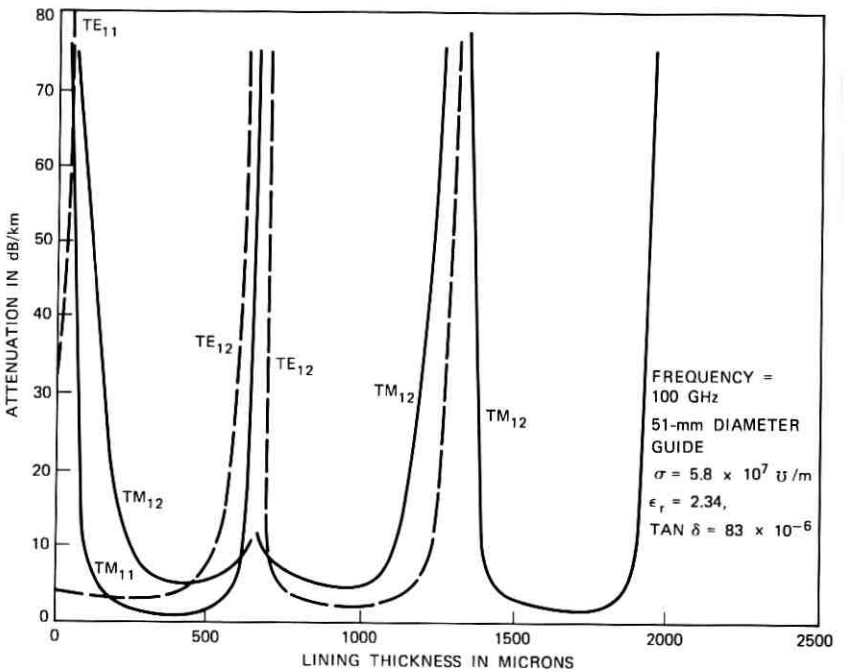

Fig. 6—Heat loss vs lining thickness for the TE_{1n} and TM_{1n} modes; induced current method.

The TM_{01} heat loss as shown in Fig. 5a increases rapidly for even very thin linings as expected for a surface wave mode. The TM_{on}, $n \neq 1$ (Figs. 5a and 5b), heat loss initially increases with lining thickness, and the initial slope of the loss curve agrees with the first-order perturbation theory of Unger,[3] but the loss then levels off (≈ 25 μm thick lining) and falls rapidly to a minimum for a quarter-wave lining (650 μm). The behavior of the TE_{1n} and TM_{1n} heat loss is similar to that of the TM_{on} and TE_{on} modes as shown in Fig. 6. The initial slope of the loss curves for the TE_{1n} and TM_{1n} modes versus thickness agrees with the perturbation theory predictions also, but from Figs. 5 and 6 it is obvious that the first-order perturbation theory is valid for predicting the normal-mode heat loss for only very thin linings (<20 μm). It is also quite clear that many modes (i.e., TE_{01}, TE_{02}, TE_{12}, TM_{11}, TM_{12}) as shown in Figs. 5 and 6 exhibit low heat loss (<5 dB/km for a 200-μm lining) in DLG.

Figure 7 is a comparison of the copper wall losses only for TE_{on} and TM_{on} modes as predicted by the induced current (the curves as shown) and the "wall impedance" method (the O's shown in the figure at

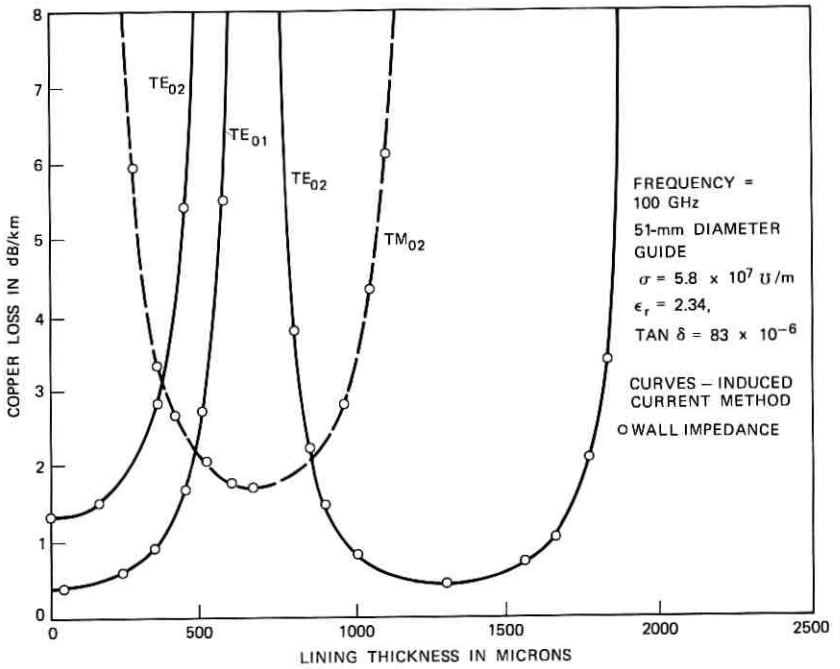

Fig. 7—Copper-loss comparison, wall impedance vs induced current method for the TE_{on} and TM_{on} modes.

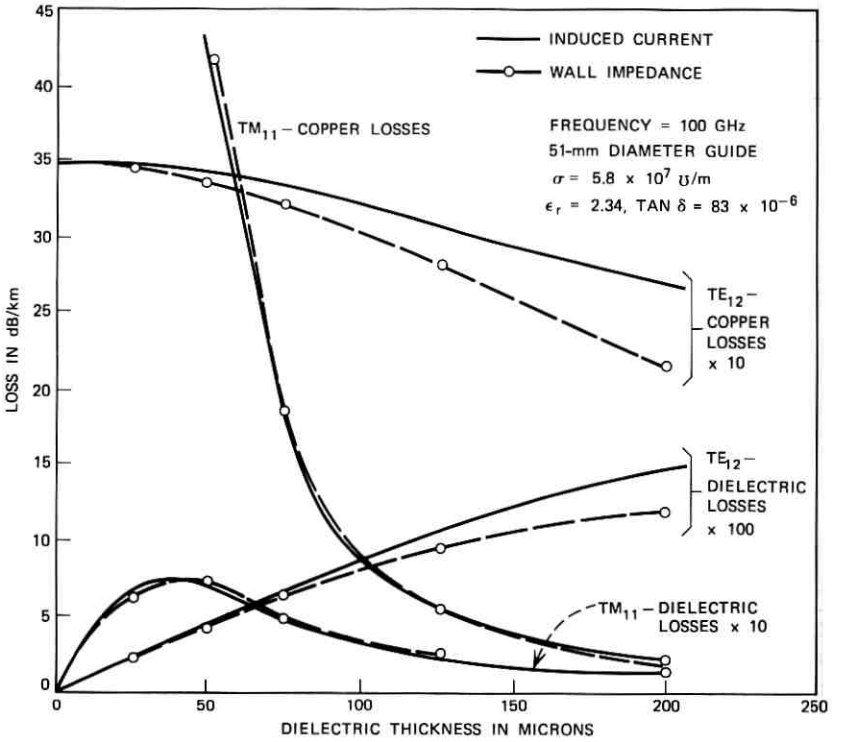

Fig. 8—Copper-loss comparison, wall impedance vs induced current method for the TE_{12} and TM_{11} modes.

selected points). The two methods are in excellent agreement over the range of thickness shown. Even in the regions where the losses are changing rapidly, the error is only on the order of 5 percent. The dielectric losses (not shown) are much smaller than the copper losses, but they also agreed to within 5 percent for the two methods.

Figure 8 is a comparison of the copper wall losses and dielectric losses predicted by the induced-current and wall-impedance methods for the TE_{12} and TM_{11} modes. The two methods differ by approximately 3 percent or less for the TM_{11} mode, but the TE_{12} difference is on the order of 10 to 20 percent for a 150 to 200-μm lining. The reason for this difference lies in the determination of the boundary conditions Z_z and Z_ϕ in (7). The approximate values of Z_z and Z_ϕ were obtained by use of a plane wave analysis as in Section III. This assumes a mode may be represented in the region local to the walls of the guide by a superposition of suitably polarized plane waves propagating in the

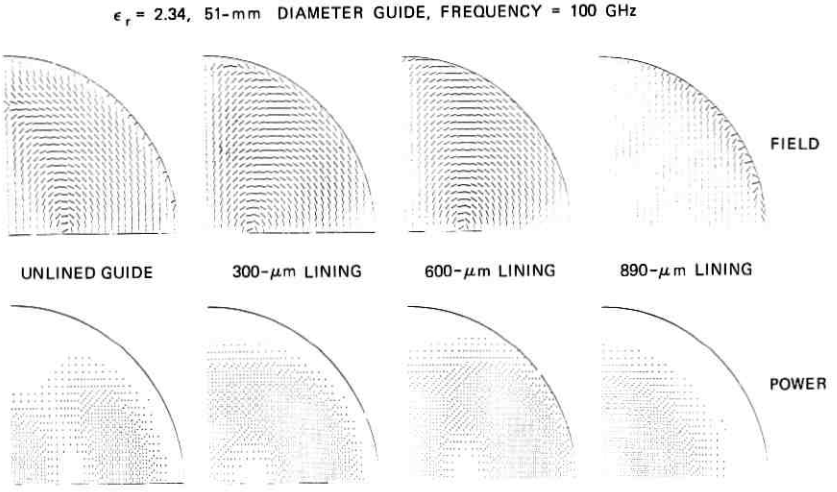

Fig. 9—TE_{12} electric field lines and power density vs lining thickness.

plane defined by the normal vector to the waveguide wall and the waveguide axis. This assumption is valid only for the circularly symmetric modes and leads to an error in the estimated boundary conditions (Z_z and Z_ϕ) for all other modes. The magnitude of the error is proportional to lining thickness to first order and has a greater effect on the loss of

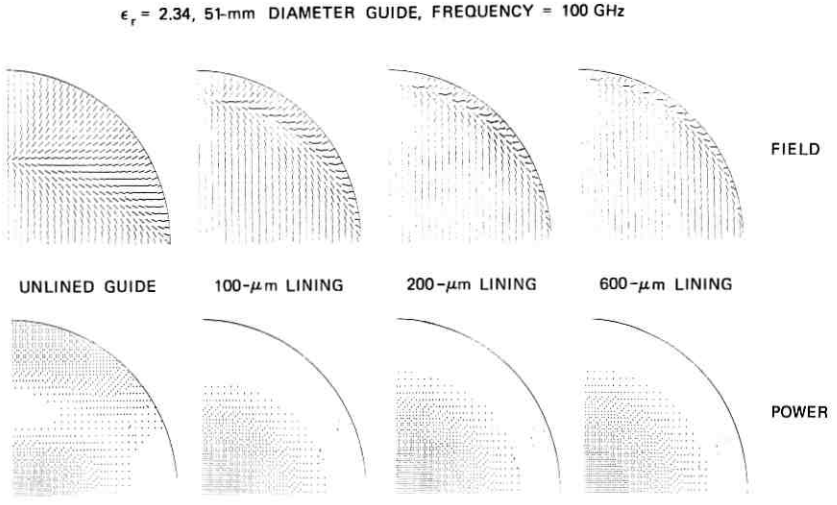

Fig. 10—TM_{11} electric field lines and power density vs lining thickness.

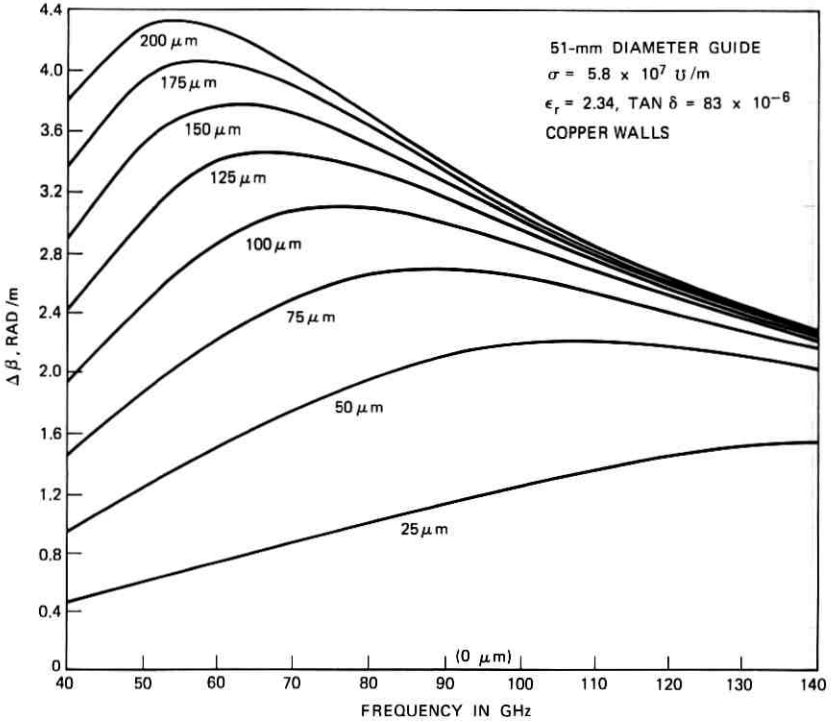

Fig. 11—TM_{11} $\Delta\beta$ vs frequency, wall-impedance model.

the TE_{pn} modes ($n \neq 1$), as they have a much lower loss than the TM_{pn} modes for thin linings.

Figures 9 and 10 are plots of the electric field lines and power distribution over a quadrant of the guide cross section for the TE_{12} and TM_{11} modes in the *free-space region* of the guide for various lining thicknesses at 100 GHz.

It can be seen that the lining drastically alters the field lines and energy distribution from those in copper waveguide. We also see that the low-loss regions of Fig. 6 are consistent with the energy distributions shown in Figs. 9 and 10 (as they correspond to the cases in Figs. 9 and 10 where the energy at the free-space dielectric interface is quite low). The TE_{12} energy density also changes much more slowly with lining thickness than does that for the TM_{11} mode as expected from the eigenvalue plots of Fig. 4. Note that the TM_{11} field lines and energy density change little over a 100- to 600-μm range of lining thickness. There would be a marked change for a lining thicker than 600 μm, as

the TM_{11} mode then assumes a surface wave behavior. The energy density distribution dependence on lining thickness was found to be similar for the circular symmetric TE_{on} and TM_{on} modes.

In the preceding discussion, we have considered the normal mode characteristics in DLG for unrealistically thick linings. (In Fig. 5, the TE_{01} heat loss is 21 dB/km at 100 GHz for the 600-μm thick polyethylene lining considered.) In the following paragraphs we will present some data for more representative lining thickness over a typical frequency band (40 to 110 GHz).

It is well known that a dielectric lining is required to break the degeneracy between the TE_{01} and TM_{11} modes in unlined guide. In Fig. 11, $\Delta\beta_{TM_{11}}$ is plotted for several lining thicknesses versus frequency. We see that for a lining thickness of greater than 150 μm the increase in $\Delta\beta$ is relatively small over the frequency range 70 to 110 GHz. We also note that $\Delta\beta$ is approximately inversely proportional to frequency for frequencies greater than 60 GHz for a 200-μm lining.

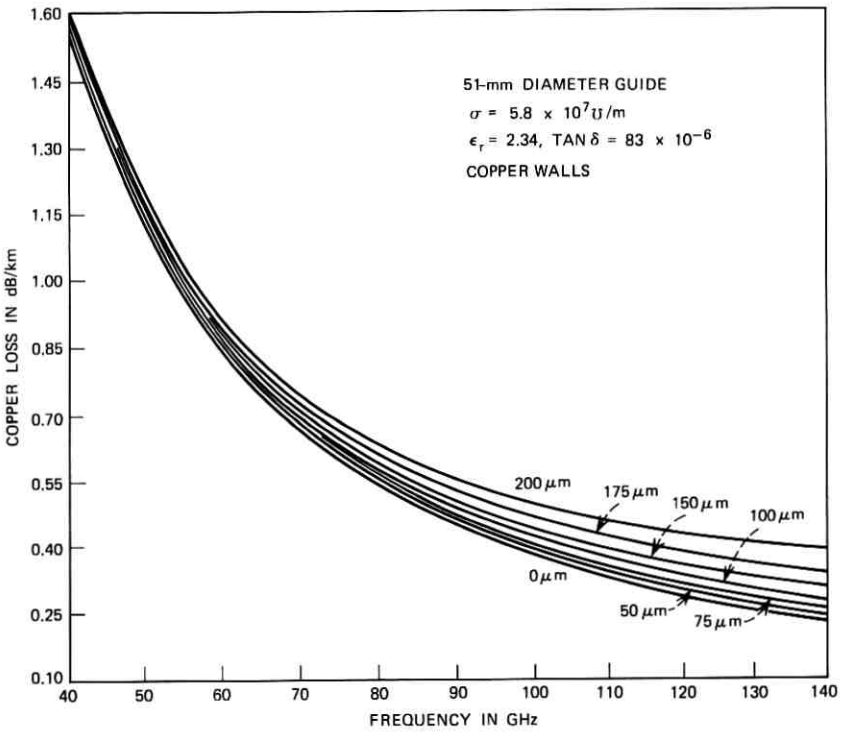

Fig. 12—TE_{01} copper loss vs frequency; wall-impedance model.

Conversely, in Figs. 12 and 13 it can be seen that a lining of 200 μm increases the TE_{01} copper loss by 0.12 dB/km at 110 GHz. The dielectric loss is negligible ($\approx$0.01 dB/km) for a 200-μm or thinner lining with a tan δ of 83 $\times$ 10^{-6}. On the other hand, the same liner with a tan δ of 10^{-3} would have a significant (0.12 dB/km) dielectric loss at 110 GHz. The preceding indicates a polyethylene liner with a thickness of 100 to 200 μm, and a tan $\delta < 10^{-4}$ is suitable for a transmission system using 51-mm diameter guide. A more precise determination of the optimal lining thickness is beyond the scope of this paper, as it depends in a complex fashion on the geometrical aberrations of the guide *in situ*.

Figures 14 through 16 are $\Delta\beta$ plots for several TE and TM modes over the 40- to 110-GHz band in 51-mm guide for 0, 100, and 200 μm polyethylene linings. The TE_{on} and $TE_{pn}(n > 1)$ mode $\Delta\beta$'s are quite close to the unlined guide $\Delta\beta$'s for a 200-μm lining and are essentially inversely proportional to frequency. The TM_{pn} mode $\Delta\beta$'s are strongly dependent on lining thickness for linings of 100 μm or less and essentially little changed for linings greater than 200 μm. The TM_{pn} mode

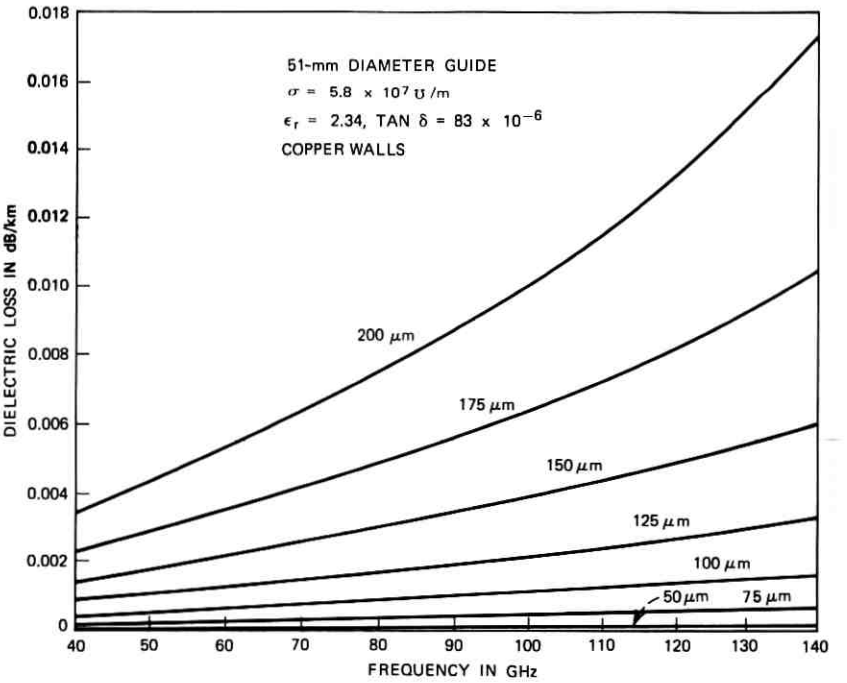

Fig. 13—TE_{01} dielectric losses vs frequency; wall-impedance model.

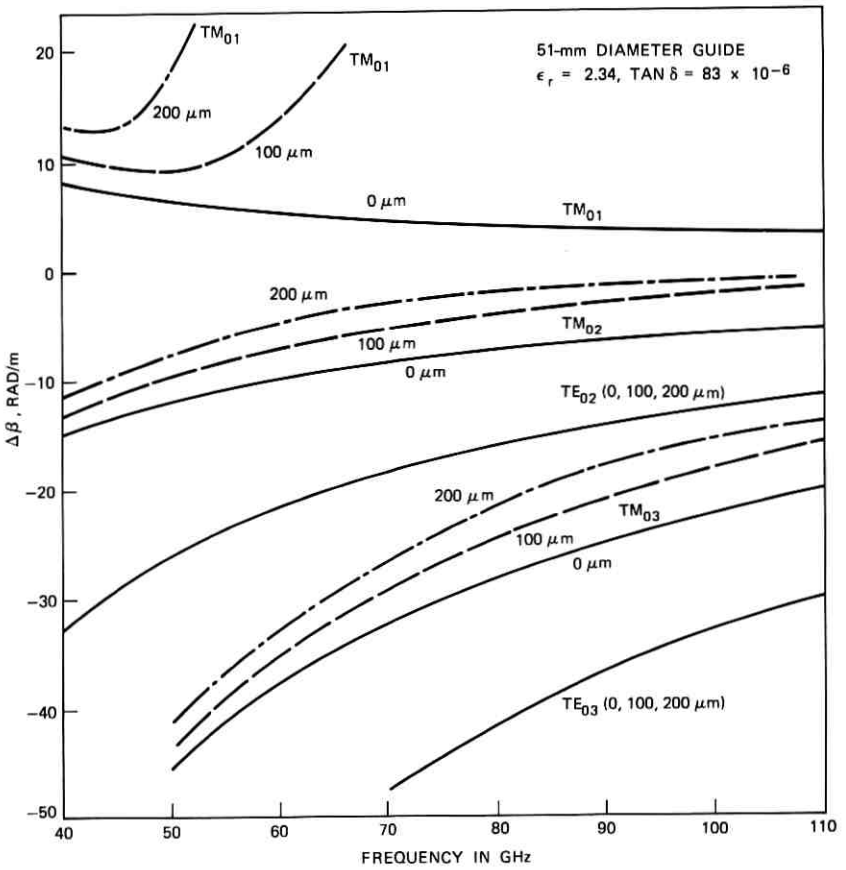

Fig. 14—TE_{on} and TM_{on} $\Delta\beta$ vs frequency; wall-impedance model.

$\Delta\beta$'s shown exhibit the $1/f$-type frequency dependence for a 200-μm lining for frequencies > 90 GHz. It should be noted that the TM_{21} mode has a very low $\Delta\beta$ for a 200-μm lining which necessitates the avoidance of long mechanical wavelength elliptical distortions over the guide cross section.

The TE_{p1} $\Delta\beta$'s are strongly dependent on lining thickness over the entire 40- to 110-GHz band, and their behavior is quite different from that of the other modes. This is due to the surface wave character of these modes. As a consequence of this, the TE_{p1} modes having eigenvalues >3.83171 (the TE_{01} eigenvalue in unlined guide) may become degenerate with the TE_{01} mode at selected frequencies within the band of interest as shown in Fig. 17. The TE_{31} mode goes through a degener-

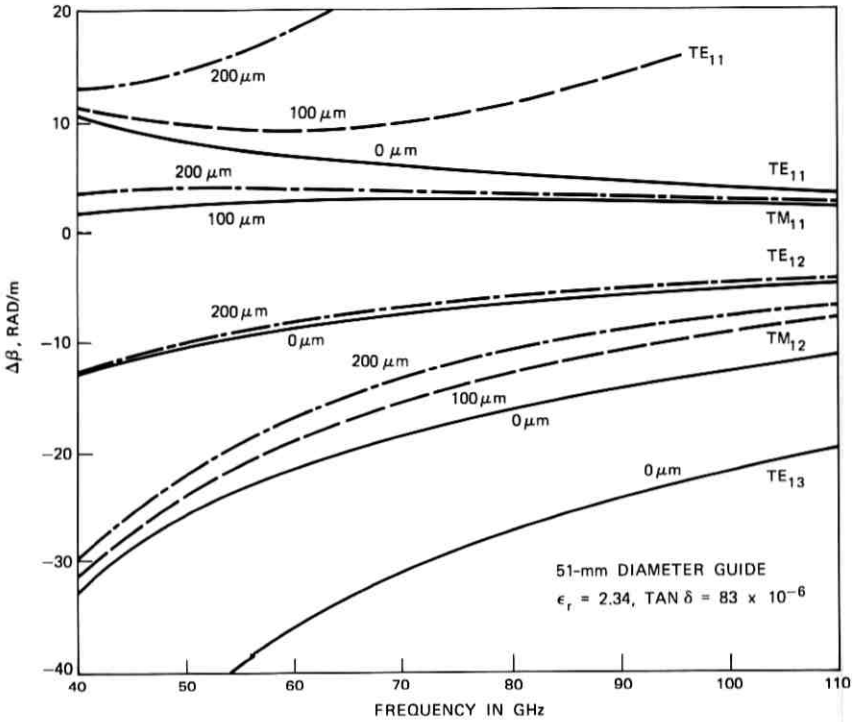

Fig. 15—TE_{1n} and TM_{1n} $\Delta\beta$ vs frequency; wall-impedance model.

acy at a frequency <40 GHz, while the TE_{21} and TE_{11} modes are never degenerate.

In Figs. 18 through 23, the heat loss of several normal modes is plotted for a 0-, 100-, and 200-μm polyethylene lining in 51-mm guide over the 40- to 110-GHz band. The characteristics may be summarized as:

(*i*) The TM_{pn} (except for the TM_{01} mode) heat loss initially increases with frequency, levels off, and then decreases by an order of magnitude for a 200-μm lining from 40 to 110 GHz. The presence of the lining substantially reduces the heat loss at 110 GHz. The losses are as low as 3 dB/km at 110 GHz.

(*ii*) The TE_{on} losses decrease with frequency, but not in the $f^{-\frac{1}{2}}$ fashion as in unlined guide. The presence of a lining leads to a frequency-dependent increase in heat loss over the 40- to 100-GHz band; approximately 5 percent and 30 percent for

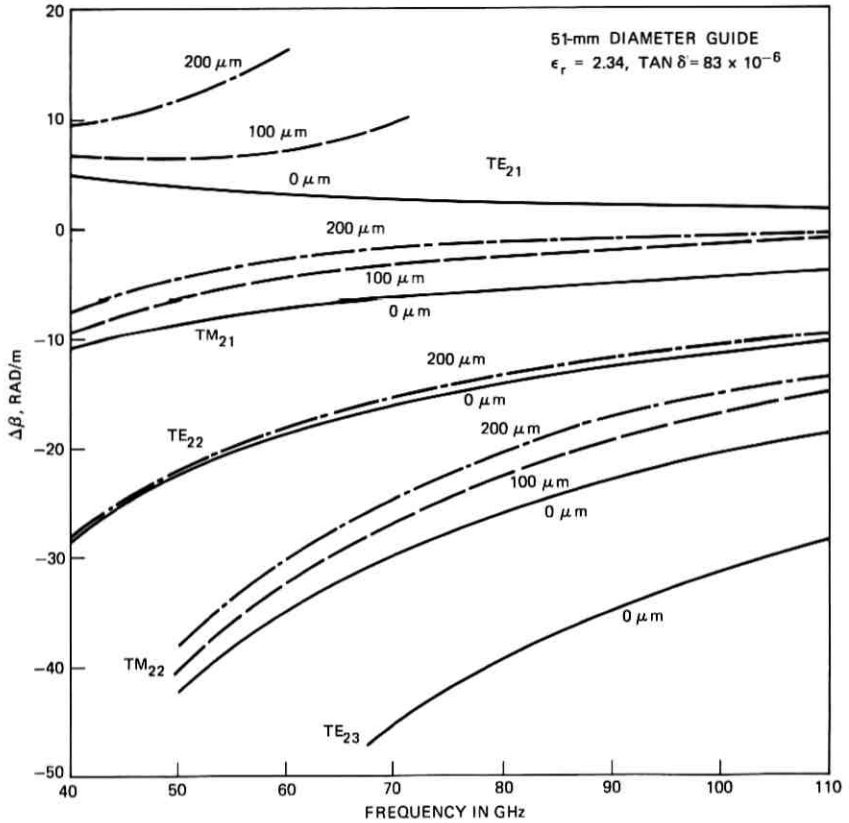

Fig. 16—TE_{2n} and TM_{2n} $\Delta\beta$ vs frequency; wall-impedance method.

the TE_{01} mode at 40 and 110 GHz, respectively, for a 200-μm lining.

(*iii*) The TE_{p1} and TM_{01} heat loss is quite high (>100 dB/km) for even thin linings, as these are surface-wave-type modes.

(*iv*) The TE_{pn} ($n \neq 1$) heat loss is low (<10 dB/km), and the presence of the lining actually reduces the heat loss over much of the 40- to 110-GHz range.

(*v*) The heat loss is substantially different from that predicted by the simple perturbation[3] theory for thin linings (for linings >10 μm) except in the case of the TE_{on} modes.

The normal mode characteristics, and in particular the low loss of the TM_{pn} modes, discussed in the previous paragraphs (Figs. 3 through 23) can be easily understood on recognizing that the lower-order modes

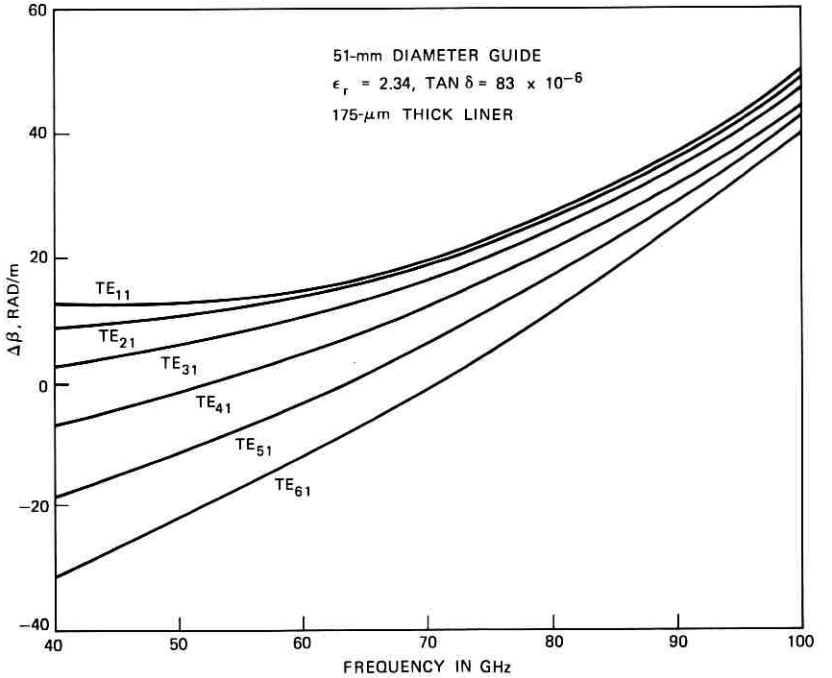

Fig. 17—TE_{n1} $\Delta\beta$ vs frequency; exact solution.

have a plane-wave-type field structure. The electric field of the plane waves may be polarized perpendicular (TE_{on} modes) to or parallel (TM_{on} modes) to the plane of incidence as shown in Fig. 24. A TE_{pn} or TM_{pn} mode (for $p \neq 0$) is necessarily a superposition of both polarizations.

In examining the interaction of these plane waves with the walls of the guide, it is helpful to consider the transmission line equivalent structures of Fig. 24. The transmission line parameters (characteristic impedance Z_c and propagation constant k_r in the radial direction) are dependent on the polarization. In the empty region of the waveguide they are given by:

Perpendicular Polarization TE_{on} Modes	Parallel Polarization TM_{on} Modes
$k_r = k_n/a$	$k_r = k_n/a$
$Z_c = \eta \dfrac{ka}{k_n}$	$Z_c = \eta \dfrac{k_n}{ka}$

(10)

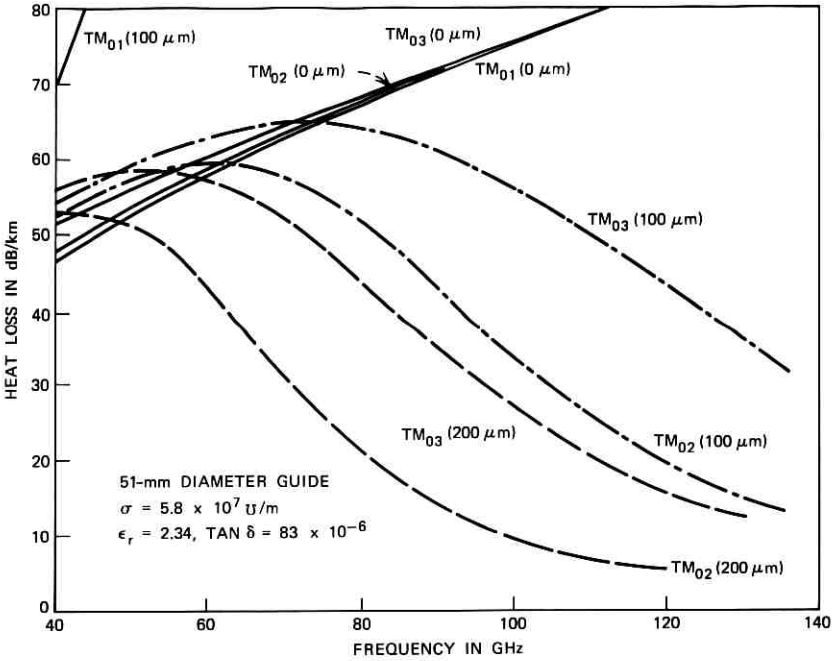

Fig. 18—TM_{on} heat loss vs frequency; wall-impedance model.

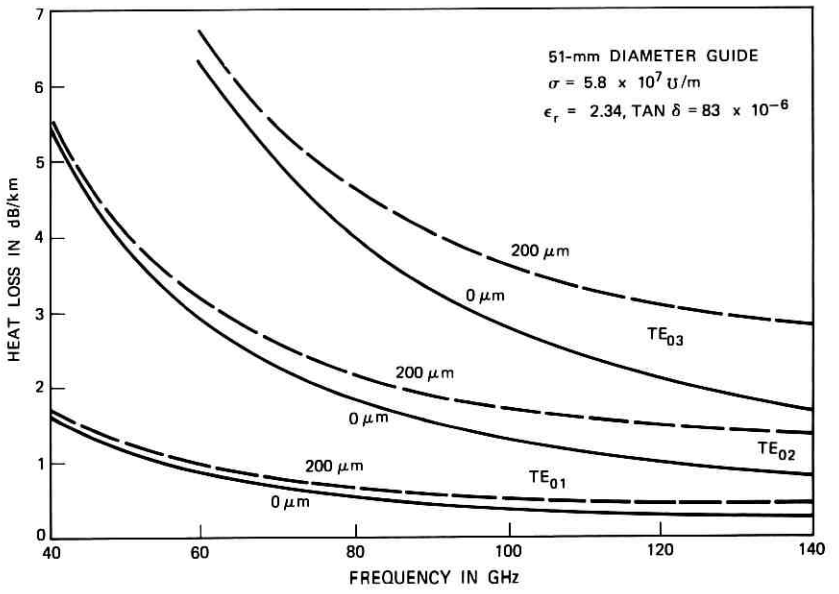

Fig. 19—TE_{on} heat loss vs frequency; wall-impedance model.

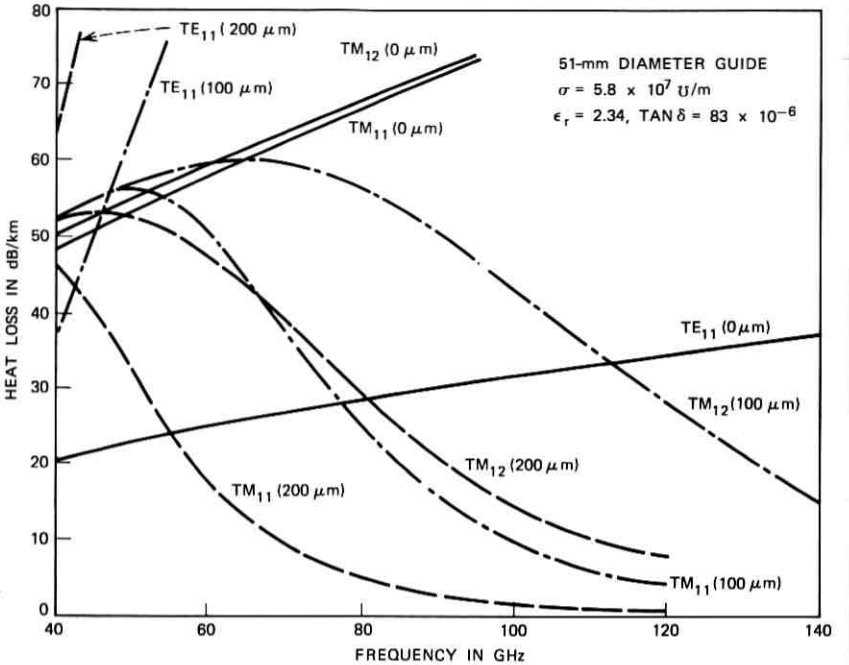

Fig. 20—TE_{11} and TM_{1n} heat loss vs frequency; wall-impedance model.

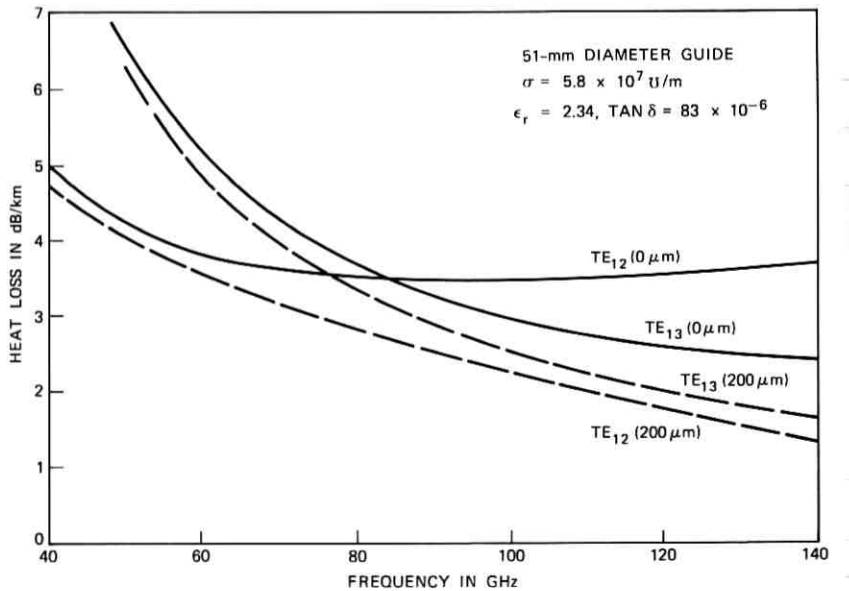

Fig. 21—TE_{12} and TE_{13} heat loss vs frequency; wall-impedance model.

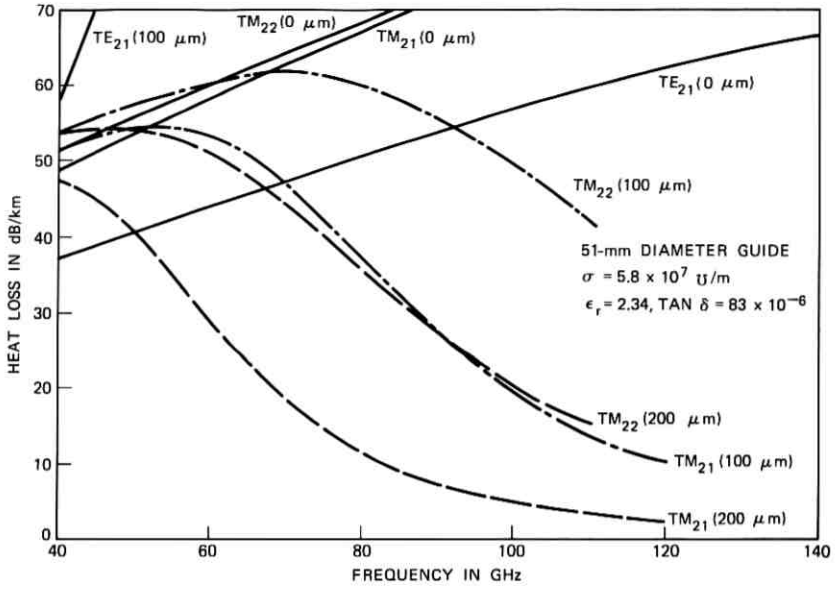

Fig. 22—TE_{21} and TM_{2n} heat loss vs frequency; wall-impedance model.

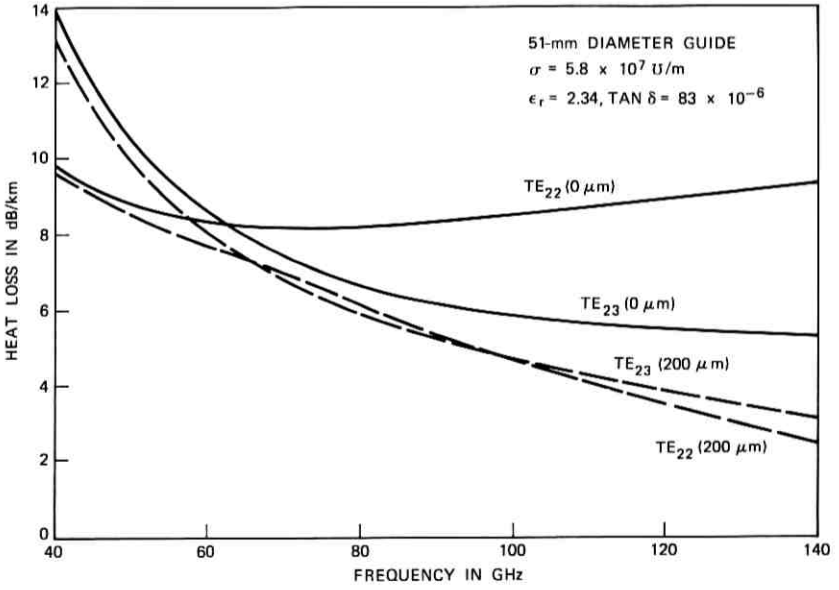

Fig. 23—TE_{22} and TE_{23} heat loss vs frequency; wall-impedance model.

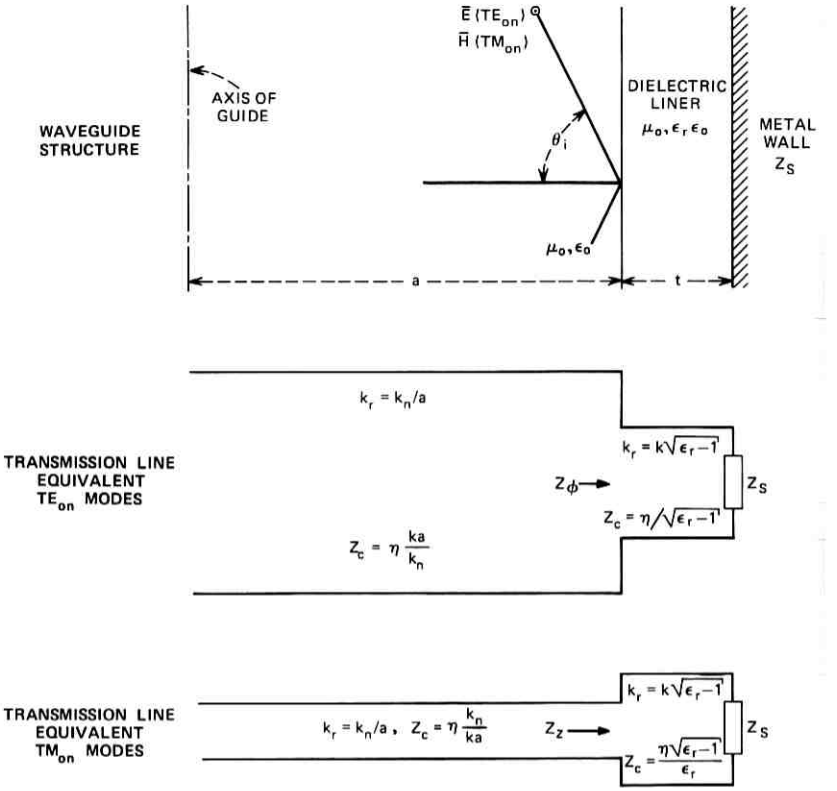

Fig. 24—Equivalent transmission line structures for the TE$_{on}$ and TM$_{on}$ modes.

where η is the impedance of free space, k_n the eigenvalue for the mode of interest, a the guide radius, and k the free-space propagation constant. In the dielectric region we have for the near grazing incidence ($\theta_i \cong 90°$) case:

Perpendicular Polarization TE$_{on}$ Modes	Parallel Polarization TM$_{on}$ Modes
$k_r \cong k\sqrt{\epsilon_r - 1}$	$k_r \cong k\sqrt{\epsilon_r - 1}$
$Z_c \cong \eta/\sqrt{\epsilon_r - 1}$	$Z_c \cong \eta\sqrt{\epsilon_r - 1}/\epsilon_r$

(11)

The single-layer case may be generalized to the multilayer case by simply adding the equivalent transmission line sections. The copper walls of the structure have a skin-effect surface impedance on the order of $0.05 + j\,0.05$ to $0.1 + j\,0.1$ ohms at 40 and 110 GHz, respectively,

for both polarizations. On transforming these impedances in the normal way through a 200-μm polyethylene liner with equivalent transmission line parameters as given by (11), we obtain the wall impedances Z_z and Z_ϕ which are seen at the dielectric free-space interface. These are on the order of $0.05 + j\,50$ to $0.1 + j\,200$ ohms. On the other hand, the characteristic impedances as given in (10) for the lower-order modes of most interest in the empty region of the guide are on the order of 10 to 80 ohms for parallel polarization (TM_{on} modes), while the perpendicular polarization (TE_{on}) characteristic impedances are on the order of 1000 to 10,000 ohms.

The reflection coefficients of the dielectric-clad copper wall for the equivalent transmission line structures in Fig. 24 are a function of Z_z/Z_c or Z_ϕ/Z_c. The propagation constant h_n or attenuation constant α for the mode of interest is related to the phase or amplitude of the appropriate reflection coefficient, respectively. In unlined copper guide, both of these ratios (Z_z/Z_c and Z_ϕ/Z_c) are $\ll 1$. For the TM_{on} modes in DLG, $|Z_z|$ is on the order of or much greater than Z_c, and thus these modes are drastically changed from their copper guide equivalents. This is shown by the large changes in $\Delta\beta$ for these modes and the drastic reduction in the heat loss. On the other hand, $|Z_\phi|$ is much less than Z_c for the TE_{on} modes, and hence the $\Delta\beta$ and heat loss of these modes is little changed from that in copper guide. The TE_{pn} modes ($n \neq 1$, $p \neq 0$) have a TE_{on}-like field structure in the region of the wall in copper guide and thus are little affected by the lining. The TM_{pn} and TE_{p1} modes have a TM_{on}-like field structure in the wall region in copper guide, and they are drastically altered by the presence of a lining as we have observed.

IV. SYSTEM IMPLICATIONS

In the preceding sections, we have seen that the TM_{11} and TE_{12} modes have very low heat losses ($\lesssim 3$ dB/km) at 110 GHz. This low heat loss leads to a reduction in route bend losses as can be seen in the following. Unger[3] shows that the TE_{01} mode undergoes added losses in a route bend due to two different effects. The first effect, called mode conversion loss, arises from the fact that a portion of the TE_{01} energy is converted into other modes (predominantly TM_{11} and TE_{12} in DLG) in the bend and is not reconverted back to the TE_{01} mode but instead emerges from the bend as a spurious mode.

The second loss mechanism, defined as the added heat loss, is associated with the energy initially converted from the TE_{01} mode to the TE_{12} or TM_{11} mode and then reconverted back to the TE_{01} mode as

the bend is traversed. Since this energy travels in the higher-loss TE_{12} and TM_{11} modes over a portion of the bend, it suffers additional attenuation over and above the TE_{01} heat loss in DLG. For a given mode, the mode conversion loss and added heat loss in a route bend of radius R and total bend length L, with a taper length ℓ (the radius of curvature is assumed to have a linear taper of 0 to R from 0 to ℓ and R to 0 for $L - \ell$ to L), is given by

$$\begin{aligned} L_{\text{RB–AH}}\ (\text{added heat loss}) &= \frac{L - \frac{3}{4}\ell}{R^2} \frac{C^2}{\Delta\beta_n^2} \Delta\alpha_n \\ L_{\text{RB–MC}}\ (\text{mode conversion loss}) &= \frac{4}{(R\ell)^2} \frac{C^2}{\Delta\beta_n^4}. \end{aligned} \tag{12}$$

Here C is the normalized coupling coefficient[4] due to curvature of the guide axis, and the other parameters are as previously defined. In Figs. 25 and 26, the total added TE_{01} loss, due to the route bend, as

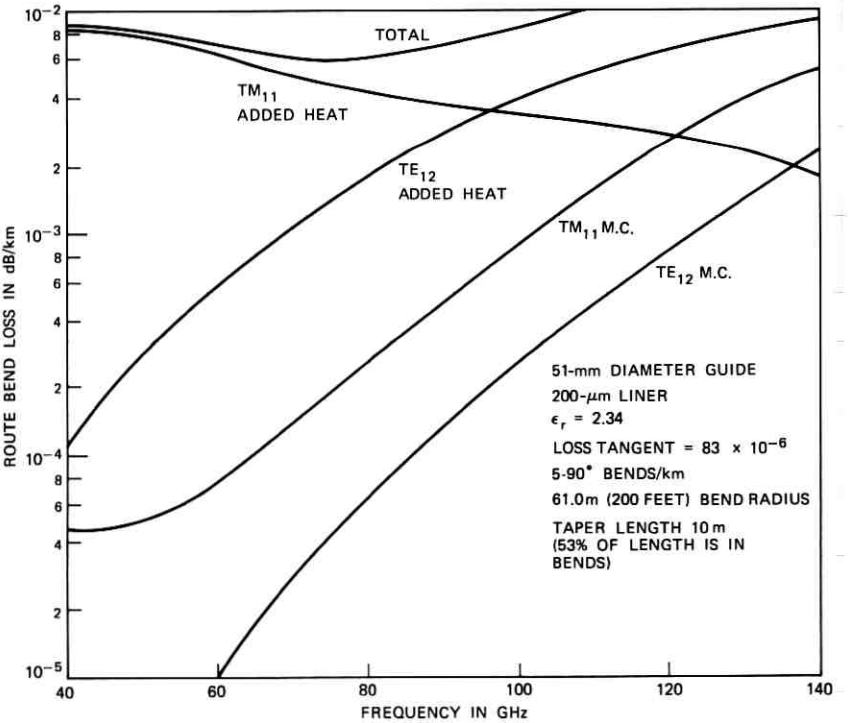

Fig. 25—Theoretical route bend losses.

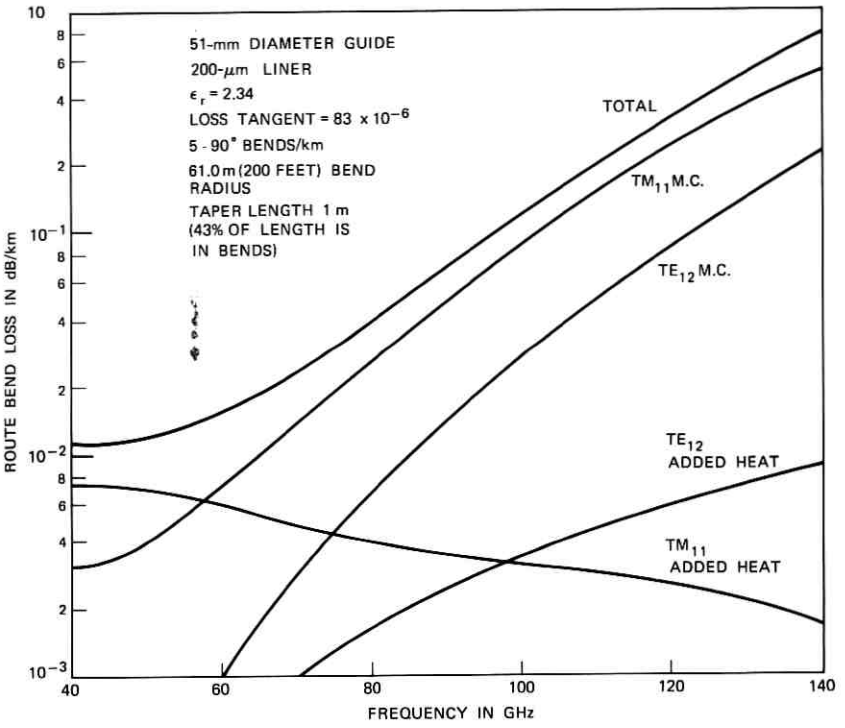

Fig. 26—Theoretical route bend losses.

well as the individual components comprising it is plotted for some idealized routes comprised of five 90-degree bends with a constant radius of curvature (61 m) and linear tapers of 1 or 10 m in length in DLG over a length of 1 km.

From Fig. 26, it is obvious that the mode conversion losses are far more significant then the added heat loss in route bends with a short taper (or zero) length. On the other hand, for a reasonable taper length (10 m), the mode conversion losses are significantly lower than the added heat loss arising from coupling to the TM_{11} and TE_{12} modes. The linear taper is probably a reasonable approximation to that expected in practice as the radius of curvature will change in a smooth fashion as the waveguide progresses from a straight run into a curve.

The added heat loss is substantially reduced by the presence of the lining due to the great reduction in the TM_{11} heat loss. For example, at 110 GHz the TM_{11} mode yields an added heat loss for the route bends of Fig. 25 of approximately 0.005 dB/km. If the TM_{11} mode

heat loss had followed the perturbation[3] theory prediction, the added heat loss for the same route bend would be ≈ 0.250 dB/km.

V. CONCLUSION

We have discussed the basic propagation characteristics of the normal modes of dielectric-lined overmoded circular waveguide. Some of the results obtained differ in surprising fashion from those predicted by the perturbation theory of Unger.[3] The most significant result is the low heat loss of the TM_{1n} modes at the upper end of the 40- to 110-GHz band discussed. This effect greatly reduces the added heat loss for the TE_{01} mode in route bends over this part of the frequency spectrum.

A simple physical explanation based on an equivalent transmission line problem derived from an analogous plane-wave problem was advanced to support the results contained here. An experimental investigation of the decreasing TM_{11} heat loss[6] vs frequency has been completed, and the data confirm the trends predicted herein.

VI. ACKNOWLEDGMENT

The authors are indebted to W. DeLang for his perspicacity and thoroughness in carrying out the numerical computations.

APPENDIX A

Exact Formulation

The waveguide structure under consideration is shown in Fig. 1 of the main text of this paper. The notation used conforms with that of Unger.[4] Much of the following is similar to that contained in Unger[3,4] but is repeated here for reasons of continuity and completeness.

The fields in the guide can be derived from two suitable scalar wave functions composed of suitable products of trigonometric and Bessel functions.

$$\begin{aligned} T_n &= N_n J_p(\chi_n r) \sin p\phi \\ T'_n &= N_n J_p(\chi_n r) \cos p\phi \end{aligned} \qquad 0 < r < a \tag{13}$$

$$\begin{aligned} T_n &= N_n \frac{\chi_n^2}{\chi_n^{e2}} J_p(k_n) \frac{Z_{pn}(\chi_n^e r)}{Z_{pn}(k_n^e)} \sin p\phi \\ T'_n &= N_n \frac{\chi_n^2}{\chi_n^{e2}} J_p(k_n) \frac{Z_{pn}(\chi_n^e r)}{Z_{pn}(k_n^e)} \cos p\phi \end{aligned} \qquad a < r < b \tag{14}$$

where

$$Z_{pn}(\chi_n^e r) = H_p^{(2)}(\chi_n^e r) - cH_p^{(1)}(\chi_n^e r)$$
$$Z_{pn}(\chi_n^e r) = H_p^{(2)}(\chi_n^e r) - c'H_p^{(1)}(\chi_n^e r).$$

Z'_{pn} and Z'_{pn} are the derivatives of the above expressions with respect to the entire argument. The constants c and c' will be defined later. The T functions satisfy the wave equation

$$\nabla_t^2 T = \frac{1}{r}\left[\frac{\partial}{\partial r}\left(r\frac{\partial T}{\partial r}\right) + \frac{\partial}{\partial \phi}\left(\frac{1}{r}\frac{\partial T}{\partial \phi}\right)\right] = -\chi^2 T \tag{15}$$

where χ^2 is given by χ_n^2 or χ_n^{e2} for $0 < r < a$ or $a < r < b$ respectively, and ∇_t^2 is the transverse Laplacian operator.

The fields in the guide may be derived from the scalar wave functions as in Unger[4]

$$E_r = \sum_n V_n\left[\frac{\partial T_n}{\partial r} + d_n\frac{\partial T'_n}{r\partial\phi}\right]$$
$$E_\phi = \sum_n V_n\left[\frac{\partial T_n}{r\partial\phi} - d_n\frac{\partial T'_n}{\partial r}\right]$$
$$H_r = -\sum_n I_n\left[\frac{\partial T_n}{r\partial\phi} - d_n\frac{h_n^2}{k^2\epsilon}\frac{\partial T'_n}{\partial r}\right]\epsilon$$
$$H_\phi = \sum I_n\left[\frac{\partial T_n}{\partial r} + d_n\frac{h_n^2}{k^2\epsilon}\frac{\partial T'_n}{r\partial\phi}\right]\epsilon \tag{16}$$

$$H_z = j\omega\epsilon\epsilon_0\sum_n V_n d_n\frac{\chi^2}{k^2}T'_n$$
$$E_z = j\omega\mu_0\sum_n I_n\epsilon\frac{\chi^2}{k^2}T_n \tag{17}$$

where μ_0 and ϵ_0 are the permeability and permittivity of free-space, respectively, ϵ is the relative permittivity (1 for $0 < r < a$, ϵ_r for $a < r < b$) over the guide cross section, $k = \omega\sqrt{\mu_0\epsilon_0\epsilon}$ is the intrinsic plane-wave propagation constant in the various regions of the guide. Here, ϵ, k, and χ have constant but different values over the guide cross section as may the separation constant d_n. A mode for the structure consists of one term in the series given in (16) and (17). For a given modal field distribution, the individual V_n and I_n's in (16) and (17) represent a *single* forward or backward traveling wave, and they are related to each other by the modal impedance Z_n

$$\frac{V_n}{I_n} = Z_n = \frac{h_n}{\omega\epsilon_0}. \tag{18}$$

Imposition of the appropriate boundary conditions at $r = a$, and $r = b$ leads to a determination of d_n, c, c', χ_n, and χ_n^e on solution of a transcendental equation. From Unger[4] we have

$$c = \frac{H_p^{(2)}(\rho k_n^e)}{H_p^{(1)}(\rho k_n^e)} \tag{19}$$

$$c' = \frac{H_p^{(2)\prime}(\rho k_n^e)}{H_p^{(1)\prime}(\rho k_n^e)} \tag{20}$$

where the $'$ associated with the Bessel function denotes differentiation with respect to the entire argument. We also define ρ, k_n, and k_n^e

$$\rho = b/a \tag{21}$$

$$k_n = \chi_n a \tag{22}$$

$$k_n^e = \chi_n^e a. \tag{23}$$

We find d_n is constant over the guide cross section and is given by

$$\frac{1}{d_n} = \frac{k_n^2 k_n^{e2}}{pk^2a^2(\epsilon_r - 1)} [y_{pn}(k_n) - Y_{pn}(k_n^e)] \tag{24}$$

where

$$\begin{aligned} y_{pn} &= \frac{J_p'(k_n)}{k_n J_p(k_n)} \\ Y_{pn}(k_n^e) &= \frac{Z_{pn}'(k_n^e)}{k_n^e Z_{pn}(k_n^e)} \\ Y_{pn}(k_n^e) &= \frac{Z_{pn}'(k_n^e)}{k_n^e Z_{pn}(k_n^e)}. \end{aligned} \tag{25}$$

k_n and k_n^e are related by

$$\begin{aligned} k_n^2 &= (\omega^2 \epsilon_0 \mu_0 - h_n^2)a^2 \\ k_n^{e2} &= (\omega^2 \epsilon_r \epsilon_0 \mu_0 - h_n^2)a^2. \end{aligned} \tag{26}$$

The individual modes may be determined by solution of the eigenvalue equation:

$$[y_{pn}(k_n) - Y_{pn}(k_n^e)] \cdot [y_{pn}(k_n) - \epsilon_r Y_{pn}(k_n^e)] = p^2(\epsilon_r - 1)^2 \frac{h_n^2 k^2 a^4}{k_n^4 k_n^{e4}} \tag{27}$$

for quasi TE_{pn}, TM_{pn} modes, $p \neq 0$. For circular electric waves ($p = 0$), the eigenvalue equation is

$$[y_{on}(k_n) - Y_{on}(k_n^e)] = 0. \tag{28}$$

For circular magnetic waves (TM_{0n}) we have

$$[y_{on}(k_n) - \epsilon_r Y_{on}(k_n^e)] = 0. \tag{28a}$$

The constant N_n is given by imposing an ortho-normalization criterion on the transverse fields given in (16). This condition is

$$\frac{1}{V_n I_m} \int\int \bar{E}_n \times \bar{H}_m \cdot d\bar{s} = \begin{cases} 1 & n = m \\ 0 & n \neq m. \end{cases}$$

On evaluation of this integral, we find that N_n is exactly given by

$$\begin{aligned} \frac{\pi}{2} N_n^2 k_n^2 J_p^2(k_n) \Bigg\{ & \left(1 + \frac{d_n^2 h_n^2}{k^2}\right)\left(1 - \frac{p^2}{k_n^2} + k_n^2 y_{pn}^2 + 2y_{pn}\right) \\ & - 2d_n \frac{p}{k_n^2}\left[\left(1 + \frac{h_n^2}{k^2}\right) - \left(1 + \frac{h_n^2}{\epsilon_r k^2}\right)\frac{\epsilon_r k_n^4}{k_n^{e4}}\right] \\ & + \left(\frac{k_n}{k_n^e}\right)^2 \frac{d_n^2 h_n^2}{k^2}\left(p^2 \frac{Z_{pn}^2(\rho k_n^e)}{Z_{pn}^2(k_n^e)}\right)\left[1 - \frac{p^2}{\rho^2 k_n^{e2}}\right] \\ & - \left[1 - \frac{p^2}{k_n^{e2}} + k_n^{e2} Y_{pn}^2 + 2Y_{pn}\right] \\ & + \epsilon_r \left(\rho^2 \frac{Z_{pn}^{2\prime}(\rho k_n^e)}{Z_{pn}^2(k_n^e)} - \left[1 - \frac{p^2}{k_n^{e2}} + k_n^{e2} Y_{pn}^2 + 2Y_{pn}\right]\right)\Bigg\} = 1 \end{aligned} \tag{29}$$

for TE_{pn} or TM_{pn} modes, $p \neq 0$ and by

$$\begin{aligned} \pi N_n^2 k_n^2 J_0^2(k_n) \frac{d_n^2 h_n^2}{k^2} & \Bigg\{(1 + k_n^2 y_{0n}^2 + 2y_{on}) \\ & + \left(\frac{k_n}{k_n^e}\right)^2 \left[\rho^2 \frac{Z_{on}^2(\rho k_n^e)}{Z_{on}^2(k_n^e)} - (1 + k_n^{e2} Y_{on}^2 + 2Y_{on})\right]\Bigg\} = 1 \end{aligned}$$

$$\begin{aligned} \pi N_n^2 k_n^2 J_0^2(k_n) & \Big\{(1 + k_n^2 y_{0n}^2 + 2y_{on}) \\ & + \epsilon_r \left(\frac{k_n}{k_n^e}\right)^2 \left[\rho^2 \frac{Z_{on}^{2\prime}(\rho k_n^e)}{Z_{on}^2(k_n^e)} - (1 + k_n^{e2} Y_{on}^2 + 2Y_{on})\right]\Big\} = 1 \end{aligned}$$

for circular electric (TE_{0m}) and magnetic (TM_{0n}) modes, respectively. We shall use a subscript n when referring to TE_{pn} or TM_{pn} modes and a subsequent m for TE_{om} modes.

Solution of the appropriate eigenvalue equation (27) or (28) leads to a value for the propagation constant (h_m or h_n) which is pure real as we have assumed a lossless dielectric-lined guide in the above analy-

sis. For an actual guide, h_n and h_m have a small imaginary part which we call the attenuation constant (α_{om} or α_{pn}) for the mode in question. The loss is due to currents flowing in the metal walls of the guide and the lossy dielectric.

The heat loss or attenuation constant (α) due to currents flowing in the metal walls of the structure is given by

$$\alpha = \frac{1}{2}\frac{P_L}{P_T}$$

where

$$P_L = \frac{1}{2}\int i\cdot i^* R_s\, dS = \frac{1}{2}\int_0^1 \int_0^{2\Pi} (H_z\cdot H_z^* + H_\phi\cdot H_\phi^*)\Big|_{r=b} R_s\, b\, d\phi\, dz,$$

i.e., power loss/unit length,

$$P_T = \frac{1}{2}\int E_t \times H_t^* ds = \frac{1}{2}\int E_t \times H_t\, dS = \tfrac{1}{2} V_n I_n,$$

i.e., the total average power flowing through the guide cross section for a given mode, and R_s is the surface resistance of the metal walls.

$$H_z(b) = j\frac{V_n d_n}{\omega\mu_0} N_n \chi_n^2 J_p(k_n)\frac{Z_{pn}(\rho k_n^e)}{Z_{pn}(k_n^e)}\cos p\phi$$

$$H_\phi(b) = I_n \epsilon_r N_n \frac{\chi_n^2}{\chi_n^{e2}}\left[\chi_n^e \frac{Z'_{pn}(\rho k_n^e)}{Z_{pn}(k_n^e)} - d_n \frac{h_n^2 p}{\omega^2\mu_0\epsilon_0\epsilon_r b}\frac{Z_{pn}(\rho k_n^e)}{Z_{pn}(k_n^e)}\right]\sin p\phi$$

noting $I_n = V_n/K_n$ where $K_n = h_n/\omega\epsilon_0$. We find

$$\alpha_{pn_{\text{copper}}} = \frac{\Pi}{2}\frac{h_n R_s}{\omega\epsilon_0}|N_n|^2\left(\frac{k_n}{k_n^e}\right)^4 J_p^2(k_n)b \times \left\{\left(\frac{\omega\epsilon_r\epsilon_0}{h_n}\right)^2\left|\chi_n^e\frac{Z'_{pn}(\rho k_n^e)}{Z_{pn}(k_n^e)} - \frac{d_n h_n^2 p}{\omega^2\mu_0\epsilon_0\epsilon_r b}\frac{Z_{pn}(\rho k_n^e)}{Z_{pn}(k_n^e)}\right|^2 + \left(\frac{|d_n|}{\omega\mu_0}\right)^2 \chi_n^{e4}\left|\frac{Z_{pn}(\rho k_n^e)}{Z_{pn}(k_n^e)}\right|^2\right\}.$$

We find for TE_{0m} modes using a similar argument that

$$\alpha_{om_{\text{copper}}} = \frac{\Pi h_m}{\omega\epsilon_0} R_s |N_m|^2\left(\frac{k_m}{k_m^e}\right)^4 J_0^2(k_m)b\left(\frac{|d_m|}{\omega\mu_0}\right)^2 \chi_m^{e4}\left|\frac{Z_{0m}(\rho k)_m^e}{Z_{0m}(k)_m^e}\right|^2$$

and for TM_{0n}

$$\alpha_{on_{\text{copper}}} = \frac{\Pi h_n R_s}{\omega\epsilon_0}|N_n|^2\left(\frac{k_n}{k_n^e}\right)^4 J_0^2(k_n)b\left(\frac{\omega\epsilon_r\epsilon_0}{h_n}\right)^2\left|\chi_n^e\frac{Z'_{pn}(\rho k_n^e)}{Z_{pn}(k_n^e)}\right|^2.$$

The dielectric heat loss may be found in a similar manner as we now have

$$P_L = \frac{1}{2}\int ii^*\rho dV = \frac{1}{2}\int \frac{\sigma E\cdot\sigma E^*}{\sigma}\,dV = \frac{1}{2}\int_0^1\int_0^{2\Pi}\int_a^b \sigma E\cdot E^* r dr d\phi dz$$

where σ is the conductivity of the dielectric. Using field approximations in the lining as given in Unger we have

$$E_r = -V_n A\left[(\chi_n^e S) + \left(\frac{d_n pC}{r}\right)\right]\cos\left[\left(\rho - \frac{r}{a}\right)k_n^e\right]\sin p\phi$$

$$E_\phi = V_n A\left[\left(\frac{pS}{r}\right) - (d_n\chi_n^e C)\right]\sin\left[\left(\rho - \frac{r}{a}\right)k_n^e\right]\cos p\phi$$

$$E_z = jI_n AS\frac{\chi_n^{e2}}{\omega\epsilon_0}\sin\left[\left(\rho - \frac{r}{a}\right)k_n^e\right]\sin p\phi$$

where

$$A \equiv N_n\left(\frac{k_n}{k_n^e}\right)^2 J_p(k_n)$$

$$S \equiv \frac{1}{\sin\delta k_n^e}; \qquad C \equiv \frac{1}{\cos\delta k_n^e}.$$

By substituting for the fields and performing the integration the power loss is obtained. Note that $\sigma = \omega\epsilon_r\epsilon_0 \times$ [lining loss tangent] where ϵ_r is the relative dielectric constant of the lining.

For the TE_{pn} or TM_{pn} modes

$$\alpha_{pn_{\text{dielectric}}} = \frac{\Pi h_n}{2}[\epsilon_r(\text{loss tangent})]|N_n|^2\left(\frac{k_n}{k_n^e}\right)^4 J_p^2(k_n)$$
$$\times\left\{\frac{(\delta+2)\delta k_n^{e2}}{4}\left[S^2\left(1+\frac{\chi_n^{e2}}{h_n^2}\right)+C^2|d_n|^2\right]\right.$$
$$+\tfrac{1}{4}[k_n^e\sin 2\delta k_n^e + \tfrac{1}{2}(1-\cos 2\delta k_n^e)]$$
$$\times\left[S^2\left(1-\frac{\chi_n^{e2}}{h_n^2}\right)-C^2|d_n|^2\right] + SC[\text{Re}\,(d_n)]p\sin 2\delta k_n^e$$
$$\left.+\frac{p^2}{2k_n^e}[\delta k_n^e(S^2+C^2|d_n|^2)+\tfrac{1}{2}\sin 2\delta k_n^e(-S^2+C^2|d_n|^2)]\right\}.$$

We find for TE_{om} modes that

$$\alpha_{om_{\text{dielectric}}} = \Pi h_m[\epsilon_r(\text{loss tangent})]|N_m|^2\left(\frac{k_m}{k_m^e}\right)^4 J_0^2(k_m)\frac{C^2|d_m|^2}{4}$$
$$\times[(\delta+2)\delta k_m^{e2} - k_m^e\sin 2\delta k_m^e - \tfrac{1}{2}(1-\cos 2\delta k_m^e)]$$

and for TM_{0n} modes

$$\alpha_{0n_{\text{dielectric}}} = \frac{\Pi h_n}{4}\left[\epsilon_r(\text{loss tangent})\right]|N_n|^2\left(\frac{k_n}{k_n^e}\right)^4 J_0^2(k_n) \times S^2\left\{(\delta+2)\delta k_n^e\left(1+\frac{\chi_n^{e2}}{h_n^2}\right)+\left[k_n^e \sin 2\delta k_n^e + \tfrac{1}{2}(1-\cos 2\delta k_n^e)\right]\left(1-\frac{\chi_n^{e2}}{h_n^2}\right)\right\}.$$

APPENDIX B

Wall Impedance Formulation

The determination of the wall impedance (Z_z and Z_ϕ) are discussed in Section III of the text. Here, we will outline the derivation of the eigenvalue equation and the subsequent determination of the propagation and attenuation constants. The boundary conditions at $r = a$ are

$$\begin{aligned} Z_\phi &= E_\phi/H_z|_{r=a} \\ Z_z &= -E_z/H_\phi|_{r=a}. \end{aligned} \tag{30}$$

As before, we can express the fields in the simplified wall impedance structure of Fig. 2 as a superposition of two scalar functions.

$$\begin{aligned} T_n &= N_n J_p(\chi_n r)\sin p\phi \\ T'_n &= N_n J_p(\chi_n r)\cos p\phi. \end{aligned} \tag{31}$$

Here p is an integer, and we must solve a suitable eigenvalue equation to determine χ_n, as before

$$\chi_n a = k_n \tag{32}$$

and γ_n the longitudinal propagation constant is given by

$$\chi_n^2 = k^2 - \gamma_n^2. \tag{33}$$

The tangential fields at the wall ($r = a$) are given by

$$\begin{aligned} E_\phi &= V_n\left[\frac{\partial T_n}{r\partial\phi} - d_n\frac{\partial T'_n}{\partial r}\right]\Bigg|_{r=a} e^{-jh_n z} \\ E_z &= j\omega u_0 I_n\frac{\chi_n^2}{k^2}T_n\Bigg|_{r=a} e^{-jh_n z} \\ H_\phi &= I_n\left[\frac{\partial T_n}{\partial r} + d_n\frac{h_n^2}{k^2}\frac{\partial T'_n}{r\partial\phi}\right]\Bigg|_{r=a} e^{-jh_n z} \\ H_z &= j\omega\epsilon_0 V_n d_n\frac{\chi_n^2}{k^2}T'_n\Bigg|_{r=a} e^{-jh_n z}. \end{aligned} \tag{34}$$

On applying the boundary conditions (30) we obtain the desired eigenvalue equation

$$j\omega\epsilon_0 aZ_z - \frac{k_n J_p(k_n)\left[J'_p(k_n) + j\omega\epsilon_0 Z_\phi \frac{\chi_n}{k^2} J_p(k_n)\right]}{-\frac{p}{k_n^2}\frac{h_n^2}{k^2} J_p^2(k_n) + J'_p(k_n)\left[J'_p(k_n) + j\omega\epsilon_0 Z_\phi \frac{\chi_n}{k^2} J_p(k_n)\right]} = 0$$

where $J'_p(k_n)$ is defined

$$J'_p(k_n) = \frac{d}{dx} J_p(x)\Big|_{x=k_n} . \tag{35}$$

The solutions (k_n) of this equation were determined numerically on a digital computer using a Newton-Raphson iterative scheme. The wall impedances Z_z and Z_ϕ are complex for the most general type of wall impedance waveguide as is the eigenvalue k_n. On determination of k_n, we obtain the complex propagation constant γ_n from (33). The attenuation constant α_n (propagation constant h_n) is then simply the imaginary (real) part of γ_n.

REFERENCES

1. Bucholz, H., "Der Hohlleiter von Kreisformischem Overschnitt mit Geschichtetem Dielektrischen Einsatz," Ann. Phys., *43*, 1943, pp. 313–368.
2. Wachowski, H. M., and Beam, R. E., "Shielded Dielectric Rod Waveguides, Final Report on Investigation of Multi-Mode Propagation in Waveguides and Microwave Optics," Microwave Laboratory, Northwestern University, Ill., 1950.
3. Unger, H. G., "Circular Electric Wave Transmission in a Dielectric Coated Waveguide," B.S.T.J., *36*, No. 5 (September 1957), p. 1253.
4. Unger, H. G., "Lined Waveguide," B.S.T.J., *41*, No. 2 (March 1962), p. 745.
5. Carlin, J. W., and D'Agostino, P., "Low-Loss Modes in Dielectric Lined Waveguide," B.S.T.J., *50*, No. 5 (May-June 1971), p. 1631.
6. Carlin, J. W., and Maione, A., "Experimental Verification of Low-Loss TM Modes in Dielectric-Lined Waveguide," B.S.T.J., this issue, pp. 487–496.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 4, April, 1973
Printed in U.S.A.

Experimental Verification of Low-Loss TM Modes in Dielectric-Lined Waveguide

By J. W. CARLIN and A. MAIONE

(Manuscript received October 30, 1972)

Experimental measurements have been carried out to verify the recent theoretical prediction that, in dielectric-lined circular waveguide, modes other than those of the TE_{on} family can exhibit a loss characteristic which decreases with increasing frequency. The TM_{11} waveguide was launched in 51-mm diameter circular waveguide lined with a 200-micron layer of polyethylene, and its loss characteristics were measured. Losses of 23 dB/km at 80 GHz and 3.5 dB/km at 110 GHz were measured, thus confirming the decreasing loss with increasing frequency trend predicted by theory.

I. INTRODUCTION

Recent analytical studies[1-3] of the heat loss characteristics of the normal modes in oversized dielectric-lined circular waveguide indicate that modes other than those of the circular electric type have low loss (<10 dB/km). Specifically, in 51-mm diameter waveguide lined with a 200-μm layer of polyethylene, the TM_{11} mode was predicted to have a theoretical attenuation coefficient which decreases with increasing frequency and achieves values less than 10 dB/km over the 80- to 110-GHz frequency range. This paper reports some recent experimental results which confirm the decreasing heat loss versus frequency trend predicted for the TM_{11} mode in lined waveguide.

The differential phase constant ($\Delta\beta$) for the TE_{01} and TM_{11} modes,

$$\Delta\beta \equiv h_n - h_{01}, \tag{1}$$

was also measured and found to be in good agreement with the theoretical predictions. Here, h_n and h_{01} are the phase constant (propagation constant in lossless guide) for the spurious mode and the TE_{01} mode, respectively.

II. MEASUREMENT METHOD

The loss and $\Delta\beta$ measurements were carried out with a CW reflectometer test set as shown in Fig. 1. Before proceeding to a discussion of the test set, it is helpful to first summarize some recently-arrived-at conclusions[4] concerning the normal-mode scattering properties at the junctions of three common types of circular waveguide.

The normal modes of helix or dielectric-lined guide, as defined herein, are characterized in terms of the equivalent wall impedance of the guide.[3] They are defined as TE_{pn} or TM_{pn} modes, although they are in fact hybrid modes. The so-called TE_{pn} or TM_{pn} modes tend to the TE_{pn} or TM_{pn} modes of metallic waveguide as the *wall impedance* (or equivalently the lining thickness for dielectric-lined guide) tends to zero. Since the modes in dielectric-lined or helix guide may differ from the modes in copper guide, we will add the appropriate superscript (Cu-copper guide, DL-dielectric lined guide, HX-helix) to the usual TE_{pn} or TM_{pn} notation.

For the helix and dielectric-lined guides used in the present study, the scattering properties of interest over the 80- to 110-GHz frequency range may be summarized as follows:

(*i*) The TE_{11}^{Cu} mode scatters $\lesssim 80$ percent of its energy into the TE_{11}^{HX} mode. Almost all of the remaining energy scatters into

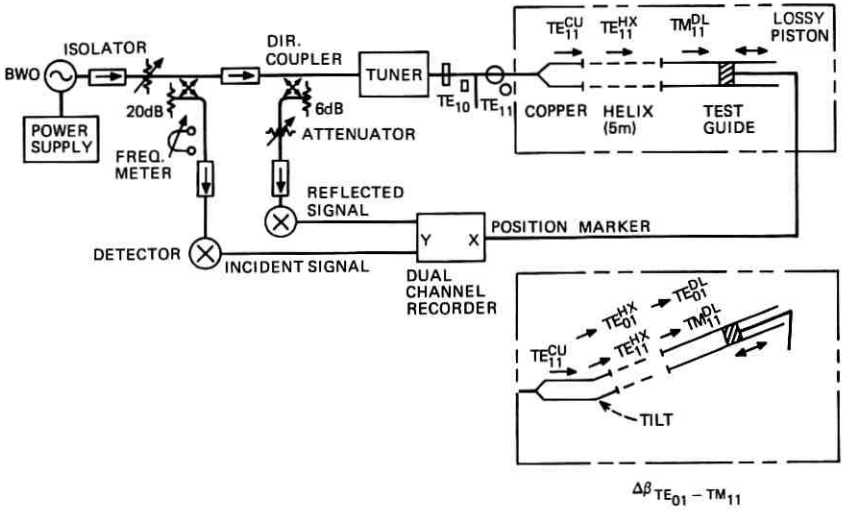

Fig. 1—CW reflectometer test set.

the TM_{1n}^{HX} modes which have much higher losses then the TE_{11}^{HX} model.

(*ii*) The TE_{11}^{Cu} mode scatters >90 percent of its energy into the TM_{11}^{DL} and TE_{11}^{DL} modes.

(*iii*) The TE_{11}^{HX} mode scatters >95 percent of its energy into the TM_{11}^{DL} mode.

The loss of the waveguide sample is measured using the circuitry in the upper portion of Fig. 1. A dominant $TE_{10}^{\square}$ mode is converted to a $TE_{11}^{\bigcirc}$ mode via a transducer and passed through a smooth copper taper to a 51-mm output diameter. At this point, the copper waveguide is connected to a helix waveguide followed by a dielectric-lined guide (as shown in Fig. 1) or simply to a dielectric-lined guide. The loss of the sample per unit length can then be determined from the slope of the reflected signal level versus piston position. A lossy reflecting piston (<−6 dB reflection coefficient) is used to minimize the effects of trapped resonances in the test set. The slide screw-tuner as shown in Fig. 1 is required to tune out the residual mismatch in the system.

The Cu-HX-DL tandem connection results in the excitation of a relatively pure high-loss TE_{11}^{HX} mode which, in turn, scatters almost all of its energy into the low-loss TM_{11}^{DL} mode as indicated by the reflectometer trace in Fig. 2. The trace in both the HX and DL sections is smooth and free of large oscillations which would be present if other modes were strongly excited. The slope of the trace in Fig. 2 may then be used to determine the loss of the TE_{11}^{HX} or TM_{11}^{DL} mode.

On the other hand, the Cu-DL arrangement results in the excitation of both the TM_{11}^{DL} and TE_{11}^{DL} modes, which leads to the oscillatory trace of Fig. 3. The period of the oscillation in Fig. 3 is one-half of the beat wavelength (λ_B) between the TE_{11}^{DL} and the TM_{11}^{DL} modes, where λ_B is defined as

$$\lambda_B \equiv |2\pi/(h_{TE_{11}^{DL}} - h_{TM_{11}^{DL}})|. \tag{2}$$

h is the phase constant for the modes. For a 51-mm guide with a 200-μm-thick polyethylene lining, the analytic prediction for the $TM_{11}^{DL} - TE_{11}^{DL}$ beat wavelength at 68 GHz is 0.38 m as compared to the experimental value of 0.36 m for the trace in Fig. 3. Note that the relative magnitude of the oscillation is decaying rapidly as the piston travels down the DL section (the TE_{11}^{DL} mode has very high loss) and thus we could use the slope of the average value of the reflected signal in Fig. 3 (the dashed curve) to obtain the TM_{11}^{DL} loss.

The test set in Fig. 1 may be modified in a simple fashion to obtain

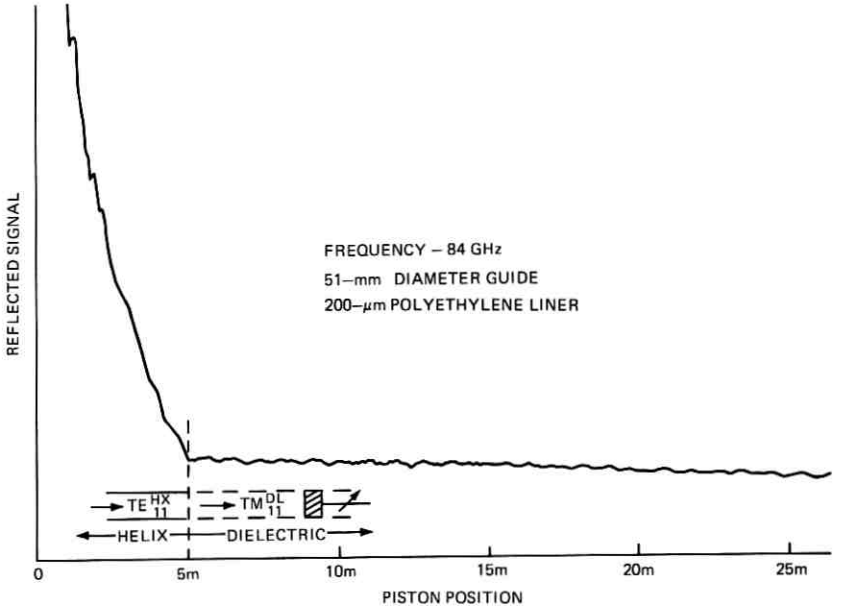

Fig. 2—Returned signal vs piston position for the helix-dielectric guide system, TM_{11}^{DL} loss measurement.

the differential propagation constant ($\Delta\beta$) defined in (1) for the TE_{01}^{DL} and TM_{11}^{DL} modes. On introduction of a small deliberate tilt at the Cu-HX interface (as shown in Fig. 1), the TE_{11}^{Cu} mode excites the

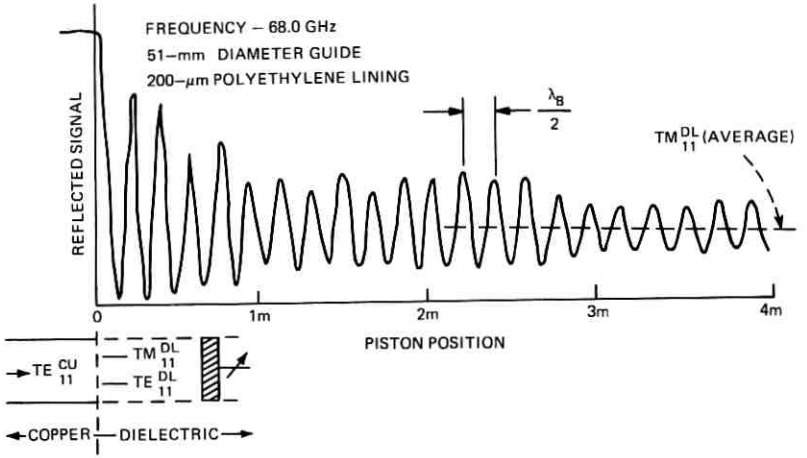

Fig. 3—Returned signal vs piston position for the copper-dielectric system, TM_{11}^{DL} loss measurement.

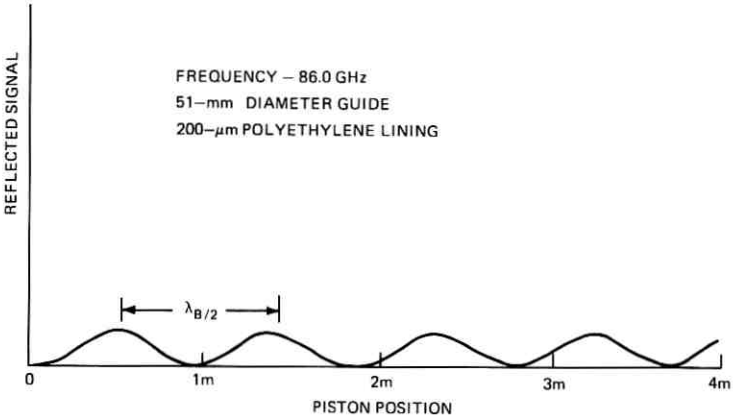

Fig. 4—Returned signal vs piston position for the helix-dielectric guide system with a deliberate tilt, $TM_{11}^{DL} - TE_{01}^{DL}$ differential phase $\Delta\beta$.

TE_{11}^{HX} and TE_{01}^{HX} modes and subsequently the TM_{11}^{DL} and TE_{01}^{DL} modes. The beat wavelength (λ_B) between the TE_{01}^{DL} and TM_{11}^{DL} modes is then twice the period of the oscillations of the moving piston trace shown in Fig. 4. The differential propagation constant $\Delta\beta$ is now related to the beat wavelength λ_B

$$\Delta\beta = 2\pi/\lambda_B. \tag{3}$$

III. RESULTS

By making the moving piston runs described above at different frequencies, it was possible to infer the loss and $\Delta\beta$ vs frequency characteristics of different waveguide samples. Figures 5 and 6 are the loss and phase characteristics obtained for the TE_{11}^{Cu} mode in a 20-m-long sample of copper waveguide. Here, the 51-mm diameter copper output taper can be connected directly to the 20-m sample of 51-mm diameter guide. This arrangement excites both the desired TE_{11}^{Cu} mode and a small amount of TM_{11}^{Cu} (−13 dB down) due to $TE_{11}^{Cu} - TM_{11}^{Cu}$ mode conversion in the taper. It was found that the undesired TM_{11}^{Cu} mode could be lowered to approximately 15 dB under the TE_{11}^{Cu} level by placing a 5-m length of helix waveguide between the taper and the 20-m sample of copper guide, and thus this configuration was used.

The TE_{11}^{Cu} measured loss appears to be 15 percent larger on the average than the theoretical prediction for the TE_{11}^{Cu} loss in perfectly straight and smooth copper guide. This increase is similar to that observed in dominant-mode rectangular guide[5] and to the measured

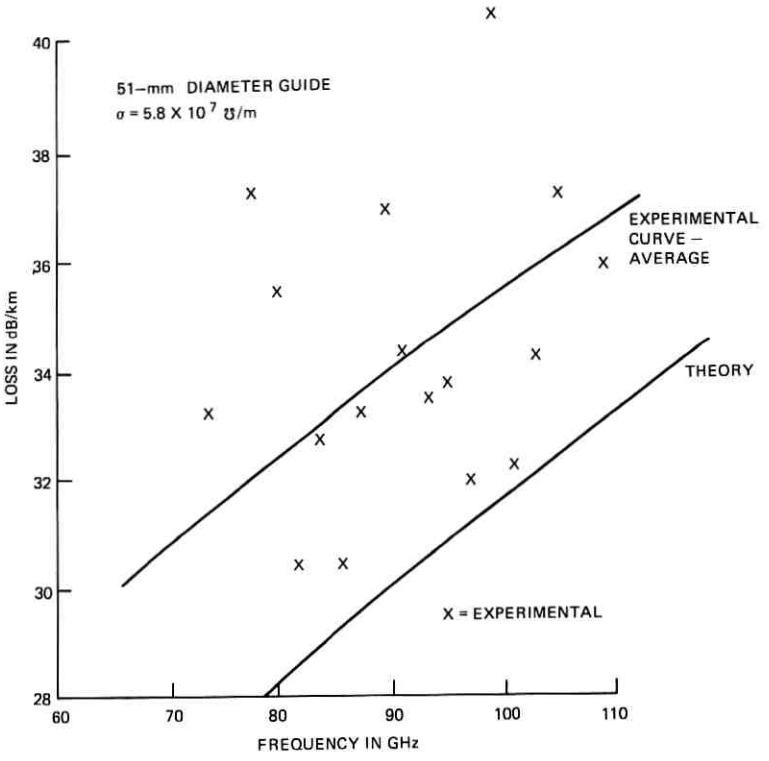

Fig. 5—TE_{11}^{Cu} loss vs frequency.

TE_{01}[6] copper loss in 51-mm copper waveguide and has been attributed to surface imperfections. The individual data points in Fig. 5 are accurate to within ±10 percent, which accounts for the ±4-dB spread about the average. The accuracy of the above CW measurement technique could be improved if a better taper were available. $\Delta\beta(TE_{11}^{Cu} - TE_{01}^{Cu})$ also agrees well with the theoretical value as shown in Fig. 6. The experimental points differ by < 3 percent from the theoretical curve. In summary, the method used yields experimental data in good agreement with the theoretical predictions for the attenuation and differential phase characteristics of the TE_{11}^{Cu} mode.

Figures 7 and 8 are comparisons of the experimental results for the TM_{11}^{DL} loss and differential phase constant, $\Delta\beta(TM_{11}^{DL} - TE_{01}^{DL})$, in dielectric-lined guide with the theoretical predictions obtained using the methods outlined in Carlin and D'Agostino.[3] The copper walls were assumed to have an effective 15-percent increase in loss due to the

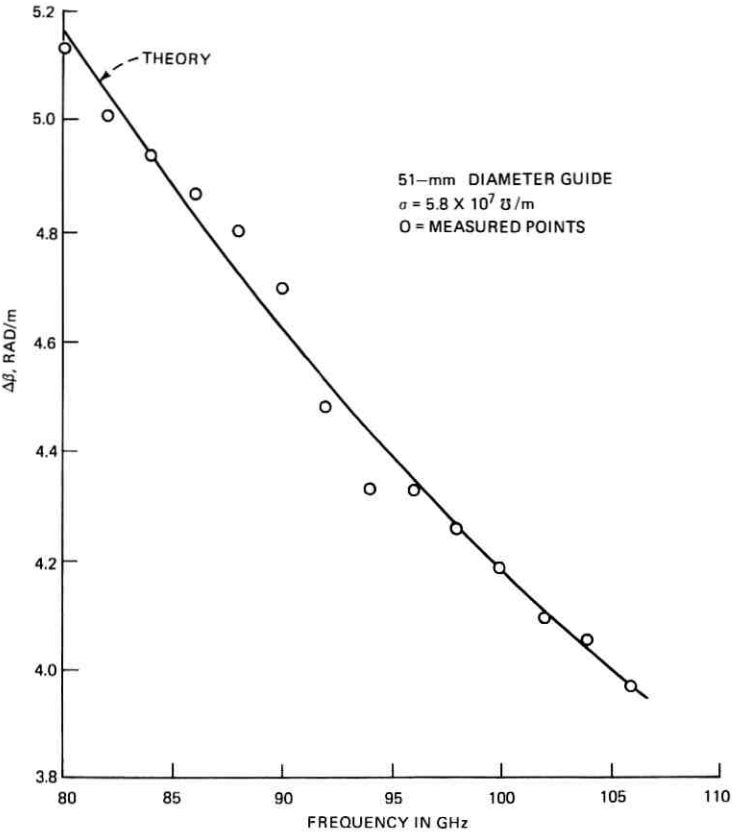

Fig. 6—$TE_{11}^{Cu} - TE_{01}^{Cu}$ differential phase $\Delta\beta$ vs frequency.

effects of surface imperfections. The actual waveguide tested had an adhesive layer of Hydropol* of thickness 12.5 μm to which a 200-μm layer of polyethylene was bonded. The electrical properties (dielectric constant, ϵ_r, and loss tangent, tan δ) of the Hydropol and polyethylene layers were measured[7] at 70 GHz and found to be

	Polyethylene	Hydropol
ϵ_r	2.28	2.5
tan δ	10^{-3}	1.4×10^{-3}

These values were used in obtaining the theoretical heat loss. A sample 30 m in length was used in the measurement of the TM_{11}^{DL} loss and differential propagation constant.

* Partially hydrogenated polybutediene manufactured by Phillips Petroleum Co.

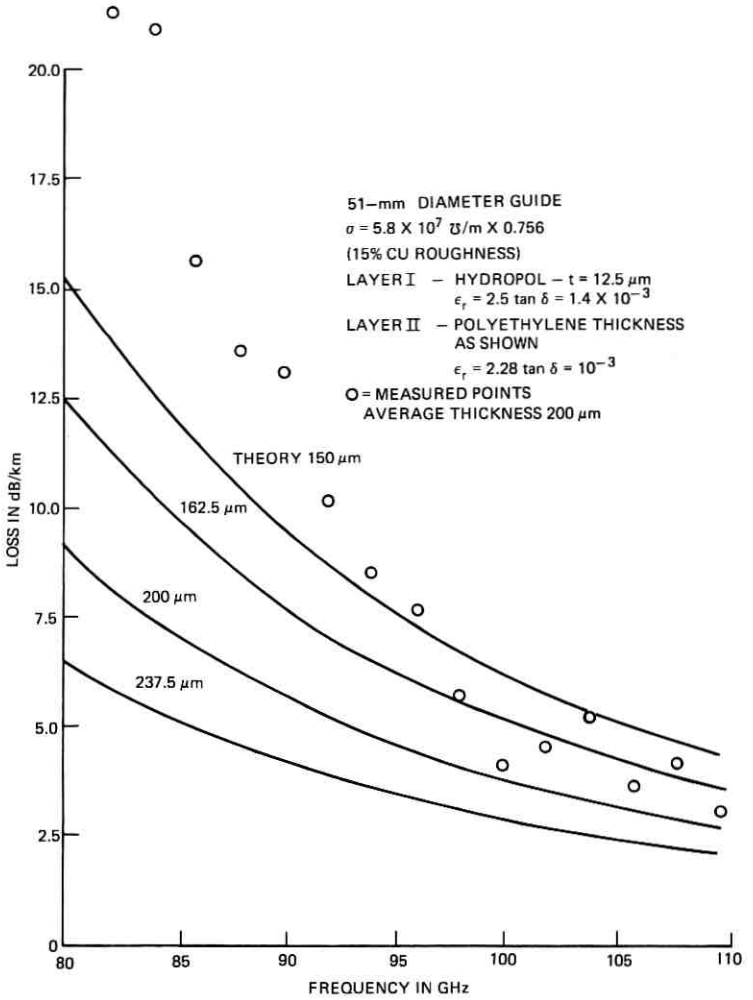

Fig. 7—TM_{11}^{DL} loss vs frequency.

The qualitative behavior of the TM_{11}^{DL} loss, decreasing with increasing frequency, is in good agreement with the theoretical prediction. The measurement accuracy is approximately $\pm$ 10 percent at the lower frequencies (80 to 90 GHz) and $\pm$20 percent at the higher frequencies (100 to 110 GHz). The quantitative agreement is also good from 100 to 110 GHz. The discrepancy at the lower frequencies (80 to 100 GHz) is probably due to the fact that the lining thickness is not a uniform 12.5 μm or 200 μm for the dielectric layers as assumed in

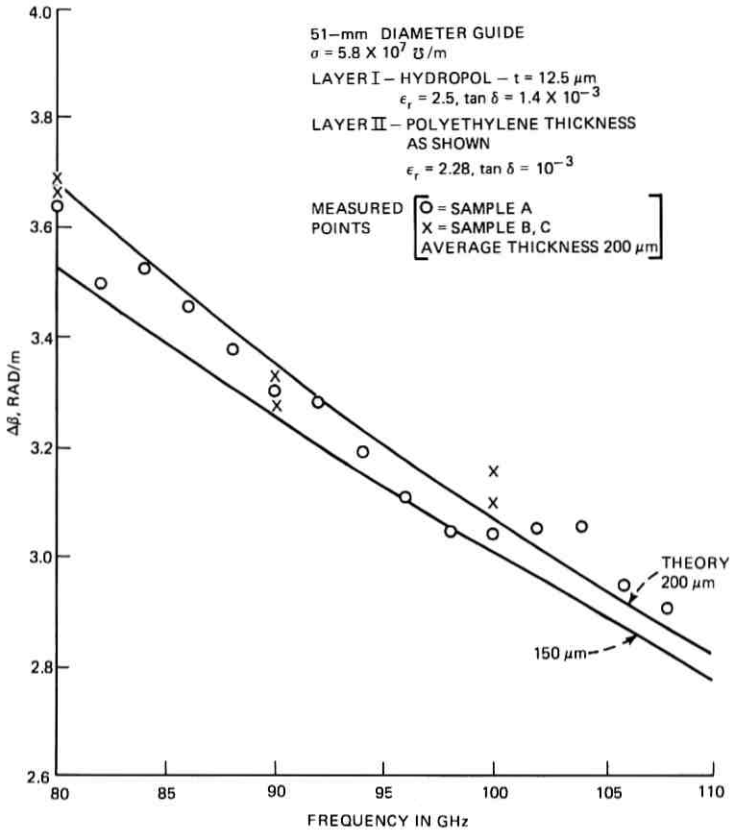

Fig. 8—$TM_{11}^{DL} - TE_{01}^{DL}$ differential phase $\Delta\beta$ vs frequency.

the theoretical models. This has two effects:

(*i*) The total heat loss as shown in Fig. 7 as well as the losses in the various layers and the copper walls is strongly dependent on the thickness of the dielectric layers. A liner with an "average" thickness of 200 μm would have an "average" heat loss higher than that expected for a smooth 200-μm liner, as the loss increased more rapidly for a decrease in lining thickness than vice versa, as seen in Fig. 7.

(*ii*) The field structure of the TM_{11}^{DL} mode fluctuates more rapidly and interacts with fluctuations in lining thickness to a greater degree as the frequency decreases. This is intuitively reasonable from the plots of the TM_{11}^{DL} mode energy density over the guide cross section as given in Carlin and D'Agostino.[3] There the

energy density at the air-dielectric interface in dielectric-lined guide decreases as frequency is increased for a fixed lining thickness, or alternatively decreases as lining thickness is increased for a fixed frequency. This suggests that mode conversion losses for the TM_{11}^{DL} mode arising from lining thickness fluctuations increase with decreasing frequency in dielectric-lined guide. A more quantitative investigation of this effect is beyond the scope of this paper.

The measured differential phase constant ($\Delta\beta$) is in excellent agreement with the theoretical prediction as shown in Fig. 8. This is further confirmation of the fact that the launching arrangement in Fig. 1 does indeed excite the TM_{11}^{DL} mode in dielectric-lined guide.

II. CONCLUSION

In summary, it has been confirmed experimentally that the TM_{11}^{DL} dielectric-lined waveguide mode is a mode which has an attenuation constant which decreases with increasing frequency, thus confirming qualitatively the behavior predicted by theory.[1-3] The measured values ($\approx$ 3.75 dB/km) are approximately 30 percent greater than the theoretical predictions from 100 to 110 GHz. At 80 GHz, the measured value (23 dB/km) is approximately 2.5 times larger than the theoretical prediction. The large discrepancy at the lower frequencies is probably due to $TM_{11}^{DL} - TE_{11}^{DL}$ mode conversion losses arising from fluctuations in the lining thickness.

V. ACKNOWLEDGMENT

The authors would like to thank R. J. Colardeau for supplying the moving piston, and M. A. Gerdine and R. D. Tuminaro for their suggestions and interest.

REFERENCES

1. Brayer, M., "Principal Results Obtained by a Program of Calculations on Guided Waves Within the Band 30–150 GHz," Conference on Trunk Telecommunications by Guided Waves, London, England, Sept.-Oct. 1970, pp. 45–50.
2. Carlin, J. W., and D'Agostino, P., "Low-Loss Modes in Dielectric-Lined Waveguide," B.S.T.J., *50*, No. 5 (May-June 1971), pp. 1631–1638.
3. Carlin, J. W., and D'Agostino, P., "Normal Modes in Overmoded Dielectric-Lined Circular Waveguide," B.S.T.J., this issue, pp. 453–486.
4. Carlin, J. W., unpublished work.
5. Maione A., unpublished work.
6. Aagesen, J., unpublished work.
7. Link, G. L., Johnson, G. E., and Anderson, W. W., "Waveguide Attenuation Caused by the Dielectric Liner," private communication.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 4, April, 1973
Printed in U.S.A.

Prefilters, Sampling, and Transmission Rates for Intraframe Codecs for *Picturephone*® Service

By R. T. BOBILIN

(Manuscript received September 21, 1972)

This study of prefilters and sampling rates analyzes the edge busyness phenomenon which is a very significant degradation introduced by intraframe video coders. For any given transmission rate, the prefilter and sampling rate can be varied to yield a tradeoff between the edge busyness and the rise time associated with a video transition. A narrower prefilter results in less edge busyness but slower rise times and vice versa. Similarly, varying the sampling rate implies an inverse variation in the number of bits per sample allowed in the quantizer, which leads to a tradeoff between aliasing and quantizing noise, both of which are components of the edge busyness phenomenon. This paper shows in detail how the prefilter, sampler, and quantizer affect both edge busyness and rise time, and optimizes the intraframe coder design as a function of transmission rate from an analysis of this two-factor tradeoff.

I. INTRODUCTION

The goal of this paper is to study the effects of filtering and sampling in the digital processing of signals for *Picturephone*® service. In doing this we are ultimately interested in choosing the optimum prefilters and sampling rates (and therefore the number of bits per sample used in the quantizer) for intraframe DPCM codecs. To simplify the analysis it is assumed that the edge busyness phenomenon is the most significant degradation introduced by these codecs. It will be shown how this edge busyness is affected by each part of the codec and how it can be controlled by trading it off against picture resolution. The main theme of this study is this tradeoff between the detrimental effects of both edge busyness and degraded rise times. Both factors are associ-

ated with video edges and hence only edges, modeled as step functions, will be considered.

Section II is a brief description of the analog *Picturephone* system, which outlines relevant short-haul system specifications. Section III defines edge busyness and shows why it is such a significant video impairment; it also describes how edge busyness is affected by each part of the codec and how it can be controlled by trading it against picture resolution. This edge busyness—resolution tradeoff is further investigated in Section IV which uses a series of subjective pair comparison tests to determine the optimum tradeoff as a function of system quality. Section V uses the results of Sections III and IV to show that optimum coder composition, in terms of prefiltering, sampling rate, and quantizer structure, is a function of the digital transmission rate. This study is concluded in Section VI with a subjective testing program which substantiates the validity of the edge busyness—rise time evaluation of video coders and rates a number of intraframe coders on a five-comment impairment scale.

II. ANALOG SYSTEM DESCRIPTION

For the purposes of this study, the important parts of the analog *Picturephone* system can be modeled as illustrated in Fig. 1. Here both the camera system and all analog links connecting the transmitter to the receiver are nominally flat to 1 MHz. If all the frequency shaping in the receiver (excluding de-emphasis) is combined into one equivalent rolloff filter, the resulting filter has a crispened Gaussian response that rolls off to -20 dB at 1 MHz. Hence the total optics-to-optics step response of the analog *Picturephone* system is dominated by that of the receiver station set.* This total optics-to-optics rolloff and overshoot characteristic is given by:†

$$|R(f)|^2 = \{[1 + K(f/T)^2] \exp [-0.5(f/T)^2]\}^2 \tag{1}$$

where f is in MHz and:

$$K = 0.5292$$
$$T = 0.35592.$$

From this the step response can be shown to be (assuming linear phase with a slope of t_0):

$$s(t) = \tfrac{1}{2} \operatorname{Erf}(B) + \frac{BK}{\sqrt{\pi}} \exp(-B^2) \tag{2}$$

* See Ref. 1, pp. 291–292.
† See Ref. 2, eq. (8), but add a set of missing parentheses.

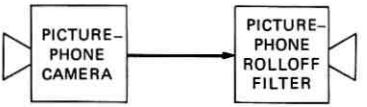

Fig. 1—Analog system model.

where

$$B = \sqrt{2}\pi T(t - t_0) \times 10^6.$$

This step response is plotted in Fig. 2. The response is symmetric about the half-amplitude point with a 4-percent single undershoot and overshoot and has a 10- to 90-percent rise time (T_R) of 0.69 μs.

III. DIGITAL SYSTEM DESCRIPTION

The essential parts of the intraframe DPCM digital system model are illustrated in Fig. 3. In addition to the analog system already described, there is a prefilter, and there are sampling circuits and a quantizer-decoder combination. The prefilter limits line interference and shapes the analog video signal to be encoded. The sample-and-hold and sampling clock circuits turn the analog video into a discrete time format. The quantizer is used to classify the sample-and-hold outputs into a finite number of values for transmission over a given digital channel. The decoder is the inverse of the DPCM quantizing algorithm and

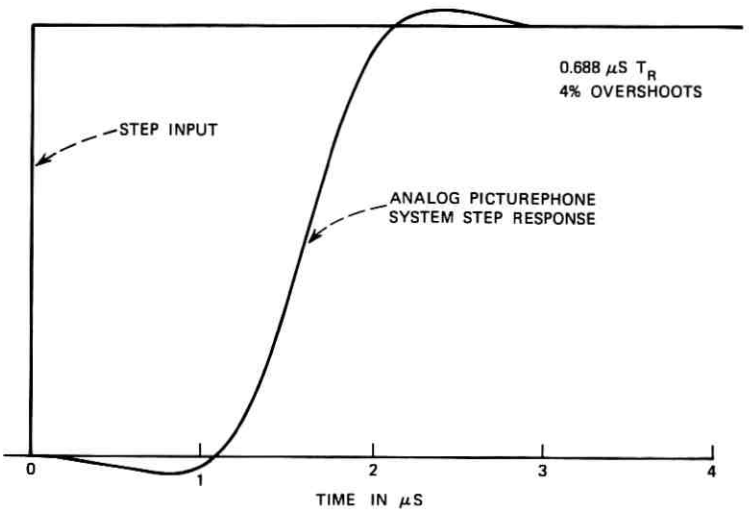

Fig. 2—Analog system step response.

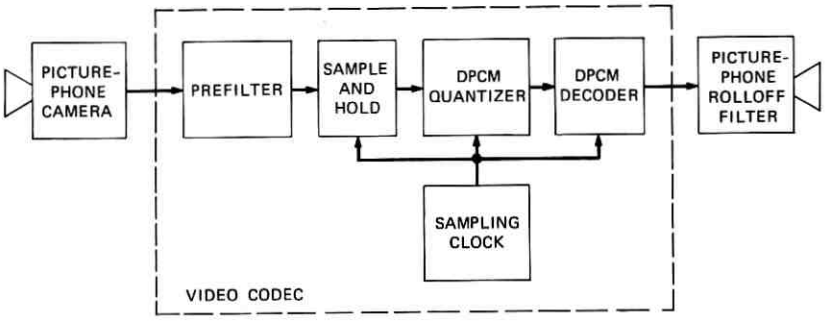

Fig. 3—Digital system model.

puts out the quantized version of the coder's sample-and-hold waveform to the receiving analog system.

In order to see what the codec does to a video transition (such as the edge of a face or a shirt or any black-white boundary), let us start with the simplest part of the codec and study its effect on the video signal. Then the remaining parts of the codec can be added one at a time and analyzed. All step responses will be measured at the output of the analog receiver, as would be seen by the user.

3.1 *Sampling Circuits*

Consider the simplest video codec consisting of only a sample-and-hold circuit driven by a sampling clock. Hence the prefilter, quantizer, and decoder circuits of Fig. 3 are deleted. If we model the ouput of the *Picturephone* transmitter as a perfect step function (the validity of this model will be discussed in the next section), the input to the codec sample-and-hold circuit is illustrated in Fig. 4a. If the sampling instants of the coder clock are given by sampling Phase 1 in Fig. 4a, then the outputs of the sample-and-hold circuit and the receiver rolloff filter are those designated Phase 1 in Figs. 4b and 4c, respectively. If the phasing of the sampling clock with respect to the step function was changed by a quarter of the sampling period, the sampling instants could be given by sampling Phase 2 of Fig. 4a. The resulting outputs of the sample-and-hold circuit and the receiver rolloff filter would be given by the Phase 2 curves in Figs. 4b and 4c, respectively. The outputs from this second phasing of the sampling clock would be the same as those of the first phasing delayed by $T/4$. If the sampling clock were delayed as given by sampling Phase 3 and 4, the corresponding outputs would be given by the Phase 3 and 4 waveforms in Figs. 4b and 4c, respectively.

In this way we can see that the step response of this digital system (always observed at the output of the receiver rolloff filter) will always have the same shape as that of the analog system. The time of occurrence of the step response will depend on the relative phasing of the sampling clock with respect to the camera step function. This indeterminacy of the edge location due to the sampling phase dependence is one specific example of aliasing and will henceforth be called sampling-induced busyness (SIB). Quantitatively, this sampling-induced busyness will be measured as the maximum indeterminacy of the edge at the 50-percent amplitude point. This measure is illustrated in Fig. 4c and in this case is equal to T, one sampling interval.

Sampling-induced busyness is detrimental because it causes the

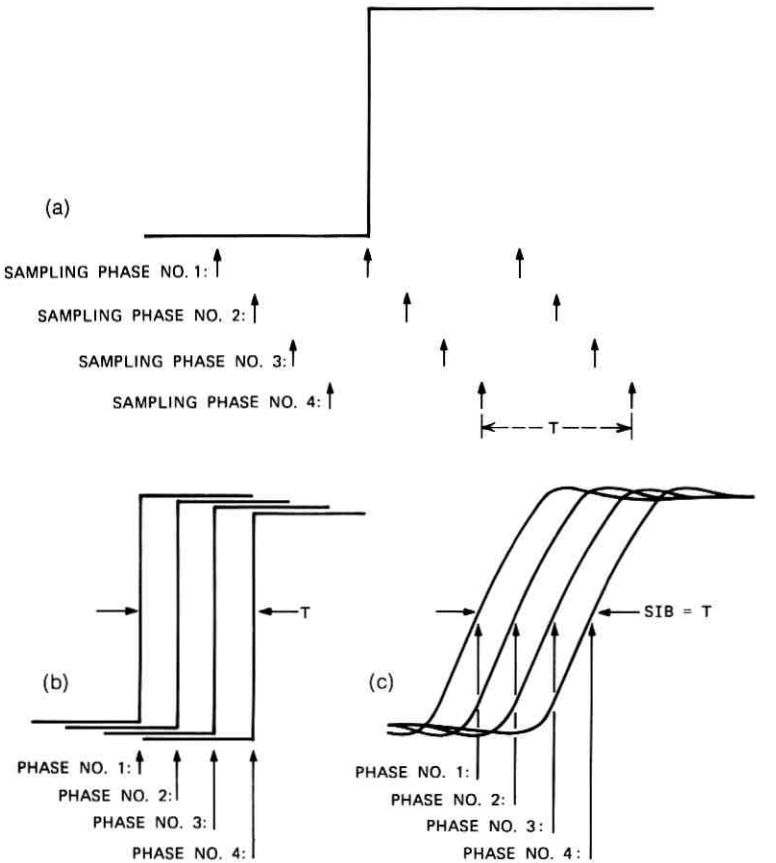

Fig. 4—(a) Step input from camera. (b) Decoder output. (c) Rolloff filter output.

breakup of edges and is easily seen in a typical *Picturephone* display. To see how this breakup occurs, turn to Fig. 5. This represents an enlargement of a small section of a *Picturephone* display. Only five (L1 to L5) of the 267 lines in a frame are illustrated, and only four sampling instants (S1 to S4) are shown. The main diagonal represents the edge of a black-white transition. In the original scene all the area to the left of the edge is black and all area to the right of the edge is white. On line L1, the black-white transition is detected by the sampler at sampling instant S4. On line L2, the transition is also detected at sampling instant S4. It is detected at S3 on line L3 and at S2 on both lines L4 and L5. Thus the transition is detected at the times indicated by circles on Fig. 5. After these sample-and-hold waveforms are passed through the receiver rolloff filter, the edge of the black-white transition would appear to the observer to be the crooked dashed curve of Fig. 5.

For a codec whose sampling clock is locked to the raster, the above curvature is motionless if the edge is still. If the codec is not locked to the raster or if the edge is moving from frame to frame, the above curvature will crawl through the picture. A straight line will be

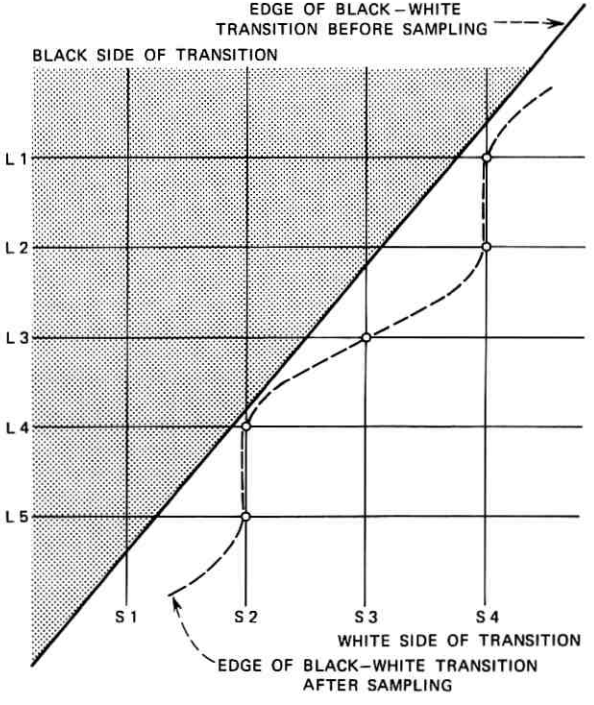

Fig. 5—Sampling-induced busyness.

turned into a moving crooked edge. This effect is extremely visible and quite annoying for an unfiltered sampled-and-held picture.

The only effect of introducing the sampling circuits into the analog system is this edge indeterminance we have called sampling-induced busyness. The rise time of the system has not been affected by the sampling circuits.

3.2 *Prefilter*

The addition of the prefilter changes the input to the sample-and-hold circuit from a perfect step function to a gentler transition with a nonzero 10- to 90-percent rise time as might be given in Fig. 6a. The

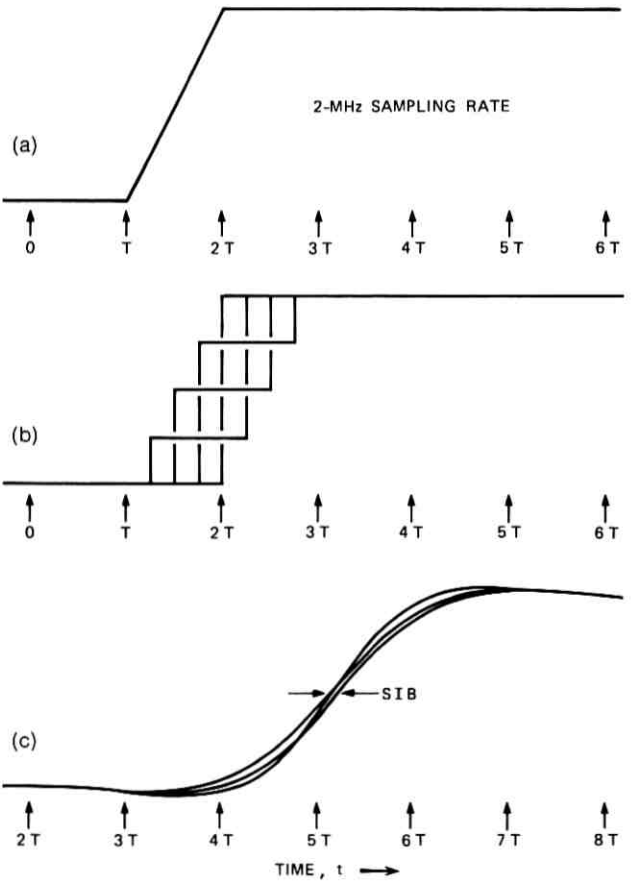

Fig. 6—Effect of prefilter on SIB. (a) Prefilter output. (b) Sample-and-hold output. (c) *Picturephone*® receiver output.

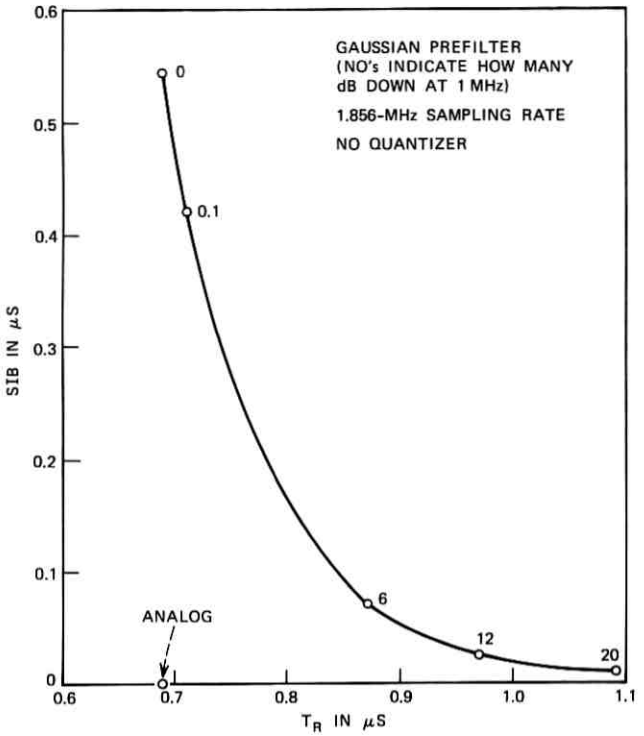

Fig. 7—Sampling-induced busyness—rise time tradeoff.

more prefiltering, the gentler the transition. The effect that this prefilter has on the decoder step response is illustrated in Fig. 6b for four particular sampling clock phasings (in particular those phasings given by $t = 0$, $T/4$, $T/2$, and $3/T4$ on the time scale of Fig. 6a). As opposed to the no prefilter case where the step functions are simply delayed in time (see Fig. 4b), the addition of the prefilter results in an overlapping of the decoder step responses. After these decoder outputs are passed through the receiver rolloff filter,* the sharp corners will be smoothed and the width of the step response envelope will be reduced from the no prefilter case, as can be seen by comparing Fig. 6c with Fig. 4c. Thus sampling-induced busyness can be reduced at the expense of degraded rise times. The exact tradeoff between sampling-induced busyness (SIB) and average rise time (T_R) is given in Fig. 7 for a Gaussian-

* See the Appendix for an outline of the computer program used to derive system step responses for any combination of transition step size, prefilter step response, sampling rate, and quantizer.

shaped prefilter. The subjective weighting between these two factors is dependent upon picture content and will be discussed in Section IV after the quantizer effects are added.

The curve of Fig. 7 corresponds to a fixed sampling rate. Higher sampling rates will shift the curve closer to the analog case, and lower sampling rates will shift the curve further away from the analog case. This effect is illustrated in Fig. 8 for no prefiltering and for a Gaussian prefilter 6 dB down at 1 MHz. The higher the sampling rate, the closer the knee of the tradeoff curve approaches the analog condition of no busyness and a 0.69-μs rise time. Other families of prefilter shapes (non-Gaussian) produce slightly different SIB-T_R curves, but all exhibit the same tradeoff, reducing one only at the expense of increasing the other.

At this point it is necessary to discuss the validity of modeling the output of the *Picturephone* camera system as an ideal step function. The camera frequency characteristic is nominally flat to 1 MHz; therefore, as explained in Section II, the optics-to-optics step response is governed by the rolloff filter that is 20 dB down at 1 MHz and the ideal step function is a good model for the analog system. Since there

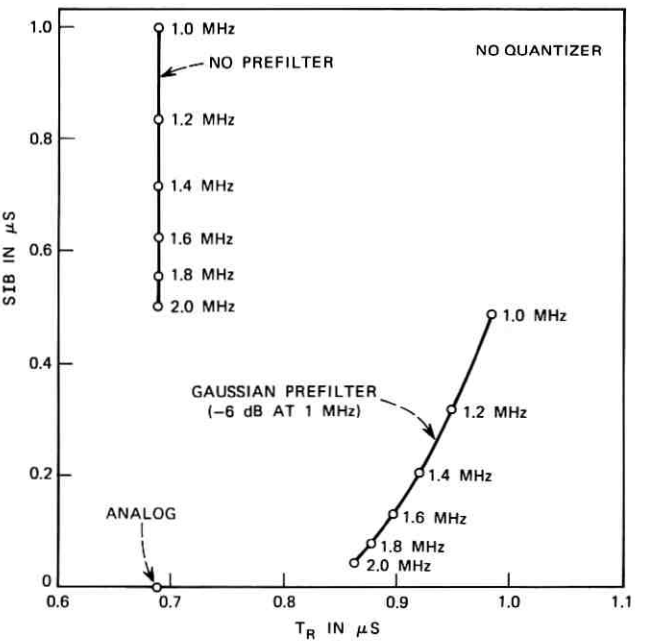

Fig. 8—SIB vs T_R for different sampling rates.

is a filter in the front end of the codec, the step-function model is also appropriate for the digital case as long as the frequency characteristic of the prefilter dominates that of the camera system. All prefiltering that will be seriously studied will be from 6 to 40 dB down at 1 MHz and will indeed dominate the camera system. In some instances, results will be given for less prefiltering (as in the case of points 0 and 0.1 in Fig. 7) and then it should be realized that these are somewhat idealized cases used only for illustrative purposes.

3.3 *Quantizer*

The addition of the quantizer-decoder combination completes the digital system given in Fig. 3. The quantizer is required in order to transmit over a finite digital channel. This necessary addition further confuses the edge indeterminacy—rise time picture given by the prefiltering and sampling circuits. The quantizer both produces its own edge indeterminance and affects average rise times. To see how a quantizer can produce edge indeterminance, refer to Fig. 9a. Here the step response of the prefilter is given by a smooth straight line. The quantizer used in this example has the standard Phase 0, companded DPCM 3-bit characteristic described in Ref. 3. With the sampling instants given at the bottom of Fig. 9a, the prefilter output could be coded as either decoder output No. 1 or No. 2, depending on slight noise variations at the beginning of the step. After passing through the receiver rolloff filter, the response to decoder outputs No. 1 and No. 2 is given in Fig. 9b by rolloff filter outputs No. 1 and No. 2 respectively. The

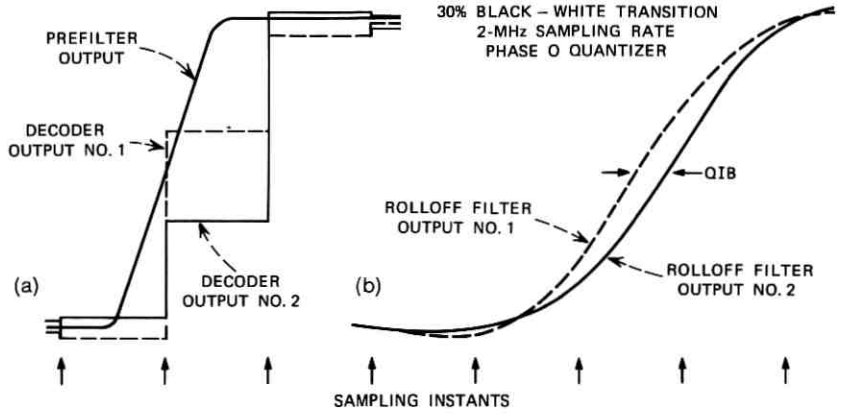

Fig. 9—Quantizer-induced busyness. (a) Prefilter and decoder outputs. (b) Resulting rolloff filter outputs.

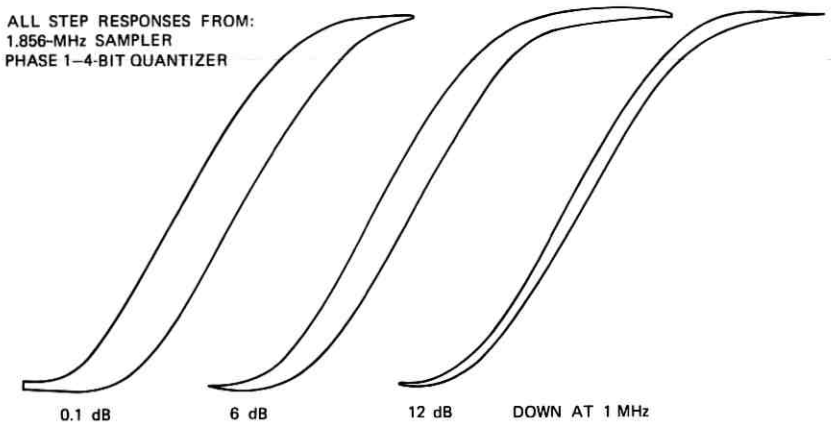

Fig. 10—Step response envelopes.

resulting edge indeterminance (again measured at the 50-percent amplitude point) is due to the quantizer and will henceforth be called quantizer-induced busyness (QIB). This quantizer-induced busyness is any resulting edge indeterminance over and above that which would be produced by just the prefiltering and sampling circuits. The total edge indeterminance resulting from a combination of both sampling- and quantizer-induced busyness will be referred to as edge busyness (EB). In all cases the edge busyness effect is calculated by varying the phase of the sampler. This phase difference can result in practice in many ways. It occurs spatially for a slanted line. It can result from a moving scene, from an unsynchronized codec, or from the residual phase jitter in a synchronized codec. It can also be shown that to some extent random noise on the video signal can be mapped into an equivalent sampling phase jitter.

The envelopes of possible step responses at the output of the receiver rolloff filter for the Phase 1 system are given in Fig. 10 for a Gaussian prefilter that is 0.1, 6, and 12 dB down at 1 MHz (see the Appendix for calculations). This Phase 1 system employs a 1.856-MHz sampler and a companded DPCM 4-bit quantizer which is described in detail in Table I. Again the prefilter yields a tradeoff between edge busyness and rise time. For any combination of sampling rate and quantizer, the prefilter yields this same type of tradeoff. To assign numbers to this characteristic, Fig. 11 plots edge busyness versus rise time for the Phase 1 system with varying amounts of Gaussian prefiltering. Also plotted for comparison purposes is the sampling-induced busyness—rise time tradeoff already given in Fig. 7.

TABLE I—PHASE 1, 4-BIT QUANTIZER CHARACTERISTIC

Output Level Number	Quantized Signal Output* $\left(\frac{\text{weight}}{128}\right)$
1, 2	± 1
3, 4	± 3
5, 6	± 7
7, 8	±15
9, 10	±23
11, 12	±31
13, 14	±39
15, 16	±47

* Weights are relative to the smallest quantum step of a 7-bit PCM system. See Ref. 4 for additional information.

The addition of the quantizer complicates the problem even more than already discussed. Up to this section the results have been independent of the magnitude of the black-white transition. Fig. 11 results from using a 40-percent black-white transition with the Phase 1 4-bit

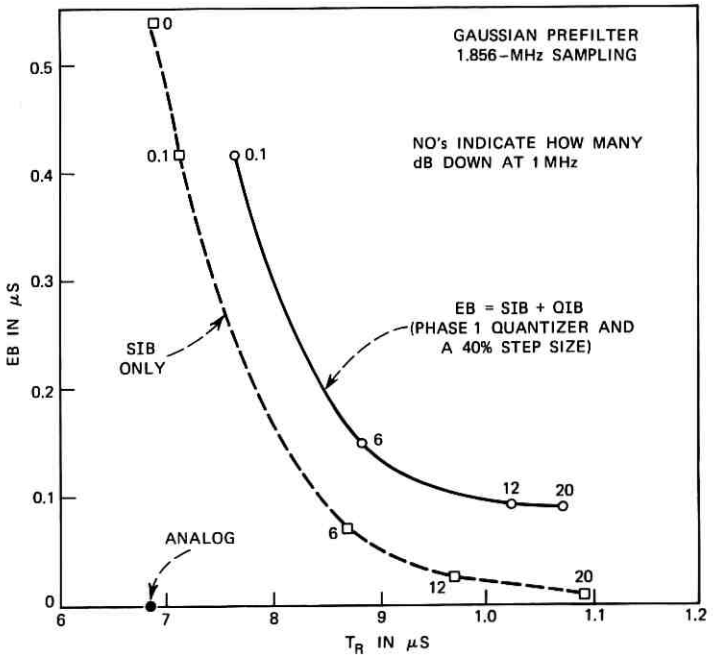

Fig. 11—Edge busyness—rise time tradeoffs.

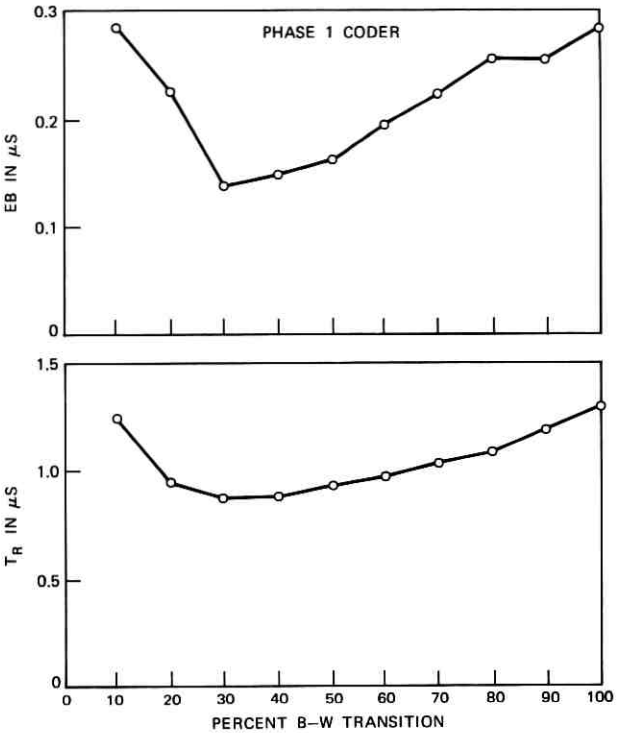

Fig. 12—Edge busyness and rise time vs step size.

quantizer. Both edge busyness and rise time are plotted versus step size in Fig. 12 for the Phase 1 coder using its 6-dB Gaussian prefilter. The shape of both curves can be explained using Fig. 13, where both the quantizing noise and the step-size-to-quantizing-noise ratio are plotted versus step size. Both edge busyness and rise time are inversely proportional to the signal-to-noise ratio and are therefore minimized at a step size of 30 to 40 percent where the signal-to-noise ratio is maximized. For the remainder of this study only three measures of system quality will be considered. These three measures are:

(*i*) the edge busyness and rise time of the 40-percent step size (M4),

(*ii*) the average edge busyness and average rise time for the 10, 20, 30, and 40 percent step sizes (M1-4),

(*iii*) the average edge busyness and average rise time for the 10 step sizes from 10 to 100 percent (M1-10).

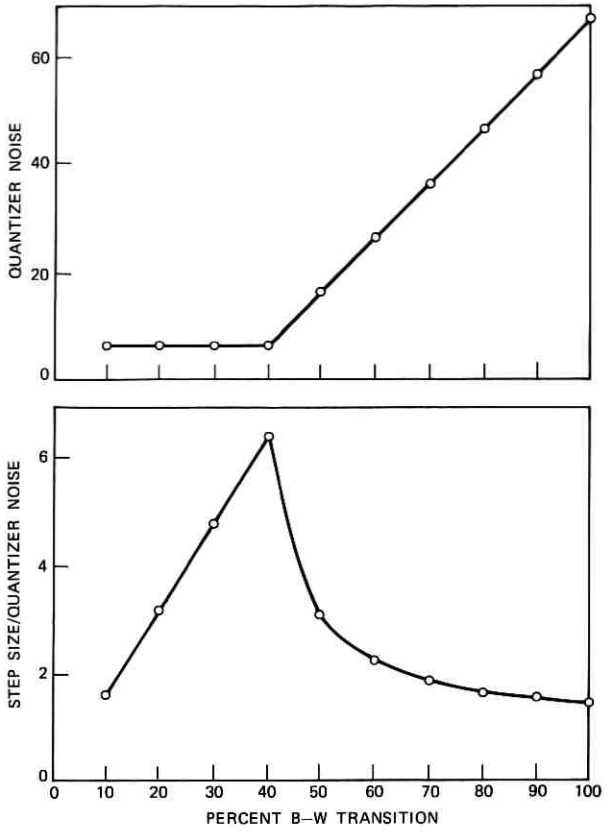

Fig. 13—Quantizer noise and SNR vs step size.

Measure M4 has the advantage of simplicity. Edge busyness can easily be seen on a transition of this size. This is also the breakpoint shown in Fig. 13 before the quantizer's slope overload characteristic takes effect. Measure M1-4 averages over the lower four step sizes. Typically 95 to 99 percent of all transitions are below a 40-percent step size for a 2-MHz sampler. Measure M1-10 uses the most information and disregards all assumed knowledge of observer subjective effects. All three measures were used in designing the coders that will be discussed in the following sections. Ultimate justification for these measures rests with the subjective evaluation of coders with quantizers as will be discussed in Section VI.

IV. SUBJECTIVE EVALUATION OF THE EDGE BUSYNESS—RISE TIME TRADEOFF

In the last section it was shown how the variation of the prefilter yielded a tradeoff between edge busyness and average rise time for any combination of sampling rate and quantizer structure. In this section a series of subjective tests will be described that were run to find the subjectively optimum edge busyness—rise time tradeoff. For these tests only sample-and-hold circuits were used in order to avoid the question of which measure to use when quantizers are involved. In Section VI a series of subjective tests will be described using coders with quantizers.

The test consisted of asking each of 30 observers to rate various systems on the basis of three different source pictures. The first camera source was a mannequin, set up as shown in Fig. 14. This mannequin,

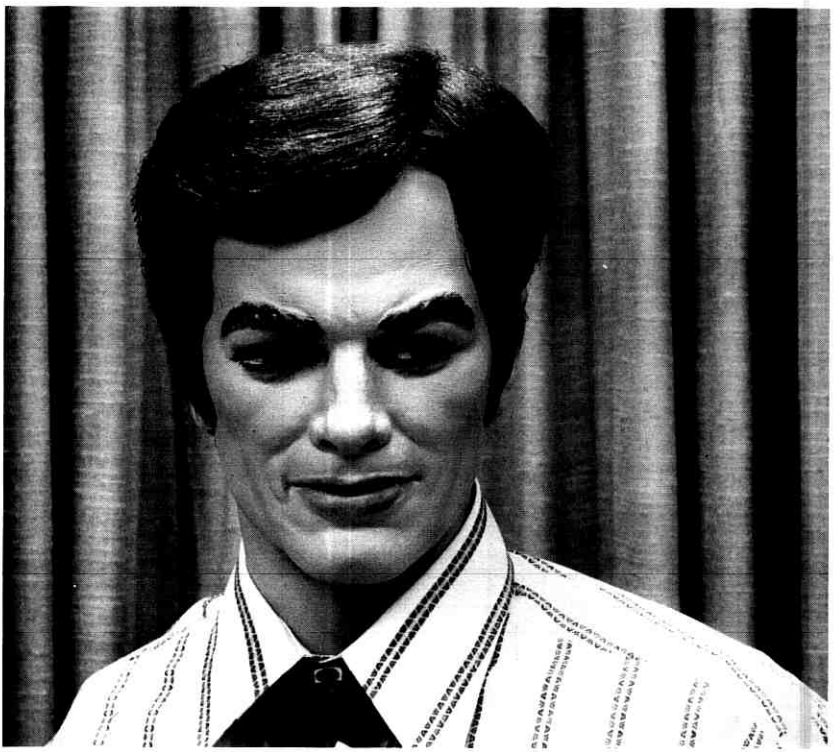

Fig. 14—Microhenry.

"Microhenry" or "Mike" for short, is mounted on a platform which tilts him in a symmetric, sinusoidal fashion from side to side. He is pivoted approximately 14 inches below his chin and was moved a maximum of ± 13 degrees from the vertical at a rate of one cycle every 4.3 seconds. The second and third camera sources were slides of Karen and a graph, as shown in Figs. 15 and 16, respectively. Both of these sources were stationary. For all three sources, each of the 30 observers was given an AB type test wherein the observer was presented with 36 different pairs of systems and asked to pick the most pleasing or the most readable of each pair (most pleasing for each test given on Mike and Karen and most readable for those tests given on the graph). These 36 pairs include all meaningful comparisons between nine different systems. If the pair (A, B) = (system 6, system 9) was given, then pair (A, B) = (system 9, system 6) would not be included. Similarly there was no AB of any system with itself. Each of these nine systems consisted only of a prefilter followed by a sample-and-hold circuit. All prefilters had a crispened Gaussian amplitude characteristic which, along

Fig. 15—Karen.

with near-linear phase, resulted in a symmetric step response with single 4-percent undershoot and overshoot. Five different prefilters were used, being designed to be 3, 6, 10, 20, and 40 dB down at 1 MHz. After tuning, these filters were measured to be 3.0, 5.4, 9.6, 18.4, and 38.1 dB down at 1 MHz with approximately linear phase. The sample-and-hold circuit was driven by one of three sampling clocks, each of which was synchronized to the camera video. The three sampling clocks had synchronized frequencies of 2.015, 1.512, and 1.200 MHz. The particular numbering sequence with system characteristics is listed in Table II. Both the predicted and the measured performance characteristics of these nine systems, as specified on an edge busyness—rise time chart, are given in Fig. 17. Each test was given using an unmodified Mod 2C *Picturephone* station set as the receiver, with the observer given complete freedom of its brightness control. Processing the resulting data for each camera source using standard-pair comparison techniques[5] results in the scale values plotted in Fig. 18. These scale values are a psychological measure of relative coder preference; the higher the scale value, the more the coder is preferred. The least

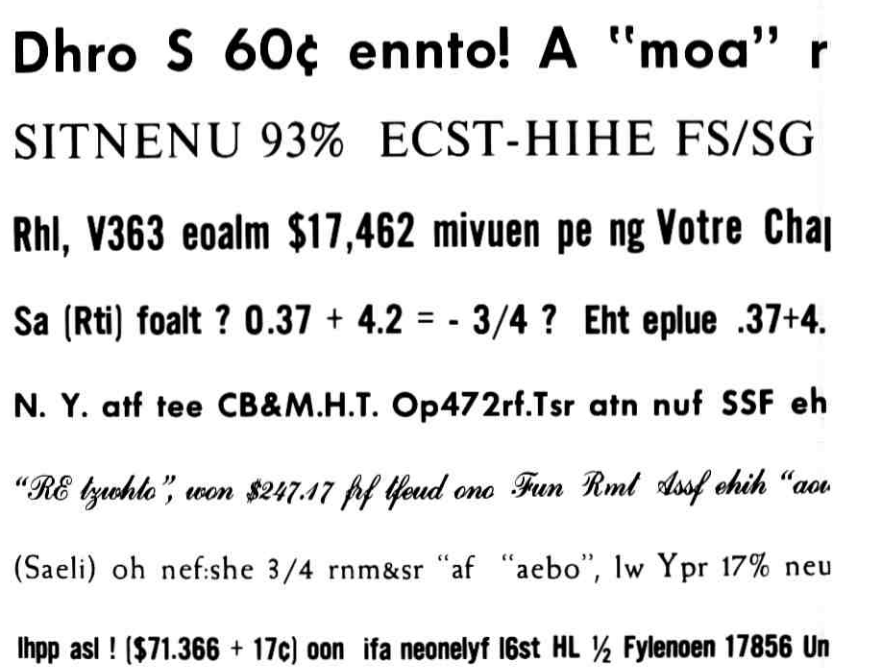

Fig. 16—Graph.

TABLE II—SYSTEM CHARACTERISTICS

System Number	Sampling Rate	Prefilter*
1	2.015 MHz	6
2	1.512 MHz	3
3	1.512 MHz	10
4	1.512 MHz	20
5	1.512 MHz	40
6	1.200 MHz	3
7	1.200 MHz	10
8	1.200 MHz	20
9	1.200 MHz	40

* Specified by the number of dB down at 1 MHz.

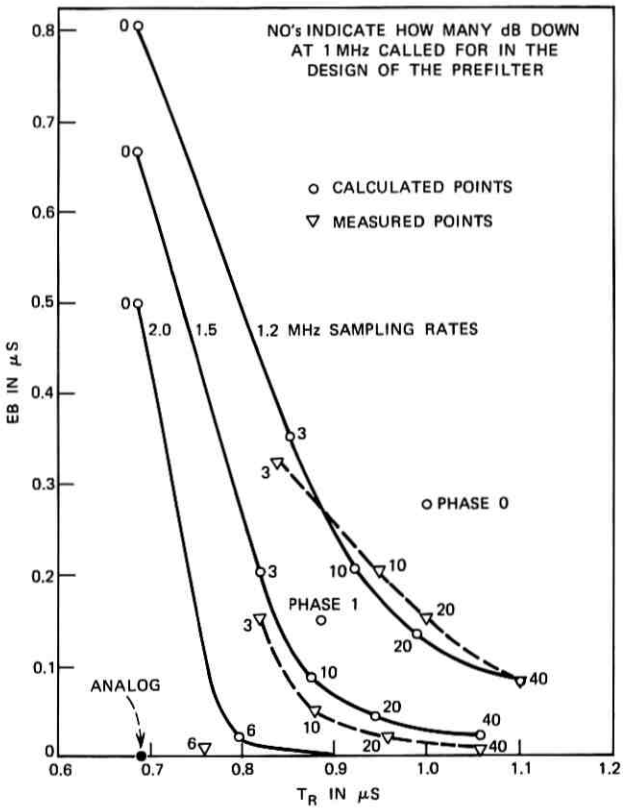

Fig. 17—Edge busyness—rise time tradeoffs for subjective tests.

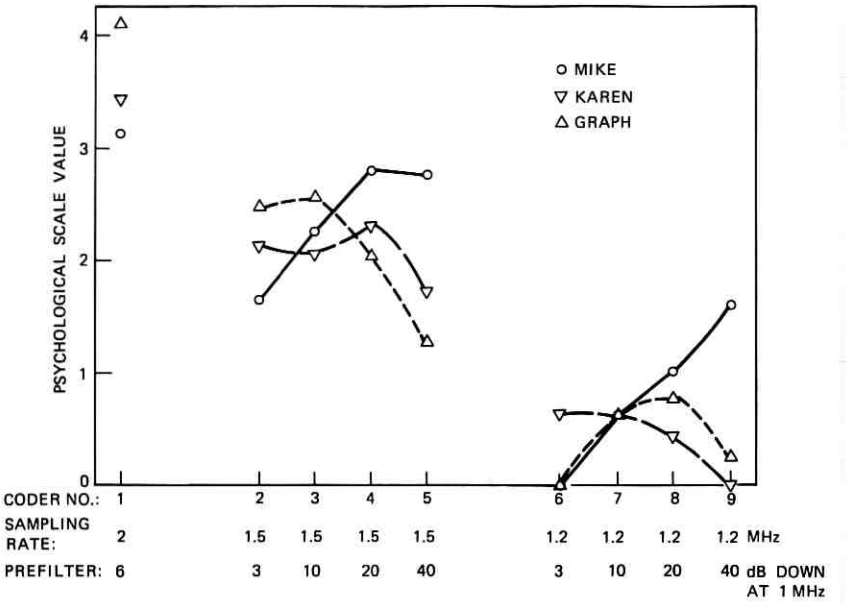

Fig. 18—Psychological scale values for Mike, Karen, and graph.

preferred coder is arbitrarily assigned the value zero with all other coders scaled relative to the worst. The following properties should be noted:

(*i*) No intersource numerical comparisons should be made. A scale value of 4.13 for the best graphics coder and 3.13 for the best coder on Mike only means that the observers were more discriminating on the graphics source. It does not mean that the best graphics coding is better than the best coding of Mike.

(*ii*) Mike is somewhat of a worst case for aliasing. His shirt was chosen for its narrow stripes because undersampling is most harmful when appreciable high-frequency content is present. This undersampling causes the straight edges to appear crooked and the moving subject imparts a distinctive motion to them. On Mike most observers found aliasing to be the only degradation caused by the different systems and hence the narrower prefilters were preferred because they yielded less aliasing at the expense of lower picture resolution. A prefilter that is 20 dB down at 1 MHz was preferred for the 1.5-MHz sampling

rate and a prefilter 40 dB down was preferred for the 1.2-MHz system.

(*iii*) Karen has been a much-used source where both edge busyness and resolution are important. The aliasing still shows up as crooked stripes in her blouse, but they are motionless for our synchronized clocks and therefore not quite as annoying. For this source a 20-dB prefilter was preferred for the 1.5-MHz system and a 3-dB prefilter was preferred for the 1.2-MHz system.

(*iv*) The graph's sole property is readability. Here the 10-dB prefilter was preferred for the 1.5-MHz system and the 20-dB prefilter was preferred for the 1.2-MHz system.

Since Mike is a worst case for aliasing, Karen a difficult source for both aliasing and resolution, and the graph a test of readability, the three sources are considered to be of equal importance, and all 90 responses will be lumped together for the final preference scale. Analyzing these 90 responses by the pair comparison analysis[5] results in the scale values shown in Fig. 19. The 2-MHz system is preferred over the best 1.5-MHz system which is in turn preferred over the best 1.2-MHz system. There is a fairly broad range of optimum prefiltering for the 1.5-MHz system with the best prefilter 20 dB down at 1 MHz. For the 1.2-MHz system, the narrower the prefilter the better, with the best choice of those studied being 40 dB down at 1 MHz.

As shown in Fig. 17, these results indicate that the best subjective tradeoff between edge busyness and rise time occurs just below the knee

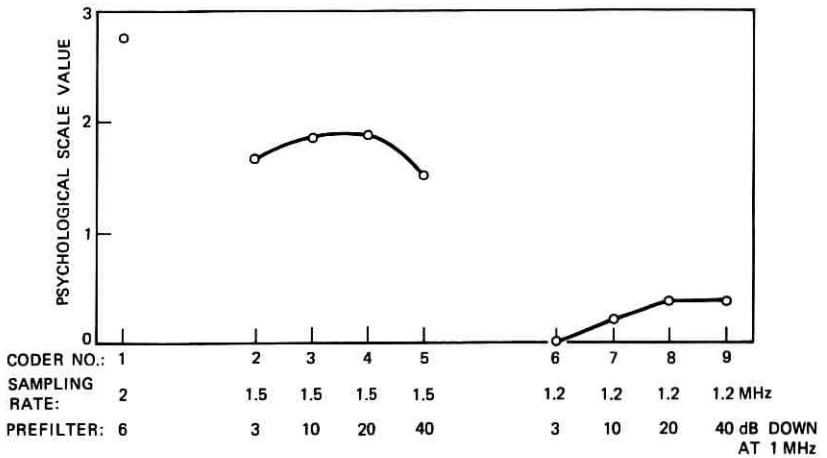

Fig. 19—Psychological scale values resulting from all 90 responses.

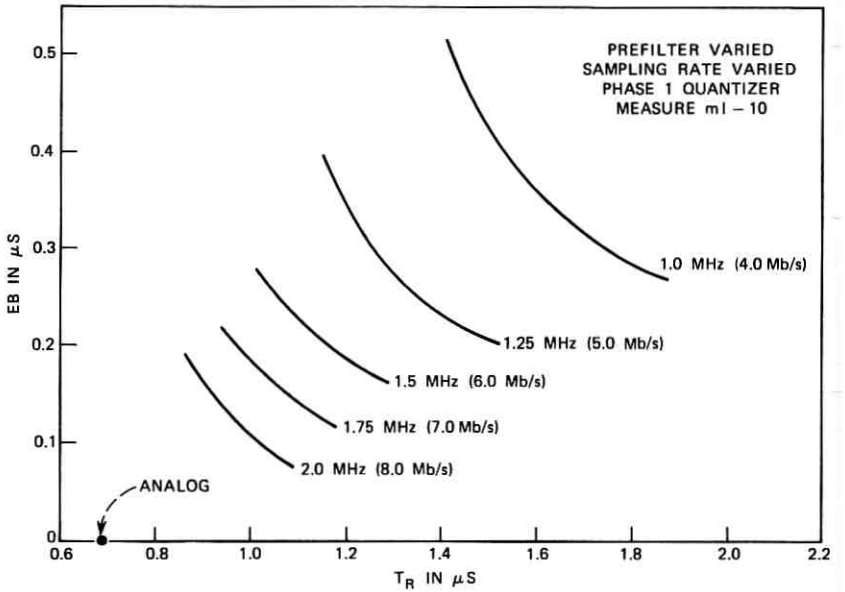

Fig. 20—Edge busyness—rise time tradeoffs for the Phase 1 quantizer.

of each tradeoff curve where the incremental reduction in edge busyness becomes small compared to the incremental increase in average rise time.

V. OPTIMIZING THE SAMPLING RATE FOR A FIXED TRANSMISSION RATE

For the Phase 1 4-bit companded DPCM quantizer given in Table I, the variation of both prefilter and sampling rate leads to the tradeoffs between edge busyness and average rise time given in Fig. 20. These curves are not unexpected; for any combination of sampling rate and quantizer, the prefilter leads to a tradeoff between edge busyness and average rise time. The higher the sampling rate (and therefore the higher the transmission rate in this case), the closer the knee of the tradeoff curve approaches the analog case and therefore the better the picture quality. Figure 20 results from using a crispened Gaussian prefilter. Many other filter shapes were studied in an attempt to find the best edge busyness—rise time tradeoff curves. At the higher sampling rates (namely 1.856 MHz and above) the prefilter shape becomes unimportant; here, less prefiltering is required and the receiver rolloff filter becomes increasingly dominant. As the sampling rate is

reduced, two things are desired in filter shaping: rapid amplitude falloff versus frequency and controlled overshoot in the time domain. The rapid falloff is required to control aliasing without removing significant video information and controlled overshoot is needed to present a pleasing picture, one without excessive ringing. The crispened Gaussian filter shape with 4-percent overshoots, as specified by eq. (1) with $K = 0.5292$ but with T a variable, yields a good compromise between these two factors.

The results given in Fig. 20 show the effects of different sampling rates with a single quantizer. The transmission rate is proportional to the sampling rate and the number of bits used to code each sample. Therefore, in the extreme case, we are comparing the performance of an 8-Mb/s transmission system (2-MHz sampler–4-bit quantizer) with that of a 4-Mb/s transmission system (1-MHz sampler–4-bit quantizer). The more interesting tradeoff results from considering a specific transmission rate. For any given transmission rate, what is the best combination of prefilter, sampling rate, and quantizer which minimizes both edge busyness and average rise time? A higher sampling rate implies fewer quantizing levels which means less SIB but more QIB. A lower sampling rate implies more quantizing levels which means more SIB but less QIB. The prefilter is used to trade off edge busyness against picture resolution to yield the best subjective picture quality for each sampling rate–quantizer combination.

To be specific, given the following three transmission rates*

(*i*) 6. Mb/s
(*ii*) 7.5 Mb/s
(*iii*) 8. – 9. Mb/s

what is the optimum combination of sampling rate and number of bits/sample?

For the 6.-Mb/s rate, the first three coders listed in Table III are considered. In order to simplify the calculations, only the first performance measure, M4, slightly modified, is calculated. A 40-percent video transition is used with the 2.0-MHz system, a 50-percent transition with the 1.5-MHz system, and a 60-percent transition with the

* The 6.-Mb/s rate results from the simplest use of a T2 line. The effective bit rate of 7.5 Mb/s results from using 300 cells of storage to expand video information into the horizontal sync pulse interval with subsequent transmission over a T2 line. An effective bit rate of 8 to 9 Mb/s can be achieved over a T2 line by using about 20,000 cells of storage (see Ref. 6) to use both the horizontal sync pulse interval and variable length coding. This variable length coding takes advantage of the statistical properties of the video signal.

TABLE III—CODER CHARACTERISTICS

Transmission Rate	Sampling Rate	Quantizer
6.0 Mb/s	2.0 MHz	3-bit DPCM
6.0 Mb/s	1.5 MHz	4-bit DPCM
6.0 Mb/s	1.2 MHz	5-bit DPCM
7.5 Mb/s	1.9 MHz	4-bit DPCM
7.5 Mb/s	1.5 MHz	5-bit DPCM
7.5 Mb/s	1.2 MHz	6-bit DPCM
8.3 Mb/s	1.9 MHz	5-bit DPCM
9.0 Mb/s	1.5 MHz	6-bit DPCM
8.4 Mb/s	1.2 MHz	7-bit DPCM

1.2-MHz system. This modification accounts for the longer sampling intervals encountered with the lower sampling rates. For each coder a wide range of prefilters was evaluated and the optimum prefilter chosen using the subjective results given in Section IV. The prefilter which results in an edge busyness—rise time tradeoff just below the knee of the curve was chosen. The three resulting systems are given by the upper curve of Fig. 21. For the effective bit rate of 7.5 Mb/s, the middle three coders listed in Table III are considered. Using similar optimization procedures, these three coders result in the three edge busyness—rise time points given by the middle curve of Fig. 21. For the 8. to 9.-Mb/s transmission rate, it is assumed that variable length coding will result in an extra bit per sample for each of the sampling rates used. Here the three coders listed in the bottom of Table III are

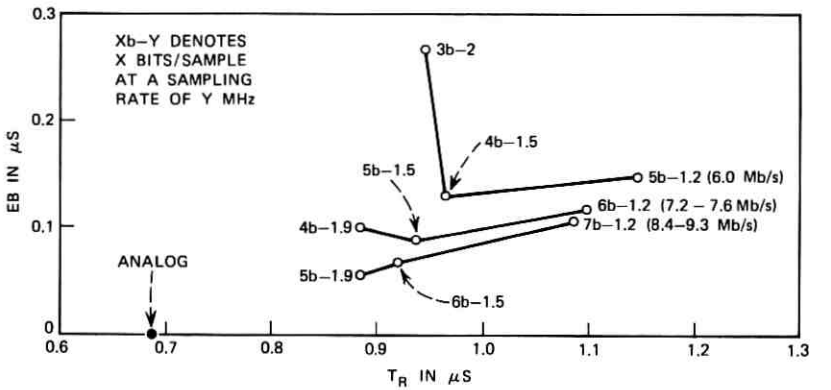

Fig. 21—Edge busyness—rise time tradeoffs for fixed transmission rates.

considered. These coders result in the three edge busyness—rise time points given by the lower curve of Fig. 21.

Examination of Fig. 21 leads to the following conclusions:

(*i*) For intraframe coders operating at an effective bit rate of 6 Mb/s (points 3b-2., 4b-1.5, and 5b-1.2 on Fig. 21), the edge busyness—rise time tradeoff is optimized by using a 1.5-MHz sampling rate and a 4-bit quantizer coupled with a *Picturephone*-type prefilter 20 dB down at 1 MHz.

(*ii*) For intraframe coders operating at an effective bit rate of 7.5 Mb/s (points 4b-1.9, 5b-1.5, and 6b-1.2), the edge busyness—rise time tradeoff is optimized by using a sampling rate of either 1.5 or 1.9 MHz, with no significant advantage associated with either of the two sampling rates.

(*iii*) Only after an intraframe coder is operating at a bit rate higher than 7.5 Mb/s (points 5b-1.9, 6b-1.5, and 7b-1.2) does it become beneficial to sample at 2 MHz.

VI. SUBJECTIVE EVALUATION OF *PICTUREPHONE* CODERS

The main point indicated in Section V was that the optimum sampling rate is some function of the transmission rate. This point is most dramatically illustrated for a transmission rate of 6. Mb/s. Here it is predicted that a 4-bit–1.5-MHz coder will outperform the Phase 0 coder (3-bit–2.0-MHz). In order to check this point subjectively, the same 30 observers used in Section IV were given an AB test between the Phase 0 coder and a coder consisting of a crispened Gaussian prefilter 20 dB down at 1 MHz, a synchronized 1.512-MHz sampling clock, and the Phase 1 4-bit quantizer. Each observer was asked to pick the most pleasing coder for Mike and Karen and the most readable coder for the graph. A summary of the results for the Phase 0–4-bit–1.5-MHz AB comparison is given below:

	Prefer Phase 0	Prefer 4-bit–1.5-MHz
Mike:	22	8
Karen:	3	27
Graph:	7	23
All:	32	58

Overall, the 4-bit–1.5-MHz coder was preferred over Phase 0 by a margin of nearly 2:1, giving some justification for the two-factor tradeoff analysis.

In order to find out where the edge busyness—rise time impairments lie on an absolute quality scale, a series of unimpaired-impaired type comparisons were used to rate six different coders on a five-comment impairment scale. For each of the three camera sources, each observer made a comparison of all six coders to the unimpaired analog video signal and was asked if the impairment added by the coder was:

(*i*) not noticeable (weight = 1)
(*ii*) just noticeable (weight = 2)
(*iii*) noticeable but not objectionable (weight = 3)
(*iv*) objectionable (weight = 4)
(*v*) extremely objectionable (weight = 5).

Six different coders were evaluated; these were:

(*i*) Phase 1 (1.856-MHz, 4-bit DPCM, 6-dB Gaussian prefilter)
(*ii*) Phase 0 (2.0-MHz, 3-bit DPCM, 6-dB Gaussian prefilter)
(*iii*) the 4-bit–1.5-MHz coder used in the previous AB comparison with Phase 0
(*iv*) a 2.0-MHz sample-and-hold system using a 6-dB crispened Gaussian prefilter (System 1 of Section IV)
(*v*) a 1.5-MHz sample-and-hold system using a 20-dB crispened Gaussian prefilter (System 4 of Section IV)
(*vi*) a 1.2-MHz sample-and-hold system using a 40-dB crispened Gaussian prefilter (System 9 of Section IV).

The detailed results of the comment scale ratings in terms of means and standard deviations for Mike, Karen, the graph, and all of the responses are listed in Table IV. When all 90 responses are lumped together, the six coder means are given below (see the next to the last row of Table IV).

	Mean Comment Scale Rating
2.0-MHz sample-and-hold:	1.63
1.5-MHz sample-and-hold:	2.12
1.2-MHz sample-and-hold:	3.46
Phase 1:	2.59
Phase 0:	3.31
4-bit–1.5-MHz:	3.30

Here a lower mean comment scale rating is preferable, indicating a lower amount of impairment added by the coder. Note that although

TABLE IV—COMMENT SCALE RESULTS

Coder: / Prefilter: / Samp. Rate: / Quantizer:	4b-1.5 20 1.5 4	Phase 0 6 2.0 3	Phase 1 6 1.9 4	2.0 S&H 6 2.0 —	1.5 S&H 20 1.5 —	1.2 S&H 40 dB 1.2 MHz — Bits
Mike						
Mean:	3.27	2.85	2.50	1.78	2.10	3.05
Standard Deviation:	0.96	0.78	0.86	0.79	0.61	0.78
Karen						
Mean:	3.23	3.38	2.48	1.28	1.60	3.32
Standard Deviation:	0.63	0.54	0.72	0.50	0.65	0.89
Graph						
Mean:	3.40	3.68	2.80	1.82	2.67	4.00
Standard Deviation:	0.85	0.66	0.88	0.76	1.06	0.96
All 90 Responses						
Mean:	3.30	3.31	2.59	1.63	2.12	3.46
Standard Deviation:	0.83	0.75	0.84	0.74	0.91	0.97

on an AB test the 4-bit–1.5-MHz coder was chosen over the Phase 0 coder by a margin of almost 2:1, when placed on a quality scale and separated in time, they are rated as being equal. In using these ratings it should be remembered that the three camera sources were picked to give all intraframe coders trouble. If Mike did not have a striped shirt or Karen a striped blouse, these tests would have shown very little difference between the coders and all would have been rated much more leniently.

In Fig. 22, the six coders used in the above subjective tests are plotted on an edge busyness—rise time chart. A 40-percent step size is used for both the Phase 0 and Phase 1 coders while a 50-percent step size is used for the 4-bit–1.5-MHz coder. All sample-and-hold coder coordinates are independent of the step size due to lack of a quantizer. With the mean scale values indicated in parentheses, it becomes apparent that the general edge busyness—rise time analysis is further substantiated. The closer the coordinates to the analog case the better, with almost equal weighting between the edge busyness and rise time axes, edge busyness being slightly more undesired than reduced rise time. (See Section IV where points just under the knee of each curve were chosen.) The two dashed circles with centers on the analog coordinates approximate equal-impairment contours with the

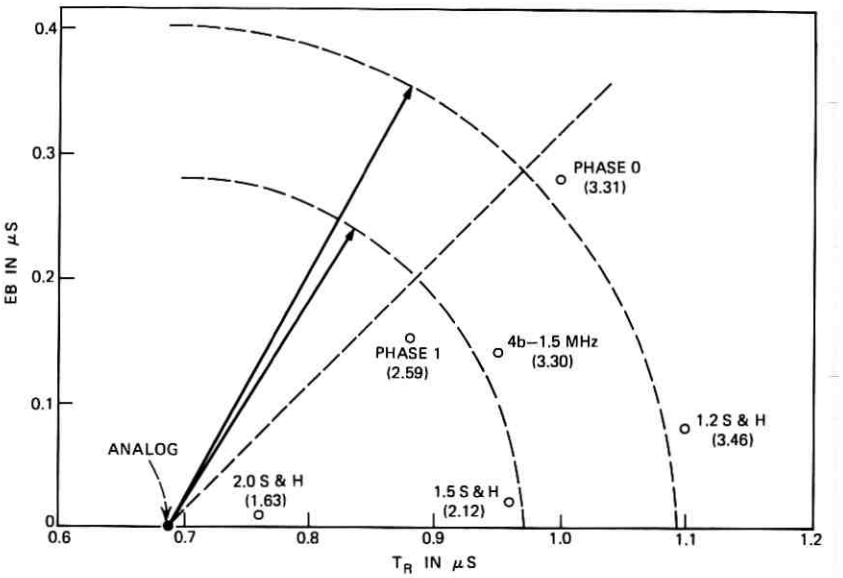

Fig. 22—Comment scale values as a function of edge busyness and rise time.

dashed 45-degree line giving equal weighting to both edge busyness and increased average rise time.

Finally it is noted that only station sets using crispened Gaussian postfilters which are 20 dB down at 1 MHz have been assumed. Other postfilters were tried, including some in the Butterworth family, and similar results were obtained. Any postfilter which affects the analog system to only a minor degree is not expected to significantly change the results presented in this paper.

VII. SUMMARY

A description of the two main degradations inherent in present intraframe video coders has been presented. Methods allowing a tradeoff between these two degradations, edge busyness and reduced rise time, have been described; and optimum coder composition, including prefiltering, sampling rate, and quantizer structure, has been presented as a function of transmission rate. A subjective testing program has been carried out which substantiates the edge busyness—rise time evaluation of video coders and rates a number of intraframe coders on a five-comment impairment scale.

APPENDIX

Edge Busyness—Rise Time Program

Program Inputs: (*i*) Step Size $[A]$
(*ii*) Prefilter Step Response $[b(t)]$
(*iii*) Sampling Rate $[f_s = 1/T]$
(*iv*) Quantizer Structure.

Program Computations:

A.1 *For each of 20 uniform phasings:*

$$(\phi_i, i = 1, 20) = (i - 1)T/20. \tag{3}$$

A.1.1 Determine a series of sample-and-hold values for the system as given in Fig. 3.

$$(X_{ij}, j = 1, 10) = (Ab[\phi_i + T(j - 1)]). \tag{4}$$

A.1.2 Transform sample-and-hold series from A.1.1 into a series of values that would be transmitted over the digital channel.

$$Y_{i1} = Q(X_{i1}) \tag{5}$$

$$(Y_{ij}, j = 2, 10) = \left(Q\left[X_{ij} - \sum_{k=1}^{j-1} Y_{ik}\right]\right) \tag{6}$$

where $Q(a)$ refers to the quantized value of a that can be calculated using the quantizer structure. All thresholds are located midway between the output levels.

A.1.3 Digitally add the quantized series from A.1.2 as would be done by the decoder to result in the quantized version of the sample-and-hold values of A.1.1.

$$(Z_{ij}, j = 1, 10) = \left(\sum_{k=1}^{j} Y_{ik}\right). \tag{7}$$

A.1.4 Since the series produced in A.1.3 represents the amplitude of a series of rectangular pulses approximating the original input video signal, summing a series of appropriately amplified and delayed *Picturephone* receiver step responses as given by eq. (2) of the text will result in the video output of the *Picturephone* station set.

$$Z(t) = \sum_{k=1}^{10} Y_{ik} s[t - \phi_i - (k - 1)T]. \tag{8}$$

A.1.5 Use a search procedure on this video output to find the instants of time when the waveform is equal to 10, 50, and 90 percent of the input step amplitude.

A.2.1 Calculate the output

$$T_R = \frac{1}{20} \sum_{i=1}^{20} (T_{i90} - T_{i10}) \tag{9}$$

as the average 10- to 90-percent rise time.

A.2.2 Calculate and output

$$EB = \max_i (T_{i50}) - \min_i (T_{i50}) \tag{10}$$

as the edge busyness measure.

The above procedure can be used to calculate the combined effects of SIB and QIB or the effect of SIB only, since replacing the quantizer with an identity equation leaves only the effects of SIB to be calculated. It is not possible to calculate the effects of QIB only, since presampling is required for the quantizer.

For all plots given in this paper, eq. (8) can be placed inside a DO loop to calculate the output waveform for a wide range of times.

REFERENCES

1. Cagle, W. B., Stokes, R. R., and Wright, B. A., "The *Picturephone*® System: 2C Video Telephone Station Set," B.S.T.J., *50*, No. 2 (February 1971), pp. 271–312.
2. Crater, T. V., "The *Picturephone*® System: Service Standards," B.S.T.J., *50*, No. 2 (February 1971), pp. 235–269.
3. Millard, J. B., and Maunsell, H. I. G., "The *Picturephone*® System: Digital Encoding of the Video System," B.S.T.J., *50*, No. 2 (February 1971), pp. 459–479.
4. Abbott, R. P., "A Differential Pulse-Code-Modulation Coder for Videotelephony Using Four Bits Per Sample," I.E.E.E. Trans. Commun. Tech., *COM-19*, No. 6, Part 1 (December 1971), pp. 907–912.
5. Edwards, A. L., *Techniques of Attitude Scale Construction*, New York: Appleton-Century-Crofts, Inc., 1957, pp. 40–46.
6. Chow, M.-C., "Variable-Length Redundancy Removal Coders for Differentially Coded Video Telephone Signals," I.E.E.E. Trans. Commun. Tech., *COM-19*, No. 6, Part 1 (December 1971), pp. 923–926.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 4, April, 1973
Printed in U.S.A.

Synthesis of Multiple-Feedback Active Filters

By G. SZENTIRMAI

(Manuscript received October 3, 1972)

A synthesis technique is developed for the active RC realization of transfer functions that have all their transmission zeros on the imaginary ($j\omega$) axis. The method leads to a realization using biquadratic blocks in a multiple-feedback arrangement that is the generalization of the structure obtained from the passive, double-terminated reactive equivalents. The realization combines the easy independent tuning properties of the cascade, with the low-sensitivity characteristics of passive ladders.

I. INTRODUCTION

In the last two decades, the question of how to realize a prescribed rational transfer function

$$T(s) = \frac{N(s)}{D(s)} = \frac{V_{\text{out}}}{V_{\text{in}}} \tag{1}$$

by RC active structures was the subject of more than a thousand learned papers. The consensus at the present time seems to be as follows:

(*i*) The active element to be used is an operational amplifier.

(*ii*) Subnetworks (building blocks) realizing biquadratic transfer functions

$$T_i(s) = \frac{n_0^i + n_1^i s + n_2^i s^2}{d_0^i + d_1^i s + d_2^i s^2} = \frac{N_i(s)}{D_i(s)} \tag{2}$$

are constructed as intermediate steps.

(*iii*) Finally, the overall transfer function is realized as a cascade connection of these leading to:

$$T(s) = \prod_{i=1}^{m} T_i(s).$$

Our objective in this paper is to challenge step (*iii*) above and introduce an alternative synthesis method valid for a large class of transfer functions. This is done by adding one or more feedback loops to the cascade configuration. The effect of this will be that, while the transmission zeros of the overall system will remain the concatenation of those of the individual biquadratic blocks, this will not be true for the poles any more.

We will assume (*i*) and (*ii*) to be correct without, however, worrying about the details of the specific configuration to be used in step (*ii*) above; that is to say, our most elementary building blocks will be "black boxes" realizing transfer functions of the form given by (2). Next we select a particular structure for our investigation and motivate this selection on the basis of prior work. This is followed by the development of a synthesis method for the selected structure that simultaneously proves the generality of it.

Finally we will illustrate the method and demonstrate its advantages by an example.

To circumscribe the class of problems that we will consider, note that $T(s)$ in eq. (1) is a real, rational function of the complex frequency variable s, and therefore both $D(s)$ and $N(s)$ are real polynomials with the further restrictions:

(*i*) $D(s)$ is a strict Hurwitz polynomial, i.e., its zeros are in the open left half of the s plane,

(*ii*) the degree of $N(s)$ is not greater than that of $D(s)$, in notation:

$$\delta N \leqq \delta D.$$

We will need the additional restriction:

(*iii*) all the zeros of $N(s)$ are on the imaginary ($j\omega$) axis of the s-plane. As a consequence, $N(s)$ is either pure even or pure odd.

This restriction, while quite serious, still leaves a very large group of useful functions to be considered, especially if the reader notes that the same restriction applies to functions realized by passive lossless ladders devoid of (magnetic) coupling.

The reason we are searching for new configurations can be traced back to a paper by Orchard,[1] where he has shown by a very simple physical argument that double-terminated lossless passive structures have very low sensitivity to changes in component values inside the passband where the loss is near zero. Furthermore, it is also known that ladders are good for maintaining high out-of-band suppression. While Orchard's argument is not completely valid for voltage transfer

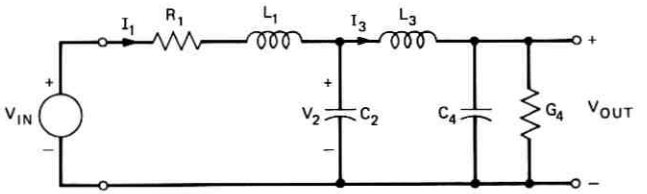

Fig. 1—Passive ladder.

functions, and therefore cannot be transferred to active realizations, improved performance *has* been obtained by simulating passive ladders by active structures.

The first results of this kind were those of Girling and Good[2] and Adams.[3] Girling and Good recognized that, say, a ladder low-pass filter of the form shown in Fig. 1 is described by the equations (also see Ref. 4):

$$I_1 = \frac{1}{R_1 + L_1 s}(V_{\text{in}} - V_2),$$

$$V_2 = \frac{1}{C_2 s}(I_1 - I_3), \tag{3}$$

$$I_3 = \frac{1}{L_3 s}(V_2 - V_{\text{out}}),$$

and

$$V_{\text{out}} = \frac{1}{C_4 s + G_4} I_3,$$

and, as such, are also the describing equations for the active RC-structure of Fig. 2. When it came to more complex filters, the structures advocated by Girling and Good become considerably more complex with nonconstant feedback and other undesirable features.

Adams,[3] working independently, went one step further and realized that the standard low-pass-to-bandpass transformation applied to both Figs. 1 and 2 would result in the very convenient realization

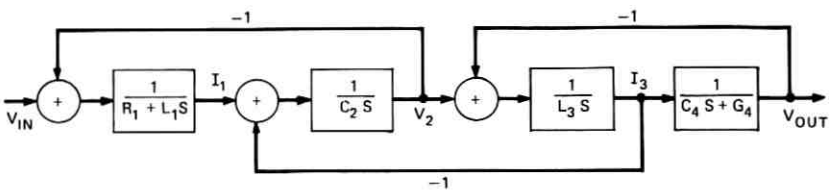

Fig. 2—Simulated passive ladder.

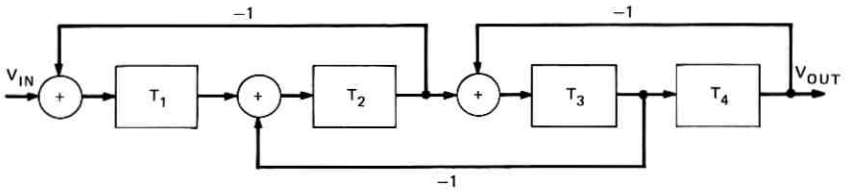

Fig. 3—Multiple feedback bandpass.

of Fig. 3 for bandpass filters. Note that all blocks have biquadratic transfer functions T_i, and further note that T_1 and T_m (the first and last ones) have poles of finite Q, but all intermediate blocks have poles on the imaginary axis of the s-plane. These blocks are therefore, strictly speaking, unstable standing alone, but the overall structure is still stable. This becomes less of a surprise if we note that *ideal* reactive elements (L's and C's) used in passive synthesis are also, strictly speaking, unstable. The only difference is that natural dissipation will always push passive elements toward the stable side of the s-plane while marginally stable active building blocks may hover over either side of the $j\omega$ axis.

Adams' numerical results were sufficiently encouraging, showing substantial reduction in sensitivities to start him and others on the road searching for similar realizations for more general filter functions.

There are many ways to handle this problem, brute-force numerical matching, flowgraph manipulation, and matrix operations on the state-variable equations being some of them.[5] All of these were used by researchers at one time or another, with mixed results. They were unable to explain the origin and uses of multiple solutions; or if they were able to, they led to nonminimal realizations or realizations with much more complex feedback and feedforward paths.[6]

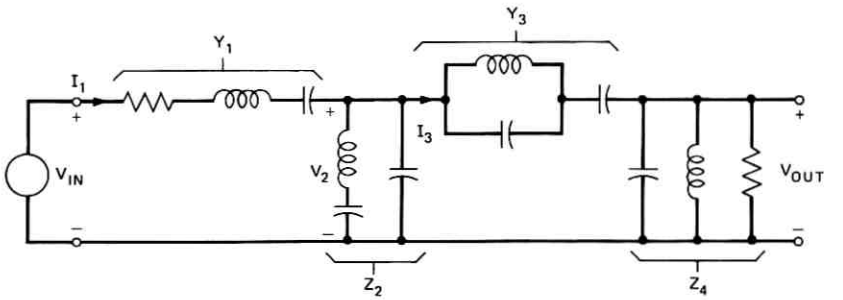

Fig. 4—Passive bandpass filter.

This author also joined in the search and, to his delight, found a group of bandpass functions where the approach of Figs. 1 and 2 was still working. Consider the passive bandpass filter shown in Fig. 4, which is a so-called minimal-inductance realization of an even degree, antimetrical filter. This network is described by the equations:

$$\begin{aligned} I_1 &= Y_1(V_{in} - V_2) \\ V_2 &= Z_2(I_1 - I_3) \\ I_3 &= Y_3(V_2 - V_{out}) \\ V_{out} &= Z_4 I_3 \end{aligned} \tag{4}$$

or, with a slight modification:

$$\begin{aligned} \frac{I_1}{s} &= \frac{Y_1}{s}(V_{in} - V_2) \\ V_2 &= sZ_2\left(\frac{I_1}{s} - \frac{I_3}{s}\right) \\ \frac{I_3}{s} &= \frac{Y_3}{s}(V_2 - V_{out}) \\ V_{out} &= sZ_4\frac{I_3}{s} \end{aligned} \tag{4a}$$

and this is identical to the equations describing the structure of Fig. 3, if we identify:

$$\begin{aligned} T_1 &= \frac{Y_1}{s}; \qquad T_3 = \frac{Y_3}{s}; \\ T_2 &= sZ_2; \qquad T_4 = sZ_4. \end{aligned} \tag{5}$$

Note that all T_i are again biquadratic, T_1 and T_4 have poles inside the left-half s-plane, while the poles of T_2 and T_3 are on the imaginary axis.

This led us to expect the general structure of the form of Fig. 3 to be in some ways canonical. Note that the restriction of all feedback coefficients to -1 does not restrict generality. For instance, the center loop feedback coefficient may be changed to $-k$ if we simultaneously multiply T_1 by k and divide T_2 by k, with no change in the overall transfer function. This procedure, in fact, can be used to adjust voltage levels inside the filter to optimize dynamic range at a later stage in the design process.

II. SYNTHESIS PROCEDURE

Consider now the general structure shown in Fig. 5 with the following assumptions:

(*i*) All T_i transfer functions are real, biquadratic functions of s. If the overall degree is odd, T_1 or T_m will become bilinear.

(*ii*) All internal block transfer functions

$$T_2, T_3, \cdots, T_{m-1}$$

are even functions of s, i.e., all their singularities are located symmetrically either on the imaginary or the real axes.

(*iii*) The poles of T_1 and T_m are inside the left-half s-plane, and their numerators are either even or odd.

The synthesis problem can now be formulated as follows. Given a suitably restricted overall rational transfer function of the form of eq. (1), find the biquadratic transfer functions

$$T_i(s) = \frac{N_i(s)}{D_i(s)} \qquad i = 1, 2, \cdots, m \tag{6}$$

satisfying (*i*) through (*iii*) above, such that these in the structure of Fig. 5 realize $T(s)$.

The outline of the proposed solution of this problem is as follows. First we recognize that our structure is equivalent to a (reciprocal) ladder network, if we restrict all the feedback coefficients to be -1. This restriction will be removed later in Appendix B. The synthesis problem can therefore be solved if we somehow derive a complete (impedance, admittance, or any other) set of parameters for this ladder. Once a set of parameters is obtained, the actual synthesis can follow along lines very similar to the well-known passive filter synthesis method.

Let us use our ladder $\leftrightarrow$ multiple-feedback analogy backwards now and construct the ladder network of Fig. 6a (assuming that m is even),

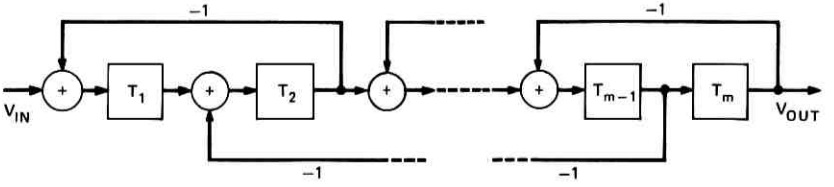

Fig. 5—General multiple-feedback structure.

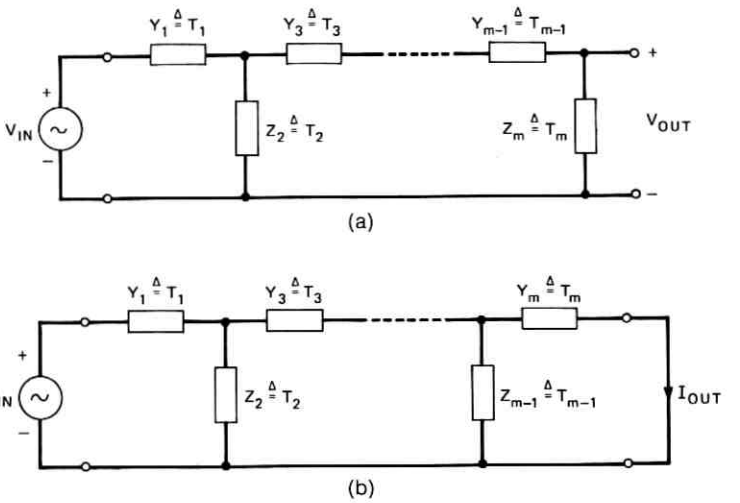

Fig. 6—(a) Equivalent pseudo-ladder for m = even. (b) Equivalent pseudo-ladder for m = odd.

where the branches need not be realizable passive immittances. However, the voltage transfer functions of the two structures are clearly identical and must be equal to $T = N(s)/D(s)$. Note first that obviously:

$$N(s) = \prod_{i=1}^{m} N_i(s). \tag{7}$$

The Z and Y matrices of our pseudo-ladder must be of the forms:

$$Z = \frac{1}{N_1 P}\begin{bmatrix} D & N \\ N & N_1 N_m R \end{bmatrix} \tag{8a}$$

$$Y = \frac{1}{N_m Q}\begin{bmatrix} N_1 N_m R & -N \\ -N & D \end{bmatrix} \tag{8b}$$

where P, Q and R are three unknown polynomials. These follow from Fig. 6a and the assumptions (*i*) through (*iii*) above. Considering degrees and parities of the polynomials in question, we see that, assuming N_1 and N_m to both be even:

$$\begin{aligned} &\delta D = 2m && \delta P = 2(m-1) \\ &\delta N = 2m \text{ and pure even} && \delta Q = 2(m-1) \\ &\delta R = 2(m-2) \text{ and pure even} \end{aligned} \tag{9}$$

where δX, as before, denotes the degree of the polynomial X. That $R(s)$ must be even follows from the fact that, with a short-circuited output

$$\left.\frac{I_2}{I_1}\right|_{V_2=0} = \frac{y_{21}}{y_{11}} = -\frac{N}{N_1 N_m R} = -\frac{\hat{N}}{R}$$

and since this parameter is independent of both Y_1 and Z_m, it must be pure even. Here we used the notation:

$$\hat{N} = \prod_{i=2}^{m-1} N_i. \tag{10}$$

Since $ZY = 1$, the 2×2 unit matrix, we get the determinantal relationship first:

$$DR - N_1 N_m \hat{N}^2 \equiv PQ. \tag{11}$$

Note that N_1 and N_m must be known at this stage. The significance of this is explored further in Appendix B.

In order to proceed further, we must consider the actual synthesis procedure. Consider, for instance, the parameter

$$y_{11} = \frac{N_1 R}{Q}.$$

We can get Y_1 by extracting the partial fraction

$$\frac{\alpha_1 + \beta_1 s}{N_1}$$

plus a constant from $1/y_{11}$ such that the remainder has a factor N_2 in its numerator. More importantly, the remainder must be a pure even function of s, since it is now independent of both Y_1 and Z_m.

Separating $Q(s)$ into even and odd parts and performing the operations outlined, we see that $Q(s)$ must be of the form:

$$Q(s) = A(s) + \frac{k_1}{2} sR(s) \tag{12}$$

where $\delta A = 2(m - 1)$ and pure even, and $k_1 = 2\beta_1$ is a constant. A similar argument applied to z_{22} shows that $P(s)$ must be of the form:

$$P(s) = B(s) + \frac{k_m}{2} sR(s) \tag{13}$$

where $\delta B = 2(m - 1)$ and pure even, and k_m is a constant. Let us

now separate $D(s)$ into even $E_2(s)$ and odd $sE_1(s)$ parts:

$$D(s) = sE_1(s) + E_2(s). \tag{14}$$

Substitute (12) through (14) into the determinantal equation (11) and separate even and odd parts:

$$2E_1 \equiv k_m A + k_1 B \tag{15}$$

$$E_2 R - N_1 N_m \hat{N}^2 \equiv AB + \frac{k^2}{4} s^2 R^2 \tag{16}$$

where we introduced $k_1 k_m = k^2$.

Considering eq. (16) a quadratic in the unknown polynomial $R(s)$, this will have a polynomial solution if and only if the discriminant:

$$E_2^2 - k^2 s^2 (AB + N_1 N_m \hat{N}^2) \triangleq G^2 \tag{17}$$

is a full square for some even polynomial G. In such a case, the solution is given by:

$$R = \frac{2}{k^2 s^2} (E_2 \pm G)$$

and, in order to get a polynomial, we must select the negative sign, since from (17) we see that the constant terms of E_2 and G will be identical and hence cancel from the difference:

$$R = \frac{2(E_2 - G)}{k^2 s^2}. \tag{18}$$

Let us next express, say, B from eq. (15) and substitute it into (17):

$$E_2^2 - k^2 s^2 N_1 N_m \hat{N}^2 - 2k_m s^2 E_1 A + k_m^2 s^2 A^2 \equiv G^2. \tag{19}$$

This is again a quadratic in A, having a polynomial solution if and only if:

$$s^2 E_1^2 - E_2^2 + k^2 s^2 N_1 N_m \hat{N}^2 + G^2 = s^2 H^2 \tag{20}$$

for some even polynomial H. Note that, by eq. (17), the left side of (20) must have a double zero at $s = 0$; hence the factor s^2 on the right.

The solution to eq. (19) is then of the form:

$$A = \frac{E_1 \pm H}{k_m} \tag{21}$$

and consequently:

$$B = \frac{E_1 \mp H}{k_1}. \tag{22}$$

Now we have exchanged our three unknown polynomials R, A, and B for two new ones, G and H, both even and

$$\delta G = 2m$$
$$\delta H = 2(m-1).$$

Let us however write eq. (20) in the form:

$$(E_2^2 - s^2E_1^2) - k^2s^2N_1N_m\hat{N}^2 \equiv (G^2 - s^2H^2). \tag{23}$$

Now introduce the notation:

$$F(s) = G(s) + sH(s)$$

and use (14) to write eq. (23) in the form:

$$\boxed{D(s)D(-s) - k^2s^2N_1N_m\hat{N}^2 \equiv F(s)F(-s)} \tag{24}$$

which is our design equation. Once we pick N_1 and N_m, the left side is known save for the constant k^2, which can be selected arbitrarily as long as the left side does not have pure imaginary roots of odd multiplicity. This condition is clearly satisfied for $k^2 = 0$ and hence, by continuity, there must be a finite range

$$0 \leqq k^2 \leqq k^2_{\max}$$

for which it is satisfied.

Our synthesis process is now straightforward and is as follows:

(*i*) Select N_1 and N_m.

(*ii*) Select k^2 arbitrarily but such that

$$D(s)D(-s) - k^2s^2N_1N_m\hat{N}^2$$

has no imaginary roots of odd multiplicity and factor this polynomial into

$$F(s)F(-s).$$

Note that $F(s)$ need not be a Hurwitz polynomial.

(*iii*) Generate G and H from:

$$F(s) = G(s) + sH(s)$$

and E_1 and E_2 from:

$$D(s) = E_2(s) + sE_1(s)$$

where E_1, E_2, G and H are all even polynomials.

(*iv*) Form A, B and R from eqs. (21), (22) and (18) respectively and P and Q from eqs. (13) and (12), giving:

$$P = \frac{(E_2 - G) + s(E_1 \mp H)}{k_1 s} = \frac{D(s) - F(\pm s)}{k_1 s} \tag{25}$$

and

$$Q = \frac{(E_2 - G) + s(E_1 \pm H)}{k_m s} = \frac{D(s) - F(\mp s)}{k_m s}. \tag{26}$$

(*v*) Select the sequence of the remaining numerators $N_i(s)$, $i = 2, 3, \cdots, m - 1$ and synthesize the network by a technique very similar to the standard "zero shifting" technique[7] used in passive reactive ladder network synthesis. The difference is merely that here we deal with even rather than odd rational fractions.

Clearly, this process has several problems.

(a) The above is only one of many cases depending on the parity of m, the parity of D, N_1, and N_m. All these cases have been analyzed, and the general design equations are summarized in Appendix A.

(b) No guarantee is available that the resulting $T_i(s)$ functions will have all positive coefficients. This is similar to the passive ladder form of realization, where all positive element values cannot, in general, be guaranteed either. However, positive coefficients for $T_i(s)$ are not necessary for realization. See Appendix D for further comments on this matter.

(c) We must make the more-or-less arbitrary choices in steps (*i*) and (*ii*) above. Our first concern is to minimize overall sensitivity. We will consider this problem in Appendix B after we have discussed the calculation of the sensitivity.

III. CALCULATION OF T_i TRANSFER FUNCTIONS

Step (*v*) above concerns the final step, that is, the calculation of the coefficients of the individual T_i transfer functions. The method is a slightly modified version of the reactive ladder synthesis, but for completeness we describe it here in some detail.

First we note that we must select an arbitrary (positive) constant k_1 and then k_m is determined from $k_1 k_m = k^2$. This constant is immaterial, since we may later multiply all odd indexed T_i by a constant and divide all even indexed T_i by the same constant. This procedure

will leave the overall transfer function unchanged if there are an even number of T_i blocks, or will multiply it by the same constant if there are an odd number of blocks in the network.

For computational simplicity, let us start by using y_{11}, which in the case of the network shown in Fig. 6a does not depend on Z_m, hence will only suffice for the calculation of $T_1, T_2, \cdots, T_{m-1}$. The last block will therefore have to be calculated from another function, say, z_{22}. From eqs. (8b), (18) and (26) we get:

$$y_{11} = \frac{N_1 R}{Q} = \frac{2N_1(E_2 - G)}{k_1 s[D(s) - F(\pm s)]}. \tag{27}$$

Inverting this, we immediately recognize that it can be written in the form:

$$(i) \quad \frac{1}{y_{11}} = \frac{k_1 s[E_2 + sE_1 - G \pm sH]}{2N_1(E_2 - G)} = \frac{k_1 s}{2N_1} + \frac{k_1}{2}\frac{s^2(E_1 \pm H)}{N_1(E_2 - G)}. \tag{28}$$

Both of these terms are known and the second is pure even; hence from this point on (that is, nearly from the beginning) we are dealing with pure even functions simplifying the work considerably. Denoting

$$(ii) \qquad z_{r1} = \frac{k_1}{2}\frac{s^2(E_1 \pm H)}{N_1(E_2 - G)} \tag{29}$$

we first note that this still has a pole pair at $s = \pm j\omega_1$, the zero of N_1, because of the factor N_1 in its denominator. We remove this by calculating:

$$(iii) \qquad \alpha_1 = N_1 z_{r1}\Big|_{s=j\omega_1} = \frac{k_1}{2}\frac{s^2(E_1 \pm H)}{E_2 - G}\Big|_{s^2=-\omega_1^2} \tag{30}$$

and calculate:

$$z_{r2} = z_{r1} - \frac{\alpha_1}{N_1} \tag{31}$$

that will not have the factor N_1 in its denominator any more. Next we select the transmission zero $\pm j\omega_2$ for the next block and calculate:

$$(iv) \qquad \beta_1 = z_{r2}\big|_{s^2=-\omega_2^2} \tag{32}$$

which leaves a remainder:

$$z_{r3} = z_{r2} - \beta_1 \tag{33}$$

that will have the factor N_2 in its numerator. At this stage we completed a full cycle and calculate the transfer function of the first block

as:

$$(v) \qquad \frac{1}{T_1} = \frac{\frac{k_1}{2}s}{N_1} + \frac{\alpha_1}{N_1} + \beta_1 = \frac{D_1}{N_1} \tag{34}$$

and have a remainder z_{r3} that is two degrees lower. We may invert this now and go back to step (*ii*) (with an admittance function this time) and repeat this loop as many times as required. Equation (34), of course, will contain only the α_i/N_i and β_i terms giving us the required even T_i transfer functions except when $i = 1$.

Finally, to obtain the last block as well as to check on the accuracy of computations, we repeat the same process from the output end using z_{22}. The resulting two structures should be identical save for a constant multiplier, that is to say, denoting the T_i transfer functions obtained from z_{22} by $\hat{T}_i$, to distinguish them from those (T_i) obtained from y_{11}, we must have:

$$\begin{aligned} \downarrow \quad & T_1 \\ & T_2 = C\hat{T}_2 \\ & T_3 = \frac{1}{C}\hat{T}_3 \\ & T_4 = C\hat{T}_4 \\ & \vdots \\ & T_{m-1} = \frac{1}{C}\hat{T}_{m-1} \\ & C\hat{T}_m \quad \uparrow \end{aligned} \tag{35}$$

T_1 and $\hat{T}_m$ have no pairs since y_{11} and z_{22} are independent of T_m and T_1 respectively. The degree to which this set of equations is satisfied with a constant C is the numerical accuracy maintained throughout the computation. The procedure varies slightly when $N(s)$ is odd and/or the degree of $D(s)$ is odd, but only in calculating the first (and last) block. These variations are self-evident and hence will not be dealt with here.

Clearly, the internal sequence of zeros (those excluding N_1 and N_m) is arbitrary at this stage.

Finally, one may note that to ease the numerical accuracy problem, the commonly used transform variable[8,9] or the product method[10] may also be used.

IV. EXAMPLE

Many numerical examples have been calculated by the aforementioned method; here we will illustrate it with one. The example is a bandpass filter with passband from $\omega_A = 0.8$ to $\omega_B = 1.25$ with a 0.5 dB passband ripple. The filter contains four pairs of transmission zeros at:

$$Z_{1,2} = \pm j0.25$$
$$Z_{3,4} = \pm j0.50$$
$$Z_{5,6} = \pm j2.0$$
$$Z_{7,8} = \pm j4.0.$$

The corresponding transfer function poles are at:

$$P_{1,2} = -0.028950107 \pm j0.79624226$$
$$P_{3,4} = -0.087386703 \pm j0.90192127$$
$$P_{5,6} = -0.10642659 \pm j1.0984326$$
$$P_{7,8} = -0.045602221 \pm j1.2542411.$$

Following our procedure, the resulting structure contains four biquadratic blocks with transfer functions of the form of eq. (2); the coefficients are tabulated in Table I.

For the synthesis, we have selected (see Appendix B):

$$N_1 = (s^2 + (0.25)^2)$$
$$N_m = N_4 = (s^2 + (4.)^2)$$

and

$$k^2 = 3.3 \times 10^{-6}$$

which is slightly below $k^2_{\max}$. Next, we selected $k_1 = k$ and $k_m = k$ and performed the synthesis in both directions. These had a ratio

$$C = 8.00488 \times 10^6$$

that was constant to the indicated six decimal digits through all the coefficients. Rescaling the coefficients by $\sqrt{C}$ (that is to say, selecting $k_1 = k/\sqrt{C}$ and $k_m = \sqrt{C}k$) resulted in identical values up to six decimal digits in the two syntheses, and these are the values tabulated in Table I. The computed performance of this structure is shown in Fig. 7. All computations were done in double precision arithmetic (≈ 16 decimal digits), and the indication is that, for higher order cases, either the transformed variable method[11] or, preferably, the product method[10] will have to be used to avoid numerical accuracy problems.

TABLE I—FINAL COEFFICIENTS OF EXAMPLE

i	1	2	3	4
n_0	0.0625	4.0	0.25	16.0
n_1	0	0	0	0
n_2	1.0	1.0	1.0	1.0
d_0	7.98786	3.08873	9.31590	11.3657
d_1	2.56983	0	0	2.56983
d_2	7.74668	3.02573	9.50989	11.6790

4.1 *Practical Results*

The filter, scaled to 1-kHz center frequency, was constructed in the laboratory, using the three-operational amplifier biquad realization[12] for the second-order building blocks. The total structure (see Fig. 8) needed two additional phase inverters to get the correct signs for the feedback loops. The component values are shown in Table II. The measured loss (save for a 3-dB flat difference) is also shown on Fig. 7. The agreement is the kind or better than the kind one usually expects in a passive realization, with no tuning.

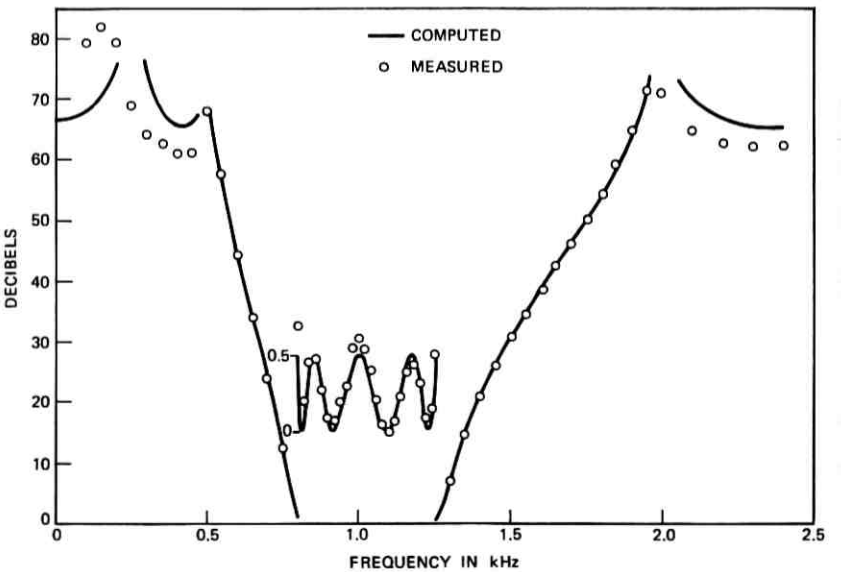

Fig. 7—Computed performance of the filter of example.

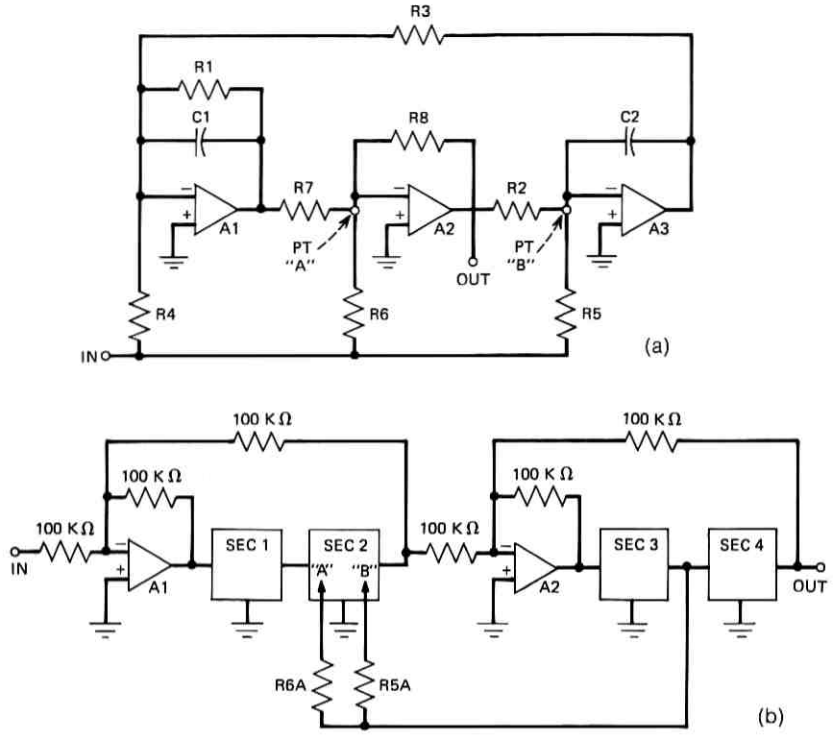

Fig. 8—(a) Basic second-order block. (b) Complete filter.

TABLE II—ELEMENT VALUES OF THE REALIZATION OF FIG. 9

	Sections			
	1	2	3	4
R1	48.1 kΩ	—	—	14.5 kΩ
R2	15.4 kΩ	15.8 kΩ	6.42 kΩ	3.29 kΩ
R3	16.7 kΩ	15.8 kΩ	6.42 kΩ	3.16 kΩ
R4	392. kΩ	—	—	169. kΩ
R5	1.97 MΩ	12.1 kΩ	237. kΩ	2.34 kΩ
R6	77.7 kΩ	6.04 kΩ	95.3 kΩ	17.2 kΩ
R7	9.53 kΩ	2. kΩ	10. kΩ	1.47 kΩ
R8	10. kΩ	2. kΩ	10. kΩ	1.47 kΩ
C1	0.01 μF	0.01 μF	0.025 μF	0.05 μF
C2	0.01 μF	0.01 μF	0.025 μF	0.05 μF
R5A	—	12.1 kΩ	—	—
R6A	—	6.04 kΩ	—	—

V. SENSITIVITY

There are many ways to analyze the sensitivity of different realizations, and the final component sensitivities will clearly depend on the particular realization used for the individual biquadratic blocks. In order to avoid this last step and provide some meaningful insight into the sensitivity properties of the multiple-feedback structure, we will compare it with a straightforward cascade realization.

If several biquadratic sections with transfer functions $T_i(s)$ are connected in cascade, the overall function is given by:

$$T(s) = \prod_{i=1}^{m} T_i(s) \tag{36}$$

and consequently the sensitivities are simple:

$$S_{T_i}^{T} = \frac{T_i}{T}\frac{\partial T}{\partial T_i} = 1 \quad \text{for all } i. \tag{37}$$

Clearly in our multiple-feedback structure, the situation is somewhat more complex. However, by the use of continuants,[13] we can derive closed-form expressions.

A continuant $K_n(X_1, \cdots, X_{n-2}, X_{n-1}, X_n)$ is defined by the recursion formula:

$$K_n(X_1, \cdots, X_{n-2}, X_{n-1}, X_n) = X_n K_{n-1}(X_1, \cdots, X_{n-2}, X_{n-1}) + K_{n-2}(X_1, \cdots, X_{n-2}) \tag{38}$$

and the two starting values:

$$\begin{aligned} K_0(\cdot) &= 1 \\ K_1(X_1) &= X_1. \end{aligned} \tag{39}$$

With these definitions, the overall transfer function of the multiple-feedback structure (with all feedback coefficients equal to -1) is given by:

$$\frac{1}{T} = K_m\left(\frac{1}{T_1}, \frac{1}{T_2}, \cdots, \frac{1}{T_m}\right). \tag{40}$$

Differentiating and rearranging, using some of the properties of continuants, we get the following expression for the sensitivity:

$$S_{T_i}^{T} = \frac{T_i}{T}\frac{\partial T}{\partial T_i} = \frac{T}{T_i}\frac{\partial K_m}{\partial\left(\frac{1}{T_i}\right)} = \frac{1}{1+X} \tag{41}$$

where

$$X = T_i \frac{K_{m-2}\left(\frac{1}{T_1}, \cdots, \frac{1}{T_{i-1}} + \frac{1}{T_{i+1}}, \cdots, \frac{1}{T_m}\right)}{K_{i-1}\left(\frac{1}{T_1}, \cdots, \frac{1}{T_{i-1}}\right) K_{m-i}\left(\frac{1}{T_{i+1}}, \cdots, \frac{1}{T_m}\right)}. \quad (42)$$

From this expression, we see that near a transmission zero, where $T_i \approx 0$, the sensitivity is about unity, that is, it is the same as for the cascade case. However, wherever $T_i \gg 1$, which will occur somewhere in the passband, the corresponding sensitivity becomes very small. No other feature is obvious from this equation, hence a specific case will be used to illustrate the numbers involved.

One additional problem arises in the case of the multiple-feedback structure, namely, the feedback coefficients can also vary from their nominal (-1) value. This can be taken into account by substituting a change

$$-1 \rightarrow -(1 + \epsilon)$$

in the kth feedback path by the changes:

$$T_{k+2i+1} \rightarrow (1 + \epsilon) T_{k+2i+1} \qquad i = 0, 1, \cdots$$

and

$$T_{k+2i} \rightarrow (1 + \epsilon)^{-1} T_{k+2i} \qquad i = 1, 2, \cdots.$$

This way we do not need to develop additional formulas for the sensitivities of the feedback coefficients.

The resulting sensitivity, neglecting the possible multiplier on the overall transfer function, can therefore be derived as follows: A change $\Delta F_k/F_k$ in the kth feedback coefficient has the same effect, ΔT, on the overall transfer function as the changes:

$$\frac{\Delta T_{k+i}}{T_{k+i}} = (-1)^{i+1} \frac{\Delta F_k}{F_k} \qquad i = 1, 2, \cdots, m - k. \quad (43)$$

In other words:

$$\Delta T = \sum_{i=1}^{m-k} \frac{\partial T}{\partial T_{k+i}} \Delta T_{k+i} = \sum_{i=1}^{m-k} (-1)^{i+1} \frac{\partial T}{\partial T_{k+i}} T_{k+i} \frac{\Delta F_k}{F_k} \quad (44)$$

or

$$\frac{\Delta T}{T} = \frac{\Delta F_k}{F_k} \sum_{i=1}^{m-k} (-1)^{i+1} \frac{T_{k+i}}{T} \frac{\partial T}{\partial T_{k+i}}$$

$$= \frac{\Delta F_k}{F_k} \sum_{i=1}^{m-k} (-1)^{i+1} S_{T_{k+i}}^{T} \quad (45)$$

and finally:

$$S_{F_k}^{T} = \sum_{i=1}^{m-k} (-1)^{i+1} S_{T_{k+i}}^{T}. \tag{46}$$

Consequently, the sensitivity to a feedback coefficient can be easily calculated in terms of the already calculated sensitivities to the second-order block transfer functions.

In order to simplify the problem of comparing the multiple-feedback structure with, say, the cascade realization, we have adopted the following procedure.

We calculated the sensitivities of the overall transfer function T with respect to all the coefficients c_i of the individual biquadratic blocks T_k. This was done for both the cascade and the leapfrog-feedback cases, and in the latter the parameters c_i also included the feedback coefficients.

Next we calculated the quantity:

$$\sigma^2 = \sum_i (\mathrm{Re}\, S_{c_i}^{T})^2 \tag{47}$$

as a function of frequency. This quantity would give a measure of the spread of the loss if all coefficients are assumed to be statistically independent variables with the same standard deviation.

Finally, the quantity

$$R = 10 \log_{10} \frac{\sigma_{\mathrm{C}}^2}{\sigma_{\mathrm{MF}}^2} \tag{48}$$

is calculated where the subscript MF indicates multiple-feedback, while the subscript C refers to the cascade configuration. R is therefore a measure of the improvement of the sensitivity of the multiple-feedback structure over that of the cascade realization.

This quantity R is plotted as a function of frequency for the example of Section IV above and is shown in Fig. 9. The curve is extremely interesting, and it shows a slight deterioration in this measure of sensitivity in the stop-band, but a spectacular improvement in the passband. The worsening in the stop-bands is minor, but the passband improvement is substantially more than an order of magnitude. This behavior is very desirable since a 1-dB spread in the stop-band is almost always immaterial, while the same spread in the passband loss can be disastrous.

In order to provide another comparison, a Monte Carlo yield study was made of this filter under the following assumptions.

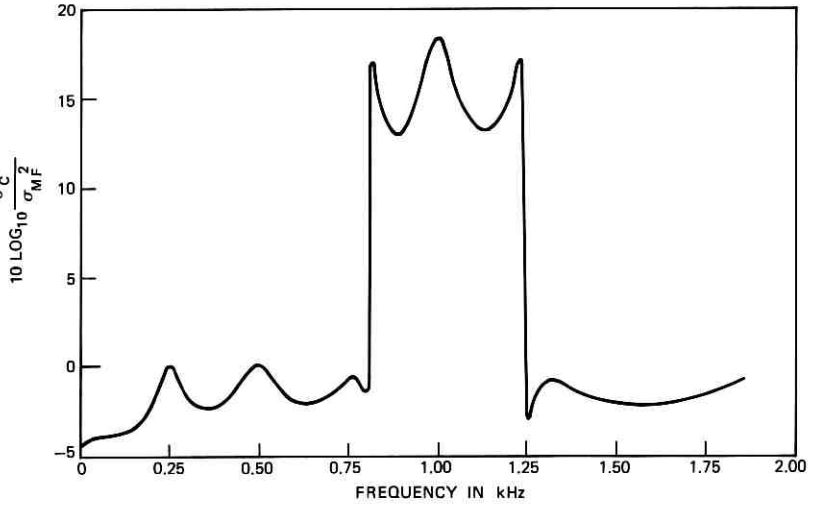

Fig. 9—Sensitivity improvement.

(*i*) Passband is acceptable if loss varies less than 1 dB from 800 Hz to 1250 Hz.

(*ii*) Stop-band is acceptable if loss is at least 60 dB from 0 to 500 Hz and from 1310 Hz up.

(*iii*) The operational amplifiers are close to ideal.

(*iv*) All passive components have the same tolerance with a flat distribution.

The resulting yields are tabulated in Table III for various tolerances. For comparison, the same calculations were performed for a cascade realization of the same filter, under the same assumptions. The results speak for themselves.

TABLE III—COMPARISON OF YIELDS OF EXAMPLE AND EQUIVALENT CASCADE

Tolerance	Yield %	
	MF Filter	Cascade
1.00%	44	—
0.50%	78.5	25
0.25%	100	67.5

Clearly, one cannot draw conclusions on the basis of a single example, but a general proof of the low sensitivity of the multi-feedback structure is lacking. Note also that no attempt has been made in the design of this example to minimize sensitivity apart from the crude heuristic arguments indicated in Appendix B.

VI. CONCLUSIONS

We have demonstrated that the multiple-feedback structure of Fig. 5 is a general one for transfer functions with pure imaginary zeros, and we developed a general synthesis procedure for this structure. The advantages of this realization are the low-sensitivity properties of passive ladders combined with the individually tunable transmission zeros of the cascade structure. The resulting configuration is minimal (no pole-zero cancellation occurs) and hence also stable.

The low-sensitivity properties of the realization have only been demonstrated through examples; a general proof of it is still to be found. Furthermore, apart from some heuristic arguments, no general guidelines are available for selecting the one continuously variable free parameter or to help us at the two additional places where discrete choices are to be made.

Apart from, conceivably, further optimizing the sensitivity properties of the structure, one could also use these choices, as well as additional scaling of the individual biquadratic transfer functions and the feedback coefficients, to optimize the dynamic range of the filter. All of these questions merit further investigations.

VII. ACKNOWLEDGMENTS

Many of the author's ideas were formed during lengthy discussions with several people. In particular, Paul Fleischer, Dan Hilberman, and Jimmy Tow contributed substantially to the developments reported here. Arthur Blake also bore the brunt of the programming of most of the computations. The author acknowledges his indebtedness to them and many others with great pleasure.

APPENDIX A

Summary of General Results

The results given in Section II of this paper were derived for a special case. The general results are as follows:

Given

$$T(s) = \frac{N(s)}{D(s)}$$

where the degree of D is n, and denoting:

$$m = \left[\frac{n+1}{2}\right] \text{(number of blocks)}$$

i.e., the integer part of $1/2(n+1)$ and

$$\hat{N}(s) = \sum_{i=2}^{m-1} N_i(s) \tag{49}$$

$$N(s) = N_1(s)N_m(s)\hat{N}(s) \tag{50}$$

where all $N_i(s)$ are, at most, second order, and $N_1(s)$ and $N_m(s)$ are either even or odd, while the others are strictly even. Now we define:

$$V_{1,m}(s) = \begin{cases} N_{1,m}(s) & \text{if } N_{1,m} \text{ is even} \\ -\left(1 + \dfrac{s^2}{\omega_{1,m}^2}\right) & \text{if } N_{1,m} \text{ is odd.} \end{cases} \tag{51}$$

The defining equation for

$$F(s) = G(s) + sH(s) \tag{52}$$

is:

$$D(s)D(-s) - (-1)^m k^2 s^2 V_1(s) V_m(s) \hat{N}^2(s) \equiv F(s)F(-s) \tag{53}$$

The resulting pseudo-ladder is of the form shown in Fig. 6a if "m" is even, while if "m" is odd, it is as Fig. 6b, where:

$$T_i(s) = \frac{N_i(s)}{D_i(s)} \qquad i = 1, 2, \cdots, m.$$

The Z and Y matrices of these networks are of the form:

$$Z = \frac{1}{N_1 P}\begin{bmatrix} D & N \\ N & N_1 N_m R \end{bmatrix} \tag{54}$$

$$Y = \frac{1}{N_m Q}\begin{bmatrix} N_1 N_m R & -N \\ -N & D \end{bmatrix} \tag{55}$$

for "m" even and

$$Z = \frac{1}{N_1 N_m R}\begin{bmatrix} N_m Q & N \\ N & N_1 P \end{bmatrix} \tag{56}$$

$$Y = \frac{1}{D}\begin{bmatrix} N_1 P & -N \\ -N & N_m Q \end{bmatrix} \tag{57}$$

where "m" is odd. In these equations, the still undefined quantities are the polynomial $R = R(s)$ that is always pure even, and the polynomials $P = P(s)$ and $Q(s)$ that are given by

$$\left.\begin{matrix} P(s) \\ Q(s) \end{matrix}\right\} = \frac{D(s) \pm F(\pm s)}{k_{1s} W_1(s)} N_{1(s)} \tag{58}$$

and

$$R(s) = \text{Ev}\left\{\frac{2N_1(s)N_m(s)[D(s) \pm F(s)]}{k^2 s^2 W_1(s) W_m(s)}\right\} \tag{59}$$

where $Ev\{\ \}$ means the even part of $\{\ \}$,

$$W_{1,m}(s) = \begin{cases} N_{1,m}(s) & \text{if } N_{1,m}(s) \text{ is even} \\ \left(1 + \dfrac{s^2}{\omega^2_{1,m}}\right) & \text{if } N_{1,m}(s) \text{ is odd} \end{cases} \tag{60}$$

and

$$k_1 k_m = k^2.$$

Here as well as above in the definition of $V_{1,m}(s)$ the factors:

$$\left(1 + \frac{s^2}{\omega^2_{1,m}}\right) = \text{Ev}\, D_{1,m}(s).$$

Hence, they are unknown initially. Note also that if n is odd, then one of the D_1 and D_m is only linear in s, i.e., the corresponding ω_1^2 or $\omega_m^2 \to \infty$.

For that case where ω_1 and/or ω_m are present, an iterative procedure must be employed, since these factors must be such that $P(s)$, $Q(s)$ and $R(s)$ are all polynomials of the correct degree.

This condition is satisfied if $(1 + s^2/\omega_1^2)$ and $(1 + s^2/\omega_m^2)$ are factors of $(E_1 + H)$ and $(E_1 - H)$ respectively, and thus can be obtained iteratively as follows. One picks the factors $(1 + s^2/\omega_1^2)$ and $(1 + s^2/\omega_m^2)$ arbitrarily, but their zeros should be in or near the passband. Then one performs the factorization of the polynomial $F(s)$ and calculates the zeros of both $(E_1 + H)$ and $(E_1 - H)$. At this stage, we can replace our arbitrary factors by the nearest factors of $(E_1 + H)$ and

$(E_1 - H)$, selecting one factor from each. A repetition of this process, or a modification of it, will converge to the desired solution.

If there is only one unknown factor involved, the iteration is somewhat simpler and any of the factors of $(E_1 \pm H)$ can be used. One must further show that, once this iteration converged, one can always select sign combinations in the expressions for $P(s)$, $Q(s)$ and $R(s)$ such that the resulting expressions reduce to simple polynomials of the correct degrees. This can be shown directly in the limiting case $k^2 = 0$ (when, of course, no iteration is needed), and the existence of a solution for a (finite) range of nonzero $k^2 > 0$ values can then be inferred from continuity again. In order to keep the length of this paper within bounds, the details of this step are left for the reader. Finally, note that not all sign combinations shown in eqs. (58) and (59) are allowed, but at least one will always work.

APPENDIX B

Selection of the Parameter k^2 as Well as the End Factors N_1 and N_m

In Sections II and III we pointed out a number of arbitrary choices the designer must make. The following comments may help in some of these decisions.

In order that eq. (53) be factorable in the form of $F(s)F(-s)$, with $F(s)$ a real polynomial, it is necessary and sufficient that the left side have no imaginary root of odd multiplicity. Since this is clearly true for the case $k^2 = 0$ and since the roots will be continuous functions of k^2, the condition will be satisfied for a range of values

$$0 < k^2 \leqq k^2_{\max}.$$

The upper limit $k^2_{\max}$ can be calculated as follows: Since $D(s)$ cannot have a pure imaginary root, $F(s)F(-s)$ will have no imaginary root of any multiplicity as long as

$$\left.\frac{F(s)F(-s)}{D(s)D(-s)}\right|_{s=j\omega} = 1 - k^2\left[\frac{(-1)^m s^2 V_1(s) V_m(s) \hat{N}^2(s)}{D(s)D(-s)}\right]_{s=j\omega} > 0 \qquad (61)$$

or as long as

$$k^2\left[\frac{(-1)^m s^2 V_1(s) V_m(s) \hat{N}^2(s)}{D(s)D(-s)}\right]_{s=j\omega} < 1$$

and therefore this condition will be satisfied as long as

$$k^2 \leqq k_{\max}^2 = \left\{ \max_{\omega} \left[\frac{(-1)^m s^2 V_1(s) V_m(s) \hat{N}^2(s)}{D(s)D(-s)} \right]_{s=j\omega} \right\}^{-1}. \quad (62)$$

In the limiting case $k^2 = k_{\max}^2$ there will be at least one $s = j\omega$ value where $F(s)F(-s)$ will be zero, but all of these will necessarily be of even multiplicity. Moreover, since this limiting case in passive filters corresponds to the maximum power transfer from source to load, we conjecture that it will also be optimum in our active case in the sense of providing us with the least sensitive results.

Going one step further, we note that in the passive case all the pure imaginary zeros of $F(s)$ fall inside the passband. In order to achieve this in the active case, as well as to make as many zeros of $F(s)$ approach the imaginary axis as close as possible, the factors $V_1(s)$ and $V_m(s)$ should be selected such that

$$\left. \frac{(-1)^m s^2 V_1(s) V_m(s)}{N_1(s)N_1(-s)N_m(s)N_m(-s)} \right|_{s=j\omega} \quad (63)$$

is as close to a *positive* constant in the passband as possible. For instance, if N_1 and N_m are both even and "m" is also even, this can be achieved by selecting $N_1(s)$ to be the lowest transmission zero below the passband and $N_m(s)$ to be the highest transmission zero above. The reason for the above requirement is simply the fact that in the passband

$$\left. \frac{N(s)N(-s)}{D(s)D(-s)} \right|_{s=j\omega} \quad (64)$$

will usually be close to a positive constant, and therefore the product of (63) and (64) will also have this property, leading to an $F(s)F(-s)$ that is "small" inside the filter passband.

Finally, we come to the question of separating $F(s)$ out of $F(s)F(-s)$. Since $F(s)$ need not be a Hurwitz polynomial and since only a few, if any, of the zeros of $F(s)F(-s)$ will be purely imaginary (and of even multiplicity), we will have a finite number of possibilities to choose from. At this time we have no guideline to offer, we simply note that in many cases we had to select roots with alternating real parts in order to reduce the size of the coefficients of $F(s)$, as this procedure seemed to be necessary to insure the positiveness of the final coefficient values.

Clearly, the selection of $F(s)$ will affect the final network sensitivity properties, but further study is needed to clarify the role $F(s)$ plays in influencing the sensitivity.

APPENDIX C

Modifying the Feedback Coefficients

As mentioned in Section I of this paper, the restriction of all feedback coefficients to -1 does not restrict the generality of our structure and can be modified to scale the individual $T_k(s)$ blocks. Such a scaling is necessary to obtain the maximum overall dynamic range for the structure and can be performed simply as follows.

The T_k transfer function can be multiplied by a constant α_k if all subsequent transfer functions are modified as

$$\left.\begin{aligned} T_{k+2i} &\rightarrow \alpha_k T_{k+2i} \\ T_{k+2i-1} &\rightarrow \frac{1}{\alpha_k} T_{k+2i-1} \end{aligned}\right\} \quad i = 1, 2, \cdots .$$

Depending on the number of blocks following T_k, this will either leave the overall transfer function unchanged, or will multiply it by α_k. Also, the -1 in the $(k-1)$th feedback loop must be replaced by $-1/\alpha_k$. This procedure works for all blocks. In the case of the first block, an alternative is to multiply T_1 by a constant α_1, multiply the second feedback loop coefficient and divide T_2 by the same α_1 constant. This way one can scale each and every one of the T_k transfer functions without any effect on the overall transfer function save for a constant multiplier.

For the purpose of adjusting the dynamic range of the individual blocks, one must be able to calculate the input and output voltage levels of each block. In the case of all -1 feedback coefficients, this can be easily done by recognizing that these voltages are numerically equal to the branch "voltages" and "currents" of our pseudo-ladders of Figs. 6a or b. As such, these can be readily computed by the use of continuants again or by any other convenient way.

APPENDIX D

Comments on the Positiveness of the Coefficients

Consider now the realizability of the multiple-feedback structure. The synthesis technique given above guarantees the existence of a set of T_k biquadratic transfer functions with the specified properties for any $0 < k^2 \leqq k^2_{\max}$ and arbitrarily specified zero sequence. However, some of the coefficients may turn out to be negative. In such a case, it is preferable that one or more of the T_k blocks have an overall

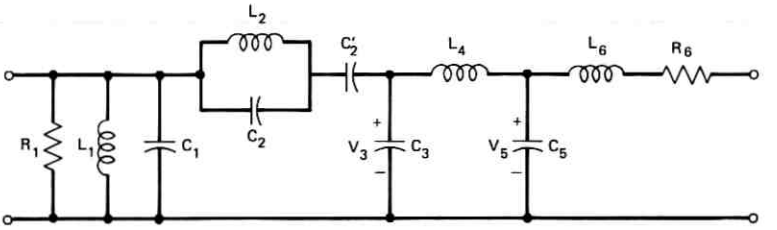

Fig. 10—Passive bandpass ladder filter.

negative sign rather than blocks having coefficients of mixed signs. We have so far been able to achieve this in all of the many examples calculated, but we had to put up with negative T_k blocks in certain cases. In particular, nonbandpass filters are likely to have negative blocks in their realization.

However, this is not a shortcoming of the method, since it appears even if the structure was derived from a passive double-terminated ladder with all positive elements. The sensitivity improvement is still realized and no instability will be generated, since our structure is minimal. Consequently, there are no pole-zero cancellations and, since the overall system poles are the zeros of $D(s)$, the system will be stable if the original requirements called for a stable transfer function.

As an example, consider the passive bandpass filter of Fig. 10. To be specific, let us select a passband from 660 Hz to 980 Hz with 0.25-dB loss ripple and transmission zeros at

$$Z_{1,2} = 0$$
$$Z_{3,4} = \infty$$
$$Z_{5,6} = \pm j1280 \text{ Hz}$$
$$Z_{7,8} = \infty .$$

The resulting poles normalized to the geometric center frequency are

TABLE IV—ELEMENT VALUES OF CIRCUIT IN FIG. 10

$R_1 = 1.0$	$C_2 = 5.46300$
$L_1 = 0.256378$	$L_4 = 0.173352$
$C_1 = 3.71840$	
$L_2 = 0.981823$	$C_5 = 72.1868$
$C_2 = 0.402048$	$L_6 = 0.0129235$
$C_2' = 0.556146$	$R_6 = 0.00327825$

TABLE V—COEFFICIENTS OF A MULTIPLE-FEEDBACK REALIZATION OF THE PASSIVE FILTER SHOWN IN FIG. 10

	1	2	3	4
n_0	—	1.0	−1.0	0.413433
n_1	—	—	—	—
n_2	0.256378	0.394775	—	—
d_0	1.0	1.98114	5.87644	1.0
d_1	0.256378	—	—	0.236646
d_2	0.953318	1.76393	5.17359	0.932908

located at

$$P_{1,2} = -0.041320302 \pm j0.81446903$$
$$P_{3,4} = -0.096338325 \pm j0.94746237$$
$$P_{5,6} = -0.085145570 \pm j1.1203830$$
$$P_{7,8} = -0.030860810 \pm j1.2239735.$$

This filter can be realized as shown in Fig. 10 with element values (normalized to 1-ohm input impedance level and 1 rad/s center frequency) given in Table IV.

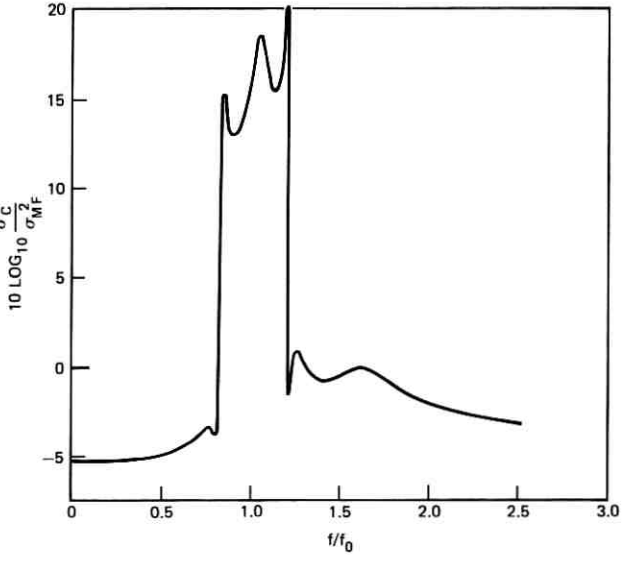

Fig. 11—Sensitivity improvement.

One can derive an active multiple-feedback equivalent of the form of Fig. 3 for this structure by writing down the branch equations as we have done in Section I, and then eliminate the variables V_3 and V_5 from them.

The resulting (normalized) coefficients are given in Table V and the most noteworthy result is that the third block has a negative sign associated with it. On the basis of Appendix C, this is equivalent to having a $+1$ (positive) feedback coefficient in the second and third loops.

Our synthesis technique was then used to generate other (equivalent) realizations, but we have failed to find one with all positive coefficients. Nevertheless, the realization shown above is quite satisfactory; the sensitivity improvement compared to the cascade realization is shown in Fig. 11.

Under what conditions are we to accept one or more negative blocks in the realization is a question that remains to be answered.

REFERENCES

1. Orchard, H. J., "Inductorless Filters," Elec. Ltrs. *2* (June 1966), pp. 242–225.
2. Girling, F. E., and Good, E. F., "The Leapfrog or Active Ladder Synthesis," Part 12, and "Applications of the Active Ladder Synthesis," Part 13, in "Active Filters," Wireless World, *76* (July 1970), pp. 341–345 and (September 1970), pp. 445–450.
3. Adams, R. L., "On Reduced Sensitivity Active Filters," Proc. 14th Midwest Symp. Circuit Theory, 1971, pp. 14.3–1/14.
4. Wing, O., "Ladder Network Analysis by Signal-Flow Graph–Application to Analog Computer Programming," IRE Trans. Circuit Theory, *CT-3* (December 1956), pp. 289–294.
5. Navot, I., "The Synthesis of Certain Subclasses of Tridiagonal Matrices with Prescribed Eigenvalues," SIAM J. Appl. Math., *15* (March 1967), p. 241–251.
6. One promising avenue is explored in Tow, J., and Kuo, Y. L., "Coupled Biquad Active Filters," Proc. Int. Symp. Circuit Theory, N. Hollywood, Cal. (April 18–21, 1972), pp. 164–168.
7. See for instance: Saal, R., and Ulbrich, E., "On the Design of Filters by Synthesis," IRE Trans. Circuit Theory, *CT-5* (December 1958), pp. 284–327.
8. Szentirmai, G., "Theoretical Basis of a Digital Computer Program Package for Filter Synthesis," Proc. 1st Annual Allerton Conf. Circuit and System Theory, Univ. of Illinois, 1963, pp. 37–49.
9. Orchard, H. J., and Temes, G. C., "Filter Design Using Transformed Variables," IEEE Trans. Circuit Theory, *CT-15* (December 1968), pp. 385–408.
10. Skwirzynski, J. K., "On Synthesis of Filters," IEEE Trans. Circuit Theory, *CT-18* (January 1971), pp. 152–163.
11. See for instance: Temes, G. C., "Filter Design in Transformed Frequency Variable," Chapter 6 in F. F. Kuo and W. G. Magnuson (Eds), *Computer Oriented Circuit Design*, Englewood Cliffs; Prentice Hall, Inc., 1969, pp. 251–303.
12. Tow, J., "Active RC Filters—A State-Space Realization," Proc. IEEE, *56* (June 1968), pp. 1137–1139.
13. Barlett, A. C., *The Theory of Electrical Artificial Lines and Filters*, London: Chapman & Hall, 1930.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 4, April, 1973
Printed in U.S.A.

Statistical Behavior of Rain Attenuation*

By S. H. LIN

(Manuscript received October 25, 1972)

Thirty-one sets of experimental data on the statistics of microwave rain attenuation at frequencies above 10 GHz, in the U.S.A., England, Japan, Italy, and Canada, indicate that: (i) the distribution of rain attenuation α, in dB, is approximately lognormal with a standard deviation σ_a of $\log_{10}\alpha$ ranging from 0.46 to 0.71 for earth-space paths, and from 0.33 to 0.86 for terrestrial paths; (ii) the distribution of the rain fade duration τ is also approximately lognormal with a standard deviation σ_τ of $\log_{10}\tau$ ranging from 0.44 to 0.76 for both earth-space paths and terrestrial paths. We propose a theory to explain this general behavior. A theoretical upper bound for the fade duration distribution in the tail region is also given.

The findings in this paper simplify the determination of rain attenuation statistics needed for the design of earth-satellite radio links and terrestrial radio links.

I. INTRODUCTION

The statistics of rain attenuation are important for the design of both terrestrial and earth-satellite radio links using frequencies above 10 GHz. Many experiments[1–33] have been performed to obtain data on rain attenuation for different frequencies, path lengths, and geographical locations. This paper presents the behavior of rain attenuation statistics found in our study to be common to the available experimental data.†

II. DEFINITIONS

Let

$V(t)$ be the time-varying amplitude of the received signal voltage normalized to its nonfaded level,

* An excerpt of this paper has been presented at the 1972 IEEE International Conference on Communications at Philadelphia, and included in the Proceedings of the Conference.

† These data are actual measured microwave rain attenuations and not the approximate attenuations calculated from rain rate data.

$\alpha(t) = -20 \log_{10} V(t)$ be the time-varying rain attenuation in dB,

$P(\alpha \geqq A)$ be the expected fraction of time that $\alpha(t)$ exceeds any specified value A,

P_o be the expected fraction of time that rain falls at the location of the radio link,

$P_c(\alpha \geqq A)$ be the expected fraction of raining time that α exceeds A,

α_m be the median value of α during the raining time, i.e., $P_c(\alpha \geqq \alpha_m) = 0.5$,

$\tau(A)$ be the duration of rain attenuation fades with α exceeding any specified threshold A in dB,

$P[\tau(A) \geqq b]$ be the probability that the fade duration $\tau(A)$ exceeds any specified duration b, and

$\bar{\tau}(A)$ be the average duration.

Notice that

$$P(\alpha \geqq A) = P_o \cdot P_c(\alpha \geqq A). \tag{1}$$

Thus, P_o may also be called the probability of rainfall and $P_c(\alpha \geqq A)$ the conditional distribution of α under the condition that the rain is falling.

III. SUMMARY OF RESULTS

3.1 *Attenuation Distribution*

(*i*) The available experimental data on both earth-space paths and terrestrial paths in the U.S.A., England, Japan, Italy, and Canada show consistently that the conditional distribution, $P_c(\alpha \geqq A)$, is approximately lognormal within the attenuation range, 1 dB $\leqq \alpha \leqq$ 50 dB, of practical interest.*

(*ii*) We propose a theory to explain the lognormal behavior of attenuation $\alpha(t)$. In essence, the value of the rain attenuation, $\alpha(t)$, at any time instant can be multiplicatively affected by a large number of random time-varying parameters of the environment such as the present states and the past histories of the weather conditions at various locations all over the world. The large number of random multiplicative components and the central limit theorem lead to the lognormal distribution of $\alpha(t)$.

* This means the distribution of signal amplitude $V(t)$ is approximately log-lognormal.

(*iii*) These results indicate that three parameters, P_o, σ_a (the standard deviation), and α_m (the median), are sufficient to determine the rain attenuation distribution $P(\alpha \geqq A)$. These parameters depend on geographic locations.

(*iv*) α_m increases almost linearly with the path length because the median rain rate usually is small and is almost uniform over the entire path.

(*v*) The dependence of α_m on frequency follows the theoretical prediction of Sezter[20] for small rain rate because the median rain rate is usually small. The rain rate data in Refs. 13 and 34 to 37 indicate that the median rain rate ranges from 0.5 to 10 mm/h, depending on location.

(*vi*) σ_a decreases slightly as the path length increases because of the averaging effect of the propagation volume.[24,25]

(*vii*) σ_a decreases slightly as frequency increases.

3.2 *Fade Duration Distribution*

(*i*) The experimental data indicate that the fade duration distributions, $P[\tau(A) \geqq b]$, are also approximately lognormal. The physical reason for the lognormal duration is the same as that for the lognormal attenuation.

(*ii*) The probability of occurrence of long fade durations has an absolute upper bound given by

$$P\left(\frac{\tau}{\bar{\tau}} \geqq X\right) \leqq \frac{1}{2}\operatorname{erfc}\left(\frac{\log_{10} X}{M}\right)^{\frac{1}{2}} \qquad (2)^*$$

for any $X \geqq 1$. For example, $P(\tau/\bar{\tau} \geqq 10) \leqq 0.0161$ means that no more than 1.61 percent of the total number of fades will have durations longer than $10 \cdot \bar{\tau}$.

IV. EXPERIMENTAL DATA

4.1 *Rain Attenuation Distribution*

Thirty-one sets of experimental data on rain attenuation distribution, each with a time base of six months or longer, are summarized in Tables I, II, and III for earth-space paths, long terrestrial paths ($>$ 10 km), and short terrestrial paths ($\leqq$ 10 km), respectively.†

* The constant M is defined in eq. (13).

† In the literature, there are many other sets of experimental data with a time base less than six months. Those data are not included because the short-term distributions of rain attenuation are fairly random.

TABLE I—RAIN ATTENUATION EXPERIMENTS ON EARTH–SPACE PATHS

Authors	Reference Number	Location	Frequency (GHz)	Path Length (km)	P_o	σ_a	α_m (dB)	Time Base
Wilson and Ruscio	4, 5, 17	Crawford Hill, N. J.	30	Earth-Space	5.23×10^{-2}	0.57	1.6	Dec. 8, 1967–Feb. 28, 1969
Wilson and Ruscio	4, 5, 17	Crawford Hill, N. J.	16	Earth-Space	5.23×10^{-2}	0.71	0.33	Dec. 8, 1967–Feb. 28, 1969
Wilson	15	Crawford Hill, N. J.	16	Earth-Space	5.9×10^{-2}	0.67	0.3	April 1, 1969–Aug. 7, 1969
Wilson	15	Sayreville, N. J.	16	Earth-Space	5.9×10^{-2}	0.60	0.48	April 1, 1969–Aug. 7, 1969
Wilson	15	Parkway, N. J.	16	Earth-Space	5.9×10^{-2}	0.70	0.38	April 1, 1969–Aug. 7, 1969
Davies	31	England	19	Earth-Space	8.5×10^{-2}	0.46	0.4	June 1968–May 1970

Note: This table does not include the results of an earth-space path at Point Reyes (30 miles north of San Francisco), California by K. O'Brien because the rain attenuation exceeded 8 dB only once (for 8 minutes) during the 9-month period of the experiment at that location.

TABLE II—RAIN ATTENUATION EXPERIMENTS ON LONG TERRESTRIAL PATHS

Authors	Reference Number	Location	Frequency (GHz)	Path Length (km)	P_o	σ_a	α_m (dB)	Time Base
Hathaway and Evans	7	Mobile, Ala.	11	43.5	3.7×10^{-2}	0.40	7.1	1956
Zimmerman	8	Mobile, Ala.	17	22.9	3.7×10^{-2}	0.46	6.4	Aug. 1, 1958–Aug. 31, 1959
Zimmerman	8	Mobile, Ala.	17	14	3.7×10^{-2}	0.49	4	Aug. 1, 1958–Aug. 31, 1959
Funakawa, et al.	11	Near Tokyo, Japan	12.62	80	10×10^{-2}	0.42	3.1	Nov. 1964–Oct. 1965
Turner, Easterbrock, et al.	12, 13, 14	Southern England	11	24	8.5×10^{-2}	0.51	0.21	Jan. 1967–Dec. 1968
Turner, Easterbrock, et al.	12, 13, 14	Southern England	18	24	8.5×10^{-2}	0.51	0.55	Jan. 1967–Dec. 1968
Turner, Easterbrock, et al.	12, 13, 14	Southern England	36	24	8.5×10^{-2}	0.52	2.3	Jan. 1967–Dec. 1968
Turner, Easterbrock, et al.	12, 13, 14	Southern England	11	24	6.4×10^{-2}	0.40	0.76	1964
Turner, Easterbrock, et al.	12, 13, 14	Southern England	11	30.6	6.4×10^{-2}	0.34	1.2	1963 and 1964
Turner, Easterbrock, et al.	12, 13, 14	Southern England	11	38.62	6.4×10^{-2}	0.38	1.0	1964
Turner, Easterbrock, et al.	12, 13, 14	Southern England	11	58	6.4×10^{-2}	0.37	1.4	1964
Blevis, et al.	6	Ottawa, Canada	15	15.78	5.8×10^{-2}	0.67	0.4	May–Oct. 1965
Stracca	33	Italy	18	20	6×10^{-2}	0.56	1.82	May–Oct. 1967
Stracca	33	Italy	11	20	6×10^{-2}	0.71	0.45	May–Oct. 1967

TABLE III—RAIN ATTENUATION EXPERIMENTS ON SHORT TERRESTRIAL PATHS

Authors	Reference Number	Frequency (GHz)	Path Length (km)	Location	Time Base	σ_a	α_m (dB)	P_o
Barnett and Bergmann	44	17.83	5.23	Palmetto, Georgia	Nov. 1970–June 1971	0.57	0.55	5.7×10^{-2}
Barnett and Bergmann	44	17.71	5.07	Palmetto, Georgia	Nov. 1970–June 1971	0.6	0.5	5.7×10^{-2}
Barnett and Bergmann	44	17.95	5.07	Palmetto, Georgia	Nov. 1970–June 1971	0.57	0.55	5.7×10^{-2}
Semplak	2	18.5	2.6	Crawford Hill, N. J.	Jan. 1–Sept. 25, 1970	0.86	0.074	4.03×10^{-2}
Semplak	2	30.9	2.6	Crawford Hill, N. J.	Jan. 1–Sept. 25, 1970	0.78	0.24	4.03×10^{-2}
Semplak	1	18.5	6.4	Crawford Hill, N. J.	1967 + 1968 + 1969	0.86	0.11	5.08×10^{-2}
Semplak	1	30.9	1.9	Crawford Hill, N. J.	1968 + 1969	0.85	0.11	5.55×10^{-2}
Delange and Dietrich	30	60	1.03	Crawford Hill, N. J.	1970	0.47	0.94	4.15×10^{-2}
Gray	3	100	0.61	Crawford Hill, N. J.	1970	0.55	0.46	4.15×10^{-2}
Kenny	16	18.4	4.3	Merrimack Valley, Mass.	April 1970–July 1971	0.44	1.1	4×10^{-2}
Hickin	9	18	10	England	March 1964–Feb. 1966	0.85	0.045	6.6×10^{-2}

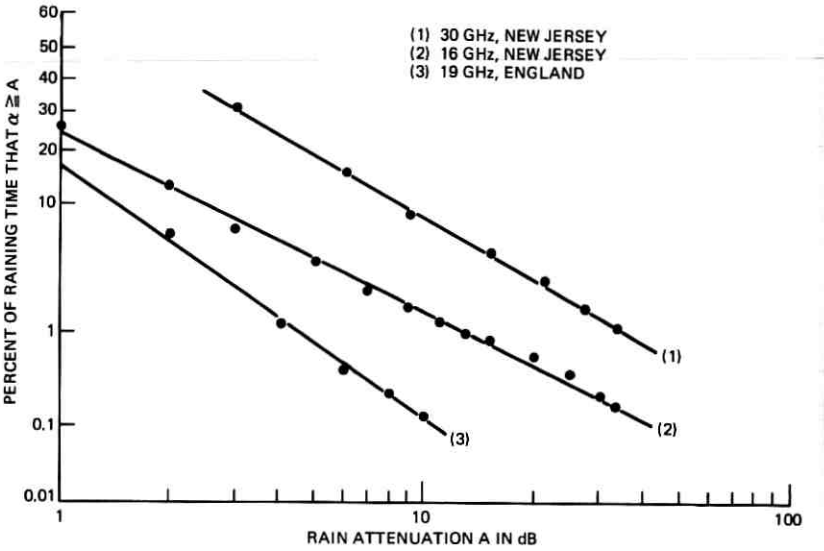

Fig. 1—Lognormal distributions of rain attenuation on earth-space paths.

The original data are all given in terms of the unconditional distribution $P(\alpha \geqq A)$. Based upon P_o (either available or estimated), we convert these unconditional distributions into the conditional distribution $P_c(\alpha \geqq A)$ by eq. (1). This conversion is done to exclude the dry periods in which the rain attenuation is identically zero.*

When plotted on a lognormal coordinate system, these conditional distributions, $P_c(\alpha \geqq A)$, are all approximately straight lines within the attenuation range, 1 dB $\leqq \alpha \leqq$ 50 dB, of practical interest. Figures 1 to 3 show nine examples. The equation describing the lognormal distribution is

$$P_c(\alpha \geqq A) = \tfrac{1}{2} \operatorname{erfc} \left[\frac{\log_{10} A - \mu_a}{\sqrt{2}\sigma_a} \right] \tag{3}$$

where erfc ($\sim$) denotes the complementary error function; σ_a is the standard deviation of $\log_{10} \alpha$ during the raining time; and

$$\mu_a = \log_{10} \alpha_m \tag{4}$$

is the mean value of $\log_{10} \alpha$ during the raining time. The estimated values of σ_a and α_m are given in Tables I, II, and III.

However, the accuracy of these estimated values of σ_a and α_m is

* Absorption by the clear atmosphere is not considered in this paper.

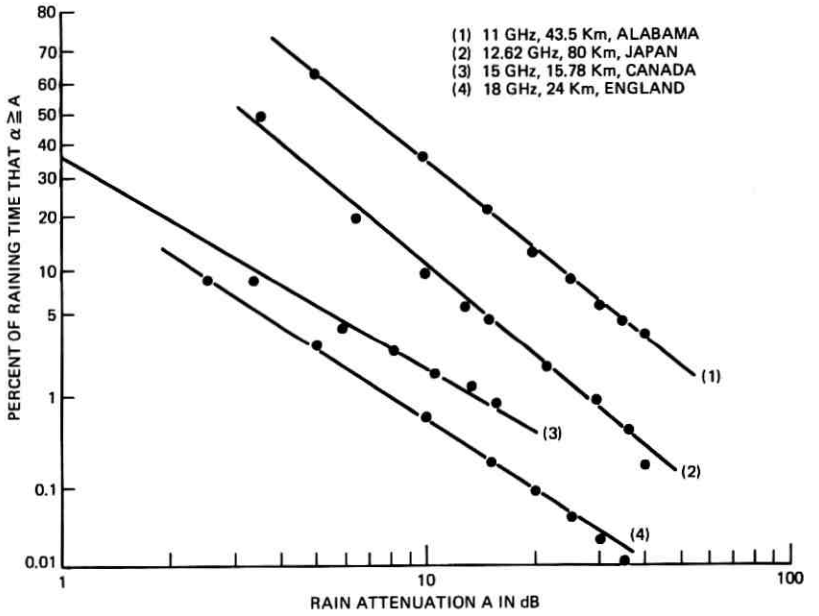

Fig. 2—Lognormal distributions of rain attenuation on long terrestrial paths.

limited by two problems:

(*i*) The time base of the experiment may not be sufficiently long to yield stable statistics, and

(*ii*) some of the published experimental data do not provide the probability P_o of rain during the experiment.*

Let

$$\beta(t) = \frac{[\log_{10} \alpha(t)] - \mu_a}{\sigma_a}. \tag{5}$$

If α is lognormally distributed, then β will be normally distributed with zero mean and unity standard deviation. This allows us to pool all the available data of β on the same graph paper for comparison. Figures 4, 5, and 6 show the pooled data of the earth-space paths, long terrestrial paths, and short terrestrial paths, respectively. It is seen that these experimental results of β are indeed normally distributed with zero mean and unity standard deviation.

* The P_o values at Crawford Hill, New Jersey, are provided by D. C. Hogg and R. A. Desmond from their rain gauge records. The P_o values in Alabama, England, Japan, and Canada are estimated from the information in Refs. 6, 8, 11, and 13. The P_o values in Italy and Massachusetts are assumed values using some judgment.

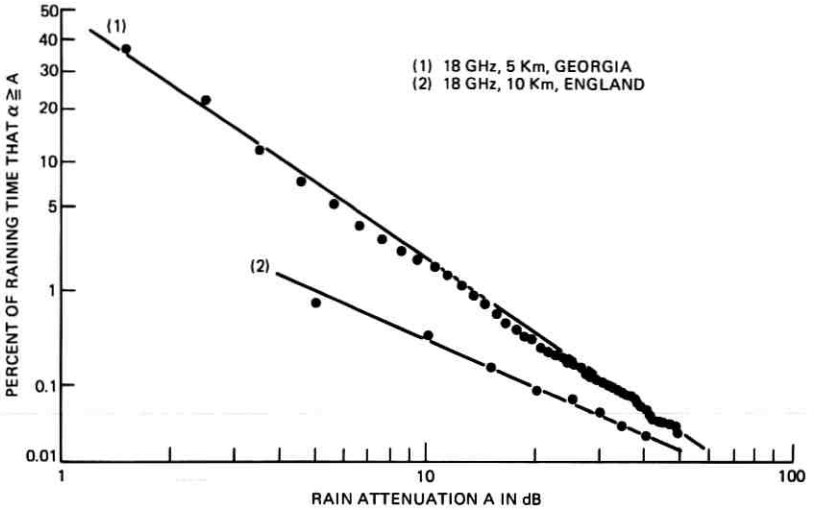

Fig. 3—Lognormal distributions of rain attenuation on short terrestrial paths.

However, Fig. 6 shows that the data of short paths have significant deviations from the normal distribution in the tail region ($\beta \geqq 2.8$). Furthermore, most of the deviations in the tail region are downward from the straight line approximation. The reason for this nonsymmetric deviation is discussed in Appendix B.

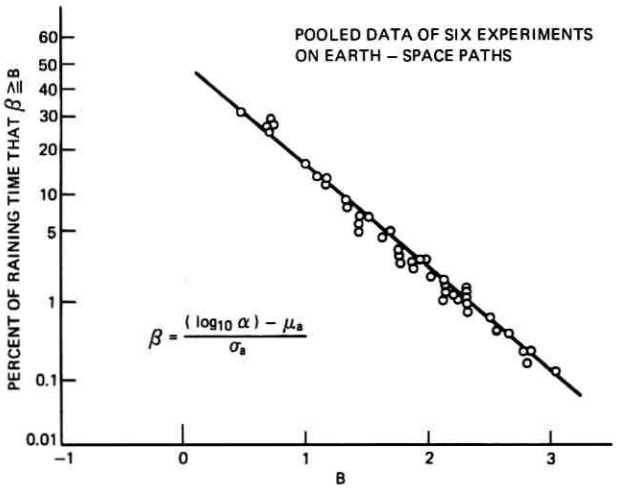

Fig. 4—Normal distribution of β of earth-space paths.

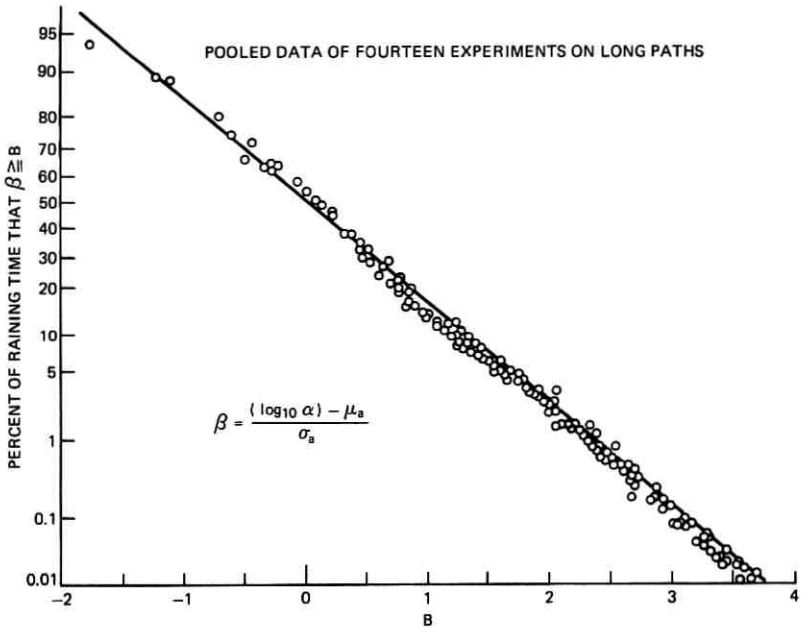

Fig. 5—Normal distribution of β of long terrestrial paths.

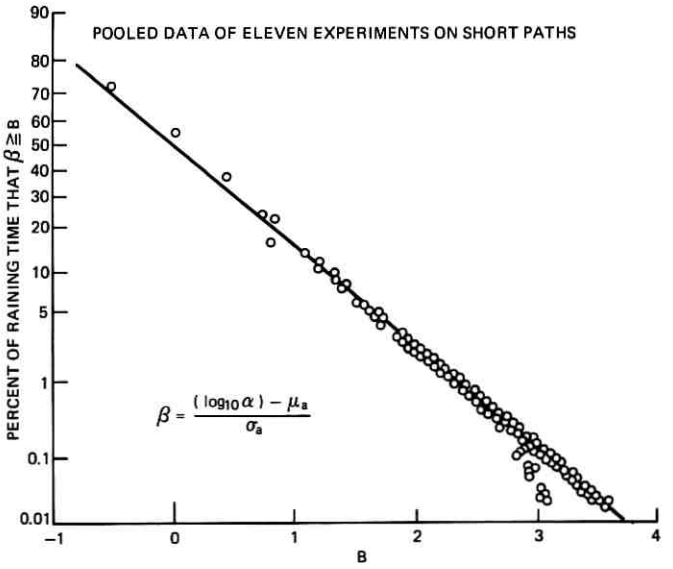

Fig. 6—Normal distribution of β of short terrestrial paths.

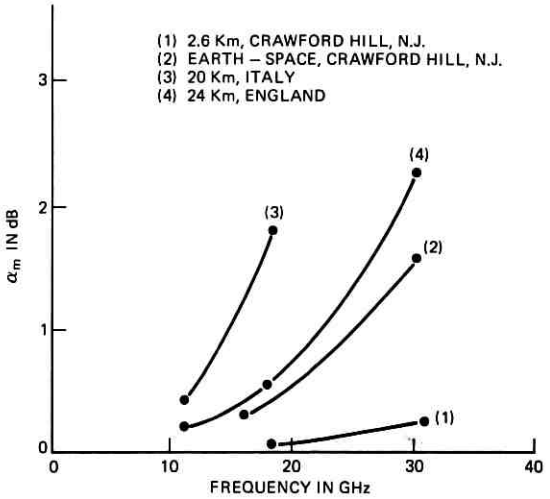

Fig. 7—Effect of frequency on median attenuation α_m.

4.2 *Dependence of σ_a and α_m on Path Length and Frequency*

The experimental results, discussed in Section 4.1, show that the three parameters, P_o, σ_a, and α_m are sufficient to determine the distribution $P(\alpha \geqq A)$. Therefore, it is important to study the dependence of σ_a and α_m on path length and frequency.

The effects of path length and frequency on σ_a and α_m are shown in Figs. 7 to 10. These experimental results indicate the effects of path

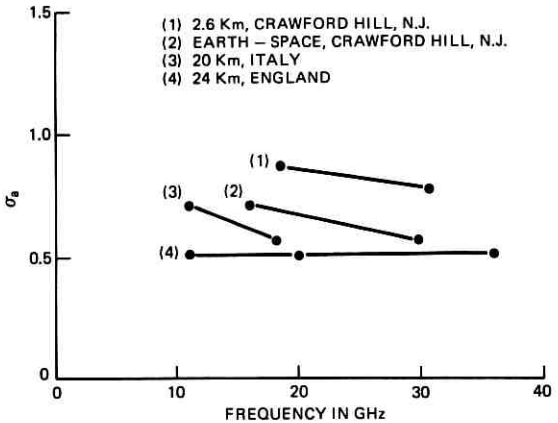

Fig. 8—Effect of frequency on standard deviation σ_a.

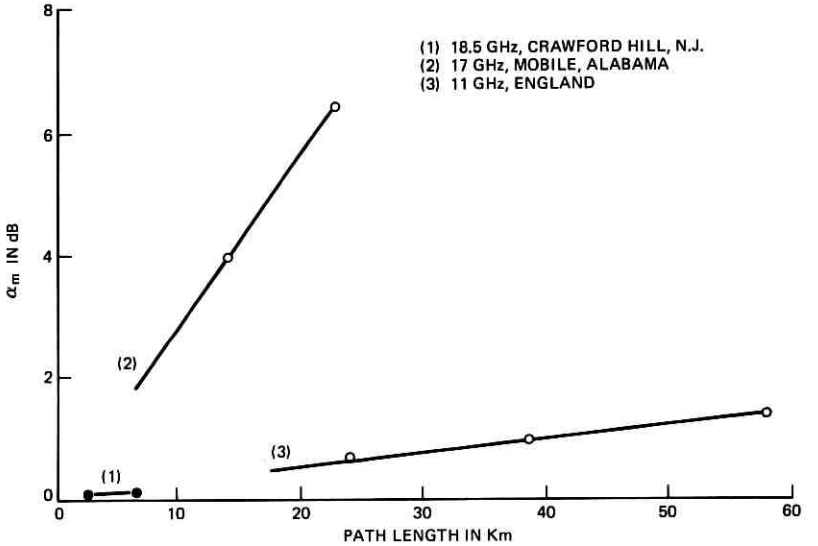

Fig. 9—Effect of path length on median attenuation α_m.

length and frequency on σ_a and α_m as stated in terms (*iv*) to (*vii*) of Section 3.1.*

The theoretical calculation[20] indicates that the increase of rain attenuation with frequency f is slightly faster than the square law in the range: 10 GHz $\leqq f \leqq$ 60 GHz, and 1 mm/h $\leqq$ rain rate $\leqq$ 15 mm/h. This is the basis of the dashed curves in Fig. 7. The slopes of straight lines in Fig. 9 are proportional to median rain rate and extinction coefficient. Curves (1) and (2) in Fig. 9 indicate that the median rain rate in Alabama is much larger than that in New Jersey.

4.3 *Fade Duration Distribution*

The available nine sets of experimental data on the histogram of the durations $\tau(A)$ of rain attenuation fades are summarized in Table IV.† We convert these histograms into the cumulative distribution $P[\tau(A) \geqq b]$. On a lognormal coordinate system, these fade duration distributions are all approximately straight lines. Figure 11 shows two examples.

* In Fig. 8, the σ in England seems to increase slightly with frequency in contrast to those in the U.S.A. and Italy. A possible reason for this inconsistency is that the time bases for the three sets of data for 11, 18, and 36 GHz measured in England are not concurrent.

† Some of the available data in the literature are not included because of short time base.

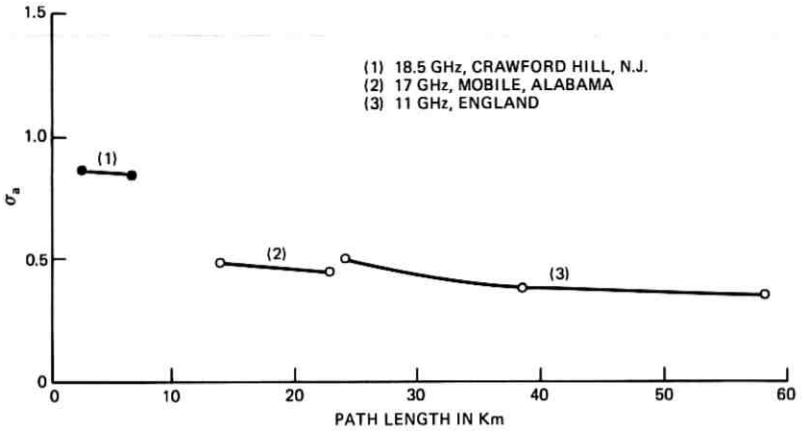

Fig. 10—Effect of path length on standard deviation σ_a.

The equation describing the lognormal distribution of $\tau(A)$ is

$$P[\tau(A) \geqq b] = \tfrac{1}{2} \operatorname{erfc}\left[\frac{\log_{10} b - \mu_\tau}{\sqrt{2}\sigma_\tau}\right] \tag{6}$$

where μ_τ and σ_τ are the mean and the standard deviation respectively of $\log_{10} \tau(A)$. The estimated values of σ_τ and the average fade duration $\bar{\tau}(A)$ are given in Table IV. Again, the unstable statistics, caused by insufficient time bases, limit the accuracy of these estimated values of σ_τ and $\bar{\tau}$.

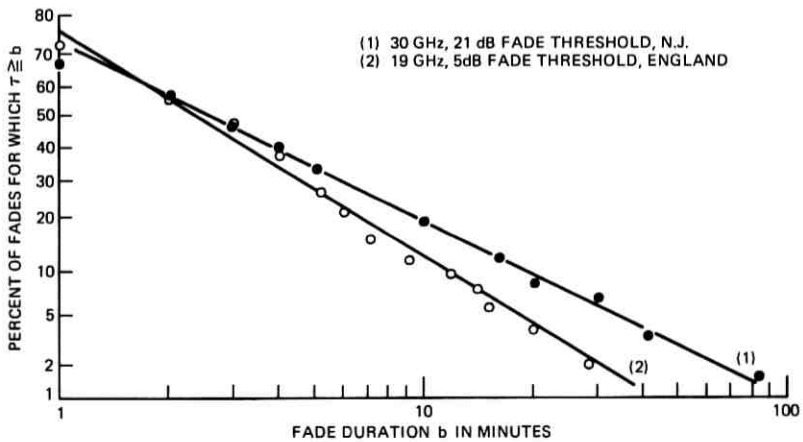

Fig. 11—Lognormal distributions of fade duration.

TABLE IV—EXPERIMENTS ON DISTRIBUTION OF FADE DURATION

Authors	Reference Number	Location	Frequency (GHz)	Path Length (km)	σ_τ	$\bar{\tau}(A)$ (min)	Fade Threshold (A dB)	Time Base
Barnett	44	Palmetto, Georgia	18	5	0.44	2.4	36	November 1970–June 1971
Wilson and Ruscio	4, 5, 17	Crawford Hill, N. J.	16	Earth-Space	0.76	8	9	Dec. 2, 1967–Feb. 28, 1969
Wilson and Ruscio	4, 5, 17	Crawford Hill, N. J.	30	Earth-Space	0.76	11.4	9	Dec. 2, 1967–Feb. 28, 1969
Wilson and Ruscio	4, 5, 17	Crawford Hill, N. J.	30	Earth-Space	0.67	7.7	21	Dec. 2, 1967–Feb. 28, 1969
Davies	31	England	19	Earth-Space	0.53	4.4	5	June 6, 1968–May 31, 1970
Semplak and Turin	18	Crawford Hill, N. J.	18.5	6.4	0.65	7.2	5	June 23, 1967–Oct. 31, 1967
Semplak and Turin	18	Crawford Hill, N. J.	18.5	6.4	0.57	5.8	10	June 23, 1967–Oct. 31, 1967
Semplak and Turin	18	Crawford Hill, N. J.	18.5	6.4	0.63	5.1	20	June 23, 1967–Oct. 31, 1967
Stracca	33	Italy	11	20	0.75	4.8	10	May–Oct 1967

V. THEORY

5.1 *Rain Attenuation Distribution*

Existing theory[19–28] and experimental data[1–19,29–33] indicate that rain attenuation is a complicated function of many parameters of the propagation medium: the total number of rain drops in the path, the drop size distribution, the fine grain spatial characteristics of the rain density along the path, wind velocity, the presence of up or down drafts, raindrop shape, raindrop cant angles, the storm cell shapes and sizes, raindrop temperature, etc. In other words, the attenuation $\alpha(t)$ is a function of many random time-varying parameters of the medium. Furthermore, through the coupling of the atmosphere, the above-mentioned parameters of the propagation medium depend on the present states and the past histories of the weather conditions at many near or faraway locations and altitudes.

We will assume that the rain attenuation can be affected by a large number of random time-varying multiplicative components:

$$\alpha(t) = S_1(t) \cdot S_2(t) \cdot S_3(t) \cdots S_n(t). \tag{7}$$

Each of $\{S_i(t)\}_{i=1}^{i=n}$ represents the random modification factor due to an environmental parameter.

Taking logarithms on both sides of eq. (7) yields

$$\log \alpha = \log S_1 + \log S_2 + \cdots + \log S_n \tag{8}$$

which shows that $\log \alpha$ is a summation of a large number of random variables. Then by the central limit theorem,[38] the distribution of $\log \alpha$ approaches a normal distribution for large n if there is no dominant component. Therefore, the distribution of α is approximately lognormal.*†

The basis for the multiplicative formulation (7) is that the environmental parameters affect the rain attenuation $\alpha(t)$ in a proportional fashion (i.e., in terms of percentage) rather than an additive fashion. For example, at 30 GHz frequency, the theoretical calculation shows that, when the rain temperature decreases from 20°C to 5°C, the rain attenuation increases by approximately 4 percent. This means the

* It is interesting to note that, in Fig. 12 of Ref. 10, the lognormal distribution appears to be a reasonable fit to the probability distribution of rainfall rate. The physical reason for this behavior probably is similar to that for rain attenuation α discussed in this section.

† A basic characteristic of a lognormal random variable y is that its value can vary in the entire semi-infinite range: $0 \leqq y < \infty$. For the rain attenuation problem, the attenuation α(dB) possesses this basic characteristic. On the other hand, the normalized signal amplitude $V(t)$ is strictly confined to the finite range: $0 \leqq V(t) \leqq 1$, which rules out the possibility of a lognormal distribution for $V(t)$.

variation of rain attenuation, due to 15°C variation of rain temperature, can be 4 dB, 0.4 dB, or 0.04 dB if the attenuation at 5°C is 100 dB, 10 dB, or 1 dB, respectively. Similar arguments apply to the effects of other parameters on the attenuation. A more general interpretation of formulation (7) is discussed in Appendix A.

We emphasize that some of the components $\{S_i(t)\}_{i=1}^{i=n}$ may be extremely slowly varying functions, which may take several months or even several years in order to show their effects. For example, the rainfall intensity-duration-frequency data of the Weather Bureau[32,39] show that in New Jersey the return period[40] for a rain rate exceeding 150 mm/h, which continues for a 5-minute duration, is about 5 years. Therefore, if the time base of a rain attenuation experiment is less than 5 years, the chance of missing these rare and extreme events is very high. The justification for considering such a long-term distribution of $\alpha(t)$ is that the microwave radio systems, which are designed based on these statistics, contains many repeaters, each sampling its own rain universe and contributing to total path outage.

5.2 *Fade Duration Distribution*

We also assume that the duration $\tau(A)$ of a rain attenuation fade with $\alpha \geqq A$ is affected by a large number of random time varying multiplicative components

$$\tau(A) = X_1 \cdot X_2 \cdot X_3 \cdots X_n. \tag{9}$$

Each of $\{X_i\}_{i=1}^{i=n}$ represents the random modification factor of an environmental parameter. Therefore, the long-term distribution of $\tau(A)$ is also approximately lognormal.

5.3 *Upper Bound for Fade Duration Distribution*

In radio system design, one is concerned with the occurrence probability of an unusually long continuous outage. It is desirable to have a "quick estimate" of the fade duration distribution, especially in the tail region of long duration. In our experimental and theoretical study[41–43] of the lognormal distributions of durations of rain and multipath fading, W. T. Barnett[41] has found an upper bound on the fade duration distribution as discussed in the following.

Let*

$$x = \frac{\tau}{\bar{\tau}}. \tag{10}$$

* The idea of normalizing fade duration τ to the average duration $\bar{\tau}$ originates from the work[48] of S. O. Rice on fade duration distributions.

The lognormal distribution of x can be written as

$$P(x \geqq X) = \tfrac{1}{2} \operatorname{erfc} \left[\frac{\log_{10} X - \mu}{\sqrt{2}\sigma} \right] \tag{11}$$

where μ and σ are the mean and the standard deviation, respectively, of $\log_{10} x$. For lognormally distributed x, it is easily shown that

$$\bar{x} = e^{\mu/M + \frac{1}{2}(\sigma^2/M^2)} \tag{12}$$

where $\bar{x}$ is the mean value of x, and

$$M = \log_{10} e \cong 0.434. \tag{13}$$

The definition (10) implies that

$$\bar{x} = \frac{\bar{\tau}}{\bar{\tau}} = 1. \tag{14}$$

Equations (12) and (14) show that

$$\mu = -\frac{1}{2}\frac{\sigma^2}{M}. \tag{15}$$

Substituting (15) into (11) yields

$$P(x \geqq X) = \tfrac{1}{2} \operatorname{erfc} \left[\frac{\log_{10} X + \sigma^2/2M}{\sqrt{2}\sigma} \right]. \tag{16}$$

Therefore, the lognormal distribution of x is completely determined by only one parameter σ.

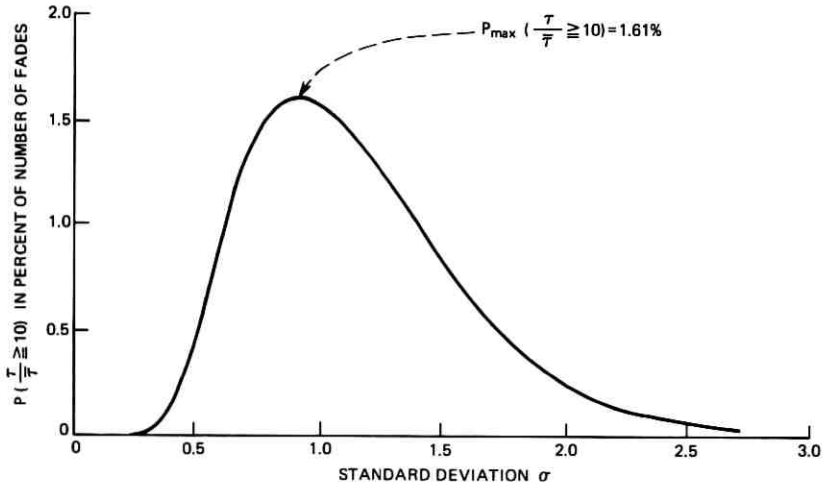

Fig. 12—Effect of standard deviation σ on the probability of long duration of fade.

By differentiating (16) with respect to σ, it is easily shown that

$$\left.\begin{aligned} \frac{dP}{d\sigma} &= 0 \\ \frac{d^2P}{d^2\sigma} &< 0 \end{aligned}\right\} \quad \text{at } \sigma = \sqrt{2M \log_{10} X} \tag{17}$$

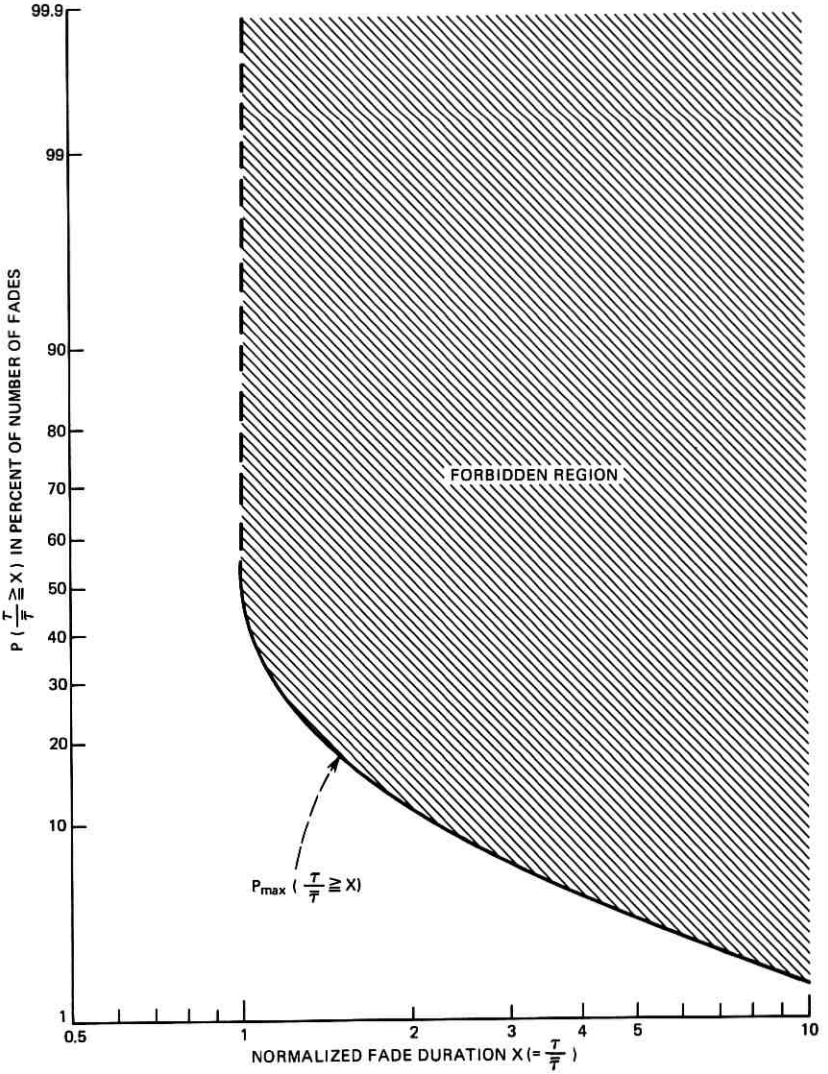

Fig. 13—Maximum probability for various normalized fade durations.

which implies that for any $X \geqq 1$ the probability $P(x \geqq X)$ as a function of σ has a maximum value:

$$P_{\max}(x \geqq X) = \tfrac{1}{2} \operatorname{erfc} \left[\sqrt{\frac{\log_{10} X}{M}} \right] \tag{18}$$

at

$$\sigma = \sqrt{2M \log_{10} X}. \tag{19}$$

For example, Fig. 12 shows $P(x \geqq 10)$ as a function of σ.

Figure 13 shows the maximum probability $P_{\max}(x \geqq X)$ as a function of X as given by eq. (18). By the use of "C-discriminant equa-

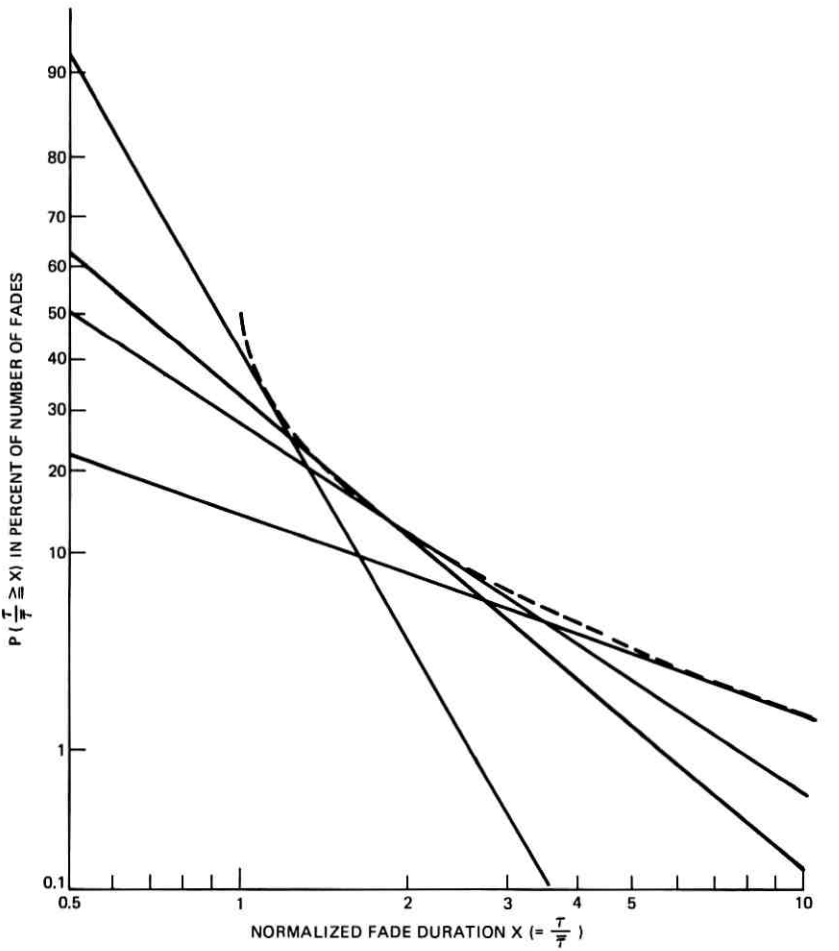

Fig. 14—The envelope of the family of lognormal duration distributions.

tion" in the theory of ordinary differential equation,[45] it can be shown that eq. (18) is, in fact, the *envelope* of the family of lognormal distributions (16) with σ as the family parameter. In other words, the family (16) are all tangent to (18) as shown in Fig. 14. The shaded area in Fig. 13 is a forbidden region where the lognormal duration distribution will never penetrate. Therefore, the upper bound for the fade duration distribution is given by eq. (2).

VI. CONCLUSION

Both the experimental data and theory indicate that both the rain attenuation distribution and the fade duration distribution are lognormal. The detailed results have already been given in Section III (Summary of Results).

VII. ACKNOWLEDGMENTS

The author wishes to thank W. T. Barnett, J. J. Kenny, D. C. Hogg, R. A. Semplak, and R. A. Desmond for providing the latest experimental data at Palmetto, Georgia; Merrimack Valley, Massachusetts; and Crawford Hill, New Jersey.

APPENDIX A

Generalized Interpretation of Formulation (7)

From a more general viewpoint, the effect of an environmental parameter $Z_i(t)$ on $\alpha(t)$ may be more complicated than the simple linear proportional relation

$$\alpha(t) \propto Z_i(t), \quad i = 1, 2, \cdots, n. \tag{20}$$

For example, $\alpha(t)$ may depend not only on the present value, but also on the past history of the environmental parameter $Z_i(t)$; then the relation between $\alpha(t)$ and $Z_i(t)$ becomes

$$\alpha(t) \propto \int_{-\infty}^{t} H_i(t, t')Z_i(t')dt' \qquad i = 1, 2, \cdots, n \tag{21}$$

where $H_i(t, t')$ is the impulse response of $\alpha(t)$ if the input $Z_i(t)$ is an impulse $\delta(t - t')$ applied at t'. In order to allow for the more general

and complicated relation between $\alpha(t)$ and $Z_i(t)$, we shall use the mathematical operator notation:

$$\alpha(t) \propto G_i[Z_i(t)], \quad i = 1, 2, 3, \cdots, n \tag{22}$$

where G_i is a transformation from $Z_i(t)$ into $\alpha(t)$. Then the formulation becomes

$$\alpha(t) = G_1(Z_1) \cdot G_2(Z_2) \cdots G_n(Z_n). \tag{23}$$

Let

$$S_i(t) = G_i[Z_i(t)] \quad i = 1, 2, 3, \cdots, n. \tag{24}$$

Substituting (24) into (23) leads to (7).

The formulation can be further generalized to include the cases where the effects of various environmental parameters on α are not completely separable. Let

$$\gamma = \log_{10} \alpha, \tag{25}$$

$$c_i \xi_i = \log_{10} \quad S_i = \log_{10} G_i(Z_i)$$
$$i = 1, 2, 3, \cdots, n - 1, \text{ and} \tag{26}$$

$$R = \log_{10} S_n. \tag{27}$$

Substituting (25), (26), and (27) into (8) gives

$$\gamma(\xi_1, \xi_2, \cdots, \xi_{n-1}) = \bar{\gamma} + C_1(\xi_1 - \bar{\xi}_1) + C_2(\xi_2 - \bar{\xi}_2) + \cdots + C_{n-1}(\xi_{n-1} - \bar{\xi}_{n-1}) + (R - \bar{R}) \tag{28}$$

where

$$\bar{\gamma} = C_1\bar{\xi}_1 + C_2\bar{\xi}_2 + \cdots + C_{n-1}\bar{\xi}_{n-1} + \bar{R}. \tag{29}$$

Equation (28) can be interpreted as the first-order Taylor series expansion of $\gamma(\xi_1, \xi_2, \cdots, \xi_{n-1})$ with a remainder term,* $R - \bar{R}$. The main point in eq. (28) is that the effects of the $(n - 1)$ random variables $\{\xi_i\}_{i=1}^{i=n-1}$ on γ do not have to be completely separable because of the remainder term $R - \bar{R}$. If $R - \bar{R}$ does not dominate the sum in the right-hand side of eq. (28), then, by central limit theorem, the distribution of γ (i.e., $\log_{10} \alpha$) is approximately normal even if the effects of various environmental parameters are not completely separable. (This argument originates from Refs. 46 and 47.) Therefore, the formulation (7) includes very general and complicated relationships between the environmental parameters and rain attenuation $\alpha(t)$.

* If higher-order derivatives of $\gamma(\xi_1, \xi_2, \cdots, \xi_{n-1})$ exist, then $R - \bar{R}$ represents the sum of all the higher-order terms; otherwise, $R - \bar{R}$ represents the difference between $\gamma(\xi_1, \xi_2, \cdots, \xi_{n-1})$ and its first-order Taylor series expansion.

APPENDIX B

Deviations of Experimental Data From Lognormal Distribution

Aside from the experimental error, there are two major factors which contribute to the deviations of experimental data from the lognormal distribution.

B.1 *Effect of Finite Number of Components*

Davenport and Root[38] have indicated that when the number, n, of components is finite, the normal distribution may well give a *poor* approximation to the *tails* of the distribution of the sum (8) even though the limiting form of the sum distribution is, in fact, normal. Therefore, in practice, we believe that the deviation of the experimental data from the lognormal distribution may increase toward the tails.

B.2 *Effect of Time Base*

Since some of the components in eq. (7) are extremely slowly varying, reducing the time base will reduce the number of contributing components since the influences of slow components are approximately constant in a short experiment. Therefore, we expect the deviation of the experimental data from the lognormal distribution to increase as the time base decreases.

The minimum required time base for the convergence of the experimental data to the lognormal distribution increases as

(*i*) path length decreases, or
(*ii*) operating frequency decreases, or
(*iii*) attenuation range of interest increases.

Furthermore, when the time base of a rain attenuation experiment is not long enough, the deviations of the short-term distribution $P_s(\alpha \geqq A)$ from the long-term distribution $P(\alpha \geqq A)$ are usually nonsymmetric in the deep fade region, i.e., $P_s(\alpha \geqq A)$ is more likely to be less than $P(\alpha \geqq A)$. This nonsymmetry is caused by:

(*i*) The distribution of fade duration $\tau(A)$ is lognormal, which is nonsymmetric with respect to the average fade duration $\bar{\tau}(A)$. Typically, about 70 percent of fade durations are shorter than the average duration.
(*ii*) The probability distribution of the number $N(A, T)$ of observed deep fades, exceeding the margin A dB in the period T, is somewhat similar to a Poisson distribution which is also

nonsymmetric with respect to the average number $\bar{N}(A, T)$ of deep fades in a period T.

Since, when the time base is too short, both $N(A, T)$ and $\tau(A)$ have a higher chance of being less than their average values $\bar{N}(A, T)$ and $\bar{\tau}(A)$, respectively, then the short-term distribution $P_s(\alpha \geq A)$ tends to deviate downward from the long-term distribution in the tail region.*

B.3 *Effect of Time Base on Standard Deviation* σ_a

Since it takes a long time base to include appreciable effects of slow components, we expect that σ_a increases slightly as the time base increases, and reaches an asymptotic value only after the time base is long enough to include the effects of all the possible slow components.

REFERENCES

1. Semplak, R. A., "The Influence of Heavy Rainfall on Attenuation at 18.5 GHz and 30.9 GHz," IEEE Trans. Ant. Prop., *AP-18*, No. 4 (July 1970).
2. Semplak, R. A., "Dual Frequency Measurements of Rain-Induced Microwave Attenuation on a 2.6-Kilometer Propagation Path," B.S.T.J., *50*, No. 8 (October 1971), pp. 2599–2606.
3. Gray, D. A., "Measurement of Attenuation by Rain Storms Using a 100 GHz FM Radar," 1971 Spring USNC/URSI Meeting in Washington D. C.
4. Wilson, R. W., "Sun Tracker Measurements of Attenuation by Rain at 16 and 30 GHz," B.S.T.J., *48*, No. 5 (May-June 1969), pp. 1383–1404.
5. Ruscio, J. T., "Attenuation Statistics of an Earth-Space Path at 16 and 30 GHz Using the Crawford Hill Sun Tracker," unpublished work, May 21, 1969.
6. Blevis, B. C., Dohoo, R. M., and McCormick, K. S., "Measurements of Rainfall Attenuation at 8 and 15 GHz," IEEE Trans. Ant. Prop., *AP-15*, No. 3 (May 1967), pp. 394–403.
7. Hathaway, S. D., and Evans, H. W., "Radio Attenuation at 11 kmc and Some Implications Affecting Relay System Engineering," B.S.T.J., *38*, No. 1 (January 1959), pp. 73–97.
8. Zimmerman, B. D., "Radio Relay Systems—Final Report of Test Results of Rainfall Attenuation at 17 Kmc," unpublished work, April 8, 1960.
9. Hickin, E. M., "Atmospheric Attenuation of Radio Signals Above 10 GHz," GEC-AEI J., *35*, No. 3, 1968, pp. 133–136 (GEC-AEI Telecommunication, Ltd., Coventry).
10. Freeny, A. E., and Gabbe, J. D., "A Statistical Description of Intense Rainfall," B.S.T.J., *48*, No. 6 (July-August 1969), pp. 1789–1851.
11. Funakawa, K., Kido, T., Kitamura, K., Otsu, Y., Katto, T., and Uratsuka, M., "Propagational Experiments of 13 GHz and 35 GHz over the Path 80 Km, Tsukuba to Kokubunji," J. Radio Res. Lab. (Japan), *14*, No. 76 (November 1967), pp. 249–265.
12. Turner, D. J. W., and Turner, D., "Attenuation Due to Rainfall On a 24-Km Microwave Link Working at 11, 18, and 36 GHz," Elec. Ltrs. *6*, No. 10, May 14, 1970, pp. 297– 298.

* This idea originates from the work of A. Vigants on the deviations of experimental data from the theoretical distribution for diversity signals subject to multipath interference fading.

13. Easterbrock, B. J., and Turner, D., "Prediction of Attenuation by Rainfall in the 10.7 to 11.7 GHz Communication Band," Proc. IEE, *114*, No. 5, May 1967, pp. 557–565.
14. Turner, D., Easterbrock, B. J., and Golding, J. E., "Experimental Investigation Into Radio Propagation of 11.0 to 11.5 Gc/s," Proc. IEE, *113*, No. 9 (September 1966), pp. 1477–1489.
15. Wilson, R. W., "A Three-Radiometer Path-Diversity Experiment," B.S.T.J., *49*, No. 6 (July-August 1970), pp. 1239–1241.
16. Kenny, J. J., private communication.
17. Wilson, R. W., "Effects of Elimination of Low Elevation Angles From 1968 Sun Tracker Data," unpublished work, December 1, 1968.
18. Semplak, R. A., and Turin, R. H., "Some Measurements of Attenuation by Rainfall at 18.5 GHz," B.S.T.J., *48*, No. 6 (July-August 1969), pp. 1767–1787.
19. Medhurst, R. G., "Rainfall Attenuation of Centimeter Waves: Comparison of Theory and Measurements," IEEE Trans. Ant. Prop. (July 1965), pp. 550–564.
20. Setzer, D. E., "Computed Transmission Through Rain at Microwave and Visible Frequencies," B.S.T.J., *49*, No. 8 (October 1970), pp. 1873–1892.
21. Kerr, D. E., *Propagation of Short Radio Waves*, MIT Radiation Lab. Series, New York: McGraw-Hill Book Company, 1951.
22. Hogg, D. C., "Millimeter-Wave Communication Through the Atmosphere," Science, *159*, No. 3810, January 5, 1968, pp. 39–46.
23. Hogg, D. C., "Statistics on Attenuation of Microwaves by Intense Rain," B.S.T.J., *48*, No. 9 (November 1969), pp. 2949–2962.
24. Hogg, D. C., "How the Rain Falls," unpublished work, October 3, 1968.
25. Ruthroff, C. L., "Rain Attenuation and Radio Path Design," B.S.T.J., *49*, No. 1, (January 1970), pp. 121–135.
26. Oguchi, T., "Attenuation of Electromagnetic Wave Due to Rain With Distorted Raindrops," J. Radio Res. Lab. (Japan), *7*, No. 33 (September 1960), pp. 467–485; and *11*, No. 53 (January 1964), pp. 19–44.
27. Oguchi, T., "Scattering and Absorption of a Millimeter Wave Due to Rain and Melting Hailstones," J. Radio Res. Lab. (Japan), *13*, No. 67 (May 1966), pp. 141–172.
28. Okamura, S., et al., "Effect of Polarization on the Attenuation by Rain at Millimeter Wave Length," J. Radio Res. Lab. (Japan), *8*, No. 36 (March 1961), pp. 73–80.
29. Semplak, R. A., "Effect of Oblate Raindrops on Attenuation at 30.9 GHz," Radio Science, *5*, No. 3 (March 1970), pp. 559–564.
30. Delange, O. E., and Dietrich, A. F., "The Effects of Rain Upon the Polarization of 60 GHz Waves," unpublished work, July 16, 1971.
31. Davies, P. G., "Radiometer Studies of Atmospheric Attenuation of Solar Emission at 19 GHz," Proc. IEE (London), *118*, No. 6 (June 1971), pp. 737–741.
32. Weibel, G. E., and Dressel, H. O., "Propagation Studies in Millimeter-Wave Link Systems," Proc. IEEE, *55*, No. 4 (April 1967), pp. 497–513.
33. Stracca, G. B., "Propagation Tests of 11 GHz and 18 GHz on Two Paths of Different Length," Alta Frequenza (Italy), *38*, 1969, pp. 345–360.
34. Funakawa, K., and Kato, J., "Experimental Studies of Propagational Characters of 8.6 mm Wave on the 24 km Path," J. Radio Res. Lab. (Japan), *9*, No. 45 (September 1962), pp. 351–367.
35. Bussey, H. E., "Microwave Attenuation Statistics Estimated From Rainfall and Water Vapor Statistics," Proc. I.R.E. (July 1950), pp. 781–785.
36. McConalogue, D. J., "On the Distribution of Rainfall Rate at some Coastal Stations in the British Isles," Meteorol. Mag., *82* (October 1953), pp. 304–306.
37. Mueller, E. A., and Sims, A. L., "Investigation of the Quantitative Determination of Point and Areal Precipitation by Radar Echo Measurements," Technical Rep. ECOM–00032–F, Illinois State Water Survey, Dec. 1966.
38. Davenport, W. B., and Root, W. L., *Random Signals and Noise*, New York: McGraw-Hill Book Company, 1958.
39. "Rainfall Intensity-Duration-Frequency Curves for Selected Stations in the United States, Alaska, Hawaiian Islands, and Puerto Rico," Weather Bureau, U. S. Department of Commerce, Tech. Paper 25, December 1955.
40. Bruce, J. P., and Clark, R. H., *Introduction to Hydrometeorology*, New York: Pergamon Press, 1966, p. 148.

41. Barnett, W. T., "On the Lognormal Duration Distribution of Signal Fluctuations," unpublished work, April 17, 1972.
42. Lin, S. H., "Lognormal Distribution of Durations of Rain and Multipath Fades," unpublished work, August 5, 1971.
43. Vigants, A., "Number and Duration of Fades at 6 and 4 GHz," B.S.T.J., *50*, No. 3 (March 1971), pp. 815–841.
44. Barnett, W. T., and Bergmann, H. J., private communication.
45. Rainville, E. D., and Bedient, P. E., *Elementary Differential Equations*, New York: The MacMillan Company, 1969, p. 258.
46. Cramer, H., *Mathematical Methods of Statistics*, Princeton: Princeton University Press, 1957, Section 17.5, pp. 218–219.
47. Misès, R., "Les lois de probabilité pour les functions statistiques," Annales De L'Institute Henri Poincaré, *6*, 1936, pp. 185–212.
48. Rice, S. O., "Distribution of the Duration of Fades in Radio Transmission," B.S.T.J., *37* (May 1958), pp. 581–635.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 4, April, 1973
Printed in U.S.A.

Simple, Low-Loss Joints Between Single-Mode Optical Fibers

By C. G. SOMEDA*

(Manuscript received September 21, 1972)

Low-loss joints between single-mode optical fibers have been made without microscopic alignment, without fusing the tips, and without monitoring the transmitted power while the joints are assembled. The fibers are tightly held in an embossed groove; an index-matching liquid is added. Average power coupling efficiencies close to 90 percent in the red and to 85 percent in the infrared have been obtained. Mediocre end faces are acceptable. Realistic discrepancies between the fiber cladding diameters (slightly in excess of twice the core diameter) do not deteriorate the results.

I. INTRODUCTION

Recently, several authors have dealt with techniques for making low-loss joints between multimode[1–4] or single-mode[5] optical fibers. For long-distance communication channels, joints between fiber cables, simply assembled in the field, would be very valuable.

The joints that were made so far with the technique described in this paper connected just one pair of fibers at a time. However, this is intended to be the first step in the development of joints between multifiber cables. Therefore, in the present experiments emphasis has been put on getting repetitive results which are not a consequence of careful laboratory adjustment. This is an important difference with respect to the methods that gave the best results reported so far.[1,5] Nonetheless, the joints between single-mode fibers described here compare favorably with previous ones as far as the losses are concerned. The demonstration with single-mode fibers (core diameters of $\approx 4\ \mu m$) means that the technique applies also to multimode fibers, the larger core diameters of which should make the alignment less critical.

* This work was performed at Bell Telephone Laboratories, Incorporated, Crawford Hill Laboratory, Holmdel, New Jersey, during a six-month internship sponsored by a NATO–CNR fellowship.

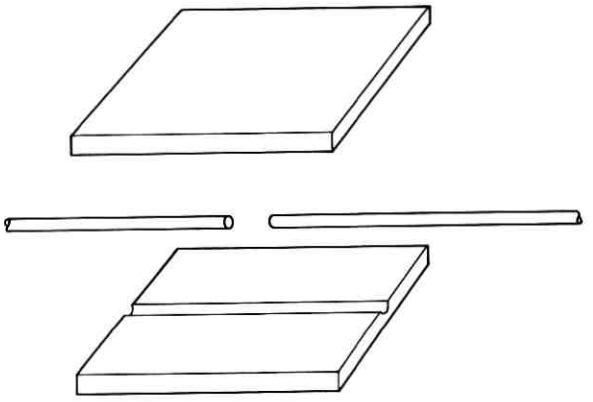

Fig. 1—View of the unassembled parts of a sandwich joint.

A complete description of the technique follows in Section II. As a rough sketch, we may say that the tips of the fibers are aligned in a groove (Fig. 1), embossed in a Plexiglas sheet.* The fiber tips are pushed against each other. An index-matching liquid is added. A "sandwich" is formed with a flat Plexiglas sample, and then squeezed by means of a small vise (Fig. 2). No fusion takes place and no microscopic alignment is needed.

It is obvious that the fiber claddings are aligned in this way and therefore the results depend on the fiber core being coaxial with the cladding. The experiments reported here were performed with single-mode fibers† having a core diameter of 3.7 μm and a cladding diameter of 254 μm; coincidence of the core and cladding axes within ≈ 1 μm had been observed preliminarily.

Experiments were performed at 6328 Å and at ≈ 9000 Å. Repetitive measurements gave average power couplings around 87 percent in the red and around 83 percent in the infrared; best results were 93 percent and 87 percent, respectively. These figures were obtained without paying any particular attention to the quality of the end surfaces of the two fibers.

The experiments were completed by testing the effects of a difference, d, between the cladding diameters, as described in Section IV. It turned out that sandwiches of good mechanical quality exhibit no deterioration of the results for $d \cong 10$ μm (i.e., more than twice the core diameter). For $d \cong 30$ μm, the results were deteriorated, but still showed

* Plexiglas (methyl methacrylate), registered trademark of Rohm and Haas.
† These fibers were manufactured by Corning Glass Works, Corning, N. Y.

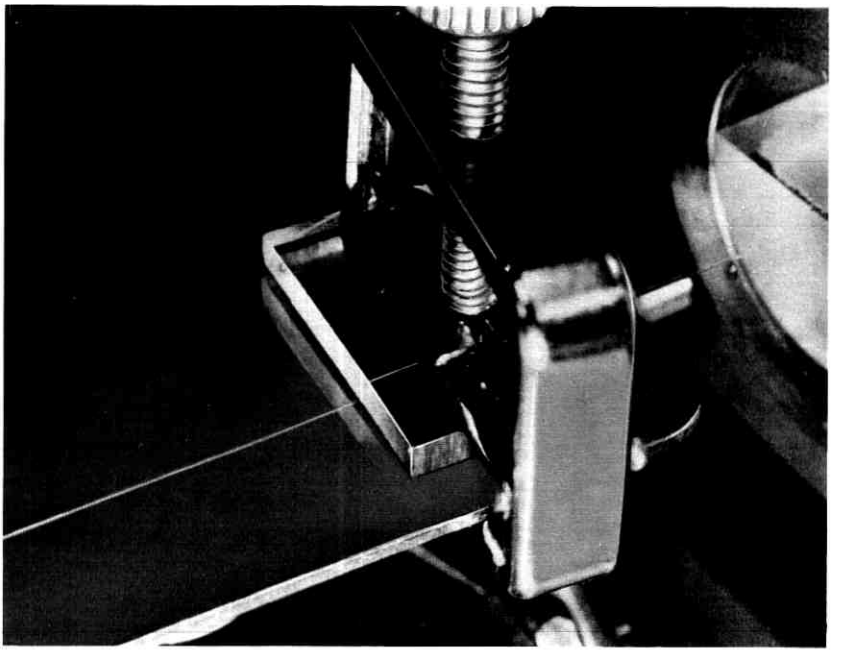

Fig. 2—View of an assembled joint. The length of the visible side of the sandwich, perpendicular to the fiber, is ≈2 cm.

the presence of an alignment mechanism, which is believed to be due to surface tension in the index-matching liquid.

II. DESCRIPTION OF THE JOINING TECHNIQUE

In order to obtain a joint of the type shown in Fig. 1, the first operation to perform is to emboss a Plexiglas sheet with a groove that fits the dimensions of the fibers to be joined.

A simple and economical technique for embossing grooves in a thermoplastic substrate by means of a glass fiber has been reported by Ulrich *et al.*[6] A further simplification of this procedure proved successful; there is no need to heat the substrate. A rectangular Plexiglas sample and a fiber piece, the length of which slightly exceeds the Plexiglas size, are put between two milled aluminum blocks (flat within ±1 mil) and then tightly pressed by means of a vise for a few minutes. When the vise is released, the fiber separates from the Plexiglas, leaving a sharply embossed groove. Any other piece of a nominally equal fiber can be introduced into the groove and fits it very closely. Figure 3

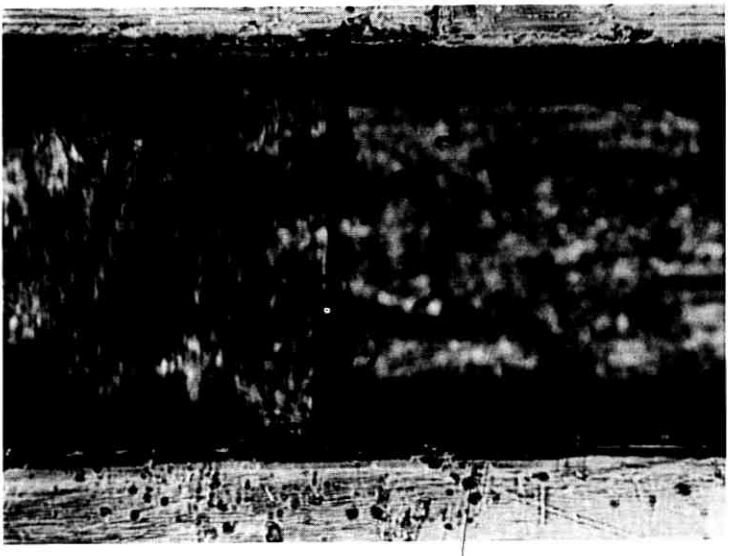

Fig. 3—Microscopic picture of a single-mode fiber tip (on the left) lying in an embossed alignment groove. Magnification 200× (1 cm in the picture is 50 μm in actual scale).

is a microscopic picture (magnification 200×) of a fiber tip lying in a groove, embossed with another fiber. Figure 4 is a further enlarged view (magnification 500×) of the groove edge in the vicinity of the fiber end; it shows no defect comparable to the size of the fiber cores to be aligned. It also shows a fairly regular pattern of longitudinal wrinkles due to the compression of the Plexiglas sample.

First, Plexiglas having a nominal thickness of $\frac{1}{16}$ inch was embossed with quartz fibers, 254 μm in cladding diameter. The samples were permanently curved with a remarkable convexity of the embossed side. The cover (Fig. 1) then tended to flip. Using a thicker Plexiglas sample (e.g., $t = \frac{1}{8}$ inch), the curvature was negligible.

There is some evidence that the deformation of the embossed Plexiglas sheets is elastic (at least partially) with a long time constant. Therefore, the joints have to be made with newly embossed samples, or with samples where a fiber has been constantly pressed in.

To assemble a joint, the fibers are put in the groove, a small amount of a suitable index-matching oil is added, and the cover is placed on top. At first, the top is not pressed against the bottom. Next, one fiber is pushed against the other in the axial direction; finally, the sandwich is squeezed with a small vise.

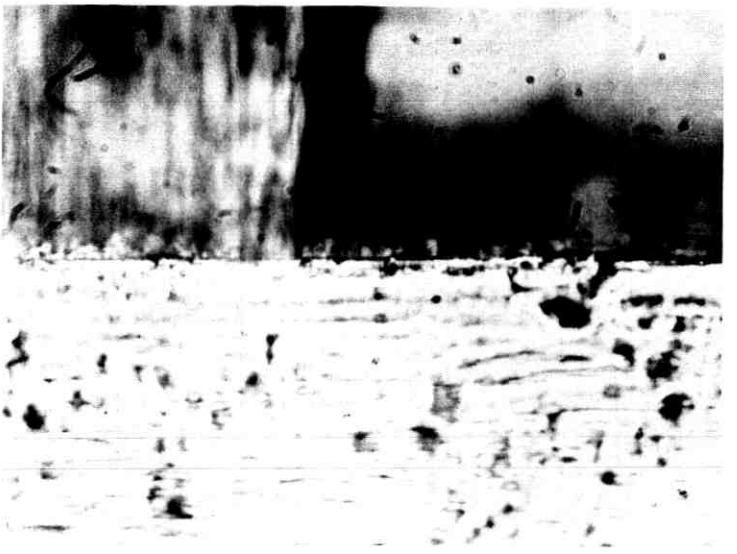

Fig. 4—Microscopic picture of the edge of a groove, with a fiber tip on the left. Magnification 500× (1 cm in the picture is 20 μm in actual scale).

It seems that the index-matching oil is rather slow in wetting the fiber tips, and some time is required for air bubbles to escape. The oil also has the purpose of lubricating the fiber longitudinal motion. Sometimes an excess of it deteriorated the joint performances, probably because an unwanted axial movement of the fiber tips was produced by the oil flux when the sandwich was squeezed. However, the presence of a small amount of liquid is needed in order to get good coupling efficiency (see Section 3.2). Furthermore, the presence of the fluid is believed to be responsible for the excellent behavior in case of unequal diameters of the two fibers (Section IV).

The amount of light transmitted by the joint does not always increase with increasing vise pressure. Often, though, a decrease in the coupling beyond a certain pressure was interpreted as due to some misalignment caused by careless movements accompanying the vise tightening. When the groove was deep enough and the top could not swing or slide, then such a decrease in transmitted power was absent or negligible. A good rule is to press the joint between surfaces that are not too stiff, so that the force is distributed on the whole area of the sandwich.

III. MEASUREMENTS

3.1 *Tests at 6328 Å*

Sandwich joints of the type described in Section II were tested first at the wavelength $\lambda = 6328$ Å. Based on manufacturer's data, the corresponding normalized frequency,[7] $V \cong 2.2$, means that single-mode propagation takes place. Direct observation of the far field at the end of a fiber confirmed that.

The set-up is illustrated in Fig. 5. The light emitted by a commercial He-Ne laser was chopped and then launched into the input fiber by means of a microscope objective (magnification 40×). A curved part of the fiber was lying in an index-matching liquid, to strip off the light launched into the cladding. The length of this mode-stripping section was such that any further addition or small subtraction did not affect the power level at the output of the fiber. This output was detected by means of an Si solar cell and the signal was sent to a lock-in amplifier. Figure 5 shows both arrangements that are needed in order to evaluate the insertion loss of a joint: a "reference" arrangement, where the detector is directly connected to the input fiber, and a "measurement" arrangement, where a joint and an output fiber are added.

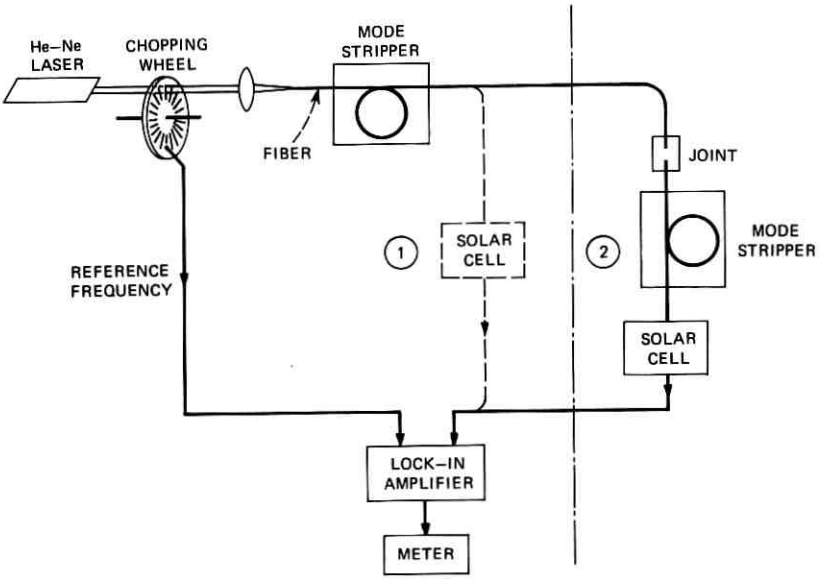

Fig. 5—Block diagram of the set-up used for the measurements at 6328 Å. The broken lines (Part 1) refer to calibration (measurement of a reference level). The solid lines (Part 2) refer to the actual measurement of the light transmitted through a joint.

TABLE I—POWER COUPLING EFFICIENCIES MEASURED AT 6328 Å AND CORRESPONDING INSERTION LOSSES OF THE JOINTS

Sample Number	1	2	3
Best results	92% (0.35 dB)	93% (0.3 dB)	92% (0.35 dB)
Worst results	74% (1.3 dB)	80% (1.0 dB)	81% (0.9 dB)
Average results	80.7% (0.9 dB)	88.7% (0.5 dB)	86.6% (0.6 dB)
Number of measurements	12	6	6
Number of different ends	4	2	2

Attenuations measured in this way result from both the joint insertion loss and the losses in the output fiber. Previous data[8] and direct observation show that the output fiber attenuation is negligible with respect to the loss in the joint, despite the need for a length of the output fiber that could insure an effective stripping of all the light transferred by the joint into the cladding.

This way to perform the measurements was preferred to that where, after taking a reference, the fiber is cut and then joined again, because frequent checks of the reference were needed. Indeed, launching into a single-mode fiber is so critical (the power drops by 3 dB if the fiber tip is misaligned by a few microns), that accidental causes can produce strong shifts in the reference level. For example, in our case the material used to bind the fiber tip on an x-y-z manipulator could deform slowly, under the strength applied by the curved and quite elastic fiber. This effect was compensated for by optimizing the alignment of the fiber with respect to the incoming beam before any meter reading, both in the reference and in the measurement arrangements. Discrepancies between optimized reference levels monitored before and after a measurement were never larger than 3 percent and usually much smaller than that. When the reference levels monitored before and after a measurement were different, the more pessimistic estimate of the joint losses was taken.

Table I contains the best, worst, and average results of three sequences of measurements. As for the "worst" results, they exclude only those instances where clearly identified man-made mistakes took place; i.e., axially separated fibers, one fiber lying out of the groove, and Plexiglas debris shaved by the fibers and accumulated between the two

ends. These errors were very rare, and were easily detected because of a large amount of scattered light shining from the sandwich. This point will be discussed further in Section 3.2.

The sequences of Table I refer to assembling, several times, joints that make use of three embossed grooves, one in a $\frac{1}{16}$-inch-thick plate and two in $\frac{1}{8}$-inch-thick plates. All the measurements were independent, in the sense that after any of them the joint was at least disassembled completely and then reassembled. The ends of the fibers were changed often, as shown in Table I. The amount of index-matching oil was changed quite often too.

An important point is that none of the measurements were of a joint where the two coupled end faces resulted from one cut of a fiber; they always resulted from two independent breaks, in order to simulate real situations.

A remarkable advantage of this technique is that there is no need for very accurate flatness of the fiber end faces. As long as there are no lips, which would prevent the two cores from getting close to one another, results are good. Figure 6 is a microscopic picture of one typical pair of ends used in the measurements. Attempts to get end surfaces of this

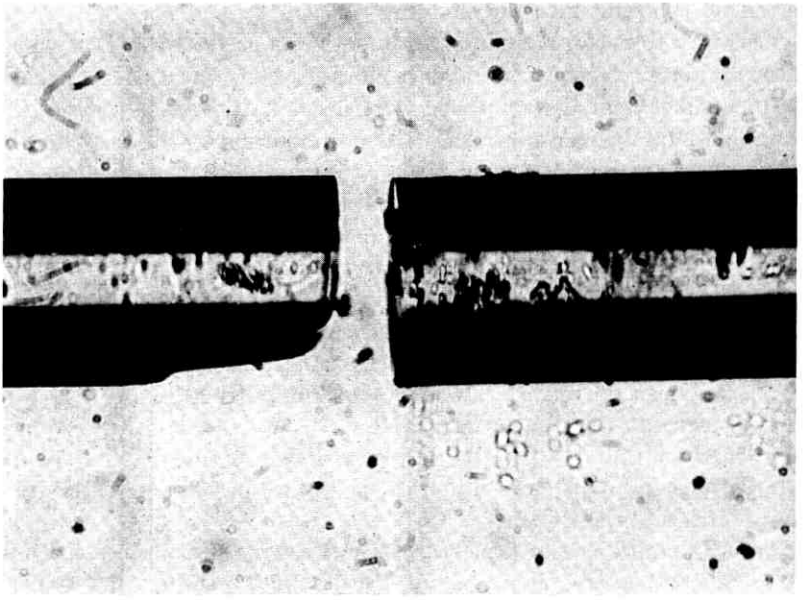

Fig. 6—Microscopic picture of a typical pair of fiber ends used in the sequence of experiments summarized in Table I. The bright zone in the center is not the fiber core, but just a consequence of the illumination.

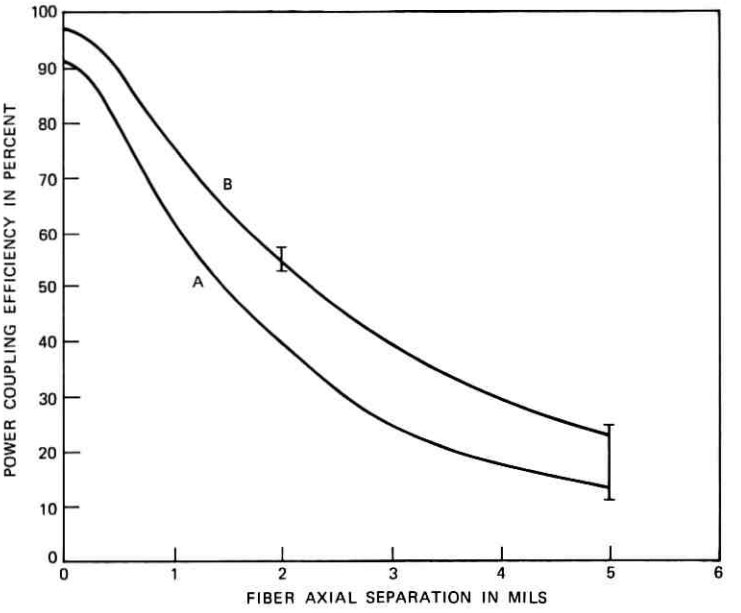

Fig. 7—Effects of the longitudinal separation between the fibers on the coupled power. Lines A and B, from Ref. 2, represent measured values (without and with index-matching oil, respectively) at 6328 Å when the fibers were aligned by means of micromanipulators. The vertical segments represent the range of results obtained when the fibers were held in a sandwich joint.

quality were almost always successful if the fibers were pulled after scoring them with a lathe tool or with an electrical discharge generated by a Tesla coil. Many successful attempts simply consisted of clamping the fibers in a Plexiglas sandwich and then pulling them. Consequences for the multifiber case will be stated in Section V.

Another interesting observation was that when a joint was tightened, bringing the fiber cores into alignment, the forward scattered light on the side surface of the output fiber decreased very remarkably.

Some additional measurements were performed in order to establish the effects of longitudinal separation between the fibers. The results are compared with those obtained by Bisbee[2] in Fig. 7. The very large spread of values observed for a 5-mil separation can be justified by an occasional shortage of index-matching oil, which could cause an air gap sometimes to show up when the fibers were separated; or by deviations of the groove from a rectilinear shape, which would cause unwanted transverse displacements of the fiber tips to accompany the longitudinal ones. The best results ($\eta \cong 56$ percent for a 2-mil separation,

TABLE II—POWER COUPLING EFFICIENCIES MEASURED AT 9000 Å AND CORRESPONDING INSERTION LOSSES OF THE JOINTS

Sample Number	2	3
Best results	87.5% (0.55 dB)	87% (0.6 dB)
Worst results	78.5% (1.05 dB)	80% (1.0 dB)
Average results	82.5% (0.85 dB)	83.7% (0.75 dB)
Number of measurements	11	13
Number of different ends	2	2

$\eta \cong 27$ percent for a 5-mil separation), which are in very good agreement with Bisbee's,[2] were observed more frequently than the bad ones.

In multifiber joints, it is unlikely that the axial separation of each pair of fibers will be as small as it was for the single pairs tested so far. However, one can think of enlarging the waist of the light beam in the fibers, by using a smaller normalized frequency[7]; the confocal length of the beam would then grow as the square of the beam waist. The effect of an axial separation upon the coupling efficiency would be smaller than that shown in Fig. 7.

3.2 *Tests at 9000 Å*

It was pointed out before that, while working with visible radiation, a misalignment or a gap between the fibers was revealed by a large amount of scattered light leaking from the joint. In a multifiber joint, this light will not identify the pair of fibers that form a leaky connection. Hence, it was necessary to check whether the very small number of mistakes in assembling the joints and the high average of their performances were independent of the information that the operator was provided by the scattered light.

A sequence of tests in the near infrared, at $\lambda \cong 9000$ Å (where $V \cong 1.55$), was then performed using a small-area GaAs LED.* The set-up differs from Fig. 5 only because the chopping of the signal is performed by an audio oscillator driving the LED. Launching spatially incoherent radiation into the fiber is much less critical than in the case of the laser; this results in a much smaller long-term drift of the reference level.

* The LED was built and provided by C. A. Burrus.[9]

Tests were performed on joints that made use of the grooved samples referred to in columns 2 and 3 of Table I. Quite repetitive results were obtained. They are summarized in Table II. Comparison with Table I shows a very slight deterioration of the best results (power coupling coefficients from 92–93 percent down to 87–87.5 percent) but even smaller changes of the averages (from 80–81 percent down to 78.5–80 percent) and almost no change in the lower ends of the range. Let us re-emphasize that the tested joints were not assembled while monitoring the transmitted power.

Comparison of both Tables I and II with previously published data on single-mode fiber joints[5] shows the desirability of the present sandwich technique. Best results given by the fusion technique[5] show power transmission efficiencies of ≈ 80 percent, at a normalized frequency $V \cong 2.0$; the average results given by the sandwich technique are better, both above and below that value of V.

The number of times when the attempts to assemble the joints were unsuccessful because of easily identified mistakes in the fiber position increased to three in a sequence of twenty-seven measurements, compared to one in twenty-five measurements at 6328 Å. This value is still low enough to insure that the technique is suitable for further developments in a multifiber system, where the risk of making such mistakes has to be negligible.

Some information on the importance of the index-matching oil has been collected, too: "dry" joints gave best power coupling efficiencies in the order of 50 percent. Comparison with available data[2] shows that the oil produces some other effect besides the elimination of end reflections. Quite likely, it compensates for irregularities of the end faces; also, it provides some alignment mechanism via surface tension. The last point is discussed further in the next section.

IV. EFFECTS OF DISCREPANCIES BETWEEN THE FIBER OUTER DIAMETERS

The usefulness of the sandwich technique would be limited if its performances were sensitive to discrepancies between the diameters of the two joined fibers, a situation that will occur in practice. The results reported in this section show that the sensitivity to this kind of imperfection is small enough to satisfy practical purposes.

A controlled difference, d, between the outer diameters of the fibers was introduced by etching one fiber with commercial hydrofluoric acid. First, a value $d \cong 10$ μm was chosen, because it would be sufficient to allow more than a complete offset of the two cores if the claddings were

TABLE III—POWER COUPLING EFFICIENCIES OF JOINTS BETWEEN FIBERS WITH A 10-μm DIFFERENCE IN CLADDING DIAMETERS. MEASUREMENTS PERFORMED AT 9000 Å.

Best result	86% (0.65 dB)
Worst result	82% (0.85 dB)
Average result	84.7% (0.7 dB)
Number of measurements	7
Number of different ends	2

aligned along one generatrix; besides which, 10 μm is 4 percent of the cladding diameter, which can be thought of as a realistic tolerance.

A new sequence of tests was performed at $\lambda \cong 9000$ Å, using the embossed sample referring to column 2 in Tables I and II. The results are summarized in Table III. Comparison with Table II shows that there is no appreciable deterioration of the joint performances. On the other hand, it was observed that the transmission was sometimes (not always) sensitive to small changes in the force applied by the vise; e.g., for one of the best joints assembled without monitoring the transmitted power ($\eta \cong 86$ percent), small changes in the tightening could cause a drop of the transmitted power to 82 percent or a raise to 89 percent (insertion loss ≈ 0.5 dB).

In all cases, releasing the vise slowly caused, first, a decrease in transmitted power; then a maximum, close to the figures quoted before, showed up; later, the signal decreased monotonically down to the noise level. The secondary maximum, which was very sensitive to vibrations and shocks, is believed to be entirely due to an alignment caused by the surface tension of the index-matching oil. The same phenomenon is believed to contribute to the excellent alignment under the strongly tightened vise, leading to the figures reported before.

Later, the difference between the cladding diameters was brought to a value $d \cong 30$ μm, i.e., about 12 percent of the outer diameter. The following behavior was observed:

(*i*) There was practically no coupling as long as the joints were not tightened very strongly. This was a remarkable difference with respect to the case of equal diameters, where the simple contact of the two faces, before squeezing the sandwich, could often produce a power transmission in the range 10 to 30 percent.

(*ii*) Several joints, assembled and very tightly squeezed without monitoring the transmitted power, had coupling efficiencies up to 40 to 50 percent and were not very sensitive to vibrations and shocks; however, the worst results ranged down to ≈ 10 percent (very seldom) or 20 percent (more frequently).

(*iii*) Repeated operations of tough tightening followed by partial releasing, while the transmitted power was monitored by a meter, led to coupling efficiencies in the range 80 to 84 percent. These joints were not tightly squeezed, and were very sensitive to vibrations and shocks, which could reduce the coupling all the way to 10 percent. Usually, though, the joints that had been disturbed, once abandoned to themselves, tended to restore spontaneously power coupling efficiencies up to 70 to 75 percent.

These observations and comparison with Bisbee's data[2] show that a remarkable alignment mechanism, due to the presence of the liquid, is still active for such a large difference between the cladding diameters, even if the performances are not any more suitable for practical applications.

V. CONCLUSION

Sandwich joints between single-mode fibers, aligned in embossed grooves, have been made and tested. This technique seems to suit very well the operation of splicing optical waveguides without using a microscope and without fusing their tips together. The average results obtained in this way–power coupling efficiencies ranging from 80 percent to almost 90 percent, i.e., insertion losses under 1 dB and down to 0.5 dB–compete successfully with the best results given by previous techniques.

The facts, that rather mediocre end surfaces of the fibers can be accepted and that cladding diameters can differ at least by a few percent without affecting the performances, induce confidence in the future use of the sandwich technique for splicing multifiber cables in the field. It seems that all the operations could be extended to fiber tapes, dealing with each of them as a whole. A fairly precise alignment of the fibers would be required as the result of the tape manufacturing process, but the final alignment should be provided by a set of precisely embossed grooves on a plate. The fibers ought to be forced into the grooves by means of carefully designed mechanical tools, but without microscopic observation and without dealing with them on an indi-

vidual scale. The operations to be performed will depend on the ratio between the diameters of the fibers and their spacing in the tape, and on the precision of the preliminary alignment obtained in the manufacturing process. The absence of stringent requirements on the quality of the end faces, proved by the available results, suggests that a simultaneous and coplanar cut of all the fibers (for instance, scoring them by means of a razor blade while pulling along their axis) should be adequate.

VI. ACKNOWLEDGMENTS

This work has been performed during a six-month internship at Bell Laboratories, Crawford Hill Laboratory, Holmdel, New Jersey, under a NATO Fellowship awarded and administered by the Italian National Research Council (C.N.R.) of Rome, Italy. I wish to thank S. E. Miller for arranging this internship in the most useful and pleasant way.

I am very grateful to E. A. J. Marcatili, who suggested the present investigation and stimulated it continuously with very fruitful discussions.

Several other people helped and encouraged me. I wish to thank in particular C. A. Burrus for providing the light-emitting diodes and for his important experimental advice.

REFERENCES

1. Bisbee, D. L., "Optical Fiber Joining Technique," B.S.T.J., *50*, No. 10 (December 1971), pp. 3153–3158.
2. Bisbee, D. L., "Measurements of Loss Due to Offsets and End Separation of Optical Fibers," ibid., pp. 3159–3168.
3. Standley, R. D., and Braun, F. A., "Some Results on Fiber Optic Connectors," unpublished work.
4. Astle, H. W., "Optical Fiber Connector with Inherent Alignment Feature," unpublished work.
5. Dyott, R. B., Stern, J. R., and Stewart, J. H., "Fusion Junction for Glass-Fibre Waveguides," Elec. Ltrs., *8*, No. 11 (June 1, 1972), pp. 290–292.
6. Ulrich, R., Weber, H. P., Chandross, E. A., Tomlinson, W. J., and Franke, E. A., "Embossed Optical Waveguides," Appl. Phys. Ltrs., *20*, No. 6 (March 15, 1972), pp. 213–215.
7. Gloge, D., "Weakly Guiding Fibers," Appl. Optics, *10*, No. 10 (October 1971), pp. 2252–2258.
8. Keck, D. B., and Tynes, A. R., "The Spectral Response of Low-Loss Optical Waveguides," Appl. Optics, *11*, No. 7 (July 1972), pp. 1502–1506.
9. Burrus, C. A., "Radiance of Small-Area High-Current-Density Electroluminescent Diodes," Proc. IEEE, *60*, No. 2 (February 1972), pp. 231–232.

Contributors to This Issue

PETER D'AGOSTINO, B.E.E., 1965, Pratt Institute; M.S. (Electrical Engineering), 1967, New York University; Bell Laboratories, 1969–1972. At Bell Laboratories, Mr. D'Agostino was concerned with determining the effects of the electromagnetic pulse associated with a thermonuclear blast and the design and evaluation of long-haul waveguide communication.

RICHARD T. BOBILIN, B.S.E.E., 1966, Union College; M.S.E.E., 1967, University of Michigan; Ph.D., 1970, Purdue University; Bell Laboratories, 1966—. Mr. Bobilin's initial activities were concerned with *Picturephone*® codec development. He is presently engaged in the development of the M34 multiplex. Member, Eta Kappa Nu, IEEE, Sigma Xi, Tau Beta Pi.

JAMES W. CARLIN, B.S.E.E., 1962, Illinois Institute of Technology; M.S.E.E., 1964, and Ph.D., 1967, University of Illinois; Philco-Ford, 1967–1968; Bell Laboratories, 1968—. Mr. Carlin has worked on electromagnetic pulse effects associated with nuclear bursts and on long-haul communication in millimeter waveguide.

SING-HSIUNG LIN, B.S.E.E., 1963, National Taiwan University; M.S.E.E., 1966, and Ph.D., 1969, University of California, Berkeley; Bell Laboratories, 1969—. Mr. Lin is working on propagation effects such as rain attenuation and multipath interference on radio communication systems. Member, IEEE, Sigma Xi.

ALBERT MAIONE, A.A.S. (E.T.), 1969, Queensborough Community College; Fairleigh Dickinson University, 1969–1971; Newark College of Engineering, 1971–; Bell Laboratories, 1969—. Mr. Maione is presently working on millimeter-wave-medium design and measurement techniques.

CARLO G. SOMEDA, Dott. Ing., 1964, University of Padua, Italy; M.S.E.E., 1966, Stanford University (as Volta Fellow of the IEEE); Libera Docenza, 1971, Ministero P. I., Rome, Italy. Faculty (E.E.),

University of Trieste, Italy, 1965–68; University of Bologna, Italy, 1968—; Bell Laboratories (Resident Visitor), March–September 1972, under a NATO exchange program. At Bell Laboratories, Mr. Someda was engaged in research on optical fiber transmission techniques. In Italy, he has been involved in millimeter-wave and optical propagation research. Member, IEEE, Italian Electrotechnical Association (AEI), Italian Physical Society (SIF).

GEORGE SZENTIRMAI, Dipl. Eng., 1951, and "Candidate of the Technical Sciences," 1955, Technical University of Budapest; Ph.D. (EE), 1963, Polytechnic Institute of Brooklyn; Bell Laboratories, 1959–1968; Consultant to Bell Laboratories and Associate Professor of Electrical Engineering, Cornell University, 1968—.Mr. Szentirmai was engaged in designing, and developing new design methods for, filters, equalizers, and other networks and was instrumental in developing a whole group of computer programs for filter synthesis. As a consultant, he is presently involved in extending these methods and programs to active RC filters. Fellow, IEEE; member, IEE (Great Britain), Sigma Xi.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 4, April, 1973
Printed in U.S.A.

B. S. T. J. BRIEF

Rain-Induced Differential Attenuation and Differential Phase Shift at Microwave Frequencies

By J. A. MORRISON, M. J. CROSS, and T. S. CHU

(Manuscript received January 15, 1973)

We give calculated results of differential attenuation and differential phase shift caused by rain, based on scattering of a plane electromagnetic wave by horizontally disposed oblate spheroidal raindrops. Two polarizations of the incident wave are considered, as depicted in Fig. 1. The factor $e^{-i\omega t}$ has been suppressed. Subscripts I and II designate electric fields parallel and perpendicular to the plane containing the axis of symmetry of the raindrop and the direction of propagation of the incident wave, respectively. The angle between the direction of propagation and the axis of symmetry is denoted by α. For terrestrial microwave relay systems we take $\alpha = \pi/2$, but for satellite systems other values of α are of interest.

The incident wave induces a transmitted field in the interior of the raindrop, and a scattered field. In the far field the quantities of primary interest are the complex forward scattering functions[1] $S_{\mathrm{I}}(0)$ and $S_{\mathrm{II}}(0)$. For the two polarizations considered, the polarization of the far scattered field is the same as that of the incident wave. However, when oblate raindrops are canted with respect to the vertical, the vertical and horizontal polarizations commonly used in radio communications systems will be neither of the two aforementioned polarizations. Then cross-polarization coupling will occur as a result of the difference between $S_{\mathrm{I}}(0)$ and $S_{\mathrm{II}}(0)$.

It is assumed that the ratio of minor to major axis of the oblate spheroidal raindrop depends linearly on the radius $\bar{a}$ (in cm) of the equivolumic spherical drop; specifically $a/b = (1 - \bar{a})$. This relationship is similar to that used by Oguchi.[2] The Laws and Parsons drop-

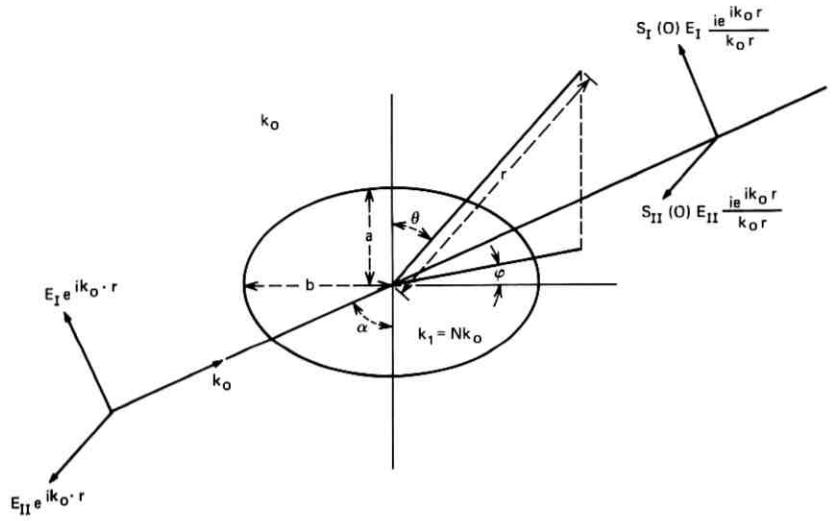

Fig. 1—Two polarizations of the incident wave.

size distribution and Best's raindrop terminal velocity are used.[3] Thus, for rain rates up to 150 mm/hour, there are 14 different drop sizes, $\bar{a} = 0.025\ (0.025)\ 0.35$, to be considered. At 20°C, the refractive indices $N = 5.581 + 2.848i$ at 30 GHz and $N = 6.859 + 2.716i$ at 18.1 GHz were obtained from an elaborate fitting equation in a recently published survey[4] of available measured data. Since the calculations at 4 GHz were made at an earlier date, the value $N = 8.77 + 0.915i$ was taken from the older literature.

Oguchi[2] has developed a perturbation theory for scattering from slightly eccentric spheroidal raindrops, but the first-order approximation is expected to be inaccurate for the larger raindrops, which have larger eccentricity. Consequently, we have used a matching theory to obtain approximate nonperturbative solutions to the problem. Although the calculations reported on here are for oblate spheroidal raindrops, the procedure may be used for axisymmetric raindrops which are not too nonspherical. Full details of the analytical and numerical procedures, together with tables of the computed values of $S_I(0)$ and $S_{II}(0)$ for each drop size, will be given at a later date, but we outline the approach here.

Spherical coordinates (r, θ, φ) are chosen with polar axis along the axis of symmetry of the raindrop, and origin at the center, as

in Fig. 1. The scattered electric field is expanded in the form $\Sigma\Sigma(a_{mn}\mathbf{M}_{mn} + b_{mn}\mathbf{N}_{mn})$, in terms of solutions of the vector wave equation[5] satisfying the radiation condition. An analogous expansion is assumed for the transmitted field, in terms of vector wave functions which are finite at the origin. The complex coefficients in the expansions are determined approximately by satisfying the boundary conditions, namely the continuity of the tangential components of the total electric and magnetic fields across the surface of the raindrop, by a fitting procedure. The index m denotes the order of the harmonic in the azimuthal angle φ. The incident field may be expressed as a Fourier series in φ, with coefficients depending on r and θ. Because of this, and the axial symmetry of the raindrop, the problem can be decomposed and the boundary fitting carried out independently for each m.

A large number of points is chosen on the curve which is the intersection of the boundary of the raindrop with the half-plane $\varphi = 0$, $0 \leqq \theta \leqq \pi$, and the coefficients are determined by requiring the boundary conditions to be satisfied at these points in the least squares sense. The advantage of using least squares fitting rather than collocation (in which the number of fitting points is equal to the number of unknown coefficients, which are then determined by solving a system of simultaneous linear equations) is that the boundary conditions are satisfied more accurately. At least twice as many fitting points as unknown coefficients were used in our calculations. The number of terms required to adequately satisfy the boundary conditions depends on both the frequency and drop size. For example, at 30 GHz it was necessary to take max $m = 8$ and max $n = 23$. In order to ensure the accuracy of the leading terms, it is necessary to take more terms in n than are really needed in the calculation of the far-field quantities such as $S_{\mathrm{I}}(0)$ and $S_{\mathrm{II}}(0)$. A convergence test was carried out for each drop size, by increasing the upper limit of n in the sums by 2 and by 4. In a number of cases more than half the capacity of a Honeywell 6070 computer was used.

The rain-induced attenuation and phase shift are obtained from the forward scattering functions as follows:[1]

$$A_{\mathrm{I,II}} = 0.434 \frac{\lambda^2}{\pi} \sum \operatorname{Re} S_{\mathrm{I,II}}(0) \mathfrak{N}(\bar{a})^{\mathrm{dB}}/\mathrm{km}$$

$$\Phi_{\mathrm{I,II}} = -36 \frac{\lambda^2}{4\pi^2} \sum \operatorname{Im} S_{\mathrm{I,II}}(0) \mathfrak{N}(\bar{a})^{\mathrm{deg}}/\mathrm{km}$$

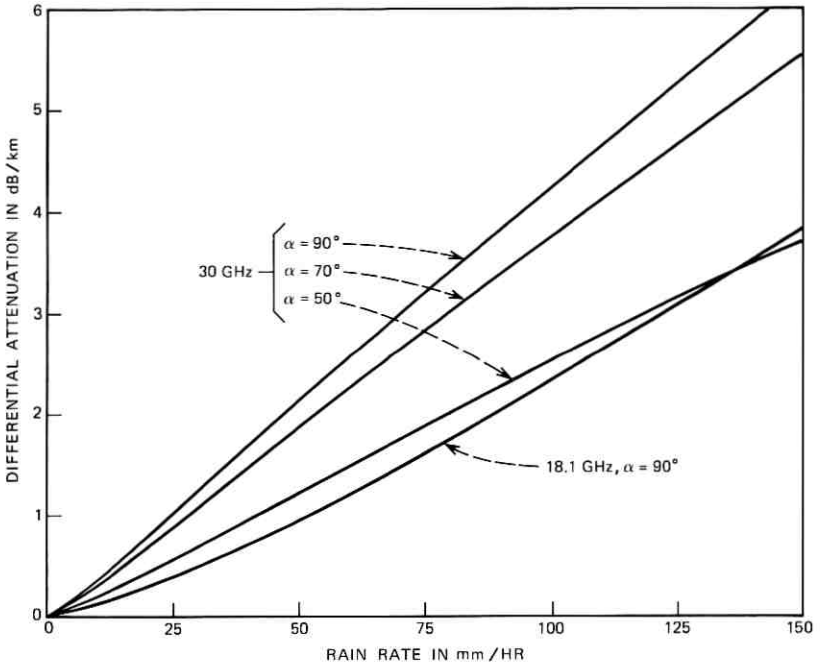

Fig. 2—Rain-induced differential attenuation.

where λ is the wavelength in centimeters, $\mathfrak{N}(\bar{a})$ is the number of raindrops with mean radius $\bar{a}$ per cubic meter, and the summation is taken over all drop sizes. For $\alpha = 90$ degrees, the attenuation and phase shift have been calculated for various rain rates at 4, 18.1, and 30 GHz for polarizations I and II. The cases $\alpha = 50$ degrees and 70 degrees have been calculated at 30 GHz. The differential attenuation $A_{II} - A_I$ and the differential phase shift $\Phi_{II} - \Phi_I$ are summarized in Figs. 2 and 3. Full numerical results will be presented later.

At 4 GHz the differential attenuation is negligibly small (0.036 dB/km at 100 mm/hr), and therefore is not plotted in Fig. 2. However, there is a differential phase shift of 5 deg/km at 100 mm/hr which can induce cross polarization as large as -10 dB over a long path of heavy rain; indeed, such a value was observed experimentally on one occasion by W. T. Barnett.[6]

The calculated differential attenuations for $\alpha = 90$ degrees in Fig. 2 are considerably larger than those obtained by first-order perturbation theory. Taking into account the canting angle distribution,[7] these

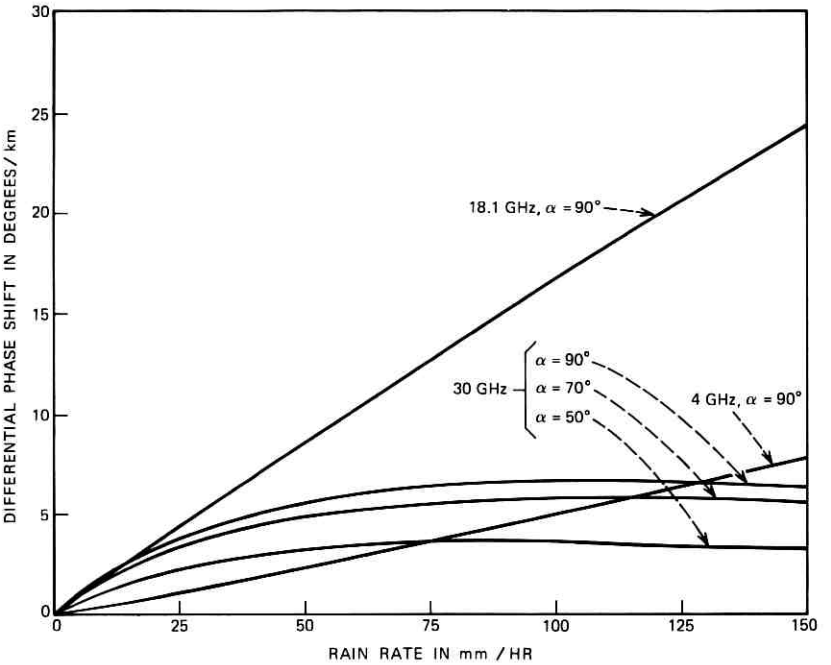

Fig. 3—Rain-induced differential phase shift.

larger values of differential attenuation are needed to explain the measured difference in attenuations for horizontal and vertical polarizations.[6,8]

At 18.1 GHz the differential attenuation and the differential phase shift will be comparable contributors to rain-induced depolarization. However, the differential phase shift remains small for heavy rain rates at 30 GHz, and hence the differential attenuation will be the dominating cause of rain-induced depolarization at this frequency and presumably also at higher frequencies. When α decreases from 90 degrees, both the differential attenuation and the differential phase shift become smaller as intuitively expected.

The authors are indebted to D. C. Hogg for bringing this problem to their attention, to J. McKenna and N. L. Schryer for suggesting the matching and least squares fitting approaches, and for several helpful discussions in relation to these, to P. A. Businger whose least squares fitting subroutine was incorporated into the main program, to Mary Ann Gatto who took over the burdensome task of running the main

program, to Susan Hoffberg who wrote the program for calculating the first-order approximation for small eccentricities, and to Diane Vitello who performed the summation over the drop size distribution. The authors are particularly indebted to J. McKenna and D. C. Hogg for their continued encouragement throughout the lengthy course of this work.

Note added in proof:

A very recent paper by Oguchi,[9] in which similar calculations are carried out for $\alpha = \pi/2$ at 19.3 and 34.8 GHz, has come to our attention. He used collocation for the expansions in terms of spherical vector wave functions, and at 34.8 GHz he also used an expansion in terms of spheroidal wave functions.

REFERENCES

1. Van de Hulst, H. C., *Light Scattering by Small Particles*, New York: Wiley, 1957.
2. Oguchi, T., "Attenuation of Electromagnetic Wave Due to Rain with Distorted Raindrops," J. Radio Res. Labs. (Tokyo), Part 1 in *7*, No. 33 (September 1960), pp. 467–485; Part 2 in *11*, No. 53 (January 1964), pp. 19–44.
3. Kerr, D. E., *Propagation of Short Radio Waves*, New York: McGraw-Hill, 1951.
4. Ray, P. S., "Broadband Complex Refractive Indices of Ice and Water," Appl. Opt., *11*, (August 1972), pp. 1836–1844.
5. Stratton, J. A., *Electromagnetic Theory*, New York: McGraw-Hill, 1941.
6. Barnett, W. T., private communication.
7. Saunders, M. J., "Cross Polarization at 18 and 30 GHz Due to Rain," IEEE Trans. Ant. and Prop., *AP-19*, No. 2 (March 1971), pp. 273–277.
8. Semplak, R. A., "Effect of Oblate Raindrops on Attenuation at 30.9 GHz," Radio Sci., *5*, (March 1970), pp. 559–564.
9. Oguchi, T., "Attenuation and Phase Rotation of Radio Waves Due to Rain: Calculations at 19.3 and 34.8 GHz," Radio Sci., *8*, (January 1973), pp. 31–38.