

THE BELL SYSTEM

Technical Journal

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

VOLUME XLI

MAY 1962

NUMBER 3

Semiconductor Surface Varactor	R. LINDNER	803
Reswitching of Connection Networks	M. C. PAULL	833
The Realizability of Multiport Structures Obtained by Imbedding a Tunnel Diode in a Lossless Reciprocal Network	I. W. SANDBERG	857
On Non-Computable Functions	T. RADO	877
On the Theory of Shrink Fits with Application to Waveguide Pressure Seals	A. J. SCHEPIS	885
Monte Carlo Solution of Bond Percolation Processes in Various Crystal Lattices	H. L. FRISCH, S. B. GORDON, V. A. VYSSOTSKY, AND J. M. HAMMERSLEY	909
Interference between Satellite Communication Systems and Common Carrier Surface Systems	H. E. CURTIS	921
Long-Term Frequency Stability for a Reflex Klystron without the Use of External Cavities	G. B. GUCKER	945
Models for Approximating Basilar Membrane Displacement—Part II. Effects of Middle-Ear Transmission and Some Relations between Subjective and Physiological Behavior	J. L. FLANAGAN	959
Approximate Solutions for the Coupled Line Equations	H. E. ROWE	1011
Transmission in Multimode Waveguide with Random Imperfections	H. E. ROWE AND W. D. WARTERS	1031
<hr/>		
Contributors to this Issue		1171

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

- H. I. ROMNES, *President, Western Electric Company*
J. B. FISK, *President, Bell Telephone Laboratories*
J. E. DINGMAN, *Executive Vice President,
American Telephone and Telegraph Company*

EDITORIAL COMMITTEE

- A. C. DICKIESON, *Chairman*
A. J. BUSCH
L. R. COOK
R. P. CROSS
R. L. DIETZOLD
L. E. DYER
J. H. FELKER
K. E. GOULD
J. R. PIERCE
M. SPARKS
W. O. TURNER

EDITORIAL STAFF

- G. E. SCHINDLER, JR., *Editor*
L. M. COLE, JR., *Assistant Editor*
A. ROSENTHAL, *Production Editor*
J. T. MYSAK, *Technical Illustrations*
T. N. POPE, *Circulation Manager*

THE BELL SYSTEM TECHNICAL JOURNAL is published six times a year by the American Telephone and Telegraph Company, 195 Broadway, New York 7, N. Y., E. J. McNeely, President; Allen G. Barry, Vice President and Secretary; L. Chester May, Treasurer. Subscriptions are accepted at \$5.00 per year, Single Copies \$1.25 each. Foreign postage is \$1.08 per year or 18 cents per copy. Printed in U.S.A.

THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XLI

MAY 1962

NUMBER 3

Copyright 1962, American Telephone and Telegraph Company

Semiconductor Surface Varactor

By R. LINDNER

(Manuscript received October 19, 1961)

The semiconductor varactor using surface space charge is analyzed and measurements made on several experimental units are described. The chief characteristics of this device are its capacity-voltage dependence and its negligible dc conduction.

The particular system used in this work is a thermally grown oxide on silicon. A theory developed from the surface charge relation is shown to agree with the experimental data over a wide range of silicon resistivity.

The theory for optimum operation for both dc and ac biasing is derived and used to compare the performance of this device with that of the p-n junction varactor. The result of this comparison shows that with careful design the semiconductor surface varactor will be able to compete favorably with the junction varactor for many possible applications, including those of ultra high frequency.

I. INTRODUCTION

In recent years semiconductor p-n junctions have found wide use in parametric amplifiers, harmonic generators, and frequency modulators. These applications result from the fact that the space charge layer capacity of the junction is voltage dependent and thus easily variable.

Another semiconductor varactor which in some instances shows a greater dependence of capacity upon voltage has been investigated recently.^{1,2} Unlike the junction varactor, the modus operandi of this new device is the change of the distribution of charge at a semiconductor surface with an applied normal field. Since the surface in a sense replaces the junction, the device has been termed a "surface varactor." The device has also been termed an "ON varactor" (for oxide-n-silicon) and an "MOS diode" (for metal-silicon oxide-silicon).

The structure is shown in Fig. 1. It consists of an insulator (across which the normal field is applied) in intimate contact with a semiconducting material. The silicon-silicon dioxide system was used in all our experiments.

II. QUALITATIVE DESCRIPTION

A qualitative description of the basic principles of the surface varactor is helpful in understanding the more detailed theory. This discussion makes use of the surface varactor as shown in Fig. 1 with p-type silicon. In this discussion, the capacity corresponds to that which would be measured with a negligibly small ac signal in conjunction with the indicated dc bias voltage.

1. Consider that a negative voltage is applied to the contact on the oxide, which we will call the "field plate." Holes from the interior of the silicon body are attracted toward the silicon-silicon dioxide interface, resulting in an enhancement or accumulation of charges. These holes, which are located close to the interface, form a charge layer of negligible thickness, equal and opposite to the charge layer on the field plate. Thus, as is shown at point A of Fig. 2, the capacity is that due to the oxide alone.

2. If, on the other hand, the field plate voltage is slightly positive, holes are repelled from the interface leaving exposed ionized acceptors. The charge layer in the silicon composed of these ionized acceptors thus extends to a relatively large depth, which effectively increases the distance between it and the charge layer at the field plate. The capacity is thereby reduced from the value of the oxide capacity as indicated by point B in Fig. 2. The charge distribution approaches that of a reverse-biased n^+p step junction.

3. Increasing the positive voltage still further will eventually begin to accumulate a substantial density of generated electrons at the interface, resulting in an inversion condition of charge; i.e., the surface charge

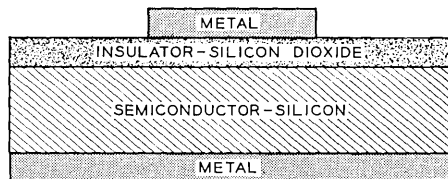


FIG. 1 — Schematic illustration of the surface varactor showing the oxide layer on one side of the silicon wafer and the metal contacts.

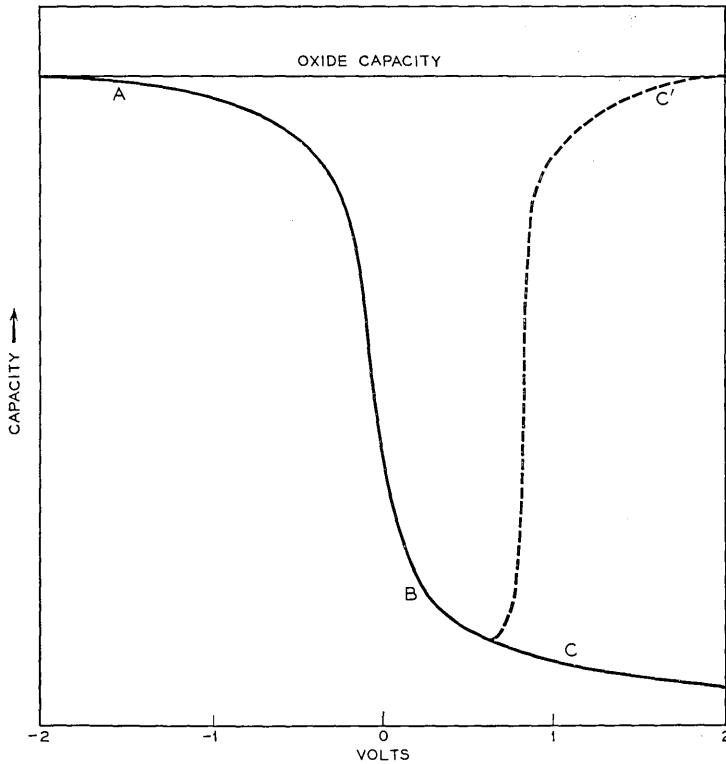


FIG. 2 — Curve ABC is a typical capacity vs voltage curve for a surface varactor. Curve ABC' is a special case.

is now opposite in sign to the normal bulk carriers. However, this effect of electrons on capacity becomes pronounced only when their density becomes as large as that of the acceptor density. At sufficiently high bias the space charge will be predominately made up of electrons, again in a relatively narrow region near the surface, and the capacity will approach the oxide capacity as a limit. This is shown by the dashed line BC' of Fig. 2. To obtain this increase in capacity, which is due to change in density and distribution of electrons (minority carrier), one must carry out the measurement at sufficiently low frequency to provide the necessary time for generation and recombination of the electrons. If, on the other hand, the measurement is carried out at higher frequencies, the above rate of the electrons and the increase in capacity with bias is not obtained. Instead, the change in charge within the space charge will be made up entirely of fixed acceptor charges. The capacity will then con-

tinue to decrease with applied bias as shown by point C of the solid line of Fig. 2. This decrease (B-C) however, is appreciably smaller than that of a reverse-biased n^+p junction. This is because, at a given dc bias, the generated minority carriers (electrons) are accumulated in the surface varactor case, while in the case of the n^+p junction they are swept away. One obtains, therefore, a narrower space-charge region for the surface varactor than for the n^+p junction with this method of biasing.

The frequency ranges in which the capacity will increase or decrease depends on the magnitude of the ac signal, the carrier lifetime, and the semiconductor resistivity. However, for silicon, using reasonable values of these parameters (ac signal > 10 mv, lifetime $> 10^{-9}$ sec., resistivity < 1000 ohm cm) at frequencies of about 10,000 cps and above the capacity will decrease.

III. METHOD OF APPLYING BIAS

In the preceding discussion the bias has been described as a dc bias. This meant that the bias was manually applied in steps and the capacity was measured with an ac signal of negligible magnitude as in Fig. 3(a). The case will be referred to as the "dc-bias case." All experimental results were obtained in this manner. However, another important method of biasing will also be discussed. This is the use of ac biasing, which is necessary when using the device as a parameter amplifier or harmonic generator. For purposes of determining the capacity variation, an additional higher frequency ac signal of small magnitude is assumed as in

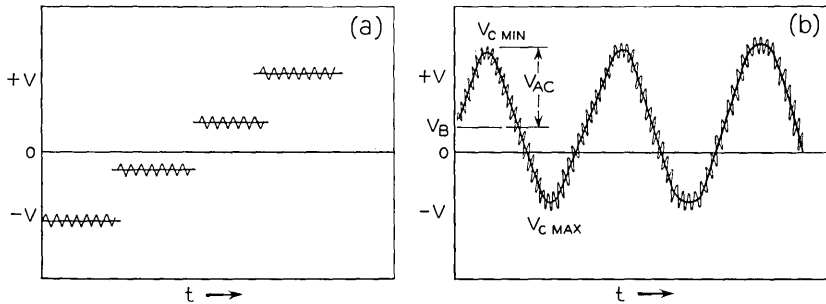


FIG. 3 — Methods used in biasing surface varactor. (a) DC biasing — dc voltage is stepped to each value and small ac signal is used to measure the capacity at each voltage step. (b) AC biasing — bias is made up of a dc component V_B (+ for p-type silicon, - for n-type silicon), and an ac component, V_{ac} ; a small higher frequency ac signal is used to measure the capacity at any point. The capacity ratio is defined as C_{max}/C_{min} .

Fig. 3(b). (Note that there can be a dc component to the ac bias.) This case will be referred to as the "ac-bias case."

With this background, the analysis of the problem is carried out as presented in the following order.

First, the relations between the equilibrium charge and the voltage across the device are presented. This leads to a general expression for the capacity.

Next, relations are derived for dc-bias case. These relations are then compared with the experimental results for several silicon resistivities. Also, curves are presented to show the highest capacity changes possible with a dc bias.

The ac-bias case is then analyzed and the proper biasing to obtain the optimum capacity changes are determined.

A relation for minimum cutoff frequency is derived and compared with experimental values. Cutoff frequencies obtainable with very thin epitaxial films are calculated. Experimental results of the effect of temperature are also shown.

In the last sections a comparison is made between the performance of the surface varactor and that of the junction varactor. Finally a brief comment is made for the case where the silicon resistivity shows a gradient from the surface.

IV. BASIC THEORY — EQUILIBRIUM CHARGE RELATIONS

The theory of the surface varactor is an extension of the fundamental theory of the space-charge region at the surface of a semiconductor

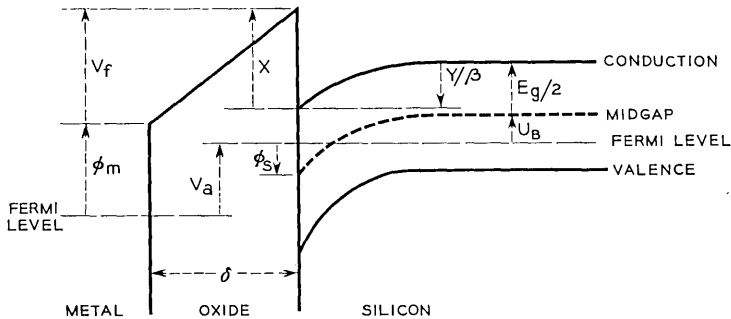


FIG. 4 — Potential diagram that exists across the metal-oxide-silicon surface. Although the equation derived from it is quite general, this diagram is specifically for a p-type silicon biased so the silicon surface at the oxide interface has inverted. The direction of the arrows refers to the positive value of the potentials.

when an external electric field is applied.³ The band picture that governs in this case is shown in Fig. 4. In general, a voltage across the device, V_a , is associated with charge in both the space-charge region and the surface states. In equilibrium these charges are given³ by the relations

$$Q_{sc} = \pm en_i \mathcal{L} [\lambda (e^{-Y} - 1) + \lambda^{-1} (e^Y - 1) + (\lambda - \lambda^{-1}) Y]^{\frac{1}{2}} \\ (-) \text{ for } Y > 0, \quad (+) \text{ for } Y < 0, \quad (1)$$

where

- e is the electron charge
- n_i is the intrinsic density of electrons
- \mathcal{L} is twice the Debye length
- λ is the doping density normalized to the intrinsic density
- Y is the surface potential relative to the bulk in kT units

and

$$Q_{ss} = eF_{ss}(Y) \quad (2)$$

where the function $F_{ss}(Y)$ includes the density of surface states and their distribution within the energy gap. A third relationship can be determined directly from Fig. 4 by balancing potentials

$$V_a = \varphi_m - (\chi + u_B + E_g/2) + V_f + Y/\beta \quad (3)$$

where

- V_a is the applied voltage
- φ_m is the work function of the metal
- χ is the electron affinity of the semiconductor
- u_B is the bulk potential difference of midgap to Fermi level
- E_g is the energy gap

and

V_f is the oxide film voltage

and also

$$Q = Q_{sc} + Q_{ss} = C_f V_f \quad (4)$$

where

- Q is the total charge and
- C_f is the capacity of the oxide film.

The general relation for the capacity of the structure under any con-

dition is

$$C = \frac{dQ}{dV_a} = \beta \frac{dQ}{dY} \frac{d(Y/\beta)}{dV_a}.$$

Then, by differentiating (3) and (4),

$$C = \frac{C_{si}C_f}{C_{si} + C_f} \quad (5)$$

where

$$C_{si} = \beta \frac{dQ}{dY}. \quad (6)$$

This relation corresponds to separate capacitors in series, one being the oxide film capacity and the other an effective silicon capacity. One must be careful, however, in applying this concept of series capacitors. It is *not*, for instance, completely equivalent to the case of a p-n junction capacitor in series with a linear capacitor. In this latter case two sets of positive and negative charges are encountered, one on the p-n junction and the other on the series capacitor. In the surface varactor, on the other hand, *only one* set of positive and negative charges is present.

V. DC-BIAS CASE

When true equilibrium is maintained, differentiation of (1) and (2) would suffice to determine the capacity explicitly. The result of this is

$$C_{si} = C_{sc} + C_{ss}$$

where

$$C_{sc} = \beta \frac{dQ_{sc}}{dY} = \frac{\epsilon_{si}[\lambda(1 - e^{-Y}) - \lambda^{-1}(1 - e^Y)]}{\mathcal{L}[\lambda(e^{-Y} - 1) + \lambda^{-1}(e^Y - 1) + (\lambda - \lambda^{-1})Y]^{\frac{1}{2}}}$$

ϵ_{si} being the dielectric constant of silicon, and

$$C_{ss} = \beta \frac{dQ_{ss}}{dY} = \beta F'_{ss}(Y).$$

Curve ABC' with the dashed portion C' of Fig. 2 was calculated using the above relation, setting $C_{ss} = 0$.

However, as discussed previously, equilibrium is not normally attained in the useful frequency range. We will now calculate the capacity-voltage characteristic of a surface varactor subjected to a dc bias and an ac signal sufficiently rapid that minority carriers cannot equilibrate

within the duration of an ac cycle. In this analysis, the surface states are also assumed to be unable to equilibrate, hence $C_{ss} = 0$ and $C_{si} = C_{sc}$. The analysis is presented in the Appendix. The results are:

for p-type

$$C_{si} = \frac{\epsilon_{si}\lambda^{\frac{1}{2}}(1 - e^{-Y})}{\mathfrak{L}(e^{-Y} - 1 + Y)^{\frac{1}{2}}}, \text{ for } Y - 1 > e^Y\lambda^{-2*} \quad (7)$$

$$C_{si} = \frac{\epsilon_{si}\lambda^{\frac{1}{2}}\mathfrak{L}^{-1}(y_1 - 1)^{\frac{1}{2}}}{y_1 - e^{-\frac{1}{2}}(Y - y_1)}, \text{ for } Y - 1 < e^Y\mu^{-2}$$

where y_1 is defined by

$$y_1 - 1 = e^{Y_1}\lambda^{-2}.$$

Physically, when the surface potential is y_1 , the charge due to minority carriers is about equal to the fixed charge.

For n-type

$$C_{si} = \frac{\epsilon_{si}\lambda^{-\frac{1}{2}}(1 - e^Y)}{\mathfrak{L}(e^Y - 1 - Y)^{\frac{1}{2}}}, \text{ for } -Y - 1 > e^{-Y}\lambda^2 \quad (8)$$

$$C_{si} = \frac{\epsilon_{si}\lambda^{-\frac{1}{2}}\mathfrak{L}^{-1}(-y_1 - 1)^{\frac{1}{2}}}{-y_1 - e^{\frac{1}{2}}(Y - y_1)}, \text{ for } -Y - 1 < e^{-Y}\lambda^2$$

where y_1 is defined by

$$-y_1 - 1 = e^{Y_1}\lambda^2.$$

The complete theoretical small-signal capacity vs voltage relation can be calculated from (1), (2), (3), (4), (5), and (7) or (8). The procedure is to list values of Y and determine the capacity and voltage common to each value of Y . It is seen from (3) and (4) that surface states, Q_{ss} , and work function difference, $\varphi_m - (\chi + u_B + E_g/2)$, can cause an additive constant to the applied voltage V_a . Nonzero values of Q_{ss} , however, can also cause more complex capacity variation with voltage. Calculations for several resistivities, assuming work function difference and Q_{ss} to be zero, are plotted as dashed lines in Figs. 5 and 6.

VI. EXPERIMENTAL COMPARISON OF THEORY

Units were fabricated using silicon of 15 ohm cm — p-type, 15 ohm cm — n-type, 0.2 ohm cm — n-type, and 0.029 ohm cm — n-type.

* For $Y = 0$, $C_{si} = \frac{\epsilon_{si}}{\mathfrak{L}} \sqrt{2\lambda}$.

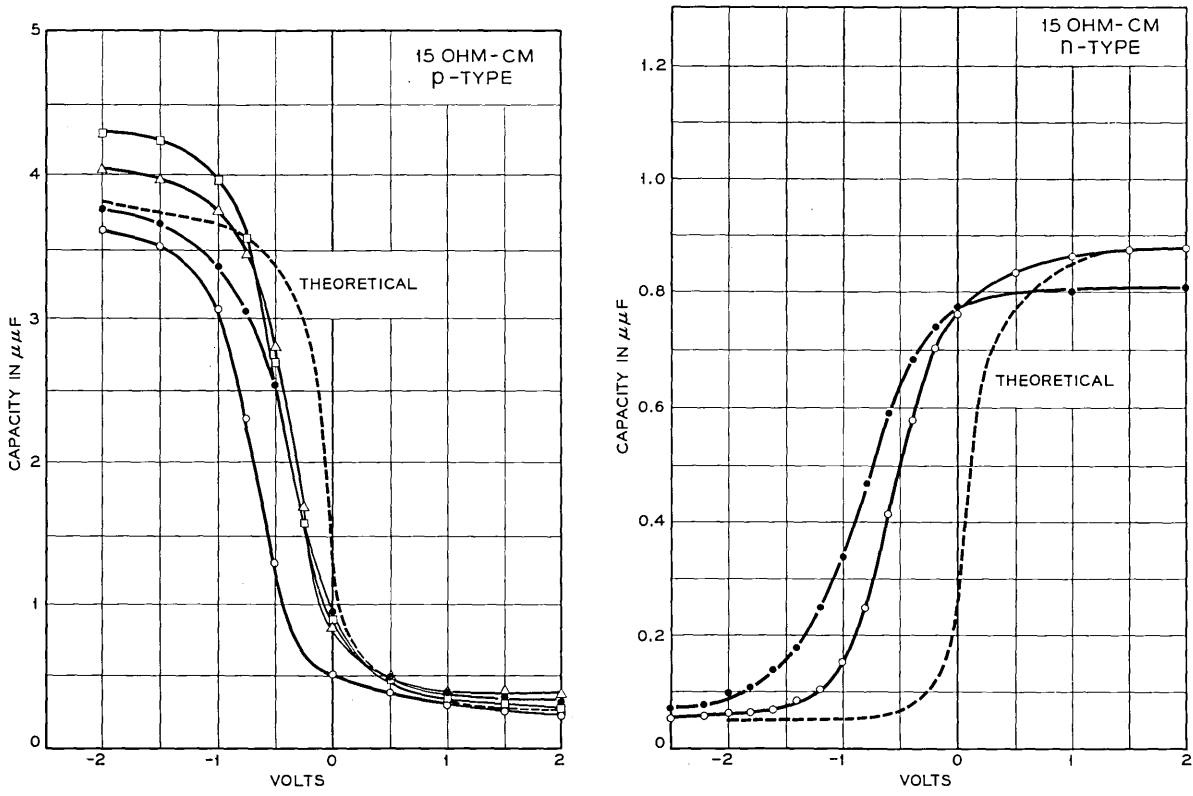


FIG. 5 — Solid lines are experimental plots of the capacity measured over a range of bias voltage. The results of four units made from 15 ohm cm p-type silicon are shown on the left and two units of 15 ohm cm n-type silicon are shown on the right. The measurements were made with a 30 mv, 100,000 cps signal. The dashed lines are the theoretical curves for the quasi-equilibrium case (assuming no surface states) described in the text.

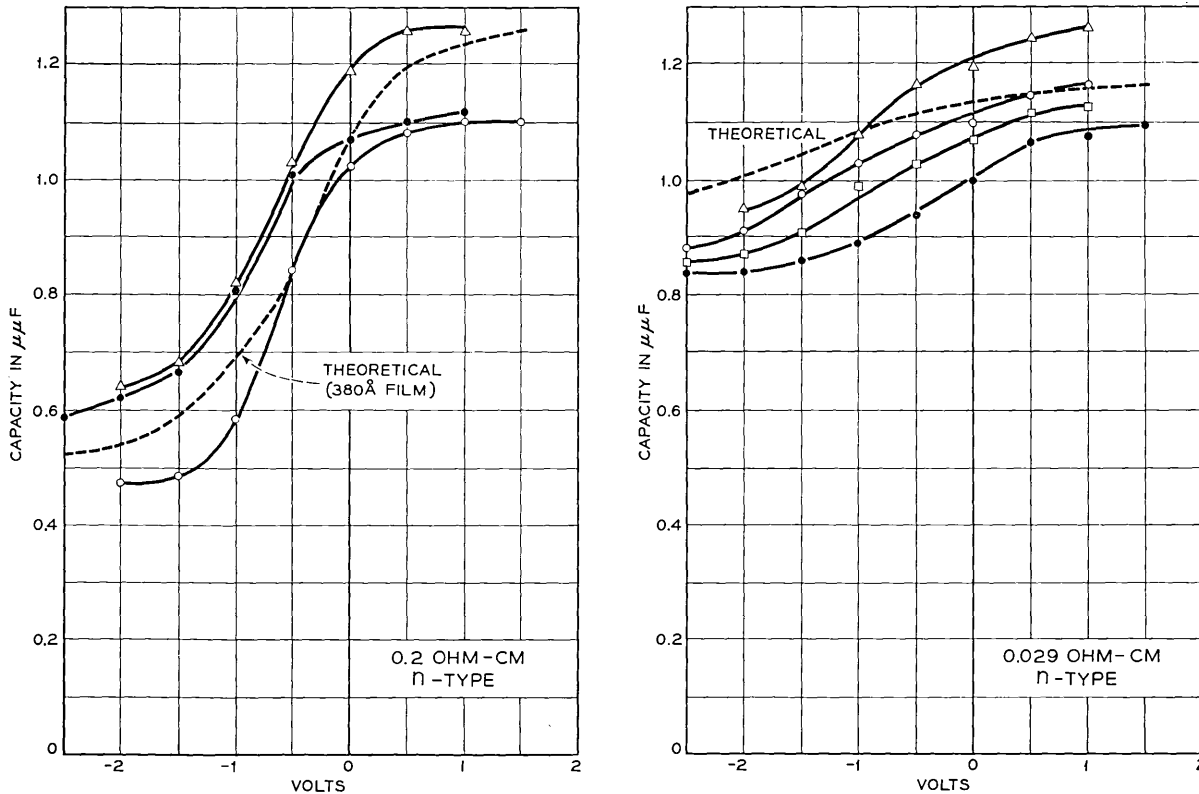


FIG. 6 — Similar to Fig. 5 except units on left were made from 0.2 ohm cm n-type silicon and those on the right are 0.029 ohm cm n-type.

The technique used was to etch the silicon slice to 0.004 inch, evaporate a metal contact on one side, and further etch the working surface to a final thickness of 0.002. The slices were then etched into 0.040 inch diameter discs and subsequently cleaned and thermally oxidized⁴ at 920°C for 30 minutes. The oxide thickness ranged from 300 Å to 400 Å, and the active or contact area, obtained by pressing a gold point on the oxide, ranged from 1 to 3×10^{-5} cm², or about a 2 mil diameter.

The capacity versus voltage measurements of several experimental units with a 100,000 cps, 30 mv signal are presented in Figs. 5 and 6, along with the theoretical quasi-equilibrium curves (by "quasi" it is meant that the minority carriers do not reach equilibrium — curve C of Fig. 2 instead of C'). The shape of the capacity-voltage curves agrees reasonably well with the theoretical curves. Except for the 0.029 ohm cm case there is a uniform voltage translation to negative values of 0.2 to 0.7 volts. Considering the surface states⁴ to be at the oxide-silicon interface, if the work function difference is zero, this voltage corresponds to a particular charge (residing in the surface states) on the oxide capacity. Thus for a voltage translation, V_{ss} , $Q_{ss} = V_{ss}C_f$, which for 0.7 volts and a 300 Å film corresponds to a donor surface state density of 5×10^{11} /cm².

VII. VARACTOR QUALITY

Two quantities of importance for the performance of a varactor are the ratio between the maximum and minimum capacities and the cutoff frequency. Various figures of merit have been proposed that usually combine these factors in various ways. The cutoff frequency is of greatest importance in high frequency amplifiers, while the capacity ratio is important throughout the frequency range.

VIII. MAXIMUM CAPACITY RATIO

8.1 DC-Bias Case

The magnitude of capacity change with bias is given by the ratio (C_{\max}/C_{\min}). C_{\max} is generally limited by and identically equal to the oxide film capacity, C_f , and C_{\min} is determined by the space-charge capacity (which is a monotonically decreasing function with bias) but is limited by the maximum field, E_f , allowed on the oxide.* The field

* This field could also be limited by Zener breakdown of the silicon for very low resistivities ($\rho < .01$) but these resistivities normally produce negligible capacity change for this structure.

must be limited to this value since dielectric breakdown of the oxide is destructive. Using (5)

$$\frac{C_{\max}}{C_{\min}} = 1 + \frac{C_f}{C_{s_{i\min}}} \tag{9}$$

The relation between the maximum field, E_f , and the maximum band bending, Y_M , is given by (1) with $e^Y \gg \lambda^2 Y$ and is

$$Y_M = 2 \ln \frac{\epsilon_f E_f \lambda^{\frac{1}{2}}}{en_i \mathcal{L}}$$

Then to find $C_{s_{i\min}}$ this value of Y_M should be substituted for Y in (7). It can be shown, however, that for $E_f \geq 10^6$ v/cm and $\lambda \leq 10^8$, we have $Y_M > y_1$. Thus, closely

$$C_{s_{i\min}} = \frac{\epsilon_{si} \lambda^{\frac{1}{2}} (y_1 - 1)^{\frac{1}{2}}}{\mathcal{L} y_1}$$

and substituting into (9),

$$\frac{C_{\max}}{C_{\min}} = 1 + \frac{\epsilon_f \mathcal{L}}{\delta \epsilon_{si} \lambda^{\frac{1}{2}} (y_1 - 1)^{\frac{1}{2}}}$$

This is plotted in Fig. 7 for a range of normalized doping density, λ , and oxide thickness, δ , using $E_f = 3 \times 10^6$ v/cm. Ratios of over a hun-

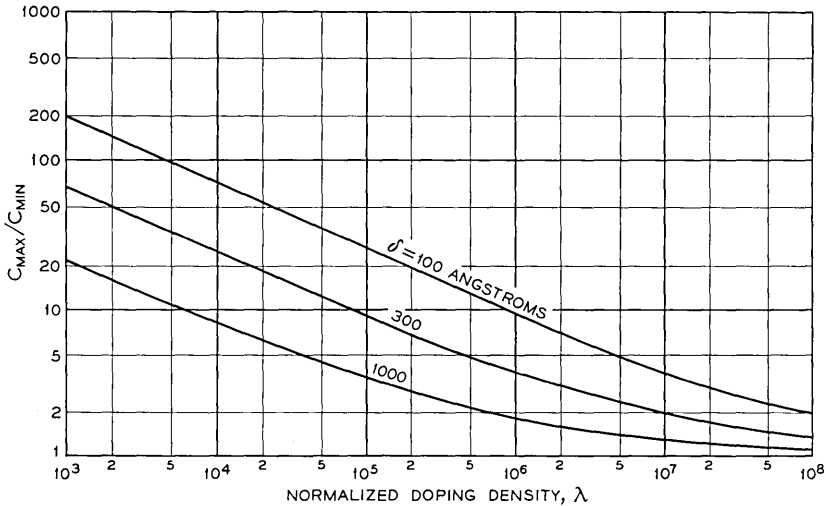


Fig. 7 — The maximum capacity ratio obtainable as a function of λ and oxide thickness for the dc-bias case. The maximum field on the oxide is 3×10^6 v/cm.

dred are possible with the high resistivity (low λ) material and decrease approximately as the one-half power of λ .

8.2 AC-Bias Case

An important mode of operation of the surface varactor is large ac bias signals superimposed on a dc bias voltage. Such conditions occur in low-noise variable capacitance amplifiers. Here we are interested in the variation of the capacity during the ac cycle rather than the variation of the capacity with dc bias as discussed for the dc-bias case (See Fig. 3). In this presentation we will assume minority carrier equilibration cannot occur during one period of the ac bias signal so that minority carrier charge is fixed at its magnitude at the dc component of the bias. Now when large ac biases are applied, the field must terminate on exposed fixed charges. Larger bending of the bands is obtainable than in the pure dc-bias case, without exceeding the oxide dielectric strength, because the density of minority carriers at the oxide-silicon interface is limited. This is shown by the following analysis.

The maximization of the dc and ac components of the bias to attain the greatest band bending is governed for either polarity of voltage by

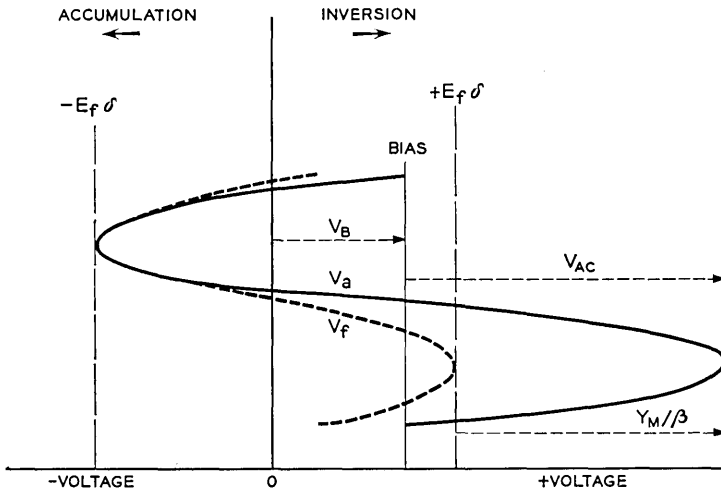


FIG. 8 — Representation of applied voltage in the ac-bias case. The solid curve is the applied voltage, V_a , which is a sine wave displaced from zero by the value of the bias voltage, V_B . The dashed curve, V_f , is the fraction of the applied voltage which is carried by the oxide film. The vertical dashed lines are the limits of voltage which can be safely carried by the oxide film. The diagram assumes p-type silicon.

the dielectric strength of the oxide. To avoid dielectric breakdown of the oxide when bending the bands toward accumulation, as seen in Fig. 8,

$$V_{ac} \leq V_B + E_f \delta, \quad (10)$$

where V_B is the applied dc bias, and V_{ac} is the zero to peak value of the ac bias. And when bending the band toward inversion,

$$V_f = V_{ac} - \frac{Y_M}{\beta} + V_B \leq E_f \delta, \quad (11)$$

where Y_M is the band bending at the peak value of the signal. For the greatest Y_M the equations are equalities. Solving (10) and (11)

$$V_{ac} = E_f \delta + \frac{1}{2} \frac{Y_M}{\beta} \quad (12)$$

and

$$V_B = \frac{1}{2} \frac{Y_M}{\beta}. \quad (13)$$

The charge in the space-charge region at the time of maximum bending is given by the sum of the minority carrier charges* plus the acceptor charges, the expressions for which can be obtained from the Appendix. Thus if $Q_{ss} = 0$

$$Q = Q_{sc} = \epsilon_f E_f = en_i \mathcal{E} \lambda^{-\frac{1}{2}} e^{Y_B/2} + en_i \mathcal{E} \lambda^{\frac{1}{2}} Y_M^{\frac{1}{2}}$$

where Y_B is the band bending at the dc bias voltage. Note that the electron charge is determined at the dc bias voltage. From which

$$Y_B = 2 \ln \left[\frac{\epsilon_f E_f \lambda^{\frac{1}{2}}}{en_i \mathcal{E}} - \lambda Y_M^{\frac{1}{2}} \right]. \quad (14)$$

The charge at the dc bias point is

$$Q = C_f V_f = \frac{\epsilon_f}{\delta \beta} \left(\frac{Y_M}{2} - Y_B \right) = en_i \mathcal{E} \lambda^{-\frac{1}{2}} e^{Y_B/2} + en_i \mathcal{E} \lambda^{\frac{1}{2}} Y_B^{\frac{1}{2}}. \quad (15)$$

Substituting (14) into (15)

$$\begin{aligned} \frac{Y_M}{2} - 2 \ln \left[\frac{\epsilon_f E_f \lambda^{\frac{1}{2}}}{en_i \mathcal{E}} - \lambda Y_M^{\frac{1}{2}} \right] + \frac{en_i \mathcal{E} \beta \lambda^{\frac{1}{2}} \delta}{\epsilon_f} \\ \times \left[Y_M^{\frac{1}{2}} - \left(2 \ln \left[\frac{\epsilon_f E_f \lambda^{\frac{1}{2}}}{en_i \mathcal{E}} - \lambda Y_M^{\frac{1}{2}} \right] \right)^{\frac{1}{2}} \right] = \beta \delta E_f. \end{aligned} \quad (16)$$

* The expression for minority carriers is true only when $e^{Y_B/\lambda^{-2}} \gg Y_B$ but when large Y_M is desired this condition will be met.

In the ac-bias case the restriction on the allowable number of minority carriers ($Y - 1 > e^Y \lambda^2$) of (7) or (8) is not necessary, since lack of sufficient generation keeps the minority carrier density low regardless of the value of Y . Thus if

$$C_{si_{min}} = \left(\frac{\partial Q}{\partial V} \right)_{Y=Y_M}$$

we have

$$C_{si_{min}} = \frac{\epsilon_{si} \lambda^{\frac{1}{2}}}{\mathcal{L}(Y_M - 1)^{\frac{1}{2}}}$$

Substituting into (9) results in

$$\frac{C_{max}}{C_{min}} = 1 + \frac{\epsilon_f \mathcal{L}(Y_M - 1)^{\frac{1}{2}}}{\epsilon_{si} \delta \lambda^{\frac{1}{2}}}$$

which with (16) determines the capacity ratio.

In Fig. 9, the C_{max}/C_{min} ratio is plotted versus λ for $E_f = 3 \times 10^6$ v/cm and a 100 Å, 300 Å and 1000 Å oxide thickness. The capacity ratio decreases with the one-half power of the normalized doping density, λ , over most of the range. Investigation of the equations show that for low λ the limiting factor is the necessity of keeping the oxide from breaking down when the bands are bent toward accumulation. For the high λ

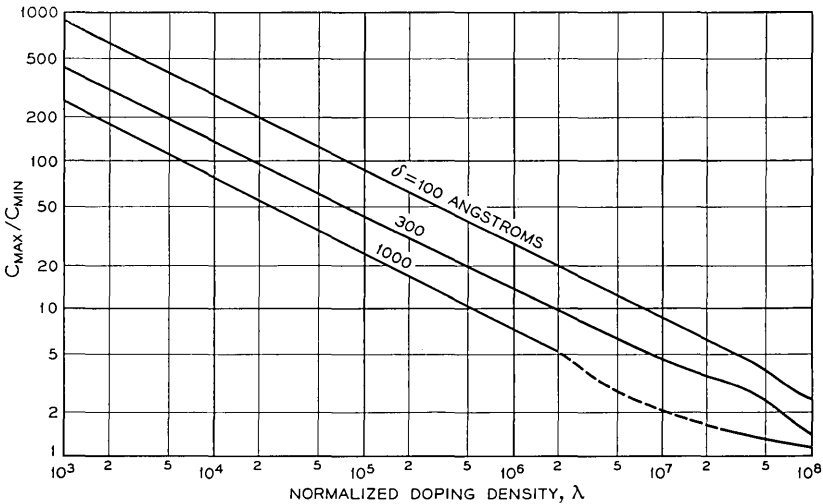


FIG. 9 — The maximum capacity ratio obtainable as a function of λ and oxide thickness for the ac-bias case. The maximum field on the oxide is 3×10^6 v/cm.

material the limiting factor is breakdown of the oxide when the bands are bent toward inversion.

In certain cases another limiting factor is encountered that is not considered in the equations, which is breakdown of the silicon itself. This occurs when the voltage carried by the silicon, Y_M/β , is greater than the avalanche breakdown voltage of a step junction whose high resistivity side is of the same resistivity as the surface varactor. An avalanche in the silicon creates sufficient minority carriers to restore true equilibrium (minority carriers terminating nearly all the field) during the part of the cycle when inversion should exist. This would increase the fraction of the applied voltage appearing across the oxide and an irreversible breakdown of the oxide might occur. Using known p⁺n step junction

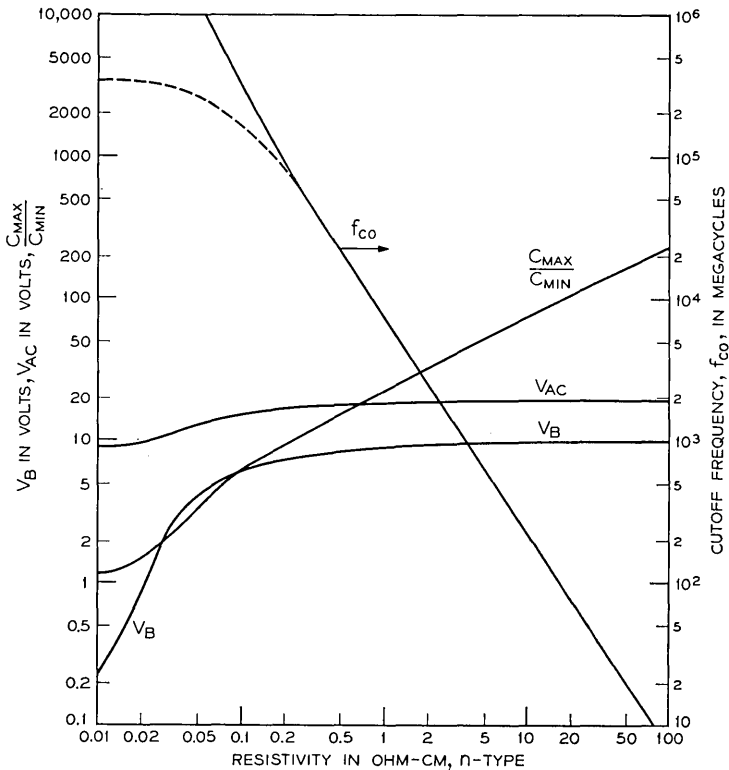


FIG. 10 — Maximum capacity ratio, C_{max}/C_{min} for ac-bias case, and the required biasing, V_B and V_{ac} , plotted against n-type resistivity. A 300 Å oxide thickness and an allowable field of 3×10^6 v/cm is used. The solid line f_{co} shows the cut-off frequency for optimum conditions of silicon thickness, while the dashed line shows the effect of adding series resistance as described in the text.

breakdown voltages⁵ it was found that this was the limiting factor for a portion of the doping density range for $\delta = 1000 \text{ \AA}$. This range is shown as a dashed line in Fig. 9.

The maximum capacity ratio obtainable with ac biasing along with the values for optimum biasing are plotted against resistivity (n-type silicon) in Fig. 10. A 300 \AA oxide thickness was assumed. The curve for cutoff frequency will be discussed in the next section.

IX. CUTOFF FREQUENCY

Assigning any one value of cutoff frequency to a varactor is difficult, since it depends upon the bias and ac signal used. Uhler,⁶ and Ueno-hara,⁷ and Pfann¹ have offered definitions which represent the varactor as a capacity in series with a resistance. For small signal operation at a bias V_B the cutoff frequency is defined as

$$f_{co} = 1/2\pi RAC_B$$

where R is the resistance, A the area, and C_B the capacity per unit area at the bias voltage. Large ac signals cause the capacity to change appreciably during a cycle, which means that if some effective time constant is sought, it must involve a type of median capacity. The simplest solution is to determine a minimum f_{co} for any mode of operation by choosing the highest possible capacity, namely the capacity, C_f , of the oxide film. The minimum cutoff frequency is then:

$$f_{co} (\text{min}) = \frac{1}{2\pi RAC_f}. \quad (17)$$

Reducing the silicon thickness increases the cutoff frequency since it decreases the resistance. However, this thickness cannot be less than $\delta(\epsilon_{si}/\epsilon_f)[(C_{\text{max}}/C_{\text{min}}) - 1]$ which is the width of the space-charge layer at maximum voltage, in order to obtain the desired capacity change. This optimum thickness may be obtained by using very thin epitaxially grown films. Table I shows the resistivity and thickness of the epitaxial

TABLE I

Epitaxial Film		$C_{\text{max}}/C_{\text{min}}$	Quality, Q at 1 kmc	
Resistivity	Thickness		Metal Base	0.6 mil-0.001 ohm cm Base +0.15 ohm
0.03 ohm cm	910 \AA	2	5300	318
0.044	1820	3	1800	281
0.060	2730	4	890	241
0.078	3640	5	510	200

film necessary for particular capacity ratios using an optimum ac bias and assuming a 300 Å oxide. Using an operating frequency of 1000 megacycles (1 kmc), the calculated quality, Q , of the unit using the *maximum* value of capacity is tabulated for each capacity ratio for two different conditions. The first is for an epitaxial film deposited on a metal base which will give the highest possible Q . The second is for a film deposited on a 0.6 mil thick wafer of 0.001 ohm cm silicon and also assumes a 0.15 ohm package resistance ($AC_f = 2 \mu\mu\text{f}$). The silicon resistivity and the package resistance were chosen as representing present practical limitations.

These results are also plotted in Fig. 10, where the solid line labeled f_{co} shows the cutoff frequency for deposition on a metal base and the dashed line shows the high frequency deviation caused by adding the silicon substrate and package resistance. The optimum silicon thickness is assumed in both cases.

Experimental measurements of the capacity of a 15 ohm cm, n-type 0.002 inch thick, surface varactor were made at frequencies up to 150 mc. The unit was biased into accumulation to give the maximum capacity, which is essentially that of the oxide. From impedance considerations, this effective capacity as measured by a capacity bridge is given by

$$C_{eff} = AC / \sqrt{1 + (2\pi f R AC)^2}$$

where f is the frequency of measurement and R is the series resistance of the device. As shown in Fig. 11 the experimental points of C_{eff} versus f are well fitted by a line corresponding to AC ($=AC_{max} \approx AC_f$) of 3.1

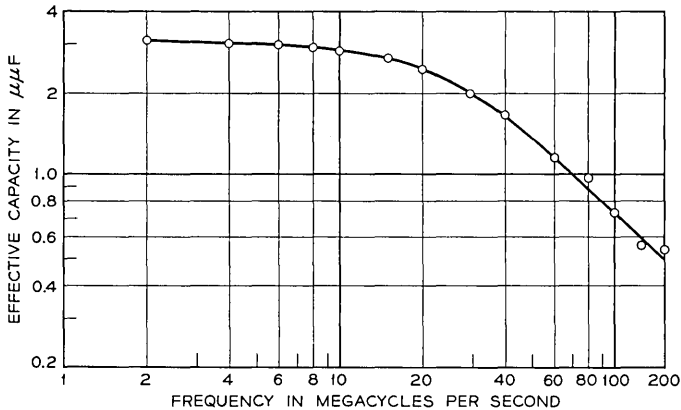


Fig. 11 — Effective capacity vs frequency for a 15 ohm cm, n-type, 0.002 inch thick, surface varactor. Experimental points lie on theoretical curve for a series RC in which $R = 2060$ ohms and $C = 3.1 \mu\mu\text{f}$.

$\mu\mu\text{f}$ and $R = 2060$ ohms. From Fig. 11 or (17), this device biased in the high capacity range exhibited a cutoff frequency of 25 mc. When the same device was biased in the minimum capacity range ($0.15 \mu\mu\text{f}$) it exhibited no change in effective capacity up to 150 mc.

Using a one kilomegacycle waveguide, two 0.2 ohm cm units 0.001 inch thick gave the results for capacity and resistance shown plotted against applied bias voltage in Fig. 12. From the *maximum* capacity and the corresponding resistance indicated, the calculated *minimum* cut-off frequency is 7000 mc.

To study possible frequency effects on the dielectric constant of the oxide films, capacity versus frequency measurements were carried out on an oxide grown on a heavily doped silicon crystal (.0013 ohm cm) where the contribution of the silicon space-charge capacity is negligible and the cut-off frequency is far in excess of the measuring frequencies.

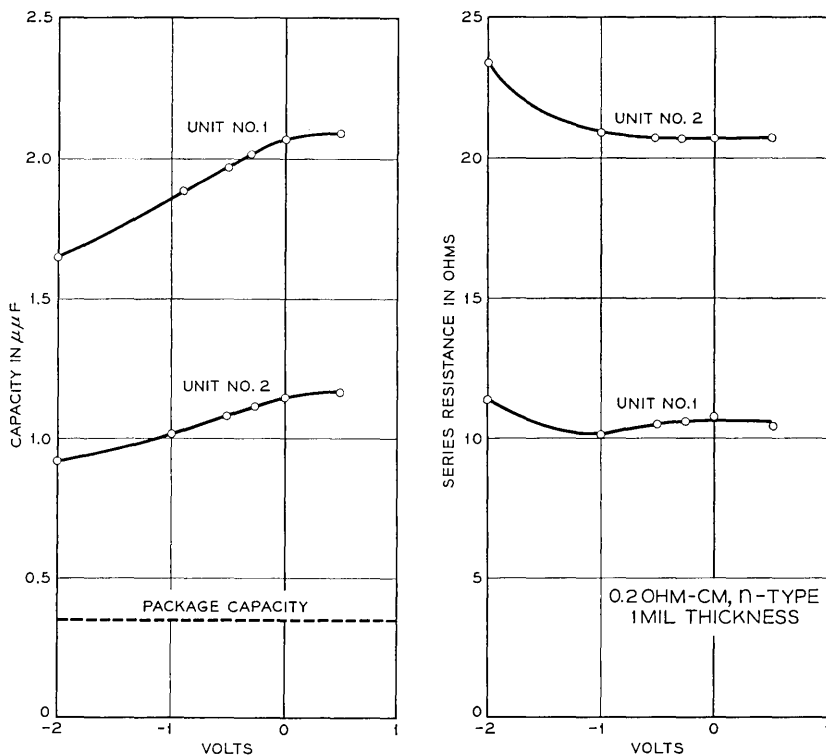


FIG. 12 — Measurements made at a 1,000 megacycle frequency for two units of 0.2 ohm cm n-type whose wafer thickness was one mil. On the left are capacity vs voltage curves and on the right are the series resistance vs voltage curves.

From the results, the dielectric constant could be represented by the relation $\epsilon = \epsilon_0(1 - 0.0465 \log_{10} f)$ where f is in megacycles and ϵ_0 is the dielectric constant at one megacycle. This relation was obtained in the range of 0.5 to 100 mc. These results are similar to those obtained for silica and other glasses.⁸ If one extrapolates the above relation, the dielectric constant at 100,000 mc would still be 77 per cent of the dielectric constant at one megacycle.

X. EFFECT OF TEMPERATURE

The effect of temperature on the characteristics of the surface varactor was examined using a 0.2 ohm cm, n-type, unit with a 15 mil diameter evaporated contact. The unit was cycled between 25°C and 90.6°C and between 25°C and -73.5°C. The capacity versus voltage curves at various temperatures are shown in Fig. 13. The results indicate two points of particular interest. (1) As the theory predicts, there is little change in both maximum and minimum capacities over the above temperature range. (2) There is a lateral shift with temperatures of the capacity curves along the voltage axis (about 1 volt between -73°C and 90.6°C). Less than one-half of this shift can be accounted for by the change in contact potential (due to the shift of the Fermi level in the silicon with

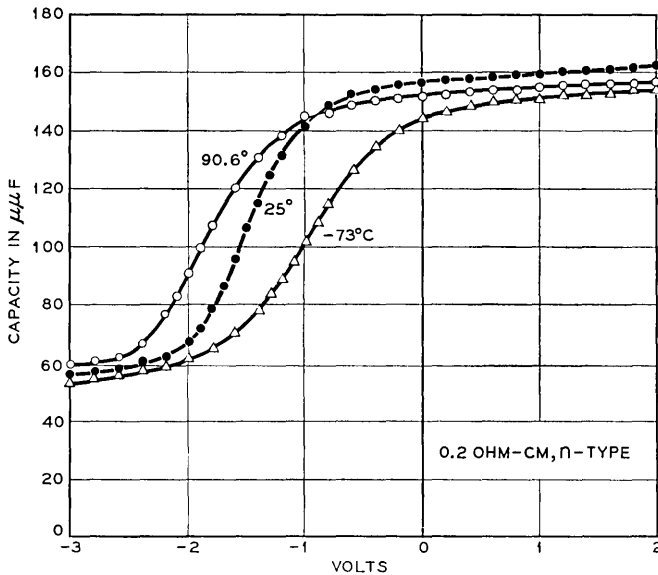


FIG. 13 — Experimental capacity vs voltage curves of a large area, 0.2 ohm cm n-type unit taken at the temperatures of 90.6°C, 25°C, and -73°C.

temperature). The remainder of the shift is not understood and is possibly associated with surface states.

XI. COMPARISON WITH STEP JUNCTION VARACTOR

Comparison of the surface varactor with the step junction varactor shows that neither device can be considered superior over the entire range of possible applications. For parametric amplifier use, the cutoff frequency, capacity ratio (amplification factor), the bandwidth, and the impedance level will effect the comparison and must all be considered. Therefore, some initial ground rules will be used to obtain a comparison, and then specific factors will be discussed which affect the comparison in various portions of the frequency range. The entire comparison is done on the basis of theory.

The ground rules are: *i.* the surface varactor is constructed of n-type silicon of optimum thickness, *ii.* the silicon dioxide layer is 300 Å thick and has a dielectric strength of 3×10^6 v/cm; *iii.* the junction varactor is constructed as a p^+n step junction whose n-type thickness is the minimum necessary to contain the space charge width; *iv.* the surface varactor is ac biased with optimum values; *v.* the junction varactor is biased from zero to the theoretical breakdown voltage; and *vi.* resistivity values refer to the n-type region of both structures. During the discussion reference will be made to Fig. 14 which shows the comparison of capacity ratio based on the cut-off frequency.

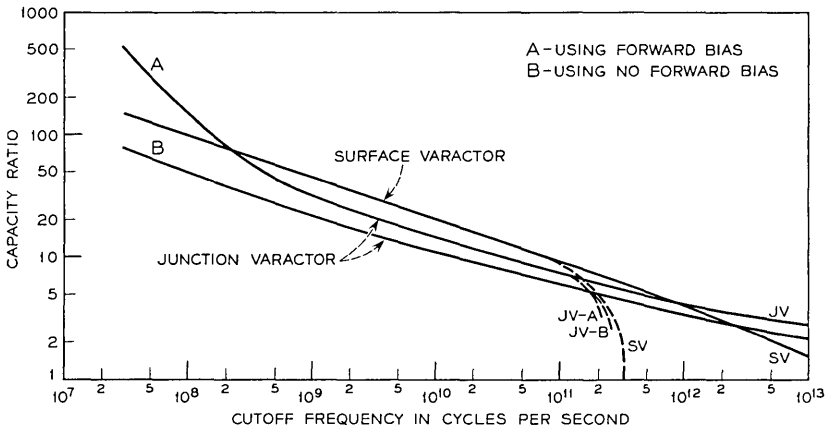


FIG. 14 — The optimum capacity ratio plotted against cutoff frequency for the surface varactor (SV) and the junction varactor (JV) using no forward bias (curve B) and using maximum allowable forward bias (curve A). The dashed lines show the deviation due to adding series resistance.

11.1 *Ultimate Frequency Range up to 10^{12} cps*

We will consider first that part of the frequency spectrum where both structures are made of resistivities above 0.1 ohm cm. For the assumptions used this includes cutoff frequencies up to 10^{12} cps. In this range the capacity ratio of the surface varactor is considerably higher than the junction varactor. This is because the zero-bias space-charge width of a junction is much wider than the oxide thickness, producing a higher C_{\max} for the surface varactor. To a lesser degree for the surface varactor, the oxide limits the voltage on the silicon to values below that needed to cause avalanche, which keeps the C_{\min} of the surface varactor from becoming as low as that of the junction varactor.

If the oxide thickness is increased to the value which allows an avalanche breakdown voltage on the silicon, it is found that the C_{\max} is the same as the zero-bias capacity of the junction varactor and the capacity ratios are identical. However, if cutoff frequency is important as well as capacity ratio, the oxide thickness should always be made as *thin* as possible. For the same capacity ratio, halving the oxide thickness allows the resistivity to be halved, thus doubling the cutoff frequency. The limit on the oxide thickness should be set by the bias voltage required, but a practical lower limit of 100 Å is recommended.

Thus, in this resistivity range, the junction varactor biased from zero to the breakdown voltage can be thought of as a surface varactor with an oxide thickness of greater than 300 Å, which puts it at a disadvantage. This is shown in Fig. 14 by comparing the surface varactor with the junction varactor curve B.

11.2 *Storage Capacity in the Junction Varactor*

The use of storage capacity in the junction varactor can be of substantial benefit in the lower frequency range if the device is tailored for its use. Storage capacity in farads is given by the relation:

$$C_{st} = (q/2kT)I\tau,$$

where I is the current in amperes, and τ the lifetime of the n-type silicon in seconds. A limitation on the lifetime is that in order for the junction to recover within a cycle, the lifetime must be equal to or less than the reciprocal of the highest frequency to be used. Allowing for this, it can be shown that if the area of the junction is kept at a minimum (2×10^{-5} cm²) and the forward current is about 5 μ a, the junction varactor under optimum lifetime conditions will attain superior performance for ultimate cutoff frequencies of less than 230 mc. In Fig. 14 compare curve A of the junction varactor with that of the surface varactor. However,

for the lowest frequencies, operation in the forward-bias direction would result in a lossy device as the resistive impedance would become less than the capacitive impedance.

11.3 High Ultimate Frequency Range

The comparison for resistivities lower than 0.1 ohm cm is of great interest because ultrahigh cutoff frequencies are possible with both structures. The capacity ratios become equal (with a value of three) at a cutoff frequency of 2×10^{12} cps. For higher cutoff frequencies the capacity ratio of the junction varactor is higher than the surface varactor. Allowing a forward bias of 0.5 v (below that required to produce appreciable storage) the cross-over point is 6×10^{11} cps, with a $C_{\max}/C_{\min} = 4.6$. However, in order to take advantage of the high ultimate f_{co} the epitaxial layer thickness would have to be a few hundred angstroms. In addition, in order to maintain a reasonable circuit impedance, due to the high capacity per unit area, the diameter of the junction area would have to be less than one mil.

11.4 Effect of Series Resistance

If practical limitations of series resistance of the silicon substrate and the package connection are included, the surface varactor has the advantage in cutoff frequency because of the lower values of C_{\max} for the surface varactor. Using a 0.6 mil thick, 0.001 ohm cm substrate and a 0.15 ohm package resistance, the cut-off frequency of junction varactor devices is maximum at 2.6×10^{11} cps for a $C_{\max}/C_{\min} \sim 3$. Under the same conditions the surface varactor has a 2.8×10^{11} cps f_{co} for $C_{\max}/C_{\min} = 3$ and a 3.2×10^{11} f_{co} for $C_{\max}/C_{\min} = 2$, as shown by the dashed lines of Fig. 14.

XII. ADDITIONAL COMPARISONS

There are some important differences between the structures of a practical nature. One is that the voltage across the oxide of the surface varactor must be controlled so as not to exceed its dielectric strength. The dielectric strength is normally between 3×10^6 and 10^7 volt/cm. If this value is exceeded, the device is permanently ruined. Thus, the junction varactor is considerably more rugged in regard to over voltages.

However, the oxide is also an aid since low-resistance ohmic contacts to the semiconductor are not required. Indeed, even the bottom contact can be applied to oxide. The characteristics are not affected since the area of the bottom contact is large compared to the field plate area.

This is especially true if the bottom surface is made degenerate. Elimination of the need for ohmic contacts could be quite useful for some III-V compounds like gallium arsenide, as well as silicon.

Another important practical advantage of the surface varactor is that the diffusion or alloying of junctions, which may be particularly difficult for very thin epitaxial films, is eliminated.

XIII. NONUNIFORM RESISTIVITY EFFECTS

The relations describing the capacity-voltage dependency all assumed a constant resistivity in the space-charge region. However, substantial changes in this dependency can be advantageously obtained using a varying doping level in the surface region. This might be accomplished by diffusion or epitaxy techniques. If the surface resistivity is high and gradually becomes lower going into the material, the $dC/dV|_{\max}$ will be lower. If the surface resistivity is low and becomes higher progressing inwards, the $dC/dV|_{\max}$ will be higher. This is analogous to a retrograde junction varactor. A case approximating the latter case has been calculated. This assumes a 1000 Å surface skin of 0.4 ohm cm ($\lambda^{-1} = 10^6$) on body material of 3 ohm cm ($\lambda^{-1} = 10^5$). A comparison of this case and that without the skin shows that the maximum slope dC/dV has been doubled. Tailoring the resistivity could also be used to attain a linear dependency over a large portion of the capacity charge.

XIV. COMPARISON OF SILICON AND GERMANIUM

To determine if germanium would make a better surface varactor than silicon it will be assumed that the thickness and dielectric of the oxide film is identical (as it could be for evaporated films) and the resistivity and optimum thickness are such as to produce identical cutoff frequencies. If this is done for large Y_M it can be found that the capacity ratio of the germanium varactor will be $(\epsilon_{si}/\epsilon_{ge})^{3/2}(\mu_{ge}/\mu_{si})^{3/2} = 1.17$ times the capacity ratio of the silicon varactor. For smaller Y_M (corresponds to higher cutoff frequencies) the advantage for germanium will be smaller but always greater than one. The same reasoning shows that n-type silicon is to be preferred over p-type silicon because of the higher mobility of electrons. For better varactors, the semiconductor should have a high mobility and a low dielectric constant.

XV. CONCLUSIONS

Using thermally grown oxide on silicon, experimental models of a surface varactor structure have been fabricated and tested. The experi-

mental results are in reasonable agreement with predictions of dc-bias performance, based on theoretical considerations. These calculations take into account nonequilibrium conditions existing when the period of the imposed signal is short compared to the time required for minority carrier generation.

For the structure used, it is shown that the usable capacitance ratio increases with the silicon resistivity, while at the same time the zero-bias cutoff frequency of the device decreases. For operation at higher frequencies, the device will have a cutoff frequency and a capacitance ratio close to that of a comparable p-n junction varactor operated on the reverse bias portion of its capacity-voltage characteristic. Detailed comparison of the surface varactor and the junction varactor is difficult because of the necessary assumptions regarding series resistance. These assumptions will be affected by the state of the technology and can even change the result of the comparison at high frequencies.

However, at medium frequencies the performance of the surface varactor is equivalent to the junction varactor, and at still lower frequencies is superior. The latter result follows from the larger usable capacitance ratio available in the surface varactor, since an equivalent capacitance ratio can be obtained with the junction varactor only at the expense of a large voltage swing, or by operation in the forward-bias direction, where the junction is a lossy element.

The most attractive feature of the surface varactor is its simplicity. The device can probably be fabricated without the need of a low-resistance ohmic contact to the semiconductor, and with a small active area and low capacitance. Thus, for high frequency use, in some cases where high contact resistance to the semiconductor or high capacitance is a limiting factor in the performance of a p-n junction varactor, the surface device may well prove to be superior.

It is noted that new materials and techniques, e.g., the use of thin epitaxially grown films and semiconductors other than silicon, may produce substantial increases in the cutoff frequency of the surface varactor. Therefore the comparative performance of the surface varactor and the junction varactor must be re-evaluated as these techniques, applicable to fabrication of both types of device, become available for use.

XVI. ACKNOWLEDGMENTS

Much of the material which appears in this paper is due to the discussions the author has had with R. M. Ryder and M. M. Atalla. They were instrumental both in the guidance of the experiment and in the presentation of the results.

The author also wants to thank A. E. Bakanowski and M. Uenohara who made the ultrahigh-frequency measurements; and G. Reich who made and measured the units as well as contributing importantly to the fabrication techniques.

APPENDIX

Derivation of DC Bias Capacity

In this appendix, the capacity relations for the dc-bias case, (7) or (8), will be derived. The difficulty in carrying out this derivation rigorously is that the minority carriers (electrons in this derivation) cannot be allowed to change in number from the value attained with the dc bias alone. Therefore, simply taking the derivative of (1) to obtain capacity is not valid over the whole range of Y . However, differentiation of (1) is valid when the value of Y is such that the number of minority carriers is so small in comparison with the total charge that any change in its magnitude can be neglected. In order to make this comparison of charge contribution, expressions for the charges on minority carriers and on exposed impurities (acceptors minus holes) will be obtained. For the purpose of this discussion, p-type ($\lambda \gg 1$) will be assumed. The exposed acceptor charge is

$$Q_a = en_i \lambda \int_0^\infty (1 - e^{-y}) dx.$$

The expression for the potential gradient dy/dx is³

$$\frac{dy}{dx} = \frac{2}{\mathcal{E}} [\lambda(e^{-y} - 1) + \lambda^{-1}(e^y - 1) + (\lambda - \lambda^{-1})y]^{\frac{1}{2}}. \quad (18)$$

For values of Y from zero to a value approximately determined by $e^Y \lambda^{-2} = Y - 1$ the second term in (18) can be neglected, and one obtains from the above relations,

$$Q_a = en_i \mathcal{E} \lambda^{\frac{1}{2}} (Y - 1 + e^{-Y})^{\frac{1}{2}} \quad (19)$$

The electron charge is $Q_e = Q - Q_a$ where Q is obtained from (1). Thus for the range of Y where (19) is valid we have,

$$Q_e = en_i \mathcal{E} \lambda^{\frac{1}{2}} (e^{-Y} - 1 + Y)^{\frac{1}{2}} \left\{ \left[1 + \frac{e^Y - 1 - Y}{\lambda^2 (e^{-Y} - 1 + Y)} \right]^{\frac{1}{2}} - 1 \right\}.$$

When $Y - 1 = e^Y \lambda^{-2}$, we have $e^{-Y} \ll 1$ and $e^Y \gg y$, so

$$Q_e = (\sqrt{2} - 1) en_i \mathcal{E} \lambda^{\frac{1}{2}} (e^{-Y} - 1 + Y)^{\frac{1}{2}}.$$

Thus at the limit of Y where (19) is valid, Q_e is $(\sqrt{2} - 1)Q_a$. Then according to the initial premise we may write for all values of Y up to a few kT of the large positive value where $e^Y = \lambda^2(Y - 1)$,

$$C_{si} = \beta \frac{d}{dY} [en_i \mathfrak{L} \lambda^{\frac{1}{2}} (e^{-Y} - 1 + Y)^{\frac{1}{2}}]$$

or

$$C_{si} = \frac{\epsilon_{si} \lambda^{\frac{1}{2}} (1 - e^{-Y})}{\mathfrak{L}(e^{-Y} - 1 + Y)^{\frac{1}{2}}}. \tag{20}$$

At higher values of Y , where $e^Y > \lambda^2(Y - 1)$, the electron charge becomes dominant and a different approach is necessary. However, it is still the exposed acceptor charge (acceptors minus holes) which is varying with the ac signal. Therefore, the only change of charge takes place at some effective distance X_o into the semiconductor where the boundary between exposed and unexposed acceptors exists. The resulting differential capacity is, therefore, given by

$$C_{si} = \frac{\Delta Q}{\Delta V} = \frac{\epsilon_{si}}{X_o}. \tag{21}$$

For positive values of Y up to $e^Y \lambda^{-2} = Y - 1$, the effective distance X_o is, from (20) and (21),

$$X_o = \frac{\epsilon_{si}}{C_{si}} = \frac{\mathfrak{L} \lambda^{-\frac{1}{2}} (e^{-Y} - 1 + Y)^{\frac{1}{2}}}{(1 - e^{-Y})} \text{ (case of depletion)}. \tag{22}$$

For larger band bending, a region near the surface will be composed predominately of minority carriers. Let the thickness of this region be X_1 , and at X_1 , $y = y_1$. y_1 is determined by $y_1 - 1 = e^{y_1} \lambda^{-2}$. Now for a band bending Y at the surface, (18) gives

$$X_1 = \mathfrak{L} \lambda^{\frac{1}{2}} (e^{-y_1/2} - e^{-Y/2}). \tag{23}$$

The width of the remainder of the space-charge region X_2 is obtained from (22) by setting $Y = y_1$. The total width of the space-charge region $X_1 + X_2 = X_o$ is then given by:

$$X_o \sim \mathfrak{L} \lambda^{-\frac{1}{2}} (y_1 - 1)^{\frac{1}{2}} + \mathfrak{L} \lambda^{\frac{1}{2}} (e^{-y_1/2} - e^{-Y/2}) \text{ (case of inversion)}. \tag{24}$$

Thus a close approximation for the capacity when $Y > y_1$ is

$$C_{si} = \frac{\epsilon_{si} \lambda^{\frac{1}{2}} \mathfrak{L}^{-1}(y_1 - 1)^{\frac{1}{2}}}{y_1 - e^{-\frac{1}{2}}(Y - y_1)}.$$

Space-charge thicknesses obtained from the above approximate relation

was compared with those obtained by the alternative and more rigorous approach of numerically integrating (18). For $\lambda = 10^5$, for instance, the maximum error for C_{si} is 1.2 per cent on the low side.

LIST OF SYMBOLS

- A — active area of device, cm^2
 C — total capacity, farads/ cm^2
 C_{si} — capacity of silicon, farads/ cm^2
 C_f — capacity of film, farads/ cm^2
 C_{sc} — capacity of space charge, farads/ cm^2
 C_{ss} — capacity of surface states, farads/ cm^2
 C_B — total capacity at dc bias, farads/ cm^2
 d — silicon thickness, cm
 E_f — maximum field to be allowed on oxide film, volts/cm
 E_g — energy gap, volts
 $F_{ss}(Y)$ — density of surface states, cm^{-2}
 f_{co} — cut-off frequency, cps
 f — frequency of measurement, cps
 $\mathcal{L} = \sqrt{\frac{2\epsilon_{si}}{en_i\beta}} = 4.8 \times 10^{-3}$ cm
 N — density of minority carriers, cm^{-3}
 n_i — density of holes and of electrons in intrinsic silicon —
 1.5×10^{10} cm^{-3}
 n_B — density of electrons in body, cm^{-3}
 Q — total charge, coul/ cm^2
 Q_{sc} — charge in space charge of silicon, coul/ cm^2
 Q_{ss} — surface state charge, coul/ cm^2
 Q_e — charge due to electrons, coul/ cm^2
 Q_a — charge due to acceptors, coul/ cm^2
 R — series resistance of silicon, ohms
 u_B — bulk potential difference of midgap to Fermi level
 V_a — applied voltage, volts
 V_B — dc component of ac bias, volts
 V_f — oxide film voltage, volts
 V_{ac} — ac bias voltage (zero to peak)
 x — distance into silicon from surface, cm
 X_o — distance from surface to effective acceptor boundary, cm
 Y/β — potential of surface compared to body, volts
 y — potential at any distance x from the surface, kT units
 y_1 — potential defined by $y_1 - 1 = e^{y_1/\lambda} - 2$
 Y_M — maximum value of surface potential, kT units

- Y_B — surface potential at bias voltage, kT units
 β — $e/kT = 38.5$ at room temperature
 δ — oxide thickness, cm
 ϵ_{si} — dielectric of silicon — 1.06×10^{-12} farads/cm
 ϵ_f — dielectric of oxide — 3.4×10^{-13} farads/cm
 $\lambda = \frac{n_i}{n_B} = \frac{p_B}{n_i}$
 μ — mobility, $\text{cm}^2/\text{volt-sec}$
 ρ — resistivity of silicon, ohm/cm
 φ_m — work function of metal, volts
 φ_s — strength of inversion, volts
 χ — electron affinity of semiconductor

REFERENCES

1. W. G. Pfann, C. G. B. Garrett, Proc. I.R.E., **47**, No. 11, November, 1959.
2. J. L. Moll, Wescon Meeting, August, 1959.
3. C. G. B. Garrett, W. H. Brattain, Phys. Rev., **99**, 1955, p. 376.
4. M. M. Atalla, E. Tannenbaum, E. J. Scheibner, B.S.T.J., **38**, 1959, pp. 749-784.
5. S. L. Miller, Physical Review, **105**, February, 1957, pp. 1246-49.
6. A. Uhlir, Jr., Proc. I.R.E., **46**, 1958, p. 1099.
7. M. Uenohara, Proc. I.R.E., **48**, No. 2, February, 1960, pp. 169-179.
8. H. E. Taylor, Jour. Glass Tech., April, 1959, p. 124.

Reswitching of Connection Networks

By M. C. PAULL

(Manuscript received November 28, 1961.)

In certain types of connection networks, it is always possible to unblock a blocked call by moving calls already set up in the network. The following results relating to these networks are derived in this article.

1. *Bounds on the number of calls which must be disturbed to unblock a blocked call.*

2. *Bounds on the relation between the number of calls which are already set up in the network, and the number of calls that must be disturbed to unblock a blocked call.*

3. *Methods of systematically changing connections to unblock a blocked call.*

I. INTRODUCTION

In a three-stage network of the type pictured in Fig. 1, it is possible that a connection between an input and an output cannot be made despite the fact that neither is already connected. This could happen if other connections already occupy at least one link in every possible path between the input and output in question. As first established by Slepian,¹ a blocked connection in such a network can be unblocked by rearranging the connections already set up in the network. Slepian further showed that such a rearrangement would never require disturbing more than $2n - 2$ calls, where the size of the switches in each stage is n by n , and there are n switches per stage. In the first sections of this article I give a proof that *to unblock a connection in such a network in no case requires disturbing more than $n - 1$ calls, and furthermore for every $n > 1$ there is at least one network state in which $n - 1$ calls must be disturbed to unblock a blocked connection.*

In subsequent sections various generalizations upon which partial results have been obtained are discussed. These include results on different network configurations, and networks having more than three stages.

As discussed in the body of this paper, the physical consequence of a

network change is to momentarily disturb network connections — to open some of these connections during the time that changes are being made. Depending on the application of the network, and the time duration of the disturbance caused by the change, this disturbance may or may not be of serious consequence. In an electromechanically operated network used to connect telephone calls such disturbances might result in disturbing conversations carried by the network. Fortunately, one can design switching networks and find change algorithms for such networks such that there will be no such disturbances. In the Appendix such a network and algorithm is described.

II. MATHEMATICAL MODEL

2.1 *The Network*

We first need a mathematical model of the network of Fig. 1 (which will simply be called “the network” from now on) in which we may conveniently represent the possible states of the network, and in which the basic properties of this particular type of network are made exact.

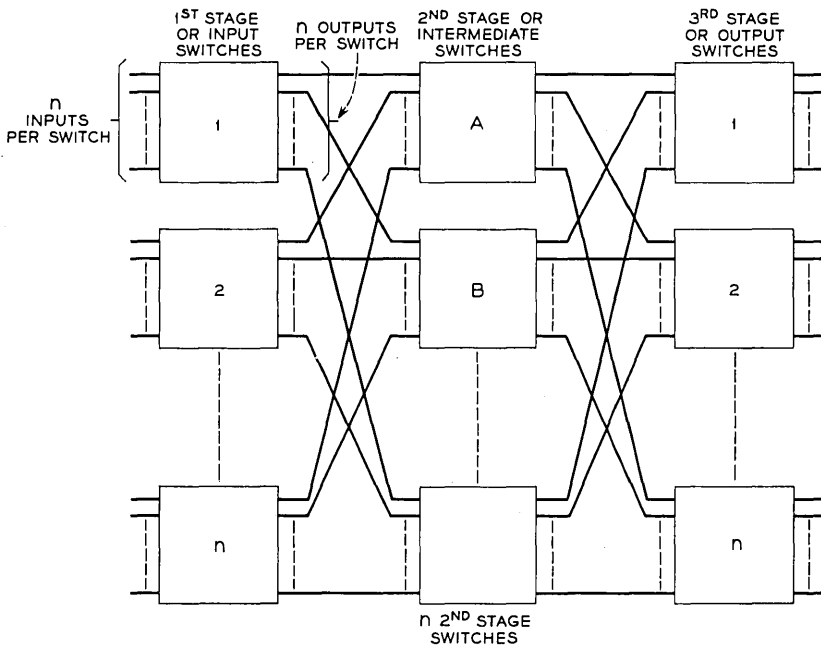


Fig. 1 — Three-stage network suitable for reswitching.

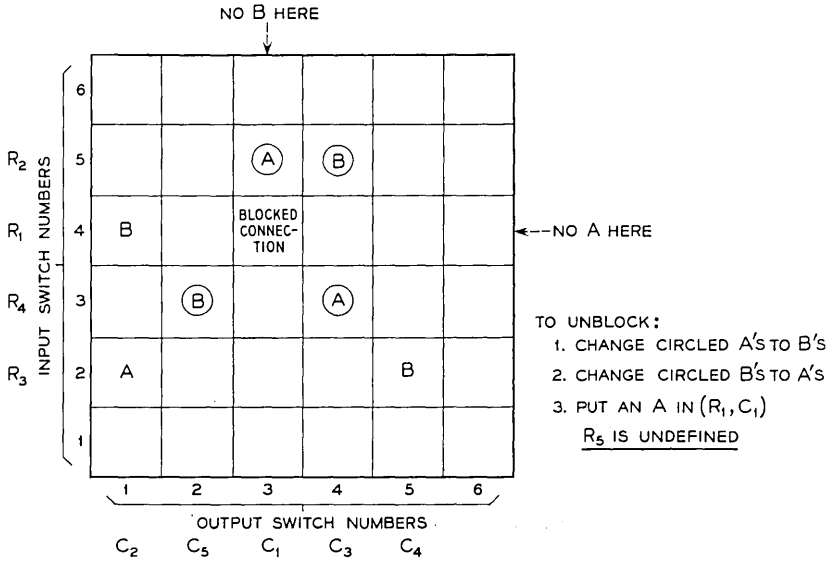


Fig. 2 — Matrix for representing the state of a three-stage network. The state pictured in the figure is illustrative of a typical blocked state as discussed in the sufficiency part of theorem 1.

We will represent the connections existing in a given network state by a set of symbols entered in a matrix (Fig. 2). The matrix has n columns and n rows, and there are n possible symbols which may be placed in any matrix position. Each position may contain from zero to n symbols. The n rows correspond to the n -input (first stage) switches; these are numbered 1, 2 \dots n . The n columns correspond to the n -output (third stage) switches, and these are numbered 1, 2 \dots n . The n symbols correspond to the n intermediate (second stage) switches. To indicate a position in the matrix we use the ordered pair (a, b) , where a is the row and b is the column. An entry, say Q in matrix position (a, b) , corresponds to a connection from input switch a through intermediate switch Q to output switch b . No entry in (a, b) indicates the absence of any connection from a to b . Although the matrix entry does not indicate which input line is connected to which output line, it does uniquely specify the links (links are the nodes in which first and second, and second and third stage crosspoints meet) involved in such a connection. For our purposes this is the important property of a connection from an input to an output line.

There are certain restrictions on the set of connections (network state)

which can exist in a network of the type of Fig. 1. These must be reflected in restrictions on the set of entries possible in our matrix:

i. There can be no more than n symbols in any row or column. This corresponds to the fact that there only n inputs to each input switch and only n outputs from each output switch.

ii. No two symbols in any row (column) may be the same. This corresponds to the fact that each input (output) switch has only one connection to each intermediate switch. If the same symbol appears more than once in a row or column, the different appearances of the symbol will be said to conflict.

A matrix with entries meeting the above restrictions will be called "legitimate," or the entries will be called "legitimate."

2.2 Blocking-Unblocking

Given a matrix with a set of entries, corresponding to a network having a corresponding set of connections, it may be impossible to make an entry in (a, b) and still have a legitimate matrix. This corresponds to the impossibility of setting up an additional connection between input switch a and output switch b . The two possible reasons for such an occurrence are:

i. There are already n symbols in row a or n symbols in column b .

ii. There are already a total of n different symbols in row a and column b , but there are less than n symbols in row a , and less than n symbols in column b .

If *i.* holds, (a, b) will be said to be trivially blocked. This corresponds to the case where either all input lines to input switch a or all output lines from switch b , or both are already connected.

If *ii.* holds, (a, b) will be said to be blocked, or legitimately, or non-trivially blocked. This corresponds to the case in which an input line on switch a cannot be connected to an output line on switch b despite the fact that neither is already connected.

Note we do not have to be specific about input and output lines, because a connection between an input and output line is legitimately blocked if and only if their corresponding switches are legitimately blocked. This is so because all switches in the network are non-blocking.

We will speak about changing connections of a network in a given state. By this we will *not* mean changing the input and output switch involved in the connection, but changing only the intermediate switch involved. That is, if a network has a connection between a certain input line and a certain output line before a *change*, it will still have a connection from the input to output line in question after the change. In terms

of our matrix a *change* corresponds to changing the symbols at various positions, but a change leaves the number of symbols in any position unchanged. A legitimate change is one which does not result in a matrix (set of network connections) which violate restrictions *i.* or *ii.* of Section 2.1.

By unblocking a blocked connection (a, b) , we mean making legitimate changes in matrix symbols (network connections) in such a manner as to provide that there are a total of at most $n - 1$ different symbols in row a and column b . In the sequel we prove a theorem on the maximum number of such changes which are sufficient and necessary to unblock any connection.

2.3 Theorem 1

In order to unblock a blocked connection in a network, no more than $n - 1$ changes are required. For any $n > 1$, there are network states in which a connection is blocked that require $n - 1$ changes to unblock that connection.

Proof. Figures 2 and 3 are provided to aid the reader (and the author) in following the proof.

2.3.1 Sufficiency

Assume (r_1, c_1) is non-trivially blocked. This implies that *there is a symbol, say A , in column c_1 which does not appear in row r_1 .* Because if

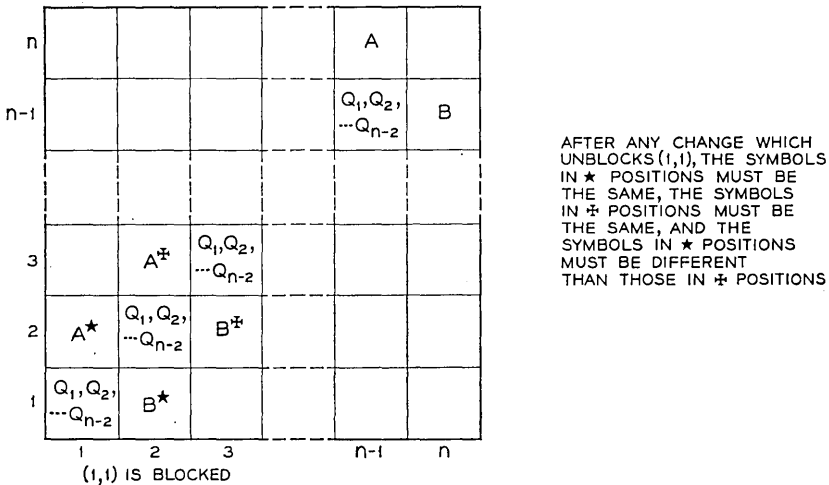


Fig. 3 — Matrix representation of the blocked network state which requires a maximum of changes to be unblocked.

there were no such symbol, then every symbol in column c_1 would also appear in row r_1 . And since all n symbols must appear in the union of column c_1 and row r_1 (condition for non-trivial blocking), it would follow that all n symbols appear in column c_1 making (c_1, r_1) trivially blocked, a contradiction of our hypothesis. Similarly *there must be a symbol, say B , in row r_1 which does not appear in column c_1 .*

Let A be in (r_2, c_1) .

Let B be in (r_1, c_2) .

Thus far we have completely defined:

- r_1 , the row in which the blocked connection appears,
- c_1 , the column in which the blocked connection appears,
- r_2 , the row in which the A in column c_1 appears, and
- c_2 , the column in which the B in row r_1 appears.

Now we wish to define other rows and columns:

- r_3 , the row in which an A appears in column c_2 if there is such a row (otherwise r_3 is undefined),
- c_3 , the column in which a B appears in row r_2 if there is such a column (otherwise c_3 is undefined),
- r_4 , the row in which an A appears in column c_3 if c_3 is defined and there is such a row (otherwise r_4 is undefined),
- c_4 , the column in which a B appears in row r_3 if r_3 is defined and there is such a column (otherwise c_4 is undefined).

In general, for all $j > 1$:

r_j is defined to be the row in which A appears in column c_{j-1} , provided c_{j-1} is defined, and provided that A does appear in column c_{j-1} . If not, r_j is undefined.

c_j is defined to be a column in which B appears in row r_{j-1} , provided r_{j-1} is defined, and provided that B does appear in row r_{j-1} . If not, c_j is not defined.

The above definition has the important property that if r_j and r_k are both defined, and $j \neq k$, then $r_j \neq r_k$. Also, if c_j and c_k are both defined, and $j \neq k$, then $c_j \neq c_k$. This is justified by the following argument: consider the sequence

$$r_1, c_1, r_2, c_2, \dots, r_i, c_i, \dots, r_n, c_n. \quad (1)$$

Assume there is a first member equal to a previous member of the sequence.

This is either a row or column.

1. Assume row r_j is the first member of the sequence which is both defined and the same as a previous defined member, say r_k , $k \neq j$. First of all k cannot be 1 since r_j , $j > 1$ is defined to have an A in it, and row r_1 has no A in it. So $r_1 \neq r_j$, $j > 1$. Now then assume $k > 1$, $j > 1$,

$j \neq k$. Then an A appears in (r_k, c_{k-1}) , and in (r_j, c_{j-1}) , (by our definition of r_j). So unless $c_{k-1} = c_{j-1}$, $j - 1 \neq k - 1$, there would be two different A 's in row $r_k = r_j$. There cannot be two different A 's in a single row. Therefore $c_{k-1} = c_{j-1}$. But this contradicts the assumption that the first member having this property is row r_j . That leaves only the possibility of column c_j being the first such member.

2. Then assume column c_j is the first member of the sequence which is both defined and the same as a previous defined member, say column c_k $k \neq j$. Then $k \neq 1$, because c_1 has no B , and $c_j, j > 1$ does by definition of c_j . If $k > 1, j > 1, k \neq j$ and $c_j = c_k$, then for similar reasons to those of the above paragraph $r_{j-1} = r_{k-1}$. Therefore our second assumption, $c_j = c_k$, is also contradicted, completing the proof.

Having shown that the defined members of (1) are distinct, we wish now to examine this sequence further. For convenience it is rewritten below.

$$r_1, c_1, r_2, \dots, r_j, c_j, \dots, r_n, c_n \tag{1}$$

There is a first member of this sequence which is defined but whose succeeding member is undefined, say c_f . Since r_{f+1} is the first member of the sequence which is undefined, it follows from the definition of r_j that there is no A in column c_f . We then know that according to our definition of c_j and r_j the following matrix positions contain A 's

$$(r_2, c_1); (r_3, c_2); \dots (r_f, c_{f-1})$$

and the following contain B 's

$$(r_1, c_2); (r_2, c_3); \dots (r_{f-1}, c_f)$$

Now in order to unblock (r_1, c_1) we make the following changes.

2.3.2 Change Algorithm

Change the original B 's to A 's in columns $c_j; j = 3, 5 \dots f$ if f is odd (or in columns $j = 2, 4 \dots$ if f is even). This involves changing B 's to A 's in rows $r_j; j = 2, 4 \dots f - 1$ if f is odd ($j = 1, 3 \dots f - 1$ if f is even).

Change the original A 's to B 's in rows $r_j, j = 2, 4 \dots f - 1$, if f is odd ($j = 3, 5 \dots f - 1$ if f is even). This involves changing A 's to B 's in columns $c_j; j = 1, 3, \dots f - 2$, if f is odd ($j = 2, 4 \dots f - 2$, if f is even). Note that the total number of changes is $f - 1$.

We see that if f is odd then after the change (r_1, c_2) will still contain a B , but (r_2, c_1) which formerly contained an A now contains a B . Therefore an A may now be legitimately placed in (r_1, c_1) . A similar argument holds if f is even.

It remains to show that the changes we have prescribed do not lead to any conflicts. For this demonstration assume f is odd. (A similar argument holds for f even.) The only conflicts possible must involve A 's and B 's, since these are the only symbols changed and resulting from the change. Furthermore, the only conflicts possible are in rows $r_1, r_2 \cdots r_f$ or columns $c_1, c_2 \cdots c_f$, since these are the only rows and columns in which changes were made. Also, at most one A has been added to any row or any column. Similarly, at most one B has been added to any row or column. Before the change there were single A 's in columns $c_j, j = 1$ to $f - 1$, and in rows $r_j, j = 2$ to f . As a result of the change single A 's were added to columns $c_i, i = 3, 5 \cdots f$ and no others, and to rows $r_i, i = 2, 4 \cdots f - 1$ and no others. So it is only these columns and rows which could possibly contain more than one A . But these columns and rows each contain only a single A , because although an A has been added to each, the original A in each has been changed to a B . For according to our prescribed changes, the original A 's in $c_i: i = 1, 3 \cdots f - 2$ were changed to B 's. This takes care of all columns to which an A was added except column c_f , and column c_f did not originally contain an A . Also, the original A 's in rows $r_i: i = 2, 4 \cdots f - 1$ were changed to B 's and this takes care of all rows to which an A was added.

Again as a result of the change, single B 's were added to rows $r_i: i = 2, 4, 6, \cdots f - 1$, and to columns $c_i: i = 1, 3, 5 \cdots f - 2$. It is therefore only those columns and rows which could have more than one B . But the original B 's in columns $c_j, j = 3, 5 \cdots f$ have been changed to A 's. This takes care of all columns to which a B was added except column c_1 , and column c_1 originally did not have a B . Also, the original B 's in rows $r_i: i = 2, 4 \cdots f - 1$ were changed A 's and this takes care of all rows to which a B was added.

If all members of sequence 1 are defined, then c_n is the last defined member, and there cannot be an A in c_n , because such an A would have to be in some row other than row r_1 . There are A 's in all rows other than r_1 , but none of these A 's are in c_n . This follows from the definition of r_j . From here, then, our argument goes on as the general case in which r_{f+1} was the first undefined member of sequence (1).

Thus the maximum number of changes required to unblock a call is $n - 1$.

2.3.3 Necessity

The network has n intermediate switches which we represent by the symbols $A, B, Q_1 \cdots Q_{n-2}$. Assume that (1, 1) is blocked by the follow-

ing network state:

$(i, i); i = 1$ to $n - 1$ each contain all the symbols Q_1, \dots, Q_{n-2} .

$(i, i + 1); i = 1$ to $n - 1$ each contain the symbol B .

$(i + 1, i); i = 1$ to $n - 1$ each contain the symbol A .

There are no other symbols in the matrix. To unblock $(1, 1)$ the symbols in $(1, 2)$ and $(2, 1)$ must be made the same because:

(a) After any change there must still be $n - 2$ different symbols in $(1, 1)$

(b) There must still be one symbol in $(1, 2)$ different from all those in $(1, 1)$

(c) There must still be one symbol in $(2, 1)$ different from all those in $(1, 1)$

(d) If then the symbols in $(1, 2)$ and $(2, 1)$ were different, there would be a total of n symbols in row 1 and column 1, leaving no symbol available to unblock $(1, 1)$.

Assume that the symbols in $(i + 1, i)$ and $(i, i + 1); i = k - 1$ must be the same, say X , in order to unblock $(1, 1)$. Now $(i, i); i = k$, which is in row k must, after the change, still contain $n - 2$ different symbols, say $Y_1, Y_2 \dots Y_{n-2}$. The symbol X in $(i + 1, i); i = k - 1$ which is also in row k must be different from $Y_1, Y_2 \dots Y_{n-2}$. Therefore the symbol in $(i, i + 1); i = k$ which is also in row k must be different than $X, Y_1, Y_2 \dots Y_{n-2}$. There is only one symbol that can be different from all $n - 1$ different symbols $X, Y_1 \dots Y_{n-2}$, say Z . So Z must appear in $(i, i + 1), i = k + 1$. Similarly as stated previously $(i, i), i = k$, which is in column k , must still have the $n - 2$ different symbols $Y_1, Y_2 \dots Y_{n-2}$. Also in column k the symbol X is in position $(i, i + i), i = k - 1$. Therefore it follows that the symbol in $(i + 1, i), i = k$, which is also in column k must be different than X, Y_1, \dots, Y_{n-2} , and must be Z .

Hence the induction is complete, proving that if $(i, i) i = 1$ to $n - 1$ each contain $n - 2$ different symbols (this must be true because of the given initial network state), and $(1, 1)$ is to be unblocked, then for each $i = 1$ to $n - 1$, the pair $(i + 1, i)$ and $(i, i + 1)$ must contain the same symbol. Since initially $(i + 1, i)$ contained a different symbol from $(i, i + 1)$ for $i = 1$ to $n - 1$, at least $n - 1$ changes are necessary to put the network in a state both equivalent to its initial state, and in which $(1, 1)$ is unblocked.

III. COMPARISON WITH SLEPIAN'S RESULT

I have been able to obtain the bound of $n - 1$ on the number of changes by considering changes of both input and output connections involved in the blocked connection, that is changes in both rows and

columns of our matrix. Slepian, on the other hand, considered, in effect, only the changes in rows. That is, he gave a method for changing the blocking symbol in a row (B in our proof) without taking advantage of the symbol in the column (A in our proof) for reducing the number of changes.

In the following sections a number of generalizations are developed.

IV. METHODS FOR RESWITCHING A NETWORK TO UNBLOCK CALLS

In the proof of Theorem 1 there is a method given for determining the changes required to unblock a blocked connection. This method involves two second-stage switches (A and B in the proof). If (r_1, c_1) is blocked, we look for a symbol in r_1 not in c_1 , and a symbol in c_1 not in r_1 , and carry out the Change Algorithm of Theorem 1 (Section 2.3.2). We will call this "method 1." We could use a slightly more complex method in which we test all symbol pairs, (A, B) , such that A is in r_1 but not in c_1 , and B is in c_1 but not in r_1 , to find which pair will require the fewest changes. The changes are then made on this pair according to the change Algorithm. We will call this "method 2." Methods 1 and 2 both involve changing only two second-stage switches. We can develop methods which are not restricted to changes of only two second-stage switches.

Assume that (r_1, c_1) is blocked. Assume that A is in r_1 , but not in c_1 and B is in c_1 but not in r_1 . As in the proof of the theorem, assume c_f is the first member of sequence (1), which is itself defined but whose succeeding member is not defined. Then by making $f - 1$ changes of A 's and B 's, we know (r_1, c_1) could be unblocked. If, however, some of the A 's or B 's which serve to define c_j , and $r_j, j < f$, could be changed *without conflict* to a symbol other than A or B , to C for example, then we could unblock (r_1, c_1) in less than $f - 1$ changes. This is best illustrated by an example (see Fig. 4). In summary, method 3 involves: first, finding a symbol in r_1 not in c_1 , say A ; a symbol in c_1 , not in r_1 , say B ; second, finding the last defined term of sequence (1); and third, examining the A 's and B 's which define sequence (1) to determine if any can be changed to a symbol other than A or B . If not, change the B 's to A 's according to the Change Algorithm. If so (say a B in column $k, k < f$, can be changed to a C without conflict) then make this change and make all changes given by the change algorithm in columns c_j , for $j < k$ rows r_j for $j < k$.

In method 4 we try method 3 on all pairs of symbols (A, B) ; A in r_1 not in c_1 ; B in c_1 not in r_1 , and actually carry it out on the pair which requires the fewest changes.

The methods discussed vary in complexity. A legitimate question to

ask is, what do we get for this complexity? The least upper bound on the number of changes required to unblock a blocked call has been established and is independent of which of the four methods is used. The greater complexity, however, does serve to decrease the average number of changes required per blocked call. We can get a quantitative idea of the value of the different methods by finding for any number of changes required to unblock a call, a lower bound on the number of calls which must already be set up in the network for each of the four methods.

These bounds are illustrated by examples in which it can be seen that the removal of any call will lower the number of changes required. These can be shown to be greatest lower bounds (Fig. 5)

x = the number of changes required

y = the number of calls already set up.

For methods 1 and 2

$$y = 2x + n - 2 \quad \dots \text{bound 1 (See Fig. 5a).}$$

For methods 3 and 4

$$y = 2x + \frac{x}{2}(n - 2) \quad \text{for } x \text{ even}$$

$$y = 2x + \frac{x + 1}{2}(n - 2) \quad \text{for } x \text{ odd} \quad \dots \text{bound 2 (See Fig. 5b).}$$

These bounds do not indicate the difference between methods 1 and 2, or between methods 3 and 4, because as far as these bounds are concerned there is no difference. However, Fig. 6 indicates a case in which

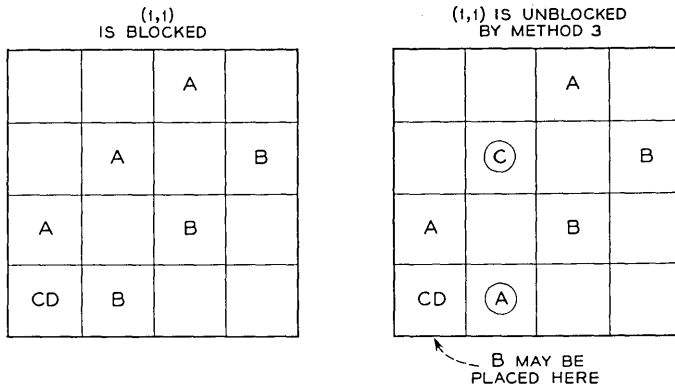


Fig. 4 — Illustration of change method 3.

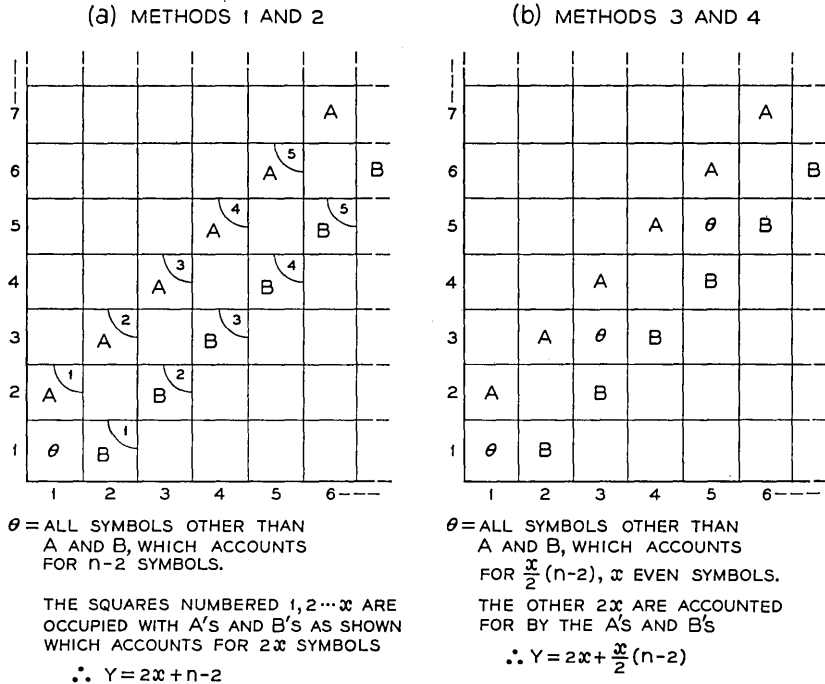


Fig. 5 — Illustrations of the smallest number of calls which must occupy a network (x) if a call is blocked and x changes are required to unblock the call.

method 2 or 4 requires (assuming A and B were changed) one change to unblock (1, 1), whereas methods 1 and 3 require five.

There are more complex methods possible, in which more complex changes are allowable than any of the four methods, discussed above. Fig. 7 indicates how a network state which requires four changes with method 4 could be unblocked with three changes. It would be interesting to obtain the general bound, equivalent to bounds 1 and 2 in the case in which no restriction was made on possible changes.

A 704 program for simulating method 2 on a simulated four-stage network is being written by J. Nervik. Also it should be fairly simple to realize circuitry for any of the four methods described.

V. GENERALIZATION TO MORE THAN THREE STAGES

In the proof of Theorem 1 we were able to show how the necessary changes to unblock a blocked connection can be made disturbing at most only two second-stage switches. This gives the following corollary.

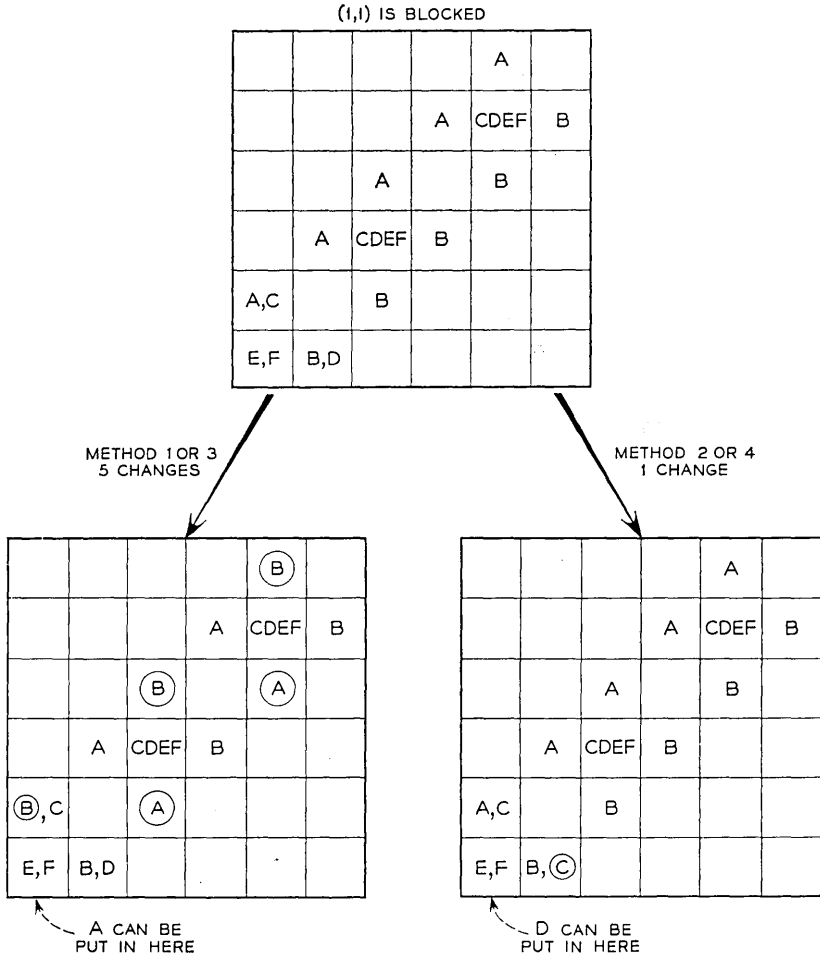


Fig. 6 — Illustration of the advantage of methods 2 and 4 over methods 1 and 3 respectively.

5.1 Corollary 1

A blocked connection may be unblocked by changing connections in such a way as to disturb no more than two second-stage switches.

This corollary will serve to obtain bounds on the number of calls which are disturbed in unblocking a blocked connection for five, seven, and nine-stage networks.

The five-stage network to which we refer is obtained by starting with

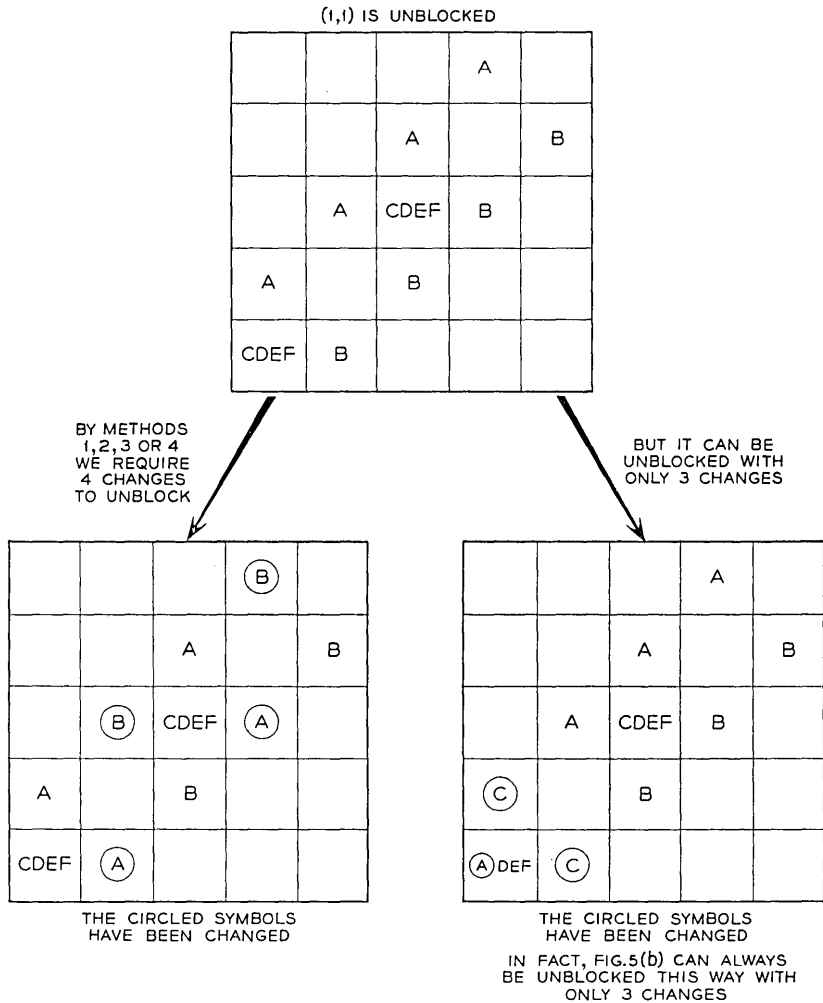


Fig. 7 — A change method more efficient than method 4.

a three-stage network as in Fig. 1. Each second-stage switch of this network is then expanded into a three-stage switch having \sqrt{n} input 1st, 2nd, and 3rd stage switches (Fig. 8).

Now suppose a three-stage switch is blocked. We can find the two switches (say *A* and *B*) in which changes can be made by the Change Algorithm. We calculate the changes which must be made in *A* and *B* by the same algorithm. When we are done we have a list of all the connections which must finally be made in second stage switches *A* and *B*. Now

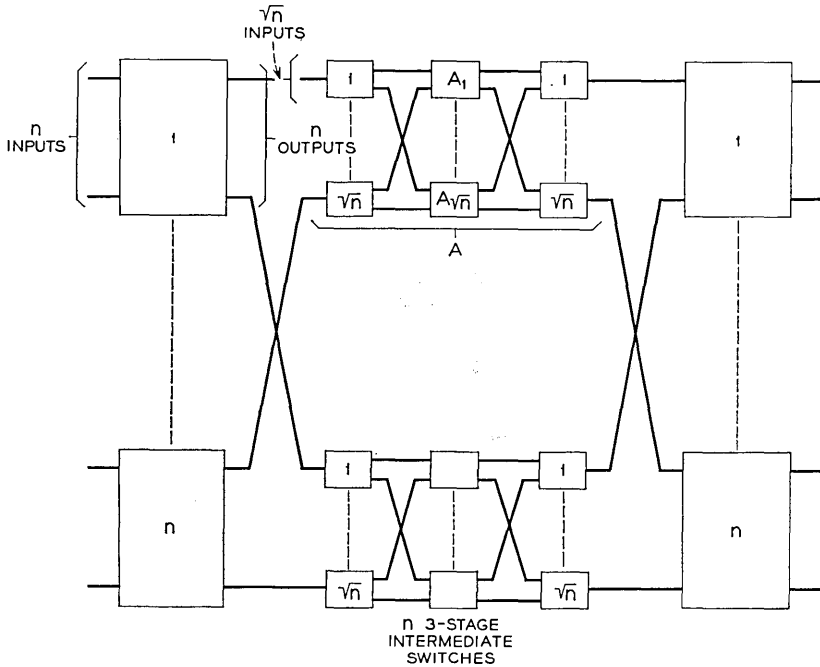


Fig. 8 — Five-stage network suitable for reswitching.

all connections which are initially set up in switches A and B may be taken down, and the new set of connections put up in their place.

If A and B are the two switches which are changed by the Change Algorithm, there can be no more than $n - 1$ connections in either of them, since by definition there is no A in c_1 , and no B in r_1 , assuming (r_1, c_1) is blocked. Since we take down all connections in A and B , no more than $2n - 2$ calls are disturbed.

If A and B were three-stage switches themselves, our network and the above argument would remain the same. The only question which might arise is whether we could make the final connections in three-stage switches A and B . These connections could be made. One would calculate exactly how to set them up by repeated application of the Change Algorithm. First one would conceptually set up one connection arbitrarily, then one would try to conceptually set up the next connection. If it were blocked, the Change Algorithm could be used to unblock it. This would continue until the final connections were decided conceptually. Then they could actually be made.

This result extends easily to seven, nine \dots stage switches.

5.2 Generalization of Theorem 1

For a q -stage network, q odd and $q > 3$, of the type described above (Fig. 8), no more than $2n - 2$ calls need be disturbed to unblock a blocked call.

This bound can probably be lowered. If A and B are the two second-stage switches in which connections are to be changed, it has been shown that no more than a total of $n - 1$ connections in both A and B need be changed. However, if A and B are themselves three-stage switches, in order to make the initial $n - 1$ changes it may be necessary to juggle other connections in both A and B . A closer study of the ways in which this juggling can be done might serve to lower the $2n - 2$ bound.

5.3 Generalizations to Other Network Configurations

So far we have discussed networks in which all stages have the same number of switches. The matrix representation, with the restrictions given in Section 2.1, is applicable to a more general class of three-stage networks than that of Fig. 1. In this more general class the numbers of

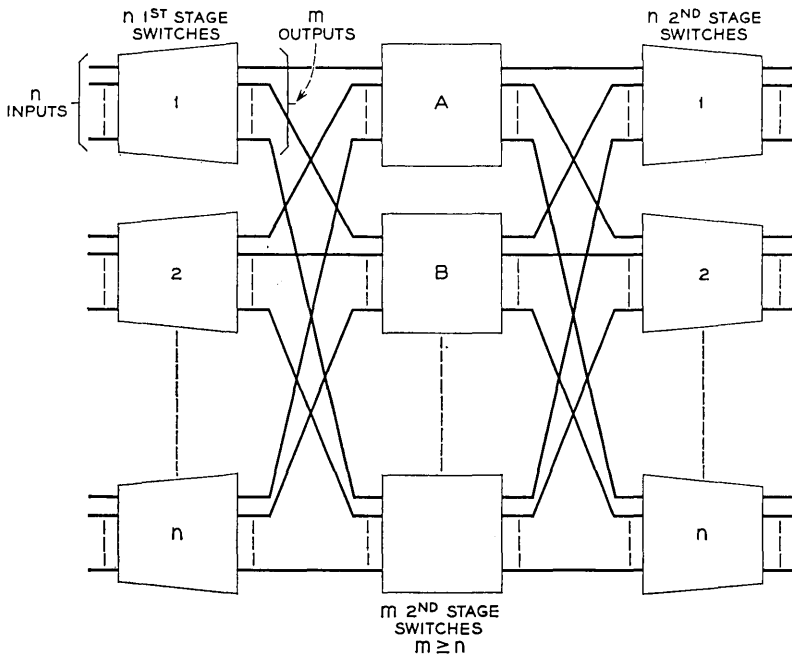


Fig. 9 — Generalization to more second-stage switches than first-stage switches.

intermediate switches may be greater than n (Fig. 9). Assume there are m intermediate switches, $m > n$. There are still n input (output) switches. There is one link from each input (output) switch to each of the intermediate switches.

For this more general class of switches the condition for legitimate blocking (Section 2.1) must be generalized to read:

(a, b) is legitimately blocked if there are a total of m different symbols in row a and column b . There are less than n symbols in row a (column b).

We have shown that if $m = n$, then no more than $n - 1$ changes are required to unblock a blocked connection. Clos² has shown that if $m = 2n - 1$, the network is nonblocking. (0 changes are required to unblock a blocked connection.) I can also prove that if $m = 2n - 2$, no more than one change is required to unblock a blocked connection. (This is justified later.) These results lead to the conjecture *that if $m = 2n - j$, no more than $j - 1$ changes are required to unblock a blocked connection.*

We will now prove a simple lemma which was useful in finding the bounds for $m = 2n - 2$, and which may prove helpful in attacking our conjecture.

5.3.1 Lemma

If $m = n + k$ and (a, b) is legitimately blocked, then there must be at least $k + 1$ symbols in row a , none of which are in column b , and there must be at least $k + 1$ symbols in column b , none of which are in row a .

5.3.2 Proof

By assumption, (a, b) is blocked; therefore there are a total of $n + k$ different symbols in row a and column b , by the blocking condition. There are less than n symbols in row a , and less than n symbols in column b , also by the blocking condition.

Let R = no. of symbols in row a , not in column b

C = no. of symbols in column b , not in row a

X = no. of symbols appearing in (a, b)

B = no. of symbols in both row a , and column b , but not (a, b)

Then we have

1. $B + C + B + X = n + K$
2. $R + B + X < n$
3. $C + B + X < n$

which by substituting for X in 2, and 3 the value obtained from 1 gives:

$$\begin{aligned} C &< k \\ R &< k \end{aligned}$$

We illustrate the use of this lemma by proving the Clos non-blocking network is non-blocking. In the Clos network $k = n - 1$. Therefore if (a, b) is legitimately blocked, there must be $n = k + 1$ symbols in row a ($C = n$). This immediately contradicts the hypothesis that (a, b) is legitimately blocked.

For the case in which $k = n - 2$, it follows from our lemma that if (a, b) is blocked, $C = n - 1$ and $R = n - 1$. If this were all the connections that were up, a single change of any of the C symbols in row a (called a -symbols) or any of the R symbols in column b (called b -symbols) would unblock (a, b) . So, in order that more than one change will be required, all the proposed unblocking changes must produce conflicts. This means that in the column of each of the symbols in row a , all $n - 1$ b -symbols must appear. Also in the row of each of the symbols in column b , all $n - 1$ a -symbols must appear. It follows that if there are to be no more than n symbols in any row or column that there must be no symbol in (a, b) , and one symbol in every other location in row a , and column b . Now we look at row k . There must be one b -symbol and all $n - 1$ a -symbols in this row. One of the a -symbols in row k must be in column p , ($p \neq a$). In column p however, there must already be an a symbol and $n - 1$ b -symbols, these together with the a -symbol in row k , column p total to $n + 1$ symbols in column p . This is not allowed. Therefore one change will always be sufficient to unblock a blocked connection of $m = 2n - 2$.

VI. CONCLUSION

There are other directions in which generalizations appear feasible with the techniques of this paper. We can deal with rectangular matrices in an analogous manner to that used for the square matrix here. These correspond to concentration networks. Triangular networks seem somewhat more difficult, but still feasible to treat. Finally, results on various network configurations can probably be generalized to more than three stages.

We have discussed here the use of the reswitching to make networks non-blocking. One might also consider a more modest goal in which provision is made for fewer than the number of reswitches or changes required to make the network non-blocking, in an attempt to improve blocking characteristics. The program being written by J. Nervik will be used to obtain some estimate of this improvement.

APPENDIX

A.1 *Introduction*

In the body of the paper there are some algorithms by which networks not originally non-blocking can be made non-blocking by rearranging connections already set up in the network. These algorithms involve the temporary disturbance of calls already set up in the network.

Here I propose to describe a slight modification of the network and of the algorithm which will allow one to make the network essentially non-blocking or rearrangeable without requiring any disturbance of existing calls set up in the network.

A.2 *Network Modification*

The basic three-stage network, each stage requiring n , $n \times n$ switches, is modified to a three-stage network in which stage 1 consists of n , $(n) \times (n + 1)$ switches, stage 2 consists of $n + 1$, $(n) \times (n)$ switches, stage 3 consists of n , $(n) \times (n + 1)$ switches. Each first (third) stage switch has one link to each second-stage switch, as pictured in Fig. 10.

As in the body, we represent the connections in this network with a $n \times n$ matrix. For each input switch, there is a row in the matrix. These are numbered 1 to n . For each output switch, there is a column in the matrix. These are numbered 1 to n . A connection between input switch j and output switch k through middle switch A is indicated by an A in position (j, k) . Middle switches are lettered A, B , etc. There are $n + 1$ letters. There cannot be more than $n + 1$ letters* in any row or column or location of the matrix.

A.3 *Algorithm Modification*

In order to make this network essentially non-blocking we use the following procedure.

We choose not to use middle switch A until we get a blocked condition. When we get a blocked condition, we are in the same situation as if the network were a three-stage network with each stage having $n \times n$ switches. From the corollary of Section V, we know that this blocked connection could be unblocked without disturbing more than two (intermediate) middle switches (not including switch A , which has not as yet been used). Suppose the two middle switches in which connections are to be changed to unblock the blocked connection are C and D . According to the change algorithm we would change a certain set of connections

* In the operation of this network there are times during which a single input lead is connected to two middle switches. Thus, although there are only n inputs per input switch there may be $n + 1$ connections in a single input switch.

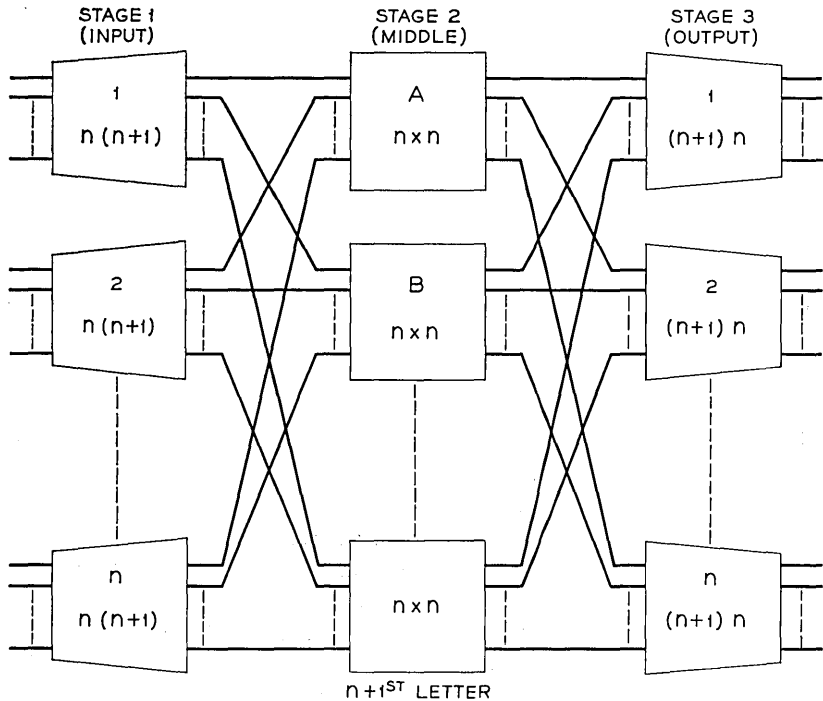


Fig. 10 — Modified three-stage network.

in middle switch C to connections in middle switch D , and a certain set of connections in middle switch D to connections in middle switch C . This would leave either switch C or D available for use for the blocked call. (Which particular switch was available depends on the exact choice of the sets of connections in C and D which are to be rearranged.) Such a rearrangement involves disturbing all the calls using the set of connections in C and D which are to be changed. In the modified network we have an extra middle switch A available which we can use to maintain all calls while connections are being rearranged. The modified algorithm is given below. The steps of the algorithm are illustrated by an example in Figure 11.

According to the algorithm, we find the set of connections in middle switches C and D which must be rearranged to unblock the blocked connections.

1. For every connection in C which we have elected to change, we add a corresponding connection in A . In the matrix representation this

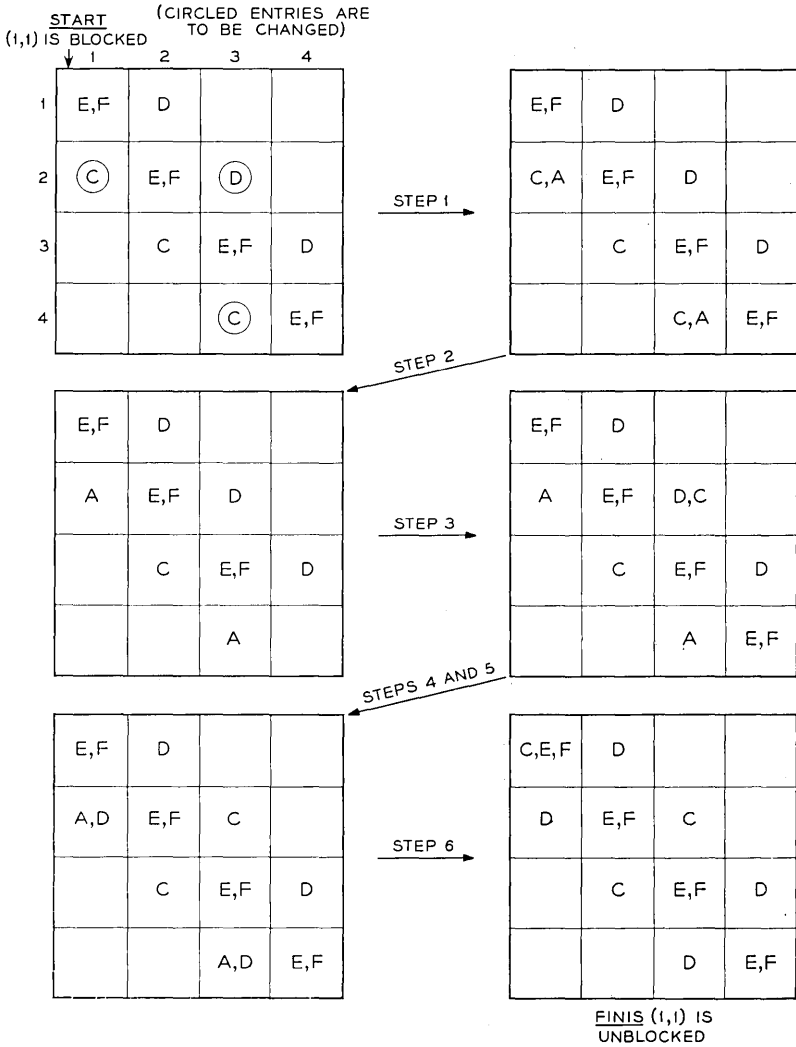


Fig. 11 — Example of rearranging without disturbing.

corresponds to adding an *A* in every position in which an elected *C* appears.

2. All elected *C* connections are taken down; the calls originally carried by these connections are now carried by switch *A*.

3. For every connection in *D* which we originally elected to change,

we add a corresponding connection in C . This means in the matrix representation an addition of a C in every position in which there is an elected D . This will always be possible because we have already taken down all the elected C connections which, according to the body of the paper, are the only connections which would prevent this addition.

4. All elected D connections are taken down; the calls carried by these D connections are now carried by C connections.

5. To every connection carried by A we add a corresponding connection carried by D . This will always be possible according to the results in the body of the paper.

6. Finally, we take down all connections in A .

We thus have carried out the change algorithm, and therefore, have unblocked the blocked connection. Switch A has no connection set up in it so it is available for use in unblocking the next blocked call.

A.4 *Making Use of Dead Time* (Time during which there is no activity in the network)

Actually, the blocked call is unblocked after step 2, at which time the desired connection can be put up using switch C (assuming that the C in the row or column of the blocked call was changed in step 1 and step 2). Since either a C or D in the row or column of the blocked call is changed in this algorithm the original choice of where to add connections in A could be made so that the call would be unblocked after step 2.

Steps 1 and 2 could have been combined in such a way as to add a connection in A , take down the corresponding connection in C , then add another connection in A , take down the corresponding connection in C , and so forth. In this case the blocked connection could be unblocked after the first addition of a connection in A and the removal of its corresponding connection in C .

If the single switch A is to serve to allow for unblocking calls without disturbing other calls, by our algorithm, A must be completely available when a blocked connection is to be unblocked. So although the blocked call may be unblocked early in the algorithm, the remainder of the algorithm must have been completed before the next blocked call is to be unblocked. The extra switch A acts as a kind of connection memory so that normal dead time in the network may be profitably used for improving blocking characteristics.

By adding additional middle switches analogously to the way A was added, we would effectively add more connection memory and increase the efficient use of dead time in the network. Thus, if there were two additional switches A and B , A would not have to be cleared before

another blocked call could be handled because B would be available. A would have to be cleared before the second blocked call after the one that engaged A were encountered.

There are some indications that the simple scheme proposed here can be improved upon.

In this scheme we are calling on A to handle no more than $n/2$ connections at any one time, and A is in use at all only during the unblocking operation. In the network A appears like any other middle-stage switch, but its function is much different from the other middle-stage switches. Perhaps the organization could be changed to share the load more symmetrically.

REFERENCES

1. Slepian, D., unpublished memorandum, March 25, 1952.
2. Clos, C., B.S.T.J., **32**, March, 1953, pp. 406-424.

The Realizability of Multiport Structures Obtained by Imbedding a Tunnel Diode in a Lossless Reciprocal Network

By I. W. SANDBERG

(Manuscript received November 9, 1961)

Necessary and sufficient conditions are presented for the realization of the short-circuit admittance matrix or open-circuit impedance matrix of the most general n -port structures characterized by such matrices obtained by imbedding a tunnel diode, represented by a parallel combination of a capacitor and a negative resistor, in a finite lossless reciprocal network. Techniques for realizing prescribed immittance matrices are included.

I. INTRODUCTION

It is generally well known that the tunnel diode possesses a small-signal equivalent circuit that can often be approximated by a parallel combination of a capacitor and a negative resistor. This model has been used extensively in the study of gain-bandwidth relations and optimum synthesis procedures for specific amplifier configurations.¹⁻⁵ It has also been used to derive bounds on the natural frequencies obtained by imbedding the tunnel diode in a passive network.^{6,7}

The purpose of this paper is to present necessary and sufficient conditions for the realization of the short-circuit admittance matrix or open-circuit impedance matrix of the most general n -port structures characterized by such matrices obtained by imbedding a tunnel diode, represented by the above mentioned model, in a lossless reciprocal network.

The properties of the short-circuit admittance matrix are considered also by another writer.⁸ With the exception of certain remarks of a tutorial nature, the arguments, results, synthesis techniques, and basic approach to the problem presented here are quite different from that in Ref. 8. In particular, it is not assumed here that the short-circuit admittance matrix of the $(n + 1)$ -port lossless network invariably exists.

Also, the necessary and sufficient conditions are stated directly in terms of the $n \times n$ short-circuit admittance matrix and its even part. They do not involve a knowledge of the short-circuit admittance matrix obtained when the tunnel diode is short-circuited.

II. DESCRIPTION OF THE STRUCTURE TO BE CONSIDERED

The basic structure under consideration is shown in Fig. 1, in which the $(n + 1)$ -port network is assumed to be a lossless reciprocal configuration containing inductors, capacitors, and ideal transformers. Port $(n + 1)$ is terminated with a unit capacitor and unit resistor in parallel. This involves no loss of generality since a similar termination with other values of positive capacitance and/or resistance (positive or negative) can be treated with the aid of simple transformations which are explicitly stated in Section VII. The overall network is restricted initially in that the symmetric positive-real short-circuit admittance matrix $\mathbf{Y}(s)$, relating the port currents and voltages at ports $(1, 2, \dots, n)$, is assumed to exist. The realizability conditions for the open-circuit impedance matrix $\mathbf{Z}(s)$ can be obtained in a manner similar to that to be described for $\mathbf{Y}(s)$ and are stated in Section VII.

The $(n + 1)$ -port lossless network is characterized by the regular para-unitary scattering matrix $\hat{\mathbf{S}}(s)$ or by the short-circuit admittance matrix $\hat{\mathbf{Y}}(s)$, when it exists. We initially assume that $\hat{\mathbf{Y}}(s)$ does exist and consider in a subsequent section the case in which $\hat{\mathbf{Y}}(s)$ does not exist.

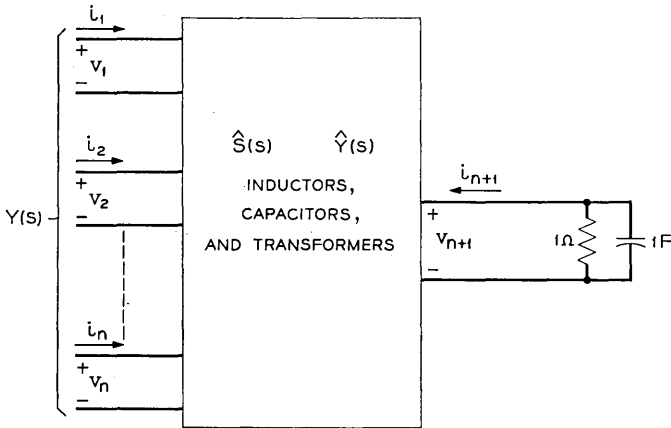


Fig. 1 — Most general structure defining $\mathbf{Y}(s)$.

III. NECESSARY CONDITIONS FOR THE REALIZATION OF $\mathbf{Y}(s)$ WHEN $\hat{\mathbf{Y}}(s)$ EXISTS AND THE EVEN PART OF $\mathbf{Y}(s)$ IS NOT A MATRIX OF CONSTANTS

The necessary and sufficient conditions for the realization of $\hat{\mathbf{Y}}(s)$ are, of course, well known.

It is also well known that†

$$\mathbf{Y} = \mathbf{Y}_{11} - \mathbf{Y}_{12}\mathbf{Y}_{12}^t \frac{1}{Y_{22} + s + 1} \tag{1}$$

where the matrices in (1) are defined by the following partition of $\hat{\mathbf{Y}}(s)$:

$$\hat{\mathbf{Y}} = \begin{matrix} n & 1 \\ \left[\begin{matrix} \mathbf{Y}_{11} & \mathbf{Y}_{12} \\ \mathbf{Y}_{12}^t & Y_{22} \end{matrix} \right] & \begin{matrix} n \\ 1 \end{matrix} \end{matrix} \tag{2}$$

The arguments to be presented center about a study of \mathbf{Y}_e , the even part of the matrix \mathbf{Y} . This matrix is given by

$$\begin{aligned} \mathbf{Y}_e &= \frac{1}{2}[\mathbf{Y}(s) + \mathbf{Y}(-s)] \\ &= -\mathbf{Y}_{12}\mathbf{Y}_{12}^t \frac{1}{[Y_{22} + s + 1][-Y_{22} - s + 1]}. \end{aligned} \tag{3}$$

It is convenient to introduce the notation: $Y_{22} = d^{-1}n_{22}$, $\mathbf{Y}_{12} = d^{-1}\mathbf{N}_{12}$ where d is an even polynomial, n_{22} is an odd polynomial and \mathbf{N}_{12} is a matrix of odd polynomials, with the understanding that d , n_{22} and every element in \mathbf{N}_{12} may have a common simple zero at the origin. In this way it is unnecessary to treat separately the cases in which d is even or d is odd. Accordingly,

$$\mathbf{Y}_e = -\mathbf{N}_{12}\mathbf{N}_{12}^t \frac{1}{[n_{22} + (s + 1)d][n_{22}(-s) + (-s + 1)d]}. \tag{4}$$

Note that the polynomial $[n_{22} + (s + 1)d]$ can be assumed to be strictly Hurwitz except possibly for a simple zero at the origin, since n_{22} and d can be assumed to be relatively prime except possibly for a simple common zero at the origin.

It is convenient to treat separately the cases in which \mathbf{Y}_e is or is not a matrix of constants.

† The superscript t denotes matrix transposition.

Consider the following

Definition:

The matrix \mathbf{Y}_e is said to be in standard form if and only if

$$\mathbf{Y}_e = -\mathbf{U}\mathbf{U}^t \frac{1}{v(s)v(-s)}$$

where $v(s)$ is a positive coefficient polynomial which is strictly Hurwitz except possibly for a simple zero at the origin and $\mathbf{U}^t = [u_1, u_2, \dots, u_n]$ is a matrix of odd real polynomials with the property that there is no factor $\eta(s)\eta(-s)$ common to all the u_i such that $\eta^2(s)\eta^2(-s)$ divides $v(s)v(-s)$ where $\eta(s)$ is a strict Hurwitz polynomial. The polynomials v_e and v_o are respectively the even and odd parts of $v(s)$.

In Section IV the following result is proved.

Theorem 1:

A rational positive-real symmetric matrix $\mathbf{Y}(s)$ with nonconstant even part is realizable as shown in Fig. 1 when $\hat{\mathbf{Y}}(s)$ exists only if \mathbf{Y}_e is expressible in standard form with $v(s)$ such that†

- i. $k = \left[\frac{v_o}{sv_e} \right]_{\infty} \geq 1$, and
- ii. If $k = 1$, $[\mathbf{Y}_e]_{\infty} = \mathbf{0}$,
- iii. If $k > 1$, $\left[\frac{1}{s} \mathbf{Y} \right]_{\infty} = \frac{k}{k-1} [\mathbf{Y}_e]_{\infty}$.

is nonnegative definite.

The case in which \mathbf{Y}_e is a matrix of constants is treated in Section VI.

IV. PROOF OF THEOREM I

We begin by observing from (4) that \mathbf{Y}_e can be expressed in standard form if \mathbf{Y} is realizable. The problem of factoring a given matrix \mathbf{Y}_e into the required form is discussed in detail in Appendix A.

Assume now that \mathbf{Y}_e is given in standard form and consider the problem of identifying \mathbf{N}_{12} , n_{22} , and d in (4). A common factor may have been canceled in the expression for \mathbf{Y}_e , and hence an unknown factor must be reinserted before \mathbf{N}_{12} , n_{22} , and d can be determined. However,

† Throughout we use the notation $\lim_{s \rightarrow \infty} [\cdot] = [\cdot]_{\infty}$.

the common factor must be of the form $a^2(s) = b(s)b(-s)$ where $b(s)$ is a strict Hurwitz polynomial. Therefore, ignoring a possible minus sign, $a(s) = \eta(s)\eta(-s)$ where $\eta(s)$ is a strict Hurwitz polynomial.

Thus, for some unknown strict Hurwitz $\eta(s)$,

$$\mathbf{N}_{12} \frac{1}{[n_{22} + (s + 1) d]} = \mathbf{U}\eta(s)\eta(-s) \frac{1}{v(s)\eta^2(s)} \tag{5}$$

and

$$\mathbf{N}_{12} = \mathbf{U}\eta(s)\eta(-s) \tag{6}$$

$$n_{22} + (s + 1) d = v(s)\eta^2(s). \tag{7}$$

In the following we shall denote by η_e and η_o the even and odd parts respectively of η . Equations (6) and (7) read

$$\mathbf{N}_{12} = \mathbf{U}[\eta_e^2 - \eta_o^2] \tag{8}$$

$$n_{22} + (s + 1) d = v_e(\eta_e^2 + \eta_o^2) + v_o(\eta_e^2 + \eta_o^2) + 2v_o\eta_e\eta_o + 2v_e\eta_e\eta_o. \tag{9}$$

Equating even and odd parts of (9) gives

$$\begin{aligned} d &= v_e(\eta_e^2 + \eta_o^2) + 2v_o\eta_e\eta_o \\ n_{22} &= 2v_e\eta_e\eta_o + v_o(\eta_e^2 + \eta_o^2) - s d \end{aligned} \tag{10}$$

and therefore

$$Y_{22} = \frac{n_{22}}{d} = \frac{v_o(\eta_e^2 + \eta_o^2) + 2v_e\eta_e\eta_o}{v_e(\eta_e^2 + \eta_o^2) + 2v_o\eta_e\eta_o} - s. \tag{11}$$

From (11) it is clear that Y_{22} is realizable provided

$$\left[\frac{1}{s} Y_{22} \right]_{\infty} \geq 0. \tag{12}$$

However,

$$Y_{22} = \frac{n_{22}}{d} = \frac{\frac{v_o}{v_e} + \frac{2\eta_e\eta_o}{\eta_e^2 + \eta_o^2}}{1 + \frac{2v_o}{v_e} \left[\frac{\eta_e\eta_o}{\eta_e^2 + \eta_o^2} \right]} - s \tag{13}$$

and since $\eta(s)$ is a Hurwitz polynomial†

$$\left[\frac{1}{s} Y_{22} \right]_{\infty} = \left[\frac{1}{s} \frac{v_o}{v_e} \right]_{\infty} \frac{1}{1 + \alpha} - 1 \tag{14}$$

† We have assumed that the degree of v_o exceeds the degree of v_e . It is easy to show that it is impossible to satisfy (12) unless this is so.

where

$$0 \leq \alpha = \left\{ \frac{2v_o}{v_e} \left[\frac{\eta_e \eta_o}{\eta_e^2 + \eta_o^2} \right] \right\}_\infty < \infty. \quad (15)$$

Clearly, it is necessary that

$$k = \left[\frac{1}{s} \frac{v_o}{v_e} \right]_\infty \geq 1. \quad (16)$$

4.1 Derivation of the Inequality Involving \mathbf{K}_∞ and $[\mathbf{Y}_e]_\infty$

Consider now the derivation of a key inequality that must be satisfied by the coefficient matrix $\mathbf{K}_\infty = [(1/s)\mathbf{Y}]_\infty$.

Let the constant matrices \mathbf{A}_{ij} be defined by

$$\mathbf{A}_{ij} = \left[\frac{1}{s} \mathbf{Y}_{ij} \right]_\infty. \quad (17)$$

Then, from (1),

$$\mathbf{K}_\infty = \left[\frac{1}{s} \mathbf{Y} \right]_\infty = \mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{12}^{-1} \frac{1}{A_{22} + 1}. \quad (18)$$

But from (8) and (10)

$$\mathbf{Y}_{12} = \frac{1}{d} \mathbf{N}_{12} = \frac{1}{v_e} \mathbf{U} \frac{\left[\frac{\eta_e^2 - \eta_o^2}{\eta_e^2 + \eta_o^2} \right]}{1 + \frac{2v_o}{v_e} \left[\frac{\eta_e \eta_o}{\eta_e^2 + \eta_o^2} \right]}$$

and thus

$$\mathbf{A}_{12} = \left[\frac{1}{sv_e} \mathbf{U} \right]_\infty \frac{\pm 1}{1 + \alpha} \quad (19)$$

where α is defined in (15) and the plus or minus sign applies according as the degree of η_e exceeds the degree of η_o or not. Recall from (14) and (16) that

$$A_{22} = \frac{k}{1 + \alpha} - 1. \quad (20)$$

Using (18), (19), and (20)

$$\mathbf{K}_\infty = \mathbf{A}_{11} - \left[\frac{1}{sv_e} \mathbf{U} \right]_\infty \left[\frac{1}{sv_e} \mathbf{U}^t \right]_\infty \frac{1}{k(1 + \alpha)} \quad (21)$$

where \mathbf{A}_{11} is unknown. However, if the $(n + 1)$ -port lossless network is to be realizable, it is necessary that the matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^t & A_{22} \end{bmatrix} \tag{22}$$

is nonnegative definite. Assume initially that $A_{22} \neq 0$. We require the following result.

Lemma 1:

A necessary and sufficient condition that \mathbf{A} is nonnegative definite with $A_{22} \neq 0$ is that \mathbf{A}' is nonnegative definite where

$$\mathbf{A}' = \begin{bmatrix} \mathbf{A}_{11} - A_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{12}^t & \mathbf{0} \\ \mathbf{0} & A_{22} \end{bmatrix}. \tag{23}$$

To prove the lemma, note that $\mathbf{A}' = \mathbf{B}\mathbf{A}\mathbf{B}^t$ with

$$\mathbf{B} = \begin{bmatrix} \mathbf{1}_n & -A_{22}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & 1 \end{bmatrix}$$

where $\mathbf{1}_n$ is the identity matrix of order n .

Thus,

$$\mathbf{A}_{11} - A_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{12}^t$$

or equivalently

$$\mathbf{A}_{11} - \left[\frac{1}{sv_e} \mathbf{U} \right]_{\infty} \left[\frac{1}{sv_e} \mathbf{U}^t \right]_{\infty} \frac{1}{(1 + \alpha)(k - 1 - \alpha)} \tag{24}$$

is required to be nonnegative definite. By combining this result with (21) we find that

$$\mathbf{K}_{\infty} - \frac{1}{k(k - 1 - \alpha)} \left[\frac{1}{sv_e} \mathbf{U} \right]_{\infty} \left[\frac{1}{sv_e} \mathbf{U}^t \right]_{\infty} \tag{25}$$

is nonnegative definite. Recalling that $(k - 1 - \alpha)$ is initially assumed to exceed zero, it is clear that (25) is nonnegative definite with $\alpha = 0$. Furthermore, with $\alpha = 0$, (25) can be expressed as

$$\mathbf{K}_{\infty} - \frac{k}{k - 1} [\mathbf{Y}_e]_{\infty} \tag{26}$$

by using the identities

$$\left[\frac{1}{sv_e} \mathbf{U} \right]_{\infty} = \left[\frac{v_o}{sv_e} \right]_{\infty} \left[\frac{1}{v_o} \mathbf{U} \right]_{\infty} \quad (27)$$

$$[\mathbf{Y}_e]_{\infty} = \left[\frac{1}{v_o} \mathbf{U} \right]_{\infty} \left[\frac{1}{v_o} \mathbf{U}^t \right]_{\infty}. \quad (28)$$

When $A_{22} = 0$, as is the case whenever $k = 1$, it is clear from (22) that every element in \mathbf{A}_{12} must vanish if \mathbf{A} is to be nonnegative definite. But from (19)

$$\mathbf{A}_{12} = k \left[\frac{1}{v_o} \mathbf{U} \right]_{\infty} \frac{\pm 1}{1 + \alpha}.$$

Thus, from (28) it is evident that $[\mathbf{Y}_e]_{\infty} = \mathbf{0}$ when $k = 1$.

This proves Theorem 1. In the next section we prove that if a positive-real matrix satisfies the conditions of Theorem 1, it is realizable as shown in Fig. 1.

V. PROOF OF SUFFICIENCY OF THEOREM I FOR $\mathbf{Y}(s)$ WITH NON-CONSTANT \mathbf{Y}_e

Assume that $\mathbf{Y}(s)$ and

$$\mathbf{Y}_e = -\frac{1}{v(s)v(-s)} \mathbf{U}\mathbf{U}^t, \quad (29)$$

in standard form, are prescribed and satisfy Theorem 1.

Let

$$Y_{22} = \frac{n_{22}}{d} = \frac{v_o}{v_e} - s \quad (30)$$

$$\mathbf{Y}_{12} = \frac{1}{d} \mathbf{N}_{12} = \frac{1}{v_e} \mathbf{U}.$$

Then from (30) and (1)

$$\mathbf{Y}_{11} = \mathbf{Y} + \frac{1}{v_e(v_o + v_e)} \mathbf{U}\mathbf{U}^t. \quad (31)$$

Hence \mathbf{Y}_{11} , \mathbf{Y}_{12} , and Y_{22} satisfy (1). We wish to prove that these submatrices defined above lead to a realizable $\hat{\mathbf{Y}}(s)$ given by

$$\hat{\mathbf{Y}}(s) = \begin{bmatrix} \mathbf{Y} + \frac{1}{v_e(v_o + v_e)} \mathbf{U}\mathbf{U}^t & \frac{1}{v_e} \mathbf{U} \\ \frac{1}{v_e} \mathbf{U}^t & \frac{v_o}{v_e} - s \end{bmatrix}. \quad (32)$$

First of all, Y_{22} is a realizable driving-point admittance function since $k \geq 1$.

The submatrix \mathbf{Y}_{11} can be expressed as follows by using (29):

$$\begin{aligned} \mathbf{Y}_{11} &= \mathbf{Y}_o - \frac{1}{v(s)v(-s)} \mathbf{U}\mathbf{U}^t + \frac{1}{v_e[v_o + v_e]} \mathbf{U}\mathbf{U}^t \\ &= \mathbf{Y}_o - \frac{v_o}{v_e} \frac{1}{v(s)v(-s)} \mathbf{U}\mathbf{U}^t. \end{aligned} \tag{33}$$

where \mathbf{Y}_o is the odd part of \mathbf{Y} . From (33) it is apparent that \mathbf{Y}_{11} is a matrix of odd functions, as it should be. Further, since from (31) \mathbf{Y}_{11} is regular in the right-half plane, it follows that \mathbf{Y}_{11} can have poles only on the $j\omega$ axis. In fact, the finite poles of \mathbf{Y}_{11} are the boundary poles of \mathbf{Y} and the zeros of v_e .

Consider now the residue matrix $\hat{\mathbf{K}}_i$ at a pole of $\hat{\mathbf{Y}}(s)$ which arises from a zero of v_e , say at $s = j\omega_i$, and let the residue matrix of \mathbf{Y} at that pole be \mathbf{K}_i . Then†

$$\hat{\mathbf{K}}_i = \left[\begin{array}{cc} \mathbf{K}_i + \frac{1}{\dot{v}_e v_o} \mathbf{U}\mathbf{U}^t & \frac{1}{\dot{v}_e} \mathbf{U} \\ \frac{1}{\dot{v}_e} \mathbf{U}^t & \frac{v_o}{\dot{v}_e} \end{array} \right]_{s=j\omega_i} \tag{34}$$

where a dot over v_e denotes the derivative of v_e with respect to s . To show that $\hat{\mathbf{K}}_i$ is nonnegative definite, we appeal to Lemma 1. Thus it is sufficient to point out that $(v_o/\dot{v}_e) |_{j\omega_i}$ is positive and that

$$\mathbf{K}_i + \frac{1}{\dot{v}_e v_o} \mathbf{U}\mathbf{U}^t \Big|_{j\omega_i} - \frac{1}{\dot{v}_e^2} \mathbf{U}\mathbf{U}^t \frac{\dot{v}_e}{v_o} \Big|_{j\omega_i} = \mathbf{K}_i \tag{35}$$

is nonnegative definite.

Finally, we must show that $\hat{\mathbf{K}}_\infty = [(1/s)\hat{\mathbf{Y}}]_\infty$ is nonnegative definite.

When $k = 1$, the proof is trivial for then

$$\left[\frac{1}{sv_e} \mathbf{U} \right]_\infty = \left[\frac{1}{v_o} \mathbf{U} \right]_\infty = \mathbf{0}$$

and

$$\hat{\mathbf{K}}_\infty = \begin{bmatrix} \mathbf{K}_\infty & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{36}$$

† When $v = su = s(u_e + u_o)$, where u is a strict Hurwitz polynomial, it is necessary to replace v_e , v_o , and \mathbf{U} respectively with u_o , u_e and the n -vector of even polynomials $s^{-1}\mathbf{U}$ before this argument is applied to verify the nonnegative definiteness of the matrix of residues associated with the pole at the origin.

When k exceeds unity,

$$\hat{\mathbf{K}}_\infty = \begin{bmatrix} \mathbf{K}_\infty + k \left[\frac{1}{v_o^2} \mathbf{U}\mathbf{U}^t \right]_\infty & k \left[\frac{1}{v_o} \mathbf{U} \right]_\infty \\ k \left[\frac{1}{v_o} \mathbf{U}^t \right]_\infty & k - 1 \end{bmatrix}. \quad (37)$$

According to Lemma 1, $\hat{\mathbf{K}}_\infty$ is nonnegative definite if and only if

$$\begin{aligned} \mathbf{K}_\infty + k \left[\frac{1}{v_o^2} \mathbf{U}\mathbf{U}^t \right]_\infty - \frac{k^2}{k-1} \left[\frac{1}{v_o^2} \mathbf{U}\mathbf{U}^t \right]_\infty \\ = \mathbf{K}_\infty - \frac{k}{k-1} \left[\frac{1}{v_o^2} \mathbf{U}\mathbf{U}^t \right]_\infty \end{aligned} \quad (38)$$

is nonnegative definite. However, from Theorem 1 [condition (iii)] and the fact that

$$[\mathbf{Y}_e]_\infty = \left[\frac{1}{v_o^2} \mathbf{U}\mathbf{U}^t \right]_\infty \quad (39)$$

it follows that (38) is indeed nonnegative definite.

Therefore, the conditions of Theorem 1 are sufficient for the realization of $\mathbf{Y}(s)$. It is of interest to note that in the preceding constructive proof it was sufficient to assume that $\eta(s)$ [defined in Section IV] is unity. All other possible matrices $\hat{\mathbf{Y}}(s)$ corresponding to a realization of $\mathbf{Y}(s)$ can be generated by exploiting the permissible choices of $\eta(s)$.

To complete the theory we consider in the next section the cases in which $\hat{\mathbf{Y}}(s)$ does not exist or \mathbf{Y}_e is a matrix of constants when $\hat{\mathbf{Y}}(s)$ does exist.

VI. NECESSARY AND SUFFICIENT CONDITIONS FOR THE REALIZATION OF $\mathbf{Y}(s)$ WHEN $\hat{\mathbf{Y}}(s)$ DOES NOT EXIST OR WHEN \mathbf{Y}_e IS A MATRIX OF CONSTANTS

The results of this section for the case in which $\hat{\mathbf{Y}}(s)$ does not exist are based on the following result which is proved in Appendix B.

Lemma 2:

If $\mathbf{Y}(s)$ in Fig. 1 exists but $\hat{\mathbf{Y}}(s)$ does not exist, then v_{n+1} , the voltage across the RC combination terminating port $(n+1)$, is related to the other port voltages by

$$v_{n+1} = \sum_{i=1}^n \beta_i v_i \quad (40)$$

where the β_i are real constants.

We wish to prove the following:

Theorem 2:

If the rational positive-real symmetric matrix $\mathbf{Y}(s)$, defined by the structure in Fig. 1, exists but $\hat{\mathbf{Y}}(s)$ does not exist, or if $\mathbf{Y}(s)$ is such that \mathbf{Y}_e is a matrix of constants, $\mathbf{Y}(s)$ can be expressed as $s\mathbf{K}_\infty + \mathbf{K}_0 + \mathbf{Y}'(s)$ where $\mathbf{Y}'(s)$ is an odd rational matrix in s such that $\mathbf{Y}'(s) \rightarrow \mathbf{0}$ as $s \rightarrow \infty$, and \mathbf{K}_0 is a real constant matrix with rank not exceeding unity such that $\mathbf{K}_\infty - \mathbf{K}_0$ is nonnegative definite. Further, if $\mathbf{Y}(s)$ satisfies the above condition, it can be realized as a reactance n -port in parallel with a network of ideal transformers that is terminated with a parallel combination of a unit resistor and a unit capacitor.

To prove the theorem for the case in which $\hat{\mathbf{Y}}(s)$ does not exist, first consider the expression for P , the average power entering the n -ports defining $\mathbf{Y}(s)$, in terms of $\mathbf{Y}_e |_{s=j\omega}$ and $\mathbf{V}^t = [v_1, v_2, \dots, v_n]$:

$$P = \mathbf{V}^t \mathbf{Y}_e |_{s=j\omega} \mathbf{V}^* \tag{41}$$

where the asterisk denotes the complex conjugate. Since P is also equal to $v_{n+1}v_{n+1}^*$, we have from (41) and Lemma 2

$$\mathbf{V}^t \mathbf{Y}_e |_{s=j\omega} \mathbf{V}^* = \sum_{i,j=1}^n \beta_i \beta_j v_i v_j^* \tag{42}$$

Because (42) is valid for arbitrary v_i , we find

$$\mathbf{Y}_e |_{s=j\omega} = \mathbf{B} \mathbf{B}^t = \mathbf{K}_0 \tag{43}$$

where $\mathbf{B}^t = [\beta_1, \beta_2, \dots, \beta_n]$. Thus $\mathbf{Y}(s)$ can be expressed as

$$\mathbf{Y}(s) = s\mathbf{K}_\infty + \mathbf{K}_0 + \mathbf{Y}'(s) \tag{44}$$

where $\mathbf{Y}'(s)$ is a matrix of odd rational functions which vanish at infinity. It is evident from (43) that \mathbf{K}_0 satisfies the rank conditions of Theorem 2.

Note that $\mathbf{Y}(s)$ has the form (44) if $\hat{\mathbf{Y}}(s)$ exists but \mathbf{Y}_e is a matrix of constants; for, in this case also, the rank of \mathbf{Y}_e cannot exceed unity.†

Next consider $\mathbf{Y}(s - 1)$, which must have a nonpositive definite real part on $s = j\omega$:

$$\mathbf{Y}(s - 1) = s\mathbf{K}_\infty + [\mathbf{K}_0 - \mathbf{K}_\infty] + \mathbf{Y}'(s - 1) \tag{45}$$

It is clear that if the real part of $\mathbf{Y}(s - 1)$ on $s = j\omega$ is to be nonpositive

† This is obvious from the form of (4).

definite for arbitrarily large values of $|\omega|$, $[\mathbf{K}_\infty - \mathbf{K}_0]$ must be a non-negative definite matrix.†

Finally, assume that $\mathbf{Y}(s)$ satisfies the conditions of Theorem 2 and consider

$$\mathbf{Y}(s) = \mathbf{C}[s\mathbf{D}_\infty + \mathbf{D}_0]\mathbf{C}^t + \mathbf{Y}'(s) \quad (46)$$

in which the real nonsingular $n \times n$ matrix \mathbf{C} is chosen so that

$$\begin{aligned} \mathbf{C}^{-1}\mathbf{K}_\infty\mathbf{C}^{-1t} &= \mathbf{D}_\infty \\ \mathbf{C}^{-1}\mathbf{K}_0\mathbf{C}^{-1t} &= \mathbf{D}_0 \end{aligned} \quad (47)$$

where \mathbf{D}_∞ and \mathbf{D}_0 are diagonal matrices.‡ Note that there can be at most one nonzero term in \mathbf{D}_0 and that this term cannot exceed the corresponding entry in \mathbf{D}_∞ for otherwise $[\mathbf{K}_\infty - \mathbf{K}_0]$ would not be non-negative definite. Hence $\mathbf{Y}(s)$ can be rewritten as

$$\mathbf{Y}(s) = (s + 1)\mathbf{CFC}^t + \mathbf{Y}''(s)$$

where $\mathbf{Y}''(s)$ is realizable as a reactance network and \mathbf{F} is a constant diagonal matrix with at most one nonzero element. This nonzero element is, of course, positive. The interpretation of the congruence transformation \mathbf{CFC}^t in terms of an ideal transformer network is well known. This proves Theorem 2.

VII. SUMMARY AND RELATED REMARKS

The principal results can be summarized as follows.

Theorem 3:

The rational positive-real $n \times n$ symmetric short-circuit admittance matrix $\mathbf{Y}(s)$ is realizable as a lossless network containing inductors, capacitors, and ideal transformers and a two-terminal element comprising a parallel combination of a unit resistor and a unit capacitor if and only if

- i. When $\mathbf{Y}_e = \mathbf{K}_0$, a matrix of constants, the rank of \mathbf{K}_0 does not exceed unity and $[(1/s)\mathbf{Y}]_\infty - \mathbf{K}_0$ is nonnegative definite.*
- ii. When \mathbf{Y}_e is not a matrix of constants, (a) \mathbf{Y}_e can be expressed in*

† This conclusion, with \mathbf{K}_0 defined as $[\mathbf{Y}_e]_\infty$, is valid when any number of unit resistor-capacitor parallel combinations are imbedded in a general lossless network.

‡ This is always possible since \mathbf{K}_∞ and \mathbf{K}_0 are nonnegative definite.

standard form (defined in Section III) with $v(s)$ such that

$$k = [v_o/sv_e]_\infty \geq 1; \quad (b) \text{ if } k = 1, \quad [\mathbf{Y}_e]_\infty = \mathbf{0};$$

$$(c) \text{ if } k > 1, \quad \left[\frac{1}{s} \mathbf{Y} \right]_\infty - \frac{k}{k-1} [\mathbf{Y}_e]_\infty$$

is nonnegative definite.

Further, if \mathbf{Y}_e is a matrix of constants and satisfies condition (i), \mathbf{Y} can be realized as a reactance n -port in parallel with a network of ideal transformers that is terminated with a parallel combination of a unit resistor and a unit capacitor. If \mathbf{Y}_e is not a matrix of constants and satisfies condition (ii), \mathbf{Y} can be realized as an $(n + 1)$ -port lossless network, characterized by the short-circuit admittance matrix $\hat{\mathbf{Y}}$, terminated at port $(n + 1)$ with a parallel combination of a unit resistor and a unit capacitor. The matrix $\hat{\mathbf{Y}}$ is given by

$$\hat{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y} + \frac{1}{v_e(v_o + v_e)} \mathbf{U}\mathbf{U}^t & \frac{1}{v_e} \mathbf{U} \\ \frac{1}{v_e} \mathbf{U}^t & \frac{v_o}{v_e} - s \end{bmatrix}.$$

For completeness, we state the following extension of Theorem 3.

Theorem 4:

The short-circuit admittance matrix $\bar{\mathbf{Y}}(s)$ is realizable as a lossless network containing inductors, capacitors, and ideal transformers and a two-terminal element comprising a parallel combination of a resistor of value R ohms ($R > 0$) and a capacitor of value TR^{-1} farads ($T > 0$) if and only if $\mathbf{Y}(s) = \bar{\mathbf{Y}}(s/T)$ is a symmetric positive-real matrix that satisfies the conditions of Theorem 3. If instead the resistor is equal to $-R$ ohms, the matrix is realizable if and only if $\mathbf{Y}(s) = -\bar{\mathbf{Y}}(-s/T)$ is a symmetric positive-real matrix that satisfies the conditions of Theorem 3.

The proof of Theorem 4 follows from two elementary transformations and is omitted.† In each case the parameter T is, of course, the time constant of the RC combination. It is convenient for some purposes to have the realizability conditions stated explicitly in terms of T . This can easily be done with the aid of the above theorem and is discussed in Appendix C.

† The fact that an $n \times n$ short-circuit admittance matrix $\mathbf{Y}(s)$ of real rational functions is realizable as a network containing only lossless elements and negative resistors if and only if $-\mathbf{Y}(-s)$ is a positive-real matrix was first established by Carlin and Youla.⁹

The following theorem states an interesting inequality involving $[(1/s)\mathbf{Y}]_\infty$ and $[(1/s)\mathbf{Y}_{sc}]_\infty$ where \mathbf{Y}_{sc} , if it exists, is the value of $\mathbf{Y}(s)$ when the RC combination, with unit capacitance, is shorted.

Theorem 5:

$$\left[\frac{1}{s} \mathbf{Y} \right]_\infty - \frac{1}{k} \left[\frac{1}{s} \mathbf{Y}_{sc} \right]_\infty \text{ is nonnegative definite.}$$

The proof follows at once from (21) and the fact that (24) is nonnegative definite.

The following theorem is of assistance in simplifying the tests indicated in Theorem 3 for the important case in which $\mathbf{K}_\infty = [(1/s)\mathbf{Y}]_\infty$ is positive definite.

Theorem 6:

If \mathbf{A} and \mathbf{B} are $n \times n$ real symmetric nonnegative definite matrices with $\det \mathbf{A} \neq 0$ and \mathbf{B} of unit rank, $\mathbf{A} - \mathbf{B}$ is nonnegative definite if and only if $\det [\mathbf{A} - \mathbf{B}] \geq 0$.

To prove this result note that $\mathbf{A} - \mathbf{B}$ can be written as $\mathbf{Q}[\mathbf{D}_a - \mathbf{D}_b]\mathbf{Q}^t$, where \mathbf{Q} is a real nonsingular matrix such that $\mathbf{A} = \mathbf{Q}\mathbf{D}_a\mathbf{Q}^t$, $\mathbf{B} = \mathbf{Q}\mathbf{D}_b\mathbf{Q}^t$, and \mathbf{D}_a and \mathbf{D}_b are diagonal matrices. Thus

$$\det[\mathbf{A} - \mathbf{B}] = \det^2 \mathbf{Q} \cdot \det[\mathbf{D}_a - \mathbf{D}_b].$$

The realizability conditions can be expressed also in terms of the open-circuit impedance matrix $\mathbf{Z}(s)$ by exploiting an approach similar to that used in treating the short-circuit admittance matrix $\mathbf{Y}(s)$. Since the ideas involved are so similar to those discussed earlier we shall omit the details and simply state the result:

Theorem 7:

The rational positive-real $n \times n$ symmetric open-circuit impedance matrix $\mathbf{Z}(s)$ is realizable as a lossless network containing inductors, capacitors, and ideal transformers and a two-terminal element comprising a parallel combination of a unit resistor and a unit capacitor if and only if the even part of the matrix \mathbf{Z} , \mathbf{Z}_e , is expressible in standard form (defined in Section III) with

$$\left[\frac{1}{s} \frac{v_e}{v_o} \right]_\infty \geq 1.$$

Further, if $v_e \neq sv_o$ and \mathbf{Z} satisfies the above conditions, \mathbf{Z} can be realized

as an $(n + 1)$ -port lossless network, characterized by the open-circuit impedance matrix $\hat{\mathbf{Z}}$, terminated at port $(n + 1)$ with a parallel combination of a unit resistor and a unit capacitor. The matrix $\hat{\mathbf{Z}}$ is given by

$$\hat{\mathbf{Z}} = \begin{bmatrix} \mathbf{Z} + \mathbf{U}\mathbf{U}^t \frac{(s + 1)}{(v_e - sv_o)(v_o + v_e)} & \mathbf{U} \frac{1}{v_e - sv_o} \\ \mathbf{U}^t \frac{1}{v_e - sv_o} & \frac{v_o}{v_e - sv_o} \end{bmatrix}.$$

If $v_e \equiv sv_o$, and \mathbf{Z} satisfies the above conditions, \mathbf{Z} can be expressed as $\mathbf{Z} = [1/(s + 1)] \mathbf{F} + \mathbf{Z}'$ where \mathbf{F} is a real symmetric nonnegative definite matrix of constants of rank not exceeding unity and \mathbf{Z}' is the open-circuit impedance matrix of an n -port reciprocal lossless network.

The simple form that the conditions assume is attributable to the fact that the impedance of the parallel RC combination is regular at infinity and that the matrix \mathbf{Z}_e is not a matrix of constants unless every element vanishes identically in s .

APPENDIX A

Factorization of $\mathbf{Y}_e(s)$

Recall that \mathbf{Y}_e is the even part of a rational symmetric $n \times n$ positive-real short-circuit admittance matrix. It is convenient to partition this matrix as follows:

$$\mathbf{Y}_e(s) = \begin{bmatrix} E_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{12}^t & \mathbf{E}_{22} \end{bmatrix} \tag{48}$$

where E_{11} , \mathbf{E}_{12} , and \mathbf{E}_{22} are respectively 1×1 , $1 \times (n - 1)$, and $(n - 1) \times (n - 1)$ submatrices of ratios of even polynomials in s . We may assume, without loss of generality, that E_{11} does not vanish identically in s .

Consider the following identity which is readily verified:

$$\begin{bmatrix} E_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{12}^t & \mathbf{E}_{22} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{E}_{12}^t E_{11}^{-1} & \mathbf{1}_{n-1} \end{bmatrix} \begin{bmatrix} E_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} - \mathbf{E}_{12}^t \mathbf{E}_{12} E_{11}^{-1} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{E}_{12} E_{11}^{-1} \\ \mathbf{0} & \mathbf{1}_{n-1} \end{bmatrix}. \tag{49}$$

It is evident that the normal rank of \mathbf{Y}_e cannot exceed unity if it is to be expressible in standard form. Accordingly we may assume that

$\mathbf{E}_{22} - \mathbf{E}_{12}^t \mathbf{E}_{12} \mathbf{E}_{11}^{-1} = \mathbf{0}$, and hence

$$\begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{12}^t & \mathbf{E}_{22} \end{bmatrix} = \mathbf{E}_{11} [1 \quad \mathbf{E}_{12} \mathbf{E}_{11}^{-1}]^t [1 \quad \mathbf{E}_{12} \mathbf{E}_{11}^{-1}]. \quad (50)$$

The right-hand side of (50) can be rewritten as $f\mathbf{P}\mathbf{P}^t$ where

$$\mathbf{P}^t = [p_1, p_2, \dots, p_n]$$

is a row matrix of even real polynomials and f is an even real rational fraction in s , analytic on $s = j\omega$ [$-\infty \leq \omega \leq \infty$], and such that

$$f(j\omega) \geq 0.$$

As is well known, $f(s)$ can be expressed as either

$$\frac{g^2}{h(s)h(-s)} \quad \text{or} \quad \frac{-l^2}{m(s)m(-s)}$$

where g and l are respectively even and odd real polynomials and $h(s)$ and $m(s)$ are real strict Hurwitz polynomials. In either case, since

$$\frac{g^2}{h(s)h(-s)} = \frac{-(sg)^2}{[sh(s)][-sh(-s)]},$$

\mathbf{Y}_e can be written as

$$\mathbf{Y}_e = \frac{-1}{w(s)w(-s)} \mathbf{W}\mathbf{W}^t,$$

in which $\mathbf{W}^t = [w_1, w_2, \dots, w_n]$ is a row matrix of real odd polynomials and $w(s)$ is a real strict Hurwitz polynomial except possibly for a simple zero at the origin.

Note that \mathbf{Y} is realizable as shown in Fig. 1 when $\hat{\mathbf{Y}}$ exists only if the degree of $w(s)$ is odd.

APPENDIX B

Proof of Lemma 2

First note that if the $(n+1)$ -port lossless network does not possess a short-circuit admittance matrix, the short-circuit admittance matrix of \bar{N} , the lossless $(n+1)$ -port with a unit capacitor added in parallel at port $(n+1)$, also does not exist.

Let $\bar{\mathbf{S}}(s)$ be the scattering matrix¹⁰ of \bar{N} and consider the circuit

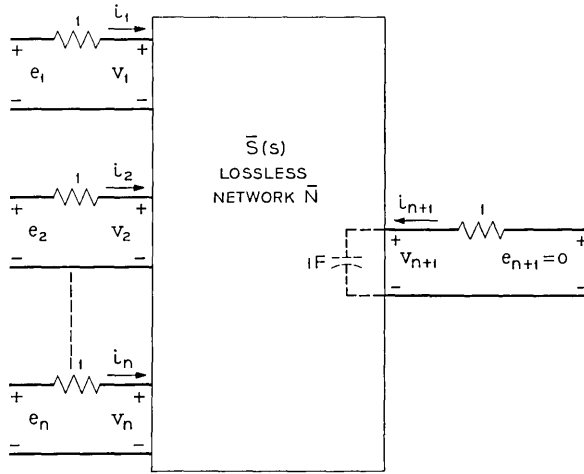


Fig. 2 — Network defining the relationship between \mathbf{E} , \mathbf{V} , \mathbf{I} , and the scattering matrix $\bar{\mathbf{S}}(s)$.

shown in Fig. 2. By definition

$$\bar{\mathbf{S}}[\mathbf{V} + \mathbf{I}] = \mathbf{V} - \mathbf{I}, \tag{51}$$

and

$$\mathbf{Y}(s) = (\mathbf{1}_n + \mathbf{S})^{-1}(\mathbf{1}_n - \mathbf{S}) \tag{52}$$

where \mathbf{S} is the matrix of elements in the first n rows and columns of $\bar{\mathbf{S}}$.

Substituting $\mathbf{E} = \mathbf{V} + \mathbf{I}$ in (51) gives

$$\mathbf{V} = \frac{1}{2}[\bar{\mathbf{S}} + \mathbf{1}_{n+1}]\mathbf{E}. \tag{53}$$

Because the short-circuit admittance matrix of $\bar{\mathbf{N}}$ exists if and only if $\det[\bar{\mathbf{S}} + \mathbf{1}_{n+1}] \neq 0$, and since \mathbf{S} is the matrix of elements in the first n rows and columns of $\bar{\mathbf{S}}$, it follows that $[\bar{\mathbf{S}} + \mathbf{1}_{n+1}]$ has normal rank† equal to n . Further, since the rank of $[\bar{\mathbf{S}} + \mathbf{1}_{n+1}]$ is invariant in the strict right-half plane, there exists, to within an arbitrary scalar multiplicative factor, one and only one real constant $(n + 1)$ -vector \mathbf{X} such that

$$[\bar{\mathbf{S}}(s_0) + \mathbf{1}_{n+1}]\mathbf{X} = \mathbf{0} \tag{54}$$

where s_0 is a fixed but arbitrarily chosen real positive constant. Let \mathbf{X} be normalized so that $\mathbf{X}'\mathbf{X} = 1$. Equation (54) then yields

$$\mathbf{X}'\bar{\mathbf{S}}(s_0)\mathbf{X} = -1 \tag{55}$$

† Since $\mathbf{Y}(s)$ exists, $\det[\mathbf{1}_n + \mathbf{S}]$ does not vanish identically in s .

Note that $\mathbf{X}'\bar{\mathbf{S}}(s)\mathbf{X}$ is a one-port passive scattering coefficient and that therefore (51) implies

$$\mathbf{X}'[\bar{\mathbf{S}}(s) + \mathbf{1}_{n+1}]\mathbf{X} = 0 \quad (56)$$

identically in s . Furthermore, since $[\bar{\mathbf{S}}(s) + \mathbf{1}_{n+1}]$ is positive semidefinite for all real positive s , it follows that

$$\mathbf{X}'[\bar{\mathbf{S}}(s) + \mathbf{1}_{n+1}] = \mathbf{0} \quad (57)$$

identically in s . Thus from (53) and (57)

$$\mathbf{X}'\mathbf{V} = \frac{1}{2}\mathbf{X}'[\bar{\mathbf{S}} + \mathbf{1}_{n+1}]\mathbf{E} = 0$$

or

$$\sum_{i=1}^{n+1} x_i v_i = 0$$

where the x_i are real constants, not all zero. However since $\mathbf{Y}(s)$ exists x_{n+1} cannot vanish. Dividing through by x_{n+1} gives an expression of the form

$$v_{n+1} = \sum_{i=1}^n \beta_i v_i$$

in which the β_i are real constants.

It is of incidental interest to note that the proof does not require that \bar{N} be lossless. It is sufficient that it be passive.†

APPENDIX C

The Realizability Conditions Stated Explicitly in Terms of the Parameter T

According to Theorem 4, $\bar{\mathbf{Y}}(s)$ is realizable with a lossless reciprocal network and a two-terminal element comprising a parallel combination of a resistor of value R ohms ($R > 0$) and a capacitor of value TR^{-1} farads ($T > 0$) if and only if $\mathbf{Y}(s) = \bar{\mathbf{Y}}(s/T)$ is a symmetric positive-real matrix that satisfies the conditions of Theorem 3. Let $\bar{\mathbf{Y}}_e(s)$ be expressed in standard form:

$$\bar{\mathbf{Y}}_e(s) = \frac{-1}{\bar{v}(s)\bar{v}(-s)} \bar{\mathbf{U}}\bar{\mathbf{U}}'. \quad (58)$$

† The original version of this proof, based also on the formulation (53) and the fact that $[\bar{\mathbf{S}} + \mathbf{1}_{n+1}]$ is of normal rank n , assumed that the network \bar{N} is lossless and hence that $\bar{\mathbf{S}}(s)$ is a regular para-unitary matrix. The final version of the proof was suggested by D. C. Youla.

In terms of $\bar{v}(s)$,

$$k = \left[\frac{1}{s} \frac{\bar{v}_o(s/T)}{\bar{v}_e(s/T)} \right]_{\infty} = \frac{1}{T} \bar{k} \tag{59}$$

$$\bar{k} = \left[\frac{1}{s} \frac{\bar{v}_o(s)}{\bar{v}_e(s)} \right]_{\infty}.$$

Also,

$$\left[\frac{1}{s} \mathbf{Y}(s) \right]_{\infty} = \left[\frac{1}{s} \bar{\mathbf{Y}} \left(\frac{s}{T} \right) \right]_{\infty} = \frac{1}{T} \left[\frac{1}{s} \bar{\mathbf{Y}}(s) \right]_{\infty},$$

and

$$[\mathbf{Y}_e(s)]_{\infty} = \left[\bar{\mathbf{Y}}_e \left(\frac{s}{T} \right) \right]_{\infty} = [\bar{\mathbf{Y}}(s)]_{\infty}.$$

Theorem 3 and the above equations yield

Theorem 8:

The rational positive-real $n \times n$ symmetric short-circuit admittance matrix $\bar{\mathbf{Y}}(s)$ is realizable as a lossless network containing inductors, capacitors, and ideal transformers and a two-terminal element comprising a parallel combination of a resistor of value R ohms ($R > 0$) and a capacitor of value TR^{-1} Farads ($T > 0$) if and only if

- i. When $\bar{\mathbf{Y}}_e(s) = \bar{\mathbf{K}}_0$, a matrix of constants, the rank of $\bar{\mathbf{K}}_0$ does not exceed unity and $[(1/s)\bar{\mathbf{Y}}(s)]_{\infty} - T\bar{\mathbf{K}}_0$ is nonnegative definite.
- ii. When $\bar{\mathbf{Y}}_e(s)$ is not a matrix of constants (a) $\bar{\mathbf{Y}}_e(s)$ can be expressed in standard form with $\bar{k} \geq T$; (b) if $\bar{k} = T$, $[\bar{\mathbf{Y}}_e(s)]_{\infty} = \mathbf{0}$; (c) if $\bar{k} > T$,

$$\left[\frac{1}{s} \bar{\mathbf{Y}}(s) \right]_{\infty} - \frac{\bar{k}T}{\bar{k} - T} [\bar{\mathbf{Y}}_e(s)]_{\infty}$$

is nonnegative definite.

Similarly, Theorem 5 can be transformed to read:

$$\left[\frac{1}{s} \bar{\mathbf{Y}}(s) \right]_{\infty} - \frac{T}{\bar{k}} \left[\frac{1}{s} \bar{\mathbf{Y}}_{sc}(s) \right]_{\infty}$$

is nonnegative definite.

The modifications necessary to treat the case in which the resistance is equal to $-R$ ohms are obvious in view of Theorem 4.

REFERENCES

1. Sommers, H. S., Jr., Proc. I.R.E., **47**, 1959, p. 1201.
2. Hines, M. E., B.S.T.J., **39**, May, 1960, p. 477.

3. Youla, D. C., and Smilen, L. I., Symposium on Active Networks and Feedback Systems, Polytechnic Institute of Brooklyn; 1960.
4. Smilen, L. I., and Youla, D. C., Proc. Nat. Elec. Conf., Oct., 1960.
5. Kuh, E. S., and Patterson, J. D., Proc. I.R.E., **49**, 1961, p. 1043.
6. Desoer, C. A., and Kuh, E. S., Symposium on Active Networks and Feedback Systems, Polytechnic Institute of Brooklyn, 1960.
7. Kinariwala, B. K., NEREM Record, **2**, 1960, p. 86.
8. Kinariwala, B. K., I.R.E.-PGCT, **CT-8**, Dec., 1961.
9. Carlin, H. J., and Youla, D. C., Proc. I.R.E., **49**, May, 1961, p. 907.
10. Youla, D. C., I.R.E. Conv. Rec., **CT-7**, August, 1960.
11. Weinberg, L., I.R.E.-PGCT, **CT-8**, March, 1961, p. 66.

On Non-Computable Functions

By T. RADO

(Manuscript received November 12, 1961)

The construction of non-computable functions used in this paper is based on the principle that a finite, non-empty set of non-negative integers has a largest element. Also, this principle is used only for sets which are exceptionally well-defined by current standards. No enumeration of computable functions is used, and in this sense the diagonal process is not employed. Thus, it appears that an apparently self-evident principle, of constant use in every area of mathematics, yields non-constructive entities.

I. INTRODUCTION

The purpose of this note is to present some very simple instances of non-computable functions. Beyond their simplicity, these examples throw light upon the following basic point. If a function $f(x)$ is to serve as an example of a non-computable function, then $f(x)$ must be *well-defined* in some generally accepted sense; hence the efforts to construct examples of non-computable functions reveal the general conviction that over and beyond the class of computable (general recursive) functions there is a much wider class, the class of *well-defined functions*. The scope of this latter class is vague; in some quarters, there exists a belief that this class will be defined some day in precise terms acceptable to all. The examples of non-computable functions to be discussed below will be well defined in an extremely primitive sense; we shall use only the principle that a non-empty finite set of non-negative integers has a largest element. Furthermore, we shall use this principle only for exceptionally well-defined sets; and thus our construction will rest upon considerations which occur constantly in every area of mathematics. It may be of interest to note that we shall not use an enumeration of computable functions to show that our examples are non-computable functions. Thus, in this sense, we do not use the diagonal process.

II. TERMINOLOGY

We shall use binary Turing machines (that is, Turing machines with the binary alphabet 0, 1), in the sense of the excellent presentation of

Kleene's *Metamathematics* (see Ref.), with the following exceptions. First, we do not permit a center shift; thus the machine must shift after the execution of an "overprint" instruction (the purpose is to simplify the following presentation). Second, we shall use the term "card" instead of "state." The reason is that the examples below were obtained as by-products of a logical game (the Busy Beaver game described below) which the writer made up to familiarize beginners with the idea of a Turing machine; and it appeared that terms such as state, internal configuration, and the like had a mysterious connotation for beginners. To illustrate some notational conventions to be used, let us consider the following example of a binary, 3-card Turing machine.

C ₁	C ₂	C ₃
0 102	0 111	0 112
1 113	1 102	1 100

Here C₁, C₂, C₃ stand for Card 1, Card 2, and Card 3. On each card, the left-most column contains the alphabet 0, 1. The next column is the "overprint by" column; the next one is the "shift" column (where 0 is the code for a left shift and 1 is the code for a right shift). The last column is the "call card" column; it contains the index of the next card to be used, or 0 (zero), where 0 is the code for "Stop." This notation was found very convenient in situations where one wanted to enumerate (serialize) Turing machines with a given number of cards.

The reader is assumed to be familiar with the meaning (in the sense of Kleene; see Ref.) of the statement that a binary Turing machine "computes" a function $f(x)$. It is understood that we consider only functions of non-negative integers with values which are again non-negative integers.

III. THE BUSY BEAVER GAME

Consider a potentially both-ways infinite tape (see Ref.), where each square contains a 0 (all-zero tape). Start the 3-card machine described in Section II (with its Card 1) under any square. The reader will find that the machine stops after a few shifts, and when it stops, there are six ones on the tape. Actually, this particular machine is one of the four highest scorers (as of today) in the international BB-3 game (the 3-card deck classification of the Busy Beaver game). The rules in this game are as follows.

i. The contestant selects a positive integer n ; and then makes up his own n -card, binary, Turing machine (using the notational conventions explained in Section II).

ii. He starts his machine (with its Card 1) on an all-zero tape, and satisfies himself that his machine stops after a certain number s of shifts.

iii. He then submits his entry, as well as the shift-number s , to any member (in good standing) of the International Busy Beaver Club.

iv. The umpire first verifies that the entry actually stops exactly after s shifts. Note that this is a decidable issue; the umpire merely operates the entry, persisting through not more than the specified number s of shifts. If the entry fails to stop after s shifts, it is rejected; if it stops after fewer than s shifts, it is returned to the contestant for correction. After the entry has been verified, its score is the number of ones on the tape when it stops.

Naturally, the BB- n champion is the contestant who achieved the highest score (so far) in the BB- n classification. For example, in the BB-3 classification, the score of 6 was first achieved by R. Hegelman (U.S. Naval Weapons Laboratory, Dahlgren, Virginia). This score has been reached since by several others; but nobody knows as yet whether 6 is the highest possible score in the BB-3 classification. The reader who tries to settle this question will soon realize the difficulties involved in this sort of problem. Beyond the enormous number of cases to survey, he will find that it is very hard to see whether certain entries do stop at all. This is the reason for the requirement that each contestant must submit the shift number s with his entry.

IV. HIGHEST SCORE

There arises now the problem of determining the highest possible score in the BB- n classification. In line with the point of view explained in the introduction, we formulate this problem with due care and caution.

Returning to rule *iv.* of the game, we see that a valid entry in the BB- n classification is a pair (M, s) , such that the following holds.

- (a) M is an n -card binary Turing machine.
- (b) s is a positive integer.
- (c) M stops after exactly s shifts if started (with its Card C_1) on an all-zero tape.

In discussing rule *iv.* above, we noted that we can actually decide whether or not an entry (M, s) is valid. Also, if (M_1, s_1) , (M_2, s_2) are valid entries such that $M_1 = M_2$, then evidently $s_1 = s_2$; hence the number of valid BB- n entries cannot exceed the number $N(n)$ of all

possible n -card, binary Turing machines. It is easy to see that

$$N(n) = [4(n + 1)]^{2^n} \quad (1)$$

Also, there exist valid BB- n entries; for example, on choosing the 0-line of Card 1 as 110, one obtains an entry which stops after one shift.

Accordingly, if we denote by E_n the set of all valid BB- n entries (M,s) , we obtain a non-empty, finite set E_n which has the following features.

- (a) We actually exhibit elements of E_n ; so E_n is non-empty as a matter of concrete observation.
- (b) We not only know that E_n is finite, but for the number $N_e(n)$ of elements of this set of valid entries we have [see (1)] the inequalities.

$$1 < N_e(n) < N(n) = [4(n + 1)]^{2^n} \quad (2)$$

- (c) For every pair (M,s) we can actually decide whether or not $(M,s) \in E_n$.

Evidently, E_n is (by current standards) an *exceptionally well-defined* non-empty, finite set. Yet, we shall show below that $N_e(n)$, the number of elements of E_n , is not a computable function of n . Next, each valid entry $(M,s) \in E_n$ has a definite score $\sigma(M,s)$ assigned to it (see Section III). Thus, for the same reasons, the set of these scores is an exceptionally well-defined non-empty finite set of non-negative integers. We denote by $\Sigma(n)$ the largest element of this set.

Thus

$$\Sigma(n) = \max [\sigma(M,s)] \quad \text{for } (M,s) \in E_n. \quad (3)$$

We shall see presently that $\Sigma(n)$ is not a computable function of n . Let us note, however, that it is entirely possible that $\Sigma(n)$ can be effectively determined for particular values of n . For example, evidently $\Sigma(1) = 1$. Also, it has been proved that $\Sigma(2) = 4$. We noted above that we know several BB-3 entries with a score of 6; hence $\Sigma(3) \geq 6$, and it seems plausible that $\Sigma(3) = 6$. Now while for low values of n it is quite hard to achieve a respectable score, Dr. C. Y. Lee observed (in a letter to the writer) that for higher values of n one can achieve very large scores. The following proof for the non-computability of $\Sigma(n)$ was obtained by developing this comment of Dr. Lee.

V. THE GROWTH OF $\Sigma(n)$

Let $f(x)$, $g(x)$ be two functions (as specified in Section II). We shall write

$$f(x) > -g(x)$$

to state that $f(x) > g(x)$ for x greater than a certain x_0 . Using this notation, we shall now prove the following theorem.

Theorem. $\Sigma(n) > -f(n)$ for every computable (that is, general recursive) function $f(n)$. Hence $\Sigma(n)$ is not computable.

Proof. Assign a computable function $f(x)$. Introduce the auxiliary function

$$F(x) = \sum_{i=0}^x [f(i) + i^2]. \tag{4}$$

Then (see Ref.) $F(x)$ is also computable. Evidently

$$F(x) \geq f(x). \tag{5}$$

$$F(x) \geq x^2. \tag{6}$$

$$F(x + 1) > F(x). \tag{7}$$

Now since $F(x)$ is computable, we have a binary Turing machine M_F , with a certain number C of cards (states) which computes $F(x)$ (in the sense described in Kleene; see Ref.). Now assign any integer $x \geq 0$. We have then a binary Turing machine $M^{(x)}$, with $x + 1$ cards (states) which prints on an all-zero tape $x + 1$ consecutive ones and stops under the right-most one of these ones. For $x = 2$, for example, $M^{(2)}$ has the 3 cards:

C ₁	C ₂	C ₃
0 112	0 113	0 101
1 110	1 ---	1 ---

Now consider the binary Turing machine $M_F^{(x)}$ given by the symbolic diagram:

$$M_F^{(x)} : M^{(x)} \rightarrow M_F \rightarrow M_F.$$

If the cards of $M_F^{(x)}$ are written out with consecutive indices, then it is seen to have $1 + x + 2C$ cards. If started on an all-zero tape, $M_F^{(x)}$ will first print (going to the right) a string of $x + 1$ consecutive ones; then, beyond a 0 to the right, it will print a string of $F(x) + 1$ consecutive ones; finally, beyond a 0 to the right, it will print a string of $F[F(x)] + 1$ consecutive ones, and then will stop (under the right-most 1 it printed). Thus evidently $N_F^{(x)}$ is a valid entry in the BB- $(1 + x + 2C)$ classifi-

cation with a score equal to

$$3 + x + F(x) + F[F(x)].$$

Hence, the maximum score $\Sigma(1 + x + 2C)$ in this classification satisfies the inequality

$$\Sigma(1 + x + 2C) \geq 3 + x + F(x) + F[F(x)]. \quad (8)$$

Now since evidently $x^2 > -(1 + x + 2C)$ and $F(x) \geq x^2$ [see (6)], it follows that

$$F(x) > -(1 + x + 2C). \quad (9)$$

Also, $F(x)$ is monotone increasing by (7); hence (9) yields

$$F[F(x)] > -F(1 + x + 2C). \quad (10)$$

From (8) and (10) we see that

$$\Sigma(1 + x + 2C) > -F(1 + x + 2C);$$

hence (since $F(x) \geq f(x)$)

$$\Sigma(1 + x + 2C) > -f(1 + x + 2C).$$

On setting $n = 1 + x + 2C$, we obtain finally

$$\Sigma(n) > -f(n)$$

and the theorem is proved.

The rate at which $\Sigma(x)$ grows is illustrated by the following intuitive observation. A Turing machine M_H for computing $H(x) = x!$ can be constructed with not more than 26 states. Let us consider the chain of Turing machines:

$$M^{(x)} \rightarrow M_H \rightarrow M_H \rightarrow M_H \rightarrow M_H.$$

It follows from (8) that the number of ones which is produced by this chain is more than $((x!)!)!$. Using the construction of the machine M_H mentioned above, we may show that by combining these machines properly, the number of states required for this chain of machines for $x = 7$, for instance, is not more than 100. Therefore, $\Sigma(100)$ is at least $((7!)!)!$. Since $\Sigma(100)$ is probably far bigger than this lower bound, it would be interesting to know how large a lower bound one can get for $\Sigma(100)$.

VI. THE FUNCTION $S(n)$

It is evident from our definitions that the set E_n of valid BB- n entries coincides with the set of the n -card stoppers, where by a stopper we

mean a (binary) Turing machine which, if started on an all-zero tape with its card C_1 , will stop after a while. Now the second coordinates s of the valid BB- n entries (M,s) constitute a finite, non-empty set of positive integers; we denote by $S(n)$ the largest element of this set. Thus $S(n)$ is the maximum of the shift-numbers of the n -card stoppers. Clearly

$$S(n) \geq \Sigma(n). \quad (11)$$

Indeed, since we do not permit center-shifts, a BB- n entry must shift after it prints a 1; thus (11) is obvious. From the theorem in Section V and from (11) we see that

$$S(n) > -f(n) \quad (12)$$

for every computable function $f(n)$. Thus $S(n)$ is non-computable (the reader will readily see that this result is equivalent to the undecidability of the so-called halting problem).

VII. THE FUNCTION $N_e(n)$

This function, defined above as the number of elements of the set E_n (that is, the number of n -card stoppers) does not grow unreasonably fast [see (2)]. However, we can discuss it as follows. Let us denote by $N(s,n)$ the number of those BB- n entries which stop after exactly s shifts. Evidently, the computation of $N(s,n)$ can be readily programmed; informally, one finds the value of $N(s,n)$ by running each one of the n -card binary Turing machines [whose number is given by (1)], persisting through not more than the given number s of shifts, and noting the number of those that stop after exactly s shifts. Let us put

$$G(s,n) = \sum_{i=1}^s N(i,n), \quad (13)$$

$$\Phi(s,n) = N_e(n) - G(s,n). \quad (14)$$

Clearly, $G(s,n)$ is the number of those BB- n entries that stop after not more than s shifts; thus $G(s,n) \leq N_e(n)$, and hence $\Phi(s,n) \geq 0$. Since evidently $G(s,n) = N_e(n)$ for $s = S(n)$, we see that $S(n)$ is the smallest value of s for which $\Phi(s,n) = 0$; in symbols:

$$S(n) = (\mu s)[\Phi(s,n) = 0], \quad (15)$$

where (μs) means "the smallest s such that." From (13)–(15) it follows (see Rcf.) that if $N_e(n)$ were computable then $S(n)$ would be computable too; since we know that $S(n)$ is not computable, it follows that $N_e(n)$ is non-computable.

7.1 *Remark*

Suppose that, for a certain integer n_0 , we somehow succeeded in determining the exact value of $N_e(n_0)$. From (13)–(15) it follows that we can then determine $S(n_0)$ also, and hence finally $\Sigma(n_0)$. Various other comments will readily occur to the reader. For example, the easily proved inequality

$$S(n) \leq (n + 1) \Sigma(5n) 2^{\Sigma(5n)}$$

gives rise to some curious observations.

VIII. SUMMARY

Inspection of the preceding presentation shows that we used in our constructions only the following “principle of the largest element”: If E is a non-empty, finite set of non-negative integers, then E has a largest element. This principle is used constantly, as a matter of course, in every field of mathematics. Our examples above show that this principle, even if applied only to *exceptionally well-defined sets* E , may take us beyond the realm of constructive mathematics. Of course, common everyday experiences may be used to illustrate this sort of phenomenon. For example, when the writer wanted to find a certain highway on an automobile trip, he received the following directions from the foreman of a construction crew: “Drive straight ahead on this road; you will cross some steel bridges; and after you cross the last steel bridge, make a left turn at the next intersection.” Luckily, the unsolvable problem implied by this advice was resolved by a member of the construction crew who volunteered the information that “after you cross the last steel bridge, there isn’t another steel bridge until you reach Richmond, 130 miles away.” The reader may find it amusing to verify, by detailed study of the excellent book of Kleene (Ref.), that this little story illustrates, in a concrete manner, some truly basic points in the theory of computable functions.

IX. ACKNOWLEDGMENT

The writer takes pleasure in thanking Dr. C. Y. Lee (of Bell Telephone Laboratories) for a number of stimulating comments.

REFERENCE

1. Kleene, S. C., *Introduction to Metamathematics*, D. Van Nostrand Co., Princeton, N. J., 1952.

On the Theory of Shrink Fits with Application to Waveguide Pressure Seals

By A. J. SCHEPIS

(Manuscript received November 15, 1961)

Waveguide antennas for missiles and space vehicles usually require a window or radome to provide a pressure seal. The technique of shrink fitting is a simple method to seal rectangular waveguides. A theoretical and experimental investigation has been undertaken to study the stresses and displacements that result from shrink fits between rectangular cross sections.

An appropriate differential equation is derived by applying simplifying assumptions to the theory of thermal stresses. The solution of this equation indicates that for certain combinations of materials and sizes of rectangles, there is a critical wall thickness below which a pressure-tight shrink fit cannot be made regardless of the temperature at which the shrink fit process is initiated. A comparison is presented of the analytical solution with experimental results of "Teflon" shrunk fit into precision fabricated X-band waveguide. Finally, tabulated results are presented to indicate the magnitude of the stresses and displacements to be expected from typical materials and rectangles of various sizes.*

I. INTRODUCTION

The derivations and conclusions reported in this paper have been stimulated by the current interest in light-weight microwave antennas for missiles and space vehicles. These antennas must be capable of maintaining a pressure seal in temperature environments from -80°F to $+400^{\circ}\text{F}$ or higher. A typical seal is intended to maintain a pressure of one atmosphere. Such pressures are required to prevent electrical breakdown of the waveguide at nominal operating power.

A common method of achieving such a seal is to provide a radome over the antenna. This scheme is usually effective but requires gaskets

* "Teflon" is a registered trademark of the E. I. du Pont DeNemours and Co. The type referred to in this paper is a polytetrafluoroethylene resin.

and screws and, moreover, is an additional structure to be carried during flight.

A design is proposed in which a suitable dielectric plug, or core, necessary to give the required antenna pattern, is inserted into the waveguide opening of the antenna. A pressure seal between the plug and the waveguide is accomplished by first chilling the plug to a temperature well below that of the waveguide piece. The plug is then inserted into the waveguide. As the plug warms up, it expands and exerts sufficient pressure on the walls of the waveguide to cause an effective seal.

The purpose of this study is to determine theoretically as well as experimentally the feasibility of such a shrink fit. The study includes the behavior of thin wall rectangular tubes into which a core is shrunk fit, where the core may or may not be of the same material as the tube. The particular solutions that are sought are the resulting stresses and deformations at the interface of the tube and the core; the stresses to indicate the effectiveness of the seal, and the deformations to indicate the degree of distortion of the tube. While it is by no means obvious that a limit on wall thickness exists, for given materials and shape of rectangles, below which no seal can be obtained, it will be shown that such a limit does indeed exist.

II. FORMULATION OF THE PROBLEM

2.1 *General Formulation*

The analysis is performed for the general case, i.e., the size of the rectangular enclosure, type of materials, and the temperature environment are all arbitrary, as shown in Fig. 1. The core material, E_1 , is subjected to a temperature change, ΔT , until it can be inserted into the tube, E_2 . Therefore, $\Delta T = T_2 - T_1$, where T_2 is the initial temperature of the tube and the core and T_1 is the temperature to which the core is reduced.

For a core with no variation of temperature or stress through the thickness we may assume a condition of plane stress. Consider, therefore, a thin wall rectangular tube of material, E_2 , in which a material, E_1 , is shrunk fit, as shown in Fig. 1. A set of coordinate axes, x - y and u - v , is located at the centers of the broad and narrow walls respectively. The deformation of the tube from its original shape is y and v . The strains perpendicular to the interface of length b_0 and a_0 are ϵ_y and ϵ_v respectively.

We assume that the corners of the tube do not move, that the stresses induced in the enclosure are due to bending only, and that the right

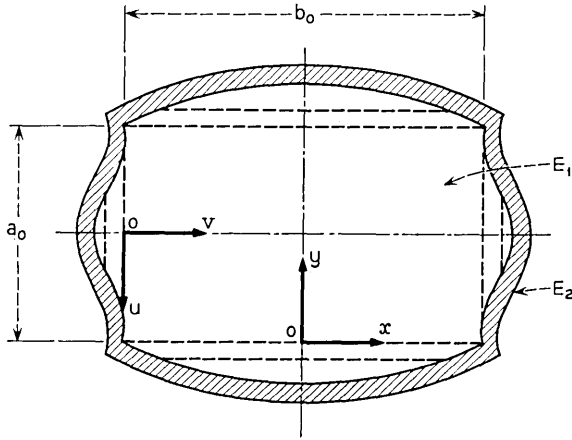


Fig. 1 — Deformation of thin wall rectangular tubing by a shrunk-fit core, general case.

angles formed by both sides remain right angles. The first two assumptions admit a contradiction since the first will result in a tension in the beam. This tension will be neglected because of the following reasons: *i.* the corners actually move and the stress due to tension introduced by the assumption of fixed corners is assumed negligible. To analyze the problem with the admission of the moveable corners would introduce undefined boundary conditions, and *ii.* the enclosure wall thicknesses are such as to consider the enclosure a beam frame where the stresses due to bending are predominant. These simple assumptions are satisfactory approximations in many practical cases where we are concerned only in the elastic range. For instance, in the case of "Teflon" inserted into aluminum or brass the solution obtained on these assumptions is in good agreement with actual measurements.¹ In other cases, the deformations of the sides of the enclosure are so great as to render the displacements of the corners negligible. Also because of symmetry the centers of the interfaces do not move parallel to the interface. We therefore make the further assumption that all strains parallel to the interfaces are zero. We note at this time we are only concerned about the interfaces.

From the theory of elasticity² as applied to restricted thermal expansion we have for the stress in the y direction:

$$\sigma_y = \frac{-E_1}{1 - \nu_1^2} (\epsilon_y + \nu_1 \epsilon_x) + \frac{\alpha_1 E_1 \Delta T}{1 - \nu_1} \tag{1}$$

where: ϵ_y and ϵ_x are strains anywhere in the region, E_1 .

α_1 is the coefficient of linear expansion of E_1 .

ΔT is the temperature change experienced by E_1 and is not a function of position or time.

ν_1 is Poisson's ratio for E_1 .

E_1 is the modulus of elasticity of material designated E_1 .

Since this stress exists throughout the core, it exists at its boundary and becomes the loading on the enclosure. In our problem for the interface of length b_o :

$$\epsilon_x = 0 \therefore \sigma_y = -\frac{E_1}{1 - \nu_1^2} \epsilon_y + \frac{\alpha_1 E_1 \Delta T}{1 - \nu_1}$$

We make the further simplified assumption that the strain ϵ_y at the interface is simply the ratio of the change in length to the original length. The original length is the initial (cold) state length: $a_o/2$, therefore $\epsilon_y = \Delta L/L_o = -(2y/a_o)$.

$$\therefore \sigma_y = \frac{2E_1}{a_o(1 - \nu_1^2)} y + \frac{\alpha_1 E_1 \Delta T}{1 - \nu_1} \quad (2)$$

Now this normal stress (2), being continuous across the boundary between E_1 and E_2 , provides an outward load on the tube. Considering E_2 of length b_o as a uniform simple beam of unit depth this stress is merely the load per unit length. Using the differential equation for the deflection of beams:³

$$E_2 I_2 \frac{d^4 y}{dx^4} = -q \quad (3)$$

where q is the intensity of the load per unit length and is considered positive acting in the negative y direction. Substituting (2) into (3):

$$\frac{d^4 y}{dx^4} + \left(\frac{E_1}{E_2 I_2} \right) \frac{2}{a_o(1 - \nu_1^2)} y = - \left(\frac{E_1}{E_2 I_2} \right) \frac{\alpha_1 \Delta T}{1 - \nu_1}$$

and I_2 is the moment of inertia of the beam. Let

$$P = \frac{E_1}{E_2 I_2}; \quad s^4 = \frac{2P}{a_o(1 - \nu_1^2)}; \quad M = \frac{P \alpha_1 \Delta T}{1 - \nu_1}; \quad \alpha^4 = \frac{s^4}{4} \quad (4)$$

$$\therefore \frac{d^4 y}{dx^4} + 4\alpha^4 y = -M.$$

Similar reasoning for the interface of length a_o leads to an analogous equation:

$$\frac{d^4 v}{du^4} + 4\beta^4 v = -M \quad (5)$$

where $t^4 = 2P/b_o(1 - \nu_1^2)$, $\beta^4 = t^4/4$ and all other terms are as previously defined.

Equations (4) and (5) are the governing differential equations which describe the general behavior of the rectangular shrink fit.

Note that both (4) and (5) have the same form as equations for beams on elastic foundations. This is understandable if we consider that the same condition exists whether the beam is embedded in a foundation that helps support the load as the beam deflects or that the intensity of the load decreases as the deflection increases, as is the case in this analysis.

2.2 A Special Case — the square shrink fit

The solutions and results of this case are carried out in Section A.2. It is determined that there is a definite wall thickness limit, determined by the choice of materials and size, below which no square enclosure can effectively be sealed by a shrink fit process.

III. RESULTS OF ANALYSIS

3.1 General

The final equations were programmed on the IBM 704 computer to include as wide a variety of combinations of materials and different size enclosures, as is practicable with materials that obey, at least in part, Hooke's Law of stress and strain.

The parameter chosen as a convenient variable to describe the merit of the shrink fit is called, here, the "shrink fit resistance," i.e., the higher the value of this parameter or "resistance" the lower the resulting compressive stresses. This parameter contains such constants of the configuration as Modulus of Elasticity of both the core and the enclosure, Poisson's ratio of the core, dimension of the enclosure, and thickness of the enclosure wall. This resistance is described by:

$$\frac{\alpha b_o}{2}$$

The initial run through the computer was for values of $\alpha b_o/2$ from 0 to 300. This range covered all possible combinations of interest. This run indicated that for values of resistance greater than 30 the expansion of the core was practically free-expansion, and the stresses between the core and the enclosure were zero at least to four decimal places. As an example, one can visualize trying to obtain a shrink fit of a block of

steel or titanium in a thin shell of cork or plastic. The block would expand as if the shell were nonexistent.

In a second run of the computer for values expanded between 0 and 30, it was determined that a seal could exist for a value of resistance lower than 2.10 for all sizes of rectangles. Above this value, separations would occur in a definite pattern for particular sizes. A "definite pattern" implies that at low values we obtain few but large separations and at progressively higher values we obtain many more but smaller separations until we approach (around 30) a "just-touching" situation of zero pressure. The results are approximately as shown in Fig. 2.

Relatively few rectangles can be sealed that possess values of resistance greater than 2.10, and less than 2.40. As was mentioned previously, these become fewer and fewer as the resistance becomes greater. Rectangles with resistance greater than 2.40 cannot be sealed.

An interesting phenomenon is that if the value of the resistance is great enough to prevent a seal, then this situation exists no matter at what temperature the shrink fit takes place. This is because the condition for a seal is independent of the temperature, while the intensity or "tightness" of the fit is directly proportional to the temperature.

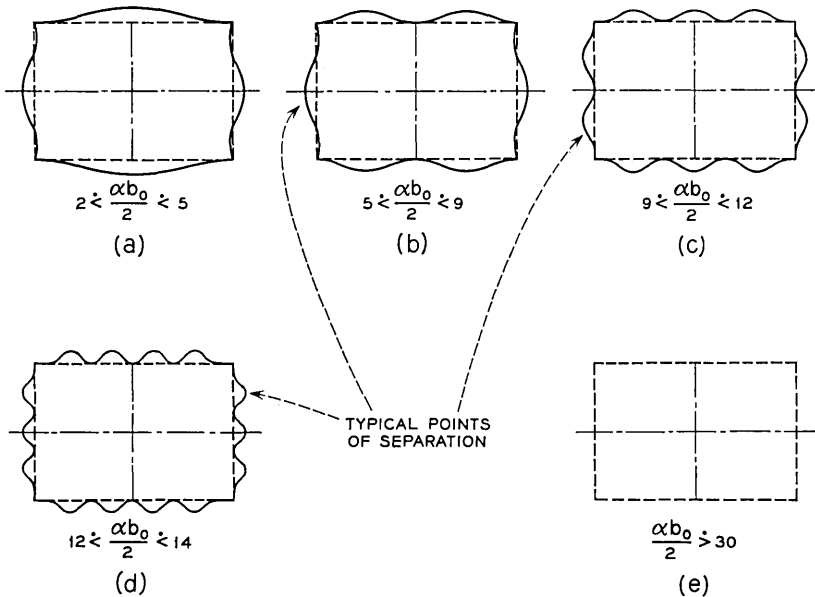


Fig. 2 — Typical separation patterns for values of $\alpha b_0/2$ between 2 and 30.

3.2 Detailed Results

An examination of Fig. 2 indicates typical modes of separation or leakage points as the resistance increases. It can be seen that separations occur in sequence. If $m = 0, 1, 2, 3 \dots$, modes of separation in any one narrow wall then there are $n = 1, 2, 3, 4 \dots$, corresponding separations in any one broad wall. i.e., $m = n - 1$ and the total number of separations in any mode is $4n - 2$.

A comparison of Fig. 3, which is a plot of stresses and displacements at the wall centers, with Fig. 2 clearly demonstrates how the wall centers have alternately positive and negative stresses. Fig. 3 also indicates the extreme reduction in the magnitude of the stresses beyond a resistance value of approximately 2.10. Therefore, although there are a few rectangles that can be sealed beyond $ab_0/2 = 2.10$, the intensity of the fit is very low. A good rule of thumb is to design a shrink fit to have a resistance less than 1.0. In this manner the lowest compressive stress is more than 75 per cent of the stress in an infinitely restricted expanded core.

For example, the resistance of teflon into small X-band waveguide is approximately 2.166. Results from the computer indicate a minimum compressive stress of approximately 10 pounds per square inch while the maximum compressive stress is over 1200 pounds per square inch.

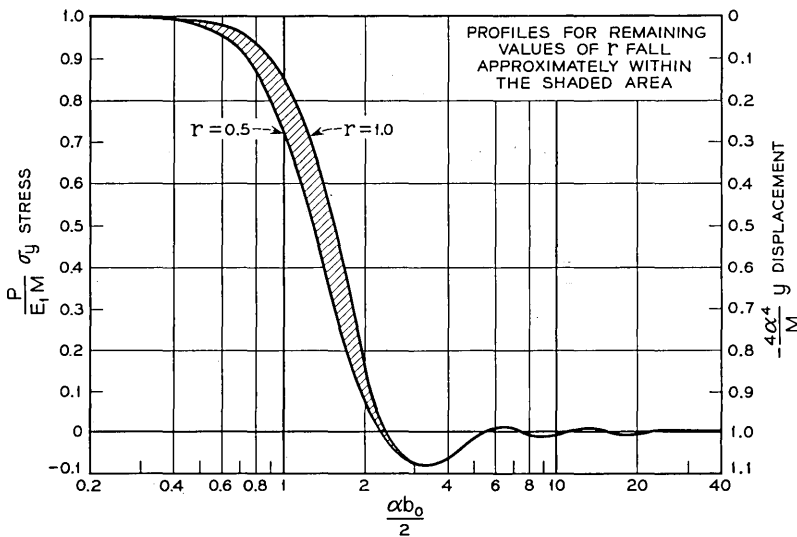


Fig. 3 — Stresses and displacements at wall centers for a square enclosure ($r = 1$) and a rectangular enclosure ($r = 0.5$).

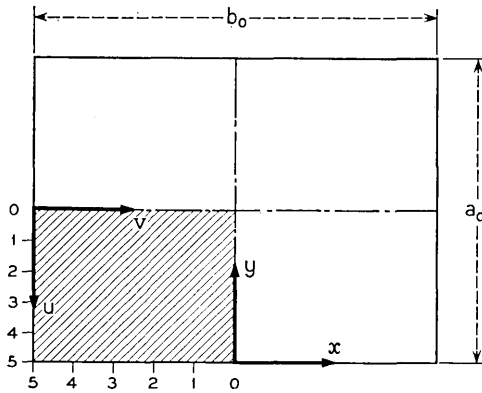


TABLE I — STRESSES AND DISPLACEMENTS FOR $\frac{1}{4}$ -SECTION OF RECTANGLE

$\frac{\alpha b_0}{2}$	$r = \frac{a_0}{b_0}$	Point	$\frac{4\alpha^4}{M} y$	$\frac{4\alpha^4}{M} v$	$\frac{P}{E_1 M} \sigma_y$	$\frac{P}{E_1 M} \sigma_v$
2.40	0.60	0	-1.00626	-0.549488	-0.00625819	0.670307
		1	-0.945438	-0.507973	0.0545624	0.695216
		2	-0.764464	-0.392764	0.235536	0.764342
		3	-0.480664	-0.232534	0.519336	0.860479
		4	-0.165034	-0.0771141	0.834966	0.953732
		5	0.000000	0.000000	1.000000	1.000000
2.40	0.70	0	-0.999453	-0.804216	0.000547044	0.437049
		1	-0.936954	-0.746459	0.0630464	0.477479
		2	-0.751725	-0.584672	0.248275	0.590730
		3	-0.463771	-0.355304	0.536229	0.751288
		4	-0.149463	-0.125321	0.850537	0.912275
		5	0.000000	0.000000	1.000000	1.000000
2.40	0.80	0	-1.00026	-0.957217	-0.000261307	0.234226
		1	-0.937961	-0.891096	0.0620387	0.287123
		2	-0.753238	-0.703685	0.246762	0.437052
		3	-0.465777	-0.432318	0.534223	0.654145
		4	-0.151312	-0.153320	0.848688	0.877344
		5	0.000000	0.000000	1.000000	1.000000

NOTE: Any negative sign appearing in the last two columns indicates a separation. All other values of r with $\alpha b_0/2 = 2.40$ cannot be sealed.

This is not considered an effective seal for the intended environment of one atmosphere.

Table I is an abstract from the computer runs. It indicates that there are a few rectangles with a resistance value greater than 2.10 that can be sealed — although not too effectively.

Therefore, although the resistance value is 2.40 and most of the sizes

of rectangles with this resistance value cannot be sealed, a rectangle with dimensions $a_o/b_o = 0.70$ can be sealed. If this intended seal is to withstand low pressures this combination may well be an effective seal. On the other hand, when $a_o/b_o = 0.6$ or 0.8 a seal cannot be made regardless of the temperature of the shrunk fit, since the quantity $(P/E_1M)\sigma_y$ is negative. In fact the larger ΔT (and therefore M) becomes, the greater the separation that results.

IV. RESULTS OF EXPERIMENT

In order to verify the preceding analytical interpretation it was necessary to experimentally perform the shrink fitting operation, and measure the subsequent deformation of the waveguide. Precision brass waveguide was selected with internal dimensions of $0.4000'' \times 0.9000''$, a wall thickness of $0.0500''$, a Modulus of Elasticity of approximately 17×10^6 p.s.i. Type 1 "Teflon" was selected for the inserts, with a Modulus of Elasticity of approximately 60,000 p.s.i. and a value of Poisson's Ratio of 0.46.

A total of 5 waveguide sections were used in the analysis. The guide was machined to 3" lengths while the "Teflon" inserts were cut in 2" lengths. The five "Teflon" inserts were machined oversize for the shrink fitting operation on the basis of the coefficients of expansion at three different temperatures. The temperatures selected were -60°F , -90°F and -120°F . This range was chosen since the proposed application of this analysis is for equipment which must operate satisfactorily over a temperature environment of -60°F to $+225^\circ\text{F}$. Since "Teflon" has a larger coefficient of expansion than brass the high temperature environ-

TABLE II

Block	Temp. °F	Coefficient Expansion—%	Calculated Size—inches	Actual Size—inches
1	-60	-1.00	0.4040 x 0.9090	0.4040 x 0.9080
2	-90	-1.18	0.4047 x 0.9106	0.4045 x 0.9098
3	-90	-1.18	0.4047 x 0.9106	0.4049 x 0.9098
4	-120	-1.28	0.4051 x 0.9115	0.4049 x 0.9107
5	-120	-1.28	0.4051 x 0.9115	0.4049 x 0.9103

Average Coefficients of Thermal Expansion⁴ for Type I "Teflon"

Temperature Range °C	Coefficient of Expansion per degree C
+25 to -50	135×10^{-6}
+25 to -100	112×10^{-6}
+25 to -150	96×10^{-6}

ment does not offer a problem. To effectively maintain a seal at the low end however, the "Teflon" blocks must be machined oversize for a temperature lower than -60°F . Each block is listed in Table II along with its associated temperature, coefficient of expansion and actual block dimensions.

To accurately measure deflections of the waveguide due to the expanding "Teflon" a recording system⁵ was devised which could provide a maximum magnification factor of 31,000 to 1. Basically it consists of a strain-gage activated transducer whose output is fed to an amplifying and recording system. The transducer itself (see Fig. 4) is essentially a cantilever beam which is deflected as the waveguide section is translated on a reference platform beneath the point probe. These deflections are picked off in strain gage outputs and fed to an amplifying system which controls a recorder chart. A deflectometer was used to insure synchronization of the chart drive and the trace of the transducer probe across the waveguide wall. The result is a magnified profile of the waveguide walls.

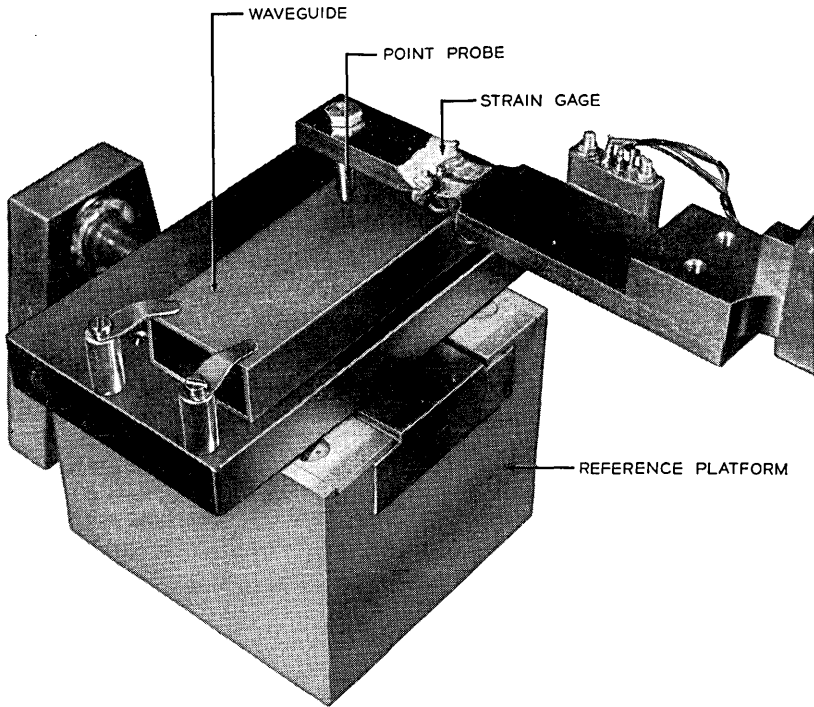


Fig. 4 — Apparatus used to measure deformation.

Prior to a discussion of the results a brief description of the shrink fit operation follows. Although the "Teflon" inserts as noted in Table I were machined oversize for three different temperatures, they were all inserted at a temperature of approximately -300°F . The coolant used in this operation was liquid nitrogen. The blocks were submerged in the liquid nitrogen for several minutes and then inserted into the waveguide allowing $\frac{1}{2}$ " opening on either end. The entire operation was performed in a controlled atmosphere of dry nitrogen to prevent the formation of frost on the "Teflon" blocks. By utilizing this extreme temperature coolant a tolerance was achieved so that upon removal of the blocks from the bath and prior to their immediate insertion into the guide, the resultant expansion was not sufficient to interfere with the placement of the cores. The specimens were allowed to stabilize for 30 minutes at the end of which time their profiles and resultant deflections were recorded. A 3-point suspension was used in mounting the waveguide sections to the reference platform so that accurate measurement of opposite sides of waveguide could be recorded after deflection had taken place. The profiles were monitored before and after insertion of the plug at various stations ($\frac{1}{2}$ inch) along the length of the waveguide, including

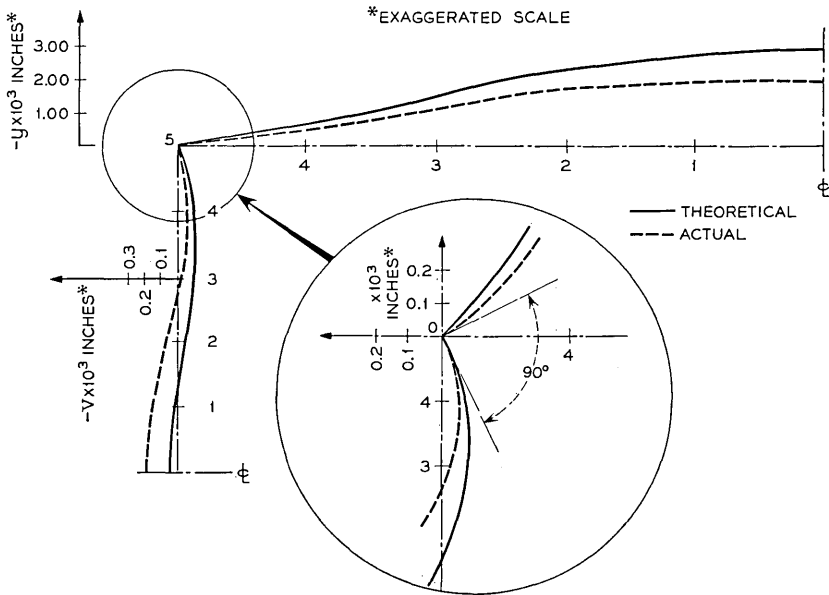


Fig. 5 — Predicted vs actual displacements at -60°F . Expanded portion (circled) illustrates preservation of right angle at corner.

the middle of the 2" "Teflon." The profiles at each station followed the same general pattern with the maximum expansion occurring at the center of the 2" "Teflon."

The curves (Figs. 5, 6, 7) were plotted for a quarter section of waveguide to illustrate the resultant expansion produced by the shrink fit operation. The predicted analytical expansion was plotted on the same scale, so a comparison could be made between the two. The analytical analysis assumed that the waveguide corners remained at right angles to each other during the expansion process. This initial assumption was borne out in the experimental data as a valid one. In order to insure the validity of this phenomenon additional profile measurements were performed on an expanded scale directly in the corner region. These measurements definitely indicated a negative or inward expansion of the narrow wall, preserving the right angle corner of the waveguide.

Although predicted and experimental expansions are not in exact agreement, the profiles of each follow the same pattern. The differences in the magnitudes of the total expansions can be attributed to several factors. Machining tolerances in both the brass waveguide and "Teflon" inserts could not be held much closer than 0.0005 inch. In addition, cold flow of the "Teflon" inserts occurring immediately after their insertion

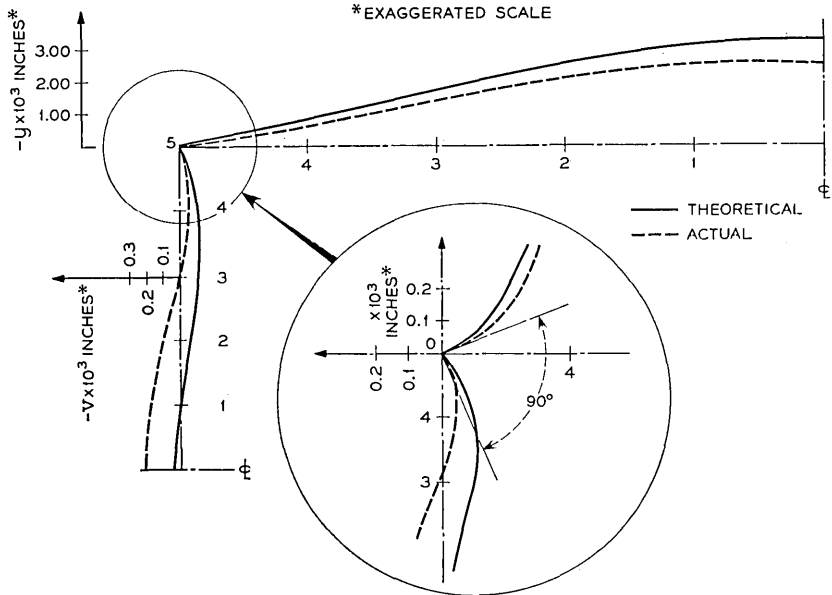


Fig. 6 — Predicted vs actual displacements at -90°F . Expanded portion (circled) illustrates preservation of right angle at corner.

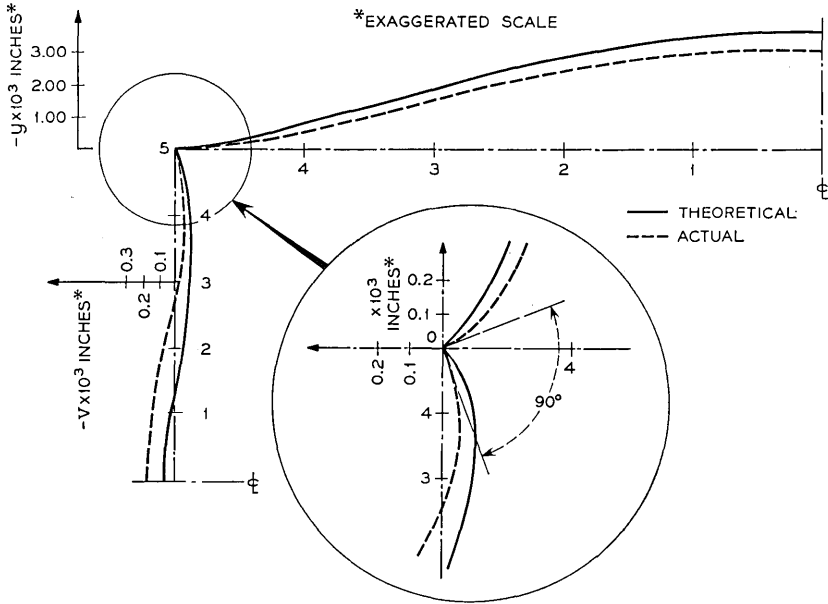


Fig. 7 — Predicted vs actual displacements at -120°F . Expanded portion (circled) illustrates preservation of right angle at corner.

and prior to remeasurement of the waveguide sections introduced an additional discrepancy.

It is apparent from the curves that because of this cold flow the loading intensity of the shrink-fit was decreased, resulting in smaller deflections in both the broad and narrow walls. Nevertheless, the predicted and experimental profiles are in relative agreement with respect to the basic assumptions.

V. GENERAL INFORMATION

Table III indicates the magnitudes of the stresses and displacements one might expect from typical materials and various sizes of rectangles.

VI. SUMMARY AND CONCLUSIONS

The stresses and deformations resulting from rectangular shrink fits can be described by general differential equations of the form:

$$\frac{d^4 y}{dx^4} + Ay = B; \quad \sigma_y = Cy + D$$

TABLE III—STRESSES AND DISPLACEMENTS AT WALL CENTERS

$\frac{\alpha b_0}{2}$	$r = \frac{a_0}{b_0}$	$\frac{4\alpha^4}{M} \gamma$	$\frac{4\alpha^4}{M} v \times 10^2$	$\frac{P}{E_1 M} \sigma_y$	$\frac{P}{E_1 M} \sigma_v$
0.5	0.1	-0.0140475	0.0369430	0.985952	1.00004
	0.2	-0.0168590	0.129932	0.983141	1.00026
	0.3	-0.0188721	0.250018	0.981128	1.00075
	0.4	-0.0200661	0.366116	0.979934	1.00146
	0.5	-0.0204413	0.444737	0.979559	1.00222
	0.6	-0.0200013	0.450008	0.979999	1.00270
	0.7	-0.0187518	0.343972	0.981248	1.00241
	0.8	-0.0167051	0.0873193	0.983295	1.00070
	0.9	-0.0138850	-0.359327	0.986115	0.996766
	1.0	-0.0103330	-1.03330	0.989667	0.989667
1.0	0.1	-0.192576	0.514591	0.807424	1.00051
	0.2	-0.224629	1.74104	0.775371	1.00348
	0.3	-0.245227	3.22422	0.754773	1.00967
	0.4	-0.255226	4.50577	0.744774	1.01802
	0.5	-0.255174	5.10577	0.744826	1.02553
	0.6	-0.245660	4.53701	0.754340	1.02722
	0.7	-0.227738	2.36752	0.772262	1.01657
	0.8	-0.203330	-1.66401	0.796670	0.986688
	0.9	-0.175387	-7.51309	0.824613	0.932382
	1.0	-0.147551	-14.7551	0.852449	0.852449
1.5	0.1	-0.603086	1.68992	0.396914	1.00169
	0.2	-0.657292	5.20538	0.342708	1.01041
	0.3	-0.682426	8.66591	0.317574	1.02600
	0.4	-0.683702	10.3389	0.316298	1.04136
	0.5	-0.664697	8.53652	0.335303	1.04268
	0.6	-0.630410	2.01635	0.369590	1.01210
	0.7	-0.588972	-9.25054	0.411028	0.935246
	0.8	-0.550402	-23.6036	0.449598	0.811171
	0.9	-0.522637	-38.2895	0.477363	0.655394
	1.0	-0.508389	-50.8389	0.491611	0.491611
2.0	0.1	-0.928988	2.91315	0.0710119	1.00291
	0.2	-0.960541	7.98587	0.0394589	1.01597
	0.3	-0.965532	11.0070	0.0344678	1.03302
	0.4	-0.950164	8.13079	0.0498358	1.03252
	0.5	-0.920893	-3.46879	0.0791066	0.982656
	0.6	-0.888056	-23.2032	0.111944	0.860781
	0.7	-0.862943	-46.0441	0.137057	0.677691
	0.8	-0.851041	-65.7452	0.148959	0.474039
	0.9	-0.850384	-78.9297	0.149616	0.289633
	1.0	-0.855904	-85.5904	0.144096	0.144096
2.5	0.1	-1.06440	4.01305	-0.0644005	1.00401
	0.2	-1.07415	9.56188	-0.0741458	1.01912
	0.3	-1.06987	8.90597	-0.0698685	1.02672
	0.4	-1.05587	-4.50505	-0.0558715	0.981980
	0.5	-1.03877	-31.5782	-0.0387744	0.842109
	0.6	-1.02657	-63.3918	-0.0265694	0.619649
	0.7	-1.02256	-88.2925	-0.0225558	0.381953
	0.8	-1.02433	-101.705	-0.0243283	0.186362
	0.9	-1.02825	-105.513	-0.0282519	0.0503840
	1.0	-1.03214	-103.214	-0.0321412	-0.0321412

TABLE III (Cont'd)

$\frac{\alpha b_0}{2}$	$r = \frac{a_0}{b_0}$	$\frac{4\alpha^4}{M} y$	$\frac{4\alpha^4}{M} v \times 10^2$	$\frac{P}{E_1 M} \sigma_y$	$\frac{P}{E_1 M} \sigma_v$
3.0	0.1	-1.08931	5.15386	-0.0893139	1.05154
	0.2	-1.09015	10.1753	-0.0901539	1.10175
	0.3	-1.08848	2.16564	-0.0884799	1.02166
	0.4	-1.08547	-26.9990	-0.0854682	0.730010
	0.5	-1.08295	-68.9590	-0.0829545	0.310410
	0.6	-1.08209	-103.274	-0.0820925	-0.0327402
	0.7	-1.08246	-120.106	-0.0824568	-0.201057
	0.8	-1.08320	-122.784	-0.0831998	-0.227840
	0.9	-1.08387	-117.430	-0.0838700	-0.174300
	1.0	-1.08436	-108.436	-0.0843641	-0.0843641
4.0	0.1	-1.04066	7.49643	-0.0406582	1.07496
	0.2	-1.04197	7.03350	-0.0419736	1.07033
	0.3	-1.04981	-28.3538	-0.0498060	0.716462
	0.4	-1.05711	-92.8190	-0.0571056	0.0718099
	0.5	-1.05871	-141.325	-0.0587114	-0.413247
	0.6	-1.05692	-156.244	-0.0569194	-0.562436
	0.7	-1.05484	-150.169	-0.0548359	-0.501691
	0.8	-1.05336	-135.702	-0.0533627	-0.357021
	0.9	-1.05238	-119.646	-0.0523759	-0.196463
	1.0	-1.05163	-105.163	-0.0516336	-0.0516336
5	0.5	-1.01252	-187.848	-0.0125207	-0.878477
	1.0	-1.00910	-100.910	-0.00910099	-0.00910099
7	0.5	-0.997154	-217.039	0.00284579	-1.17039
	1.0	-0.997427	-99.7427	0.00257313	0.00257313
10	0.5	-1.00014	-207.423	-0.000136673	-1.07423
	1.0	-1.00013	-100.013	-0.000125572	-0.000125572
15	0.5	-1.000000	-199.256	0.223517 $\times 10^{-7}$	-0.992557
	1.0	-1.000000	-100.000	-0.596046 $\times 10^{-7}$	-0.596046 $\times 10^{-7}$
20	0.5	-1.000000	-199.981	-0.	-0.999813
	1.0	-1.000000	-100.000	-0.	-0.
30	0.5	-1.000000	-199.999	-0.	-0.999986

where y is the deflection at any point x along the interface, σ_y is the stress at that point, and A, B, C and D are constants.

For any rectangular enclosure and core there exists a practical limit above which no effective seal can be obtained. This limit is defined as the shrink fit resistance and is a function of dimensions and materials. The limiting value of resistance is 2.40. All rectangles with resistance values less than 2.10 can be sealed. Rectangles with resistances between 2.10 and 2.40 may or may not be capable of being sealed, depending on the details of configuration and materials.

Pressure sealing of waveguide is a typical application of shrink fits between rectangular connections. Usually the designer may choose the material and thickness of the waveguide. The choice should be made to limit the resistance to a value of 2.0 or less, regardless of the size of the enclosure or the temperature at which the shrink fit takes place. Therefore, for given materials and size of rectangle a pressure seal will exist if the wall thickness is

$$t \geq \frac{1}{3.5} \left[\left(\frac{E_1}{E_2} \right) \frac{b_o^4}{a_o(1 - \nu_1^2)} \right]^{\frac{1}{3}}$$

where E_1 , E_2 are the modulus of elasticity of the core and enclosure, respectively

ν_1 is Poisson's ratio for the core

a_o , b_o are the narrow and broad wall dimension, respectively.

VII. ACKNOWLEDGMENT

The author wishes to acknowledge the helpful and stimulating discussions held with G. J. Bareille. The experimental analysis was performed under his direction.

APPENDIX

A.1 The General Solution

The general solution to (4) is:

$$y = A \sin \alpha x \sinh \alpha x + B \sin \alpha x \cosh \alpha x + C \cos \alpha x \sinh \alpha x + D \cos \alpha x \cosh \alpha x - (M/4\alpha^4) \quad (6)$$

where

$$4\alpha^4 = s^4.$$

Taking the origin of the coordinates in the middle of beam b_o as in Fig. 1, we conclude from symmetry that $B=C=0$.

$$\therefore y = A \sin \alpha x \sinh \alpha x + D \cos \alpha x \cosh \alpha x - (M/4\alpha^4). \quad (7)$$

From the boundary conditions at the ends of the beam: $y = 0$ at $x = b_o/2$

$$\therefore A = \frac{\frac{M}{4\alpha^4} - D \cos \alpha \frac{b_o}{2} \cosh \alpha \frac{b_o}{2}}{\sin \frac{\alpha b_o}{2} \sinh \frac{\alpha b_o}{2}}. \quad (8)$$

Similar reasoning leads to the deflection equation for beam a_o with coordinate axes u and v in the middle of beam a_o :

$$v = E \sin \beta u \sinh \beta u + H \cos \beta u \cosh \beta u - (M/4\beta^4). \tag{9}$$

where

$$\beta^4 = \frac{t^4}{4} \quad \text{and} \quad t^4 = \frac{2P}{b_o(1 - \nu_1^2)}$$

$$\therefore E = \frac{\frac{M}{4\beta^4} - H \cos \frac{\beta a_o}{2} \cosh \frac{\beta a_o}{2}}{\sin \frac{\beta a_o}{2} \sinh \frac{\beta a_o}{2}} \tag{10}$$

We assume that the corners of the tube, E_2 , are much stiffer than the "beams," and that the right angles formed by both sides remain right angles. We also assume that any deformations to cause the corners to open would exceed the elastic limit of the tube and would result in permanent deformation. We are not concerned with this situation.

In view of the above, the slopes at the corners are equal:

$$y'_{x=-(b_o/2)} = \dot{v}_{u=a_o/2} \text{ as indicated in Fig. 8} \tag{11}^*$$

The last boundary condition necessary is that the moments at the ends of the beams are equal:

$$y''_{x=-(b_o/2)} = \ddot{v}_{u=a_o/2} \tag{12}$$

Now

$$y' = \alpha A [\sin \alpha x \cosh \alpha x + \sinh \alpha x \cos \alpha x]$$

$$+ \alpha D [\cos \alpha x \sinh \alpha x - \sin \alpha x \cosh \alpha x]$$

\therefore substituting (8) and evaluating y' at $x = -(b_o/2)$

$$y'_{x=-(b_o/2)} = \alpha D \left[\frac{\sin \alpha b_o + \sinh \alpha b_o}{2 \sin \frac{\alpha b_o}{2} \sinh \frac{\alpha b_o}{2}} \right] - \frac{M}{4\alpha^3} \left(\cot \frac{\alpha b_o}{2} + \coth \frac{\alpha b_o}{2} \right)$$

Likewise for beam a_o : \dot{v} at $u = a_o/2$

$$\dot{v}_{u=a_o/2} = -\beta H \left[\frac{\sin \beta a_o + \sinh \beta a_o}{2 \sin \frac{\beta a_o}{2} \sinh \frac{\beta a_o}{2}} \right] + \frac{M}{4\beta^3} \left(\cot \frac{\beta a_o}{2} + \coth \frac{\beta a_o}{2} \right)$$

* Note: Primes indicate derivatives with respect to x and dots indicate derivatives with respect to u .

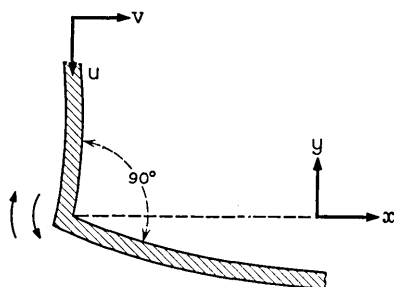


Fig. 8 — Conditions at corner of tube.

From boundary condition (11) we can solve for D :

$$D = \frac{1}{\alpha E} \left[\frac{M}{4} \left(\frac{K}{\beta^3} + \frac{F}{\alpha^3} \right) - H\beta G \right] \quad (13)$$

where

$$E = \frac{\sin \alpha b_o + \sinh \alpha b_o}{2 \sin \frac{\alpha b_o}{2} \sin \frac{\alpha b_o}{2}}$$

$$G = \frac{\sin \beta a_o + \sinh \beta a_o}{2 \sin \frac{\beta a_o}{2} \sinh \frac{\beta a_o}{2}}$$

$$F = \cot \frac{\alpha b_o}{2} + \coth \frac{\alpha b_o}{2}$$

$$K = \cot \frac{\beta a_o}{2} + \coth \frac{\beta a_o}{2}.$$

Now $y'' = 2\alpha^2 [A \cos \alpha x \cosh \alpha x - D \sin \alpha x \sinh \alpha x]$

$$\therefore y''_{x=-b_o/2} = \frac{-2\alpha^2 D \left(\cos^2 \frac{\alpha b_o}{2} + \sinh^2 \frac{\alpha b_o}{2} \right) + \frac{M}{2\alpha^2} \left(\cos \frac{\alpha b_o}{2} \cosh \frac{\alpha b_o}{2} \right)}{\sin \frac{\alpha b_o}{2} \sinh \frac{\alpha b_o}{2}}$$

Likewise for beam a_o :

$$\ddot{v}_{u=a_o/2} = \frac{-2\beta^2 H \left(\cos^2 \frac{\beta a_o}{2} + \sinh^2 \frac{\beta a_o}{2} \right) + \frac{M}{2\beta^2} \left(\cos \frac{\beta a_o}{2} \cosh \frac{\beta a_o}{2} \right)}{\sin \frac{\beta a_o}{2} \sinh \frac{\beta a_o}{2}}$$

∴ from boundary condition (12) we again solve for D :

$$D = \frac{\frac{\beta^2 H}{\alpha^2} \left(\cos^2 \frac{\beta a_o}{2} + \sinh^2 \frac{\beta a_o}{2} \right) - \frac{M}{4\beta^2 \alpha^2} \left(\cos \frac{\beta a_o}{2} \cosh \frac{\beta a_o}{2} \right)}{\frac{\sin \frac{\beta a_o}{2} + \sinh \frac{\beta a_o}{2}}{\left(\frac{\cos^2 \frac{\alpha b_o}{2} + \sinh^2 \frac{\alpha b_o}{2}}{\sin \frac{\alpha b_o}{2} \sinh \frac{\alpha b_o}{2}} \right)}} + \frac{\frac{M}{4\alpha^4} \left(\cot \frac{\alpha b_o}{2} \coth \frac{\alpha b_o}{2} \right)}{\left(\frac{\cos^2 \frac{\alpha b_o}{2} + \sinh^2 \frac{\alpha b_o}{2}}{\sin \frac{\alpha b_o}{2} \sinh \frac{\alpha b_o}{2}} \right)} \tag{14}$$

∴ to evaluate H , we equate Equations (13) and (14); and let

$$\frac{\beta^4}{\alpha^4} = \frac{a_o}{b_o} = r, \quad \text{then} \quad \frac{4\alpha^4 H}{M} = \frac{B'}{A'}$$

where,

$$B' = 2 \left[\frac{\left(\cot \frac{\alpha b_o}{2} + \coth \frac{\alpha b_o}{2} \right) + r^{-1} \left(\cot \frac{\alpha b_o}{2} r^{\frac{1}{2}} + \coth \frac{\alpha b_o}{2} r^{\frac{1}{2}} \right)}{\sin \alpha b_o + \sinh \alpha b_o} \right] + \left[\frac{1}{\sqrt{r}} \frac{\left(\cot \frac{\alpha b_o}{2} r^{\frac{1}{2}} \coth \frac{\alpha b_o}{2} r^{\frac{1}{2}} \right) - \left(\cot \frac{\alpha b_o}{2} \coth \frac{\alpha b_o}{2} \right)}{\cos^2 \frac{\alpha b_o}{2} + \sinh^2 \frac{\alpha b_o}{2}} \right]$$

and

$$A' = r^{\frac{1}{2}} \left[\frac{\sin \alpha b_o r^{\frac{1}{2}} + \sinh \alpha b_o r^{\frac{1}{2}}}{\left(\sin \frac{\alpha b_o}{2} r^{\frac{1}{2}} \sinh \frac{\alpha b_o}{2} r^{\frac{1}{2}} \right) (\sin \alpha b_o + \sinh \alpha b_o)} \right] + \sqrt{r} \left[\frac{\cos^2 \frac{\alpha b_o}{2} r^{\frac{1}{2}} + \sinh^2 \frac{\alpha b_o}{2} r^{\frac{1}{2}}}{\left(\sin \frac{\alpha b_o}{2} r^{\frac{1}{2}} \sinh \frac{\alpha b_o}{2} r^{\frac{1}{2}} \right) \left(\cos^2 \frac{\alpha b_o}{2} + \sinh^2 \frac{\alpha b_o}{2} \right)} \right]$$

With this value of H , D can be determined from either (13) or (14). Finally:

$$\frac{4\alpha^4}{M} y = \frac{4\alpha^4 D}{M} \left[\cos \alpha x \cosh \alpha x - \left(\cot \frac{\alpha b_o}{2} \coth \frac{\alpha b_o}{2} \right) \cdot \sin \alpha x \sinh \alpha x \right] - \left(1 - \frac{\sin \alpha x \sinh \alpha x}{\sin \frac{\alpha b_o}{2} \sinh \frac{\alpha b_o}{2}} \right) \quad (16)$$

where

$$-\frac{b_o}{2} \leq x \leq \frac{b_o}{2}$$

and

$$\begin{aligned} \frac{4\alpha^4}{M} v = \frac{4\alpha^4 H}{M} \left[\cos \alpha r^{\ddagger} u \cosh \alpha r^{\ddagger} u \right. \\ \left. - \left(\cot \frac{\alpha b_o}{2} r^{\ddagger} \coth \frac{\alpha b_o}{2} r^{\ddagger} \right) \sin \alpha r^{\ddagger} u \sinh \alpha r^{\ddagger} u \right] \\ - \frac{1}{r} \left(1 - \frac{\sin \alpha r^{\ddagger} u \sinh \alpha r^{\ddagger} u}{\sin \frac{\alpha b_o}{2} r^{\ddagger} \sinh \frac{\alpha b_o}{2} r^{\ddagger}} \right) \end{aligned} \quad (17)$$

where

$$-\frac{r b_o}{2} \leq u \leq \frac{r b_o}{2}$$

or

$$-\frac{a_o}{2} \leq u \leq \frac{a_o}{2}$$

and the stresses for interfaces b_o and a_o are respectively:

$$\frac{P}{E_1 M} \sigma_y = \frac{4\alpha^4}{M} y + 1 \quad \text{and} \quad \frac{P}{E_1 M} \sigma_v = \frac{4\alpha^4 r}{M} v + 1.$$

A.2 A Special Case: The Square Shrink Fit

As a point of interest: if $a_o = b_o$ (square enclosure); $r = 1$

$$H = D = \frac{M}{2\alpha^4} \left(\cot \frac{\alpha b_o}{2} + \coth \frac{\alpha b_o}{2} \right) \sin \frac{\alpha b_o}{2} \sinh \frac{\alpha b_o}{2} \quad (18)$$

and

$$y = v = \frac{M}{2\alpha^4} \left[\left(\frac{\cot \alpha \frac{b_o}{2} + \coth \frac{\alpha b_o}{2}}{\sin \alpha b_o + \sinh \alpha b_o} \right) \left(\sin \frac{\alpha b_o}{2} \sinh \frac{\alpha b_o}{2} \cos \alpha x \right. \right. \\ \left. \left. \cdot \cosh \alpha x - \cos \frac{\alpha b_o}{2} \cosh \frac{\alpha b_o}{2} \sin \alpha x \sinh \alpha x \right) \right. \\ \left. - \frac{1}{2} \left(1 - \frac{\sin \alpha x \sinh \alpha x}{\sin \frac{\alpha b_o}{2} \sinh \frac{\alpha b_o}{2}} \right) \right] \tag{19}$$

The maximum deflection is at $x = 0$:

$$\therefore y = v |_{\max} = \frac{M}{2\alpha^4} \left[\frac{\left(\cos \frac{\alpha b_o}{2} - \cosh \frac{\alpha b_o}{2} \right) \left(\sinh \alpha \frac{b_o}{2} - \sin \alpha \frac{b_o}{2} \right)}{\sin \alpha b_o + \sinh \alpha b_o} \right] \tag{20}$$

It can also be shown that the slopes are zero at the end of the interfaces and at the four midpoints, when $r = 1$.

The condition for zero stress at $x = u = 0$ is obtained from (2) as

$$y \leq -\frac{\alpha \Delta T a_o (1 + \nu_1)}{2} = -\frac{M}{4\alpha^4} \tag{21}$$

The solution breaks down wherever a tension at the interface is indicated, since the load becomes zero over the separated portion of the beam. The calculated tension is that which would be required to prevent separation of E_1 from the computed deflection curve. Substituting (21) into (16) when $x = 0$ gives $D = 0$; then from (18) we have the necessary condition that

$$\cot \frac{\alpha b_o}{2} = -\coth \frac{\alpha b_o}{2}$$

therefore from Fig. 3 the eigenvalues of this equation are determined. (Although these values are for a square and indicate resistance values which result in zero stress at the wall centers, they are approximately the boundaries for the ranges of the modes of separation for any rectangle.) (Compare Fig. 9 with Fig. 3.)

$$\therefore \text{eigenvalues of } \frac{\alpha b_o}{2} = 2.365; 5.498; 8.639; 11.781 \dots$$

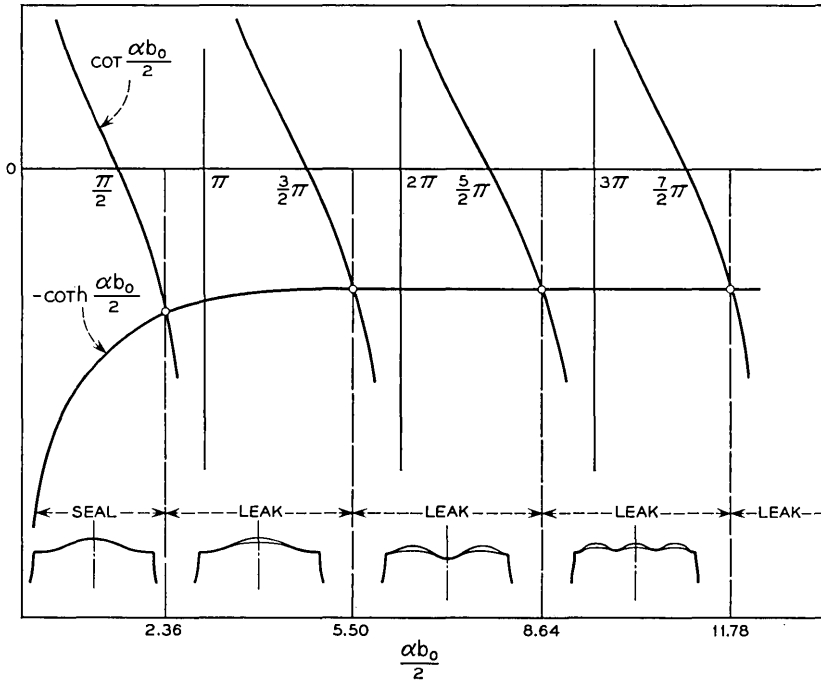


Fig. 9 — Square shrink fit: boundaries for modes of separation. The values shown are also approximately valid for any rectangle.

These eigenvalues after the first (2.365) are for all practical purposes separated by π .

To extend this to the more general case of rectangular enclosures we can determine from (16), (17), (2) and the equivalent of (2) the following at $x = 0$ and $u = 0$ respectively (midpoints of the enclosure walls);

$$\sigma_y = \frac{E_1}{P} 4\alpha^4 D \quad \text{and} \quad \sigma_v = \frac{E_1}{P} 4\alpha^4 r H.$$

$\therefore \sigma_y$ and σ_v have the same sign (\pm) as the constants D and H respectively. When these constants are negative the stress on the insert is a pseudo-tension or in reality a separation of the insert and the enclosure at the midpoint of the enclosure walls. Care must be exercised in determining whether a seal exists or not. If the sign of either constant D or H is negative or if either constant is zero then a seal is not possible. A seal will exist only if both constants are positive.

REFERENCES

1. Bareille, G. J., unpublished memorandum.
2. Wang, C. T., *Applied Elasticity*, McGraw-Hill, N. Y., 1953.
3. Timoshenko, S., *Strength of Materials, Part II, Advanced Theory and Problems*, 3rd Edition, Van Nostrand, Princeton, N. J., 1956.
4. Kirby, R. K., U. S. Bur. Standards J. Research **57**, No. 2, August, 1956.
5. Benson, R. J., unpublished memorandum.

Monte Carlo Solution of Bond Percolation Processes in Various Crystal Lattices

By H. L. FRISCH, S. B. GORDON, V. A. VYSSOTSKY, and
J. M. HAMMERSLEY†

(Manuscript received November 29, 1961)

We present the outline of an IBM 7090 machine program for the Monte Carlo estimation of the percolation probability for a variety of space lattices. The underlying theory is briefly summarized.

I. STATEMENT OF THE PROBLEM

Percolation processes deal with the transmission of a “fluid” (disturbance, signal, etc.) through a “medium” (material, region, etc.) against impediment by random irregularities situated in the medium.¹ This paper considers the case where the medium is a regular crystal lattice in two or three dimensions, consisting of “atoms” (the vertices or sites of the lattice) and “bonds” joining specified pairs of atoms. The next section will specify the structure of the lattice more completely. The fluid originates at one or more atoms of the lattice, called the source atoms, and flows from atom to atom along the connecting bonds. However, each bond (independently of all other bonds) has a fixed probability p of being able to transmit fluid and a probability $q = 1 - p$ of being blocked: these randomly situated blocked bonds constitute the random impediments to the spread of the fluid. We write $P_N(p)$ for the probability that the fluid will reach (or “wet”, as we shall say) more than N other atoms besides the source atoms; and the problem is to estimate $P(p) = \lim_{N \rightarrow \infty} P_N(p)$. We do this by estimating $P_N(p)$ for a suitably large value of N : it turns out that $N \sim 2000$ is sufficient in many cases. The present paper describes the general organization of an IBM 7090 program for obtaining a Monte Carlo estimate of $P_N(p)$. The numerical results appear elsewhere.²

† Oxford University, Oxford, England.

II. STRUCTURE OF THE LATTICE

We consider below the three-dimensional problem in such a way that it contains as a special case the two-dimensional problem. A regular three-dimensional lattice consists of a number of fundamental cells, all identical apart from their position in space. Each cell is specified by integer coordinates (x, y, z) representing its position in space $(x, y, z = 0, \pm 1, \pm 2, \dots)$. Each cell contains a finite number of atoms, limited in our IBM 7090 program to a maximum of eight atoms per cell, and denoted by A, B, \dots, H .

In the problem, as originally posed in the previous section, some of the bonds may be one-way (i.e., only able to transmit fluid in a specified direction) while others may be two-way (i.e., able to transmit fluid in either direction). It can, however, be proved theoretically that $P_N(p)$ is unaltered if a two-way bond is replaced by two one-way bonds of opposite directions. It is therefore enough to consider the case where all bonds are one-way bonds, and we confine ourselves to this case hereafter.

Let us write $T(x, y, z)$ for the atom of type T ($T = A, B, \dots, H$) in cell (x, y, z) . The lattice is regular in the sense that, if there is a bond from $T(x, y, z)$ to $T^*(x^*, y^*, z^*)$, then there is a bond from $T(x + l, y + m, z + n)$ to $T^*(x^* + l, y^* + m, z^* + n)$ for any $l, m, n = 0, \pm 1, \pm 2, \dots$. Therefore we need only specify the terminal atoms reached by each bond from each atom of the cell $(0, 0, 0)$. The program limits the number of bonds from a given atom to a maximum of 12. There is no restriction that distinct bonds from a given atom shall lead to distinct terminal atoms; and thus we may, if we wish, have two or more bonds in the same direction between a given pair of atoms.

The machine receives information about the lattice structure from a series of input cards, having the format described below. To each atom of the cell $(0, 0, 0)$ there is a "structure" card, followed perhaps by one or two "continuation" cards. The format of a structure card is:

Columns	5-13	15-23	24-26	27-35	36	37-45	46	47-55	56	57-65	70-72
Contents	STRUCTURE	T(+0+0+0)	T \emptyset	T(xy z)	,	T(xy z)	,	T(xy z)	,	T(xy z)	CT1

In such a card, T (in columns 15, 27, 37, 47, 57) stands for one of the letters A, B, \dots, H (not necessarily the same in each case); and each of the symbols x, y, z is a *signed* integer. The entries in columns 5-13, 24-26, 36, 46, 56, 70-72 are in BCD (binary coded decimal). Instead of the whole word *STRUCTURE* in columns 5-13, the single letter S in column 5 will suffice. Columns 15-23 specify the atom of cell $(0, 0, 0)$ from which bonds lead to the atoms appearing in columns 27-35, 37-45, 47-55, and 57-65 respectively. Thus the structure and card provide for

up to four bonds from the given atom of cell $(0, 0, 0)$. If there are fewer than four bonds, some of the $T(xyz)$ will be left blank. The next four (or fewer) bonds are similarly specified on a first continuation card with format:

Columns	5-7		27-65		70-72
Contents	CT1		Same format as structure card		CT2

The last four (or fewer) bonds appear on a second continuation card with format:

Columns	5-7		27-65		70
Contents	CT2		Same format as structure card		S

In both continuation cards, columns 5-7 and 70-72 are in BCD. If the atom specified in columns 15-23 of the structure card has four or fewer bonds from it, both continuation cards are omitted and we replace CT1 in columns 70-72 by *S* in column 70 (in BCD); if it has between 5 and 8 (both inclusive) bonds from it, the second continuation card is omitted and we replace CT2 in columns 70-72 by *S* in column 70. The final structure on continuation card is followed by an "end" card with *END* (in BCD) in columns 5-7.

For example, the simple cubic lattice with a pair of one-way bonds (one in each direction) between each pair of nearest-neighbor atoms is specified by:

```

STRUCTURE A(+0+0+0) TØ- A(+1+0+0), A(+0+1+0), A(+0+0+1), A(-1+0+0) CT1
CT1       A(+0+0+0) TØ- A(+0-1+0), A(+0+0-1) S
END

```

Similarly, the tetrahedral lattice (diamond crystal) with a pair of one-way bonds in each direction between nearest neighbors is given by:

```

STRUCTURE A(+0+0+0) TØ- B(+0+0+0), B(+1+0+0), B(+0+1+0), B(+0+0+1) S
S         B(+0+0+0) TØ- A(+0+0+0), A(-1+0+0), A(+0-1+0), A(+0+0-1) S
END

```

The two-dimensional cases arise when $z = 0$ identically on all cards.

III. INPUT TO THE MACHINE

The complete input to the machine consists of the program deck, followed by (i) an identification card, followed by (ii) a parameter data card, followed by (iii) a set of structure and continuation cards (as described above), followed by (iv) an end card (as described above).

The format of the identification card is:

Columns	5-15		18-25		26-66		68-72
Contents	PERÇOLATIØN		CRYSTAL		Name of crystal		PUNCH

These entries are all in BCD. If *PUNCH* appears in columns 68–72, the output from the machine will appear on the printout and will also be punched onto cards; if *PUNCH* is omitted from columns 68–72, the output will appear on the printout but cards will not be punched.

The format of the parameter data card is:

Columns	5-8	10-12	13	14-16	19-23	25-28	31-37	39-47	48	49-57
Contents	<i>RUNS</i>	***	-	***	<i>LINES</i>	****	<i>SOURCES</i>	<i>T(xyz)</i>	,	<i>T(xyz)</i>

Columns 5–8, 13, 19–23, 31–37, 48 are in BCD. Columns 10–12 contain a first run number; columns 14–16 contain a last run number: the purpose of these two run numbers will be described presently. All run numbers must be *positive* integers. Columns 25–28 contain the number N , appearing as a suffix in the desired function $P_N(p)$. The maximum value of N permitted by the program is 7000.

IV. GENERAL CONDUCT OF THE CALCULATION

The calculation carried out by the machine is at first sight rather different from the problem posed in the first section of this paper. The change of formulation simplifies the calculation without affecting the ultimate numerical answer.

The calculation consists of a number of separate runs. In each individual run, independent random numbers η are assigned to each bond of the lattice. Each η is uniformly distributed between 0 and 1. This process of assigning η -values to the bonds replaces the original process of blocking bonds, so that the question of a bond being blocked or not does not arise in the reformulated process. Consider a connected path of bonds $\beta_1, \beta_2, \dots, \beta_k$ on the lattice, where the (necessarily one-way) bond β_i leads *from* the atom *to* which the preceding bond β_{i-1} led ($i > 1$). Let $\eta_1, \eta_2, \dots, \eta_k$ be the η -values of the respective bonds $\beta_1, \beta_2, \dots, \beta_k$. Define the η -value of this path to be the minimum of $\eta_1, \eta_2, \dots, \eta_k$. Next consider any atom other than a source atom. Define the γ -value of this atom to be the supremum of all path η -values, taken over all paths which lead from some source atom to the given atom. Finally define c_n to be the largest number such that there are more than n atoms, other than source atoms, whose γ -values are equal to or greater than c_n .

Now c_n for any fixed n is clearly a random variable, depending upon the several η -values assigned to bonds of the lattice. $P_n(p)$, regarded as a function of p , is the cumulative distribution function of the random variable $1 - c_n$.³

The machine is programmed to calculate c_n for each run. Thus the set of all runs provides a sample of values of $1 - c_n$, and the empiric

distribution of this sample can be taken as an estimate of the required function $P_n(p)$.

So far we have regarded n as a fixed integer. Actually, the output of the machine on any one run is a table of c_n as a function of n for all $n \leq N$, where N is the number set on the parameter data card. Thus the complete calculation provides estimates for $P_n(p)$ for all $n = 1, 2, \dots, N$ and all $0 \leq p \leq 1$.

Successive runs are numbered serially $R_1, R_1 + 1, \dots, R_2$, where R_1 is the first run number specified on the parameter data card, and R_2 is the final run number specified on the parameter data card. Hence $R_2 - R_1 + 1$ is the sample size for each empiric distribution. The run number $R_1 + i$ of the $(i + 1)$ th run is used to trigger off the generation of random numbers η assigned to bonds in this run. Hence a run may be repeated for checking purposes by repeating the run number; but, if a fresh and independent sample of γ_n is desired, the parameter data card must specify a set of run numbers which does not overlap the set previously used. Since three decimal digits are available for run numbers, the maximum sample size is 999. A sample of size about 100 is usually adequate.

V. OUTPUT FROM THE MACHINE

The printout from the machine, also punched onto cards if ordered on the identification card during input, is as follows.

The printout begins with a copy of all input data (excluding the program deck). Thereafter follows a table, whose columns are headed:

<i>RUN</i>	<i>NØ.</i>	<i>STATE</i>	<i>c VALUE</i>
------------	------------	--------------	----------------

Each row of the table has an entry in each of these four columns. An entry under *RUN* is the current run number R being computed ($R_1 \leq R \leq R_2$). The entries under *NØ.* and *c VALUE* are respectively n and c_n , tabulated for $0 \leq n \leq N$ omitting any values of n such that $c_n = c_{n-1}$ ($0 \leq n \leq N$). Thus the only values of c_n printed are new values less than all preceding values in the run. (Clearly c_n is a non-increasing function n by virtue of its definition.) Such new values are indicated by the prefix *NEW*. The value of c_N is, on the other hand, prefixed by *FINAL* and also suffixed by an asterisk to help in reading the output quickly. The entry under *STATE* is either *FINISHED* or *TØ BE CØNT.*, according to whether the run is complete or not. The only other possible entry under *STATE* is *INHIBITED*: this is a safety device, to be explained later. It means that certain technical circumstances

TABLE I—EXTRACT OF PRINTOUT

<i>RUN</i>	<i>NØ.</i>	<i>STATE</i>	<i>c VALUE</i>
23	518	<i>TØ BE CØNT.</i>	<i>NEW c = .748</i>
23	1000	<i>FINISHED</i>	<i>FINAL c = .748*</i>
24	0	<i>TØ BE CØNT.</i>	<i>NEW c = .906</i>
24	1	<i>TØ BE CØNT.</i>	<i>NEW c = .803</i>
24	5	<i>TØ BE CØNT.</i>	<i>NEW c = .791</i>
24	11	<i>TØ BE CØNT.</i>	<i>NEW c = .783</i>
24	17	<i>TØ BE CØNT.</i>	<i>NEW c = .770</i>
24	19	<i>TØ BE CØNT.</i>	<i>NEW c = .762</i>
24	20	<i>TØ BE CØNT.</i>	<i>NEW c = .715</i>
24	21	<i>TØ BE CØNT.</i>	<i>NEW c = .711</i>
24	1000	<i>FINISHED</i>	<i>FINAL c = .711*</i>

(whose occurrence is extremely unlikely) have arisen to prevent completion of the run. In an inhibited run all values of c_n in the printout are valid: all that has happened is that n has been prevented from rising beyond a certain value, at which instant the run is automatically discontinued.

The extract of the printout shown as Table I will help clarify matters. It gives the end of the twenty-third and the whole of the twenty-fourth run for a computation on the simple cubic crystal with $N = 1000$. To find the value of c_n for a value of n not printed in the Table, take the value of c for the largest n less than the required n . In the above example, $c_8 = 0.791$ in run 24.

The machine stores values of c as 9-bit binary decimals. Hence the rounding error in c is about 0.001 and not 0.0005.

VI. OUTLINE OF THE PROGRAM

What has been said so far contains everything that a user of this program needs to know. What follows in the remainder of this paper is an explanation of how the machine carries out the program, and is intended for those who are interested in programming techniques.

The two main entities in any given run are (i) a number denoted by c , and (ii) a so-called "wet list" of atoms. An atom is qualified for membership on the wet list if it is a source atom or if its γ -value is not less than c . Normally the value of c is held constant and the machine recruits new members of the wet list. However, if a stage is ever reached during a run where no further recruits can be found with the existing value of c , then the machine reduces c by an amount just sufficient to ensure the existence of at least one fresh recruit. The run begins with $c = 1$ and a wet list containing just the source atoms. Since a given

atom in a given run has a given γ -value (depending upon the η -values assigned to bonds in that run), reduction of c can never disqualify existing membership of the wet list. Successively recruited members of the wet list (other than source atoms) are numbered $0, 1, \dots$. A little reflection will show that at the moment, when the member numbered n is added to the wet list, the current value of c must be c_n . The run is terminated as soon as the wet list contains $N + 1$ members. Thus the machine output is simply the result of printing any freshly reduced value of c against the number of the next member to be added to the wet list.

This procedure would be unworkable if the machine had to examine all atoms of the lattice for this qualification to belong to the wet list. What makes the procedure workable is the observation that there can exist no qualified fresh recruits if the existing value of c exceeds the η -values of all bonds, which lead from some atom of the wet list to some atom not in the wet list. Let us call such bonds the *outgoing* bonds of the wet list. If the wet list has an outgoing bond whose η -value is at least c , then this bond leads to an atom which is qualified for membership of the wet list. If the value of c has to be reduced, the new value of c is the highest η -value of all existing outgoing bonds. Hence, at any stage of the calculation the machine need only examine the η -values of the outgoing bonds of the currently existing wet list.

Information about the current status of the wet list resides in a block of registers in the core storage of the machine, with three registers (denoted by $E, E + 1, E + 2$) allocated to each atom of the list. We write

$$\begin{aligned} E &= (x, y, z, T) = [11, 11, 11, 3]; \\ E + 1 &= (X) = [36]; \\ E + 2 &= (\psi, \phi, \pi) = [9, 12, 15]; \end{aligned} \tag{1}$$

to indicate that E contains four different quantities (denoted by x, y, z , and T respectively) occupying 11, 11, 11, and 3 bits of the 36 bits available in a single register. The contents of $E + 1$ and $E + 2$ are exhibited in a similar manner. If we wish to emphasize that we are talking about the n th atom A_n ($n = 0, 1, \dots, N$) in the wet list (excluding the source atoms of the wet list), we place n as a suffix to any of the above quantities: thus

$$(E + 2)_n = (\psi_n, \phi_n, \pi_n). \tag{2}$$

The symbols have the following meanings: (x_n, y_n, z_n) are the three integer coordinates of the cell containing the atom A_n , and T_n is the type of the atom A_n ($T_n = A, B, \dots, H$ in the notation for crystal structure). X_n is a 36-bit pseudo-random number, generated according to the multiplicative congruential recursive relation (low multiplication)

$$X_n \equiv \beta^4 X_{n-1} \pmod{2^{36}}, \quad (3)$$

where β is a permanently fixed odd integer, and the recursion is triggered from the initial state

$$X_{-1} = f(R) \quad (4)$$

where $f(R)$ is a fixed function of R , the run number of the run under consideration. ($R = R_1 + j$ in the $(j + 1)$ th run). The meanings of ψ , ϕ , and π will be stated in a moment.

The lattice structure allows us to have up to 12 bonds leading from each atom; and for the sake of exposition it is convenient to suppose in the first place that each atom has a full complement of 12 bonds from it. The atom A_n thus requires 12 η -values for its 12 bonds: we denote these by $(\eta_n^{(1)}, \eta_n^{(2)}, \dots, \eta_n^{(12)})$. The quantity ϕ consists of 12 independent bits $\phi_n = (\phi_n^{(1)}, \phi_n^{(2)}, \dots, \phi_n^{(12)})$ also corresponding to the 12 bonds. If an atom has fewer than 12 bonds from it, we put the corresponding $\phi^{(i)} = 0$ to signalize the absence of a bond. For a bond actually present, we also put $\phi^{(i)} = 0$ if it is *known* that this bond leads to some atom already in the wet list. In all other cases, $\phi^{(i)} = 1$. Thus all outgoing bonds from A_n have $\phi_n^{(i)} = 1$: the converse is not necessarily true, since we may have a bond, which exists and which leads to an atom already in the wet list (and is therefore not an outgoing bond), although at the current stage of the calculation we have not yet discovered that this bond leads to an atom in the wet list. Thus the quantity ϕ represents a state of current knowledge. We define

$$\psi_n = \max (\eta_n^{(1)} \phi_n^{(1)}, \dots, \eta_n^{(12)} \phi_n^{(12)}). \quad (5)$$

To validate this definition we require that each $\eta_n^{(i)}$ be a 9-bit number. We achieve this by means of

$$\left. \begin{aligned} (X_n) &\equiv (\eta_n^{(1)}, \eta_n^{(2)}, \eta_n^{(3)}, ?) = [9, 9, 9, 9,] \\ (\beta X_n) &\equiv (\eta_n^{(4)}, \eta_n^{(5)}, \eta_n^{(6)}, ?) = [9, 9, 9, 9,] \\ (\beta^2 X_n) &\equiv (\eta_n^{(7)}, \eta_n^{(8)}, \eta_n^{(9)}, ?) = [9, 9, 9, 9,] \\ (\beta^3 X_n) &\equiv (\eta_n^{(10)}, \eta_n^{(11)}, \eta_n^{(12)}, ?) = [9, 9, 9, 9,] \end{aligned} \right\} \pmod{2^{36}}, \quad (6)$$

where ? denotes a 9-bit number which is not used (because the terminal

digits of pseudo-random numbers are too regularly distributed), and where the congruential notation in (6) indicates low multiplication as in (3).

VII. PROGRESSIVE CONSTRUCTION OF THE WET LIST

We are now able to describe recursively how the wet list is compiled. In what follows, it is important to remember that the wet list consists of both source atoms as well as recruited atoms in this list.

Suppose that the wet list is already partly compiled and that we have reached a stage at which the current c value has just been reduced to a new value. Starting at the beginning of the wet list, we successively scan each atom in the order of the list. For each atom scanned we ask first if its ψ -value is less than c . If $\psi < c$, we pass to the next atom on the list. If $\psi \geq c$, we determine all values of i such that $\eta^{(i)}\phi^{(i)} \geq c$: these represent the only bonds which can lead to an atom at present qualified for membership of the wet list. Call such an atom a target atom of the scanned atom. Noting the cell coordinates and the type number of the scanned atom, we compute the cell coordinates and the type number of each target atom of the scanned atom. We then look through the wet list to see which of the target atoms do not belong to the wet list, and we add to the end of the wet list all target atoms not already on the wet list. With these new additions to the wet list, every bond from the scanned atom to a target atom leads to an atom of the wet list, and therefore we now set $\phi^{(i)} = 0$ for *all* values of i for which we had $\eta^{(i)}\phi^{(i)} \geq c$. Next we recompute ψ from (5). Of course the new value of ψ is less than c . Therefore in the scanning procedure we go to the next atom in the wet list. Ultimately the scanning procedure will reach the end of the wet list. At this stage, all ψ in the wet list are less than c . We therefore reduce c to the largest ψ in the wet list, and restart the scanning from the beginning of the wet list. We continue this procedure until we have recruited $N + 1$ atoms to the wet list.

To assist in computing the new value of c required at any reduction of c , we make a small modification of the foregoing procedure. We define a number c^* , called the c -candidate. At the beginning of any scan, we set $c^* = 0$. Before leaving any scanned atom and proceeding to the next one, we take $\max(\psi, c^*)$ to be the new value of c^* . Thus when we reach the end of the scan, c^* equals the required new value of c .

VIII. WET LIST SEARCH

As described above, we have to search through the wet list to decide if a target atom is already in the list. To expedite this search we define

the modular type of an atom A_n in the wet list to be the least non-negative residue

$$t_n = x_n + y_n + z_n \pmod{256}, \quad (7)$$

and we dissect the wet list into 256 equivalence classes according to their modular type. A target atom can only be in the wet list if it is its own equivalence class, and hence it is sufficient to search just one equivalence class in the wet list.

Suppose that $A_{n(1)}, A_{n(2)}, \dots, A_{n(k)}$ are the atoms in the wet list currently belonging to the equivalence class to be searched. We define π_n , appearing in (1), by

$$\pi_{n(j)} = n(j - 1). \quad (8)$$

Thus we can search the equivalence class backwards; for at any stage of the search, the π -value for the atom currently examined gives us the address of the next atom to be examined. To start off this iteration, the core storage holds a table, called the H -table, with 256 entries providing the values of $n(k)$ for each equivalence class. When a new atom has to be added to the wet list, we must write up its π -value: its π -value is simply the entry (for the appropriate equivalence class) that is in the H -table immediately before adding the new atom to the wet list; and, immediately after adding this new atom to the wet list into address $n(k + 1)$, say, we enter $n(k + 1)$ into the appropriate position of the H -table in place of $n(k)$.

IX. INHIBITION OF RUNS

Each of the coordinates x, y, z of a lattice cell is represented by an 11-bit integer, treated by the machine modulo 2048. Thus, effectively, the lattice lies on a four-dimensional torus instead of the required three-dimensional flat of four-dimensional Euclidean space. To remedy this defect, we cut the torus on each of the three two-dimensional flats defined by $x = 0, y = 0$, and $z = 0$ respectively. We place the source atoms in, or in the immediate neighborhood of, the cell (1024, 1024, 1024); and we set an inhibition flag if any cell coordinate becomes zero modulo 2048.

The run is allowed to proceed as before, after the inhibitor flag has been set, up to such time as the machine calls for a new value of c . At this instant, however, the inhibitor flag prevents the new value of c being set, and instead terminates the run with the printout comment *c INHIBITED* together with a point of the *old* value of c .

The net effect of this procedure is to allow the fluid to pass across the

two-dimensional cuts in the torus, and even to complete circuits which are not homotopic to zero. However, if such a circuit occurs, it may imply an unnecessary reduction in c , or an unnecessarily large reduction; and in that event the run must be terminated as a safety measure. Nevertheless it is very unlikely, with the values of N used, that inhibition will be invoked; and in fact it has not been invoked on any run calculated to date.

X. GENERATION OF PSEUDO-RANDOM NUMBERS

There is no point in calculating pseudo-random numbers which are not going to be used. At the start of a run, all entries X , defined in (1), are set with a negative sign. When the machine comes to scan any atom in the wet list to look for possible outgoing bonds, it first asks if X is negative. If X is negative, it replaces X by a positive pseudo-random number X_n , generated there and then by means of (3). If X is positive, the machine knows that a pseudo-random number has already been calculated for this atom, and it does not change X_n . Thus, as the problem requires, each individual pseudo-random number remains fixed throughout a run.

XI. DELTA AND PHI TABLES

Consider the stage of the calculation when the machine is scanning the wet list and looking for the possible outgoing bonds from a particular atom A to the corresponding target atoms. Let (x, y, z, T) denote the cell coordinates and type number of A ; and suppose that we are considering the i th bond from A ($i = 1, 2, \dots, 12$) as a possible outgoing bond to a target atom, whose cell coordinates and type number we denote by (x', y', z', T') . Then

$$E' - E = (x', y', z', T') - (x, y, z, T) = \Delta(T, i) \quad (9)$$

is a function of T and i only. We store $\Delta(T, i)$ in relative location $16T + i$ of a block of 128 locations called the delta table. We can thus calculate the coordinates of the target atom by entering the delta table and using the addition

$$E' = E + \Delta(T, i). \quad (10)$$

Similarly, the value of ϕ a target atom is a function $\phi(T, i)$, stored in relative location $16T + i$ of a block of 128 locations called the phi table; and this enables us to write ϕ' into $E' + 2$ by a straight look-up procedure.

The contents of the delta and phi tables are permanent settings computed from the lattice structure data cards before starting the first run.

REFERENCES

1. Broadbent, S. R., and Hammersley, J. M., Percolation Processes, Crystals and Mazes. Proc. Camb. Phil. Soc., **53**, 1957, p. 629.
2. Vyssotsky, V. A., Gordon, S. B., Frisch, H. L., and Hammersley, J. M., Phys. Rev. **123**, 1961, p. 1566; Frisch, H. L., Sonnenblick, E., Vyssotsky, V. A., and Hammersley, J. M., Phys. Rev. **124**, 1961, p. 1021; Frisch, H. L., Hammersley, J. M., and Welsh, D. J. A., Phys. Rev., to be published.
3. Hammersley, J. M., Monte Carlo Solution of Percolation in the Cubic Crystal, *Computational Methods in the Physical Sciences* (book), to be published.

Interference between Satellite Communication Systems and Common Carrier Surface Systems

By HAROLD E. CURTIS

(Manuscript received December 6, 1961)

Various published papers have discussed in quite general terms the problem of interference between satellite systems and ground systems. These studies have been largely qualitative, rather than quantitative, in nature. The magnitude of the interference between a satellite system and ground system, however, depends greatly on the frequency plans involved, the character and degree of modulation used, and the parameters of the equipment. Bell Telephone Laboratories has under construction experimental satellite equipment designed to operate in the heavily used 4- and 6-kmc common carrier bands, and the present paper is directed to the potential interference between this satellite equipment and ground point-to-point systems.

Interference involving a satellite station and the TD-2 and TH systems is analyzed specifically, and it is shown that the separation between systems must be of the order of 100 to 120 miles or more when the antenna of the common carrier transmitter or receiver is pointed directly at the satellite ground station. If the antenna is beamed 90 degrees or more from the satellite site, the minimum distance may be of the order of 10 miles even when line-of-sight propagation exists between locations. This assumes the use of the Bell System's horn-reflector antenna on the terrestrial system. With a parabolic dish antenna the latter distance must be increased to about 40 miles and adequate blocking must exist in the interference path. These distances provide adequate freedom from mutual interference for both telephone and television modulating signals.

I. INTRODUCTION

Satellite systems will of necessity use ground transmitter powers of several thousand watts. Present microwave systems in the United States operating in the common carrier band utilize transmitters of the general

order of one watt output power and in this sense are consequently only about one-thousandth as interfering as a satellite ground transmitter. Furthermore, the inherent noise per cycle of bandwidth at the input of the satellite ground receiver will be about 20 db less than that of the present-day commercial common carrier receiver, thus making it correspondingly more sensitive to interference from other systems.

Several general studies of interference^{1,2} have used criteria of interference intended to encompass all varieties of ground radio relay systems, but inevitably the decision as to whether interference between particular sites is tolerable or intolerable must be made on the basis of the specific radio systems involved and the frequency bands in which they operate.

The F.C.C., as a result of its studies, has recommended for consideration a number of bands between 3700 and 8400 mc, including the two common carrier bands 3700 to 4200 mc for the spacecrafts and 5925-6425 mc for the earth stations.³

The experimental satellite equipment presently under construction by Bell Telephone Laboratories will operate in the top 100 megacycles of the 4- and 6-kmc common carrier frequency bands mentioned above and thus, potentially, interference may occur between the satellite system and the many ground commercial systems operating in these bands.

The Western Union Company has a radio relay system operating in the 4-kmc band, and a transcontinental system in the 6-kmc band is under construction. There are also a relatively large number of short haul common carrier systems in the 6-kmc band. However, the most extensive user of each of these two bands in the United States is undoubtedly the Bell System which had in operation at the beginning of 1961 approximately 300,000 one-way broadband channel miles of microwave systems. A large fraction of this service utilizes the TD-2 system⁴ operating in the 3700- to 4200-mc band, and, at present, a small fraction utilizes the recently developed TH system⁵ operating in the 6-kmc band. For this reason the interference study described herein is directed quite specifically to the TD-2 and TH systems, but the basic philosophy is readily applicable to other microwave systems.

Such a complicated network of microwave routes as presently exists in the United States and Canada with its numerous sources of inter- and intra-system radio interferences has necessitated most careful attention to this problem in order that the interference at baseband at the end of a long circuit would be in reasonable balance with other sources of system impairment, such as intermodulation between elements of the system load and the noise arising in the converters. This problem has been discussed at some length in an earlier paper,⁶ and the philosophy

developed therein will be applied in the present paper to interference involving satellite systems.

The experimental equipment under construction by Bell Laboratories will transmit from ground to spacecraft in the 6-kmc frequency band, and in the reverse direction in the 4-kmc band, so that the only appreciable interferences between ground stations are those from satellite ground transmitters into TH receivers and from TD-2 transmitters into satellite ground receivers. For completeness, consideration is also given to the two complementary interferences that would exist if the frequencies were interchanged. Possible interference from the spacecraft into ground receivers is discussed briefly.

Contours of permissible minimum separation between satellite ground station and TH and TD-2 microwave stations are presented in this paper. It should be emphasized at this point that the results are based on values of parameters for the three systems that are pertinent at the present time. Changes that may be made in the future such as increases in transmitter power, improvements in receiver noise figure or change in frequency plan would, of course, alter the conclusions reached herein.

While the contours are based on propagation under "average" terrain conditions, it is believed that they should be of considerable value in the early phases of site selection. However, in any particular case, if the profile of the path so indicates, the power of the interference should be calculated and compared with the objectives given later in Table III.

II. OBJECTIVES

Microwave systems with which we are concerned may handle television or multichannel data and telephone signals, and in the latter case the signal load may range from busy-hour full load to light early-morning load. Interference objectives must be sufficiently stringent to protect the systems under all normal conditions; moreover, interference powers should be sufficiently less than the total receiver input noise so as not to impair significantly the fading margin of the interfered-with system.

Basically the amount of RF interference that can be tolerated depends on the interference that it produces at baseband frequencies. The spectrum of the interference at baseband frequencies resulting from RF interference between two FM or PM waves is made up of beats between each frequency component of the spectrum of one RF wave and each frequency component of the other. The frequency of any baseband component is that of the frequency difference between the two RF com-

ponents that produced it; and finally the power of the baseband component depends on the powers of the RF components.

Therefore the baseband interference spectrum can, for convenience, be thought of as the result of (a) a tone resulting from the beat between the two carrier frequencies, (b) the sidebands of one wave beating against the carrier of the second and vice-versa, and (c) the sidebands of one beating against the sidebands of the other. The beat tone ideally may appear as sinusoidal interference in a video signal, or as a tone in some particular telephone channel. Actually, because of the very low frequency noise normally present on FM transmitters using klystron deviators, the tone is more like a "burble" spread over a number of channels, the particular channels affected at any time depending on the difference frequency between the carriers at that instant. The second and third classes of interference appear normally as unintelligible cross-talk.

The relative magnitudes of carrier and sidebands at any time depend on the degree and type of modulation applied to the radio transmitter. Consequently the RF interference objectives must be sufficiently restrictive that the baseband interference is adequately low for all conditions of modulation. The procedure here will be to develop objectives on the basis of full load telephone considerations and then to make certain that they are adequate for all other loads and signals, whether telephone or television. In general, the interference objective set by telephone considerations is sufficiently controlling so that it is satisfactory for other types of signals.

All long-haul microwave systems are subject to a number of sources of transmission impairments. For example, a 4000-mile TD-2 system may have approximately 140 sources of thermal-type noise due to the converters, an equal number of sources of cross modulation due to repeater phase and amplitude nonlinearity, 280 sources of waveguide echoes, 280 sources of intersystem co-channel interference, together with a number of somewhat less important contributions.

Good engineering practice indicates that for telephone service, the rms sum of these impairments should, during busy hours, be not over 38 dba0, i.e., 38 dba* at a point of zero-db transmission level. This is equivalent to -43 dbm in a 4-ke band, and the signal-to-noise ratio is 27 db where the signal in each telephone channel is random noise equal in power to an rms talker, one quarter active, or -15.8 dbm in a 4-ke band, using values obtained from the Holbrook and Dixon paper.⁷

* The unit dba identifies a particular weighting characteristic for which 82 dba is equivalent to one milliwatt of thermal noise in a 3-ke band.

The division of this total permissible impairment into all the various individual sources cannot be done by any set rule. The total number of significant sources of noise impairment in a 4000-mile system, as enumerated above, totals very roughly one thousand. Therefore if each were given an equal share of the total, the individual allotment would be 8 dba0. Since interference and cross modulation are generally more annoying than thermal-type noise, it is normally a desirable goal that the baseband noise due to the converters in a repeater should be slightly greater than inter-modulation, which in turn should exceed the baseband interference from RF crosstalk. Then during a fade of the desired carrier, the converter noise and interference in any telephone channel will rise together and the interference will not predominate over the noise.

The distance between any satellite system site and a potentially involved interfering or interfered-with TD-2 and TH station may vary from a few miles to more than 100 miles. Thus, propagation between them may range from line-of-sight, in which free-space propagation normally exists for a very large fraction of the time over a full year, to tropospheric scatter propagation, in which the long-term distribution of path loss is normal in db and in which the chance of the received signal being, let us say, 32 db or more above the median value is about 0.01 per cent.

For this reason two interference objectives are proposed. The first applies to line-of-sight interference paths in which the propagation is very close to that of free-space nearly 100 per cent of the time. Up-fades greater than about 5 db occur less than about 0.3 per cent of a year's time, and down-fades of the interfering carrier simply decrease the baseband interference. This is referred to herein as the "100 per cent" objective, and it is intended that it be met for line-of-sight paths during free-space transmission.

For interference signals which are constantly fading both up and down, such as on tropospheric scatter paths, a second objective is proposed, which should be exceeded only 0.01 per cent of the time. This objective in terms of baseband noise may obviously be higher than the "100 per cent" objective, and it appears reasonable to let the 0.01 per cent objective be 15 db more lenient than the "100 per cent" objective. The "100 per cent" objective is so chosen that the unintelligible crosstalk type of interference in the worst telephone channel in the full-load telephone case may be expected to be 9 dba0 during nonfading periods. It may be 24 dba0 or greater 0.01 per cent of the time when the interfering signal path is well beyond line-of-sight. This may be compared with the contribution of about 10 dba0 per repeater due to noise arising in

the converters of a 4000-mile system during nonfading conditions. Interference may also manifest itself as tones in certain telephone channels or in television transmission. The magnitudes of these effects will be discussed in a subsequent section.

III. FREQUENCY PLAN

Interference between two FM or PM systems depends upon such parameters as frequency deviation, top baseband frequency, and upon the frequency separation between the carrier frequencies of the systems involved.

The CCIR recommends for the 6-kmc common carrier band a plan based on a spacing of 29.65 mc starting at 5945.20 mc, and this is identical with the plan used in the United States by the Bell System for the TH system and also by Western Union for its 6-kmc system. Thus eight satellite assignments, each about 50-mc wide, can be obtained in the same band with a minimum of mutual interference by placing the satellite carriers midway between the common carrier assignments as shown on Fig. 1.

Coordination in the 4-kmc common carrier range is less satisfactory. The TD-2 system, when a route is fully developed, will have a channel

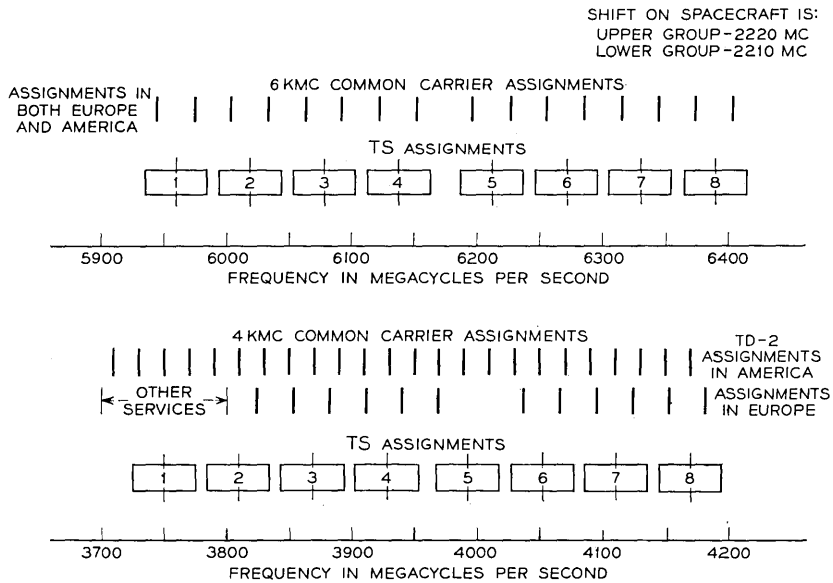


Fig. 1 — Frequency plan.

every 20 mc from 3710 to 4170 mc, while in Europe a spacing of exactly 29 mc is used.

A frequency shift from the 6-kmc to the 4-kmc band must obviously be made on the spacecraft, the optimum value of the shift for each of the channels depending to a considerable extent on problems outside the scope of this paper.

However, one possibility might be to use a shift of 2220 mc for the upper four channels, and 2210 mc for the lower four channels. The satellite assignments in the 4-kmc band would then be as shown on Fig. 1, and this plan is assumed in the present study. It will be noted that the satellite carrier frequencies would, in all cases, be very close to certain of those used in the TD-2 system. The effect on interference of moderate departures in frequency from the plan shown on Fig. 1 is discussed below.

IV. INTERFERENCE BETWEEN TWO FM OR PM WAVES

Signal and interference are customarily specified in units of watts per cycle of bandwidth. Since there is a linear relationship between signal power per cycle and carrier deviation in mean square radians per cycle of bandwidth, it is convenient in this paper to express both signal and interference in the latter units.

The method used herein to compute quantitatively the baseband interference due to the presence of a weak interfering FM or PM wave can be demonstrated by the limiting case when neither interfered-with nor interfering carrier is modulated. Let the peak amplitudes of the two carriers be E and kE , respectively, where k is small relative to unity, and let the frequency difference between them be f . The interfering carrier phase modulates the stronger carrier by $k \cos 2\pi ft$ radians, and the baseband interference power is proportional to $k^2/2$ mean square radians.

However, the present problem is most closely approximated by two carriers separated an appropriate amount in frequency and modulated with random noise so as to simulate a number of telephone channels arranged in frequency division multiplex. This problem has been treated for the case of pure frequency modulation by several writers.^{8,9}

The Appendix gives an expression for the interference in integral form that is valid when the interfering carrier is substantially weaker than the interfered-with carrier. This integral is difficult to deal with numerically. However, following the argument used in the Appendix a practical, but basically exact, method of evaluating the interference has been developed, subject to the same premise of weak interference. This method

consists simply of normalizing the level of the unmodulated power or voltage of the stronger carrier to a reference value of 0 db, and expressing the spectrum of each of the two waves in db below this reference value. The two spectra thus described are convoluted (adding the db values). The interference thus computed in a very narrow band, such as one cycle, is varying with time, and in this case to obtain the mean square value, or power, the result must be decreased by 3 db.

This procedure, then, gives the distributed interference spectrum in mean square radians per cycle of bandwidth, together with a sine wave component at the difference frequency, whose power is expressed in mean square radians. The noise signal simulating a typical talker can be expressed in the same units as the distributed baseband interference, and thus the baseband signal-to-interference ratio can be obtained. Since the noise signal, for the loading constants assumed herein, is 65 dba0, the interference can be expressed in dba0 by subtracting the signal-to-interference ratio in db from 65. Baseband signals are usually pre-emphasized so that the higher frequencies phase modulate the carrier and the lower ones frequency modulate it. In this study, pure phase modulation is assumed for simplicity of analysis for all baseband frequencies.

Nominal values of the parameters determining the spectra for the three systems considered here are given in Table I.

For both the TD-2 and the TH carriers the frequency deviation is sufficiently low that, for purposes of this study, sidebands of order greater than unity are sufficiently small that they may be neglected.

The spectral power per cycle of bandwidth for first order sidebands

TABLE I

Item	Symbol	System		
		TD-2	TH	TSX-1*
Top baseband frequency.....	f_b	2 mc**	10 mc	2 mc**
RMS frequency deviation due to noise load.....	f_d	0.71 mc	0.71 mc	5 mc

* This is the designation for the Telstar experimental satellite equipment being constructed at Bell Telephone Laboratories. The ground station will be located in Andover, Maine. The spacecraft will have active repeater equipment for one broadband channel. Transmission to the spacecraft will be at a frequency of about 6390 mc. Transmission from spacecraft to ground will be at about 4170 mc. The parameters given in Table I above for the TSX-1 equipment represent values that appear to be reasonable at the time of preparation of the paper.

** The top frequency of 2 mc is appropriate to 480 telephone channels arranged in frequency division multiplex.

only, relative to the unmodulated carrier, for a low index of modulation is¹⁰

$$P_i = \frac{\exp[-3f_a^2/f_b^2]}{2f_b} \frac{3f_a^2}{f_b^2} \quad \text{for} \quad |f| < f_b \quad (1)$$

and zero for $|f| > f_b$

where f is any frequency relative to that of the unmodulated carrier.

In the case of the TSX-1 system, the deviation with noise load is sufficiently great that the wide deviation approximation may be used. Assume the power of the carrier when unmodulated is unity. When the phase deviation, as defined in (4) below, is substantially greater than unity, the power per cycle of the PM wave at a frequency $\pm f$ from that of the carrier frequency is very closely

$$P = \frac{1}{\sqrt{2\pi}f_d} \exp(-f^2/2f_d^2). \quad (2)$$

There is also a carrier spike present whose power relative to that of the unmodulated carrier of unity power is

$$P(\text{spike}) = \exp(-3f_a^2/f_b^2). \quad (3)$$

Fig. 2 shows the spectra for the three systems with noise loading as computed using the parameters and formulae given above.

The mean square phase deviation, D , of the carrier is related to the above defined constants in the following way:

$$D = 3f_a^2/f_b^2 \text{ mean square radians.} \quad (4)$$

The applied signal power, S , per cycle of bandwidth is

$$S = D/f_b = 3f_a^2/f_b^3 \text{ mean square radians per cycle of bandwidth.} \quad (5)$$

The signal power in db is given by $10 \log S$ and is tabulated in Table II for the three systems of interest, using the constants from Table I. The symbol dbR/cbw will be used to denote mean square radians per cycle of bandwidth expressed in decibels.

The application of the method of determining the interference in a specific case will illustrate the procedure. Let us consider, as an example,

TABLE II

Signal Power	TD-2	TH	TSX-1	Units
$10 \log S$	-67.3	-88.2	-50.3	dbR/cbw

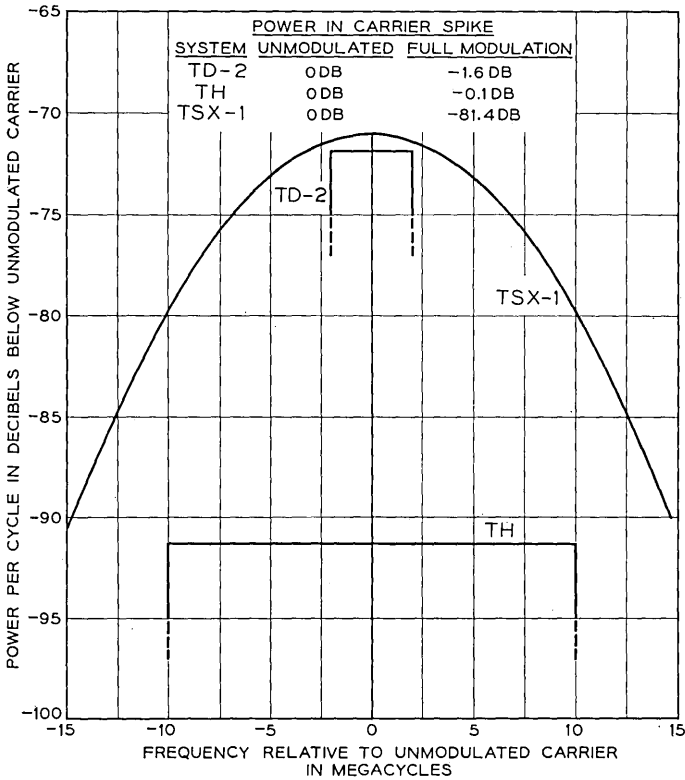


Fig. 2 — Spectra of phase-modulated waves.

interference into the satellite system from the TD-2 system and compute, specifically, the dba value of interference corresponding to a particular ratio of unmodulated TD-2 carrier power to unmodulated satellite carrier power at the input to the TSX-1 receiver, such as -35 db. The unmodulated carrier frequencies in this case will be nearly co-channel.

Fig. 2 shows that in the case of the satellite system modulated with random noise, the sideband power per cycle near the carrier will be 71 db below the unmodulated satellite carrier power. The power of the TD-2 carrier in turn was assumed to be 35 db below the carrier power of the interfered-with TSX-1 carrier.

The TD-2 carrier spike then beats with each component of the TSX-1 spectrum to produce baseband interference as described above. Consider specifically the beat between the unmodulated TD-2 carrier and

the TSX-1 spectral power one kilocycle removed therefrom. The resulting interference that will fall in a bandwidth of one cycle at a baseband frequency of one kilocycle can be obtained, as described above, by adding the db values of the appropriate points on the two spectra and subtracting 3 db from the result. This gives an interference value of $-71 + (-35) - 3$, or -109 dbR/cbw.

From Table II we find that the TSX-1 signal power is -50.3 dbR/cbw. The signal-to-noise ratio in a narrow band then is 55.7 db, which is equivalent to approximately $+9$ dba0. This establishes the position of the linear relationship shown on Fig. 3, for interference from the TD-2 system into the TSX-1 system.

If the speech load on the satellite system is very low, then the continuous portion of baseband interference of the satellite system due to the TD-2 system can be obtained by convoluting the TSX-1 carrier spike and the TD-2 spectrum. By following the above procedure it can be shown that the baseband interference is 0.9 db less than when the TSX-1 system was fully modulated.

Interference from the satellite system into the TH system behaves quite differently. With the frequency plan shown on Fig. 1, the unmodulated carriers are approximately 15 mc apart, and only when the satellite system is substantially fully loaded will there be appreciable interference into the TH system.

Assume a ratio of TH carrier to satellite carrier power of 68 db. The maximum interference into TH falls at 10 mc and is the result of a beat between the TH carrier and satellite sidebands 10 mc removed, i.e., 5

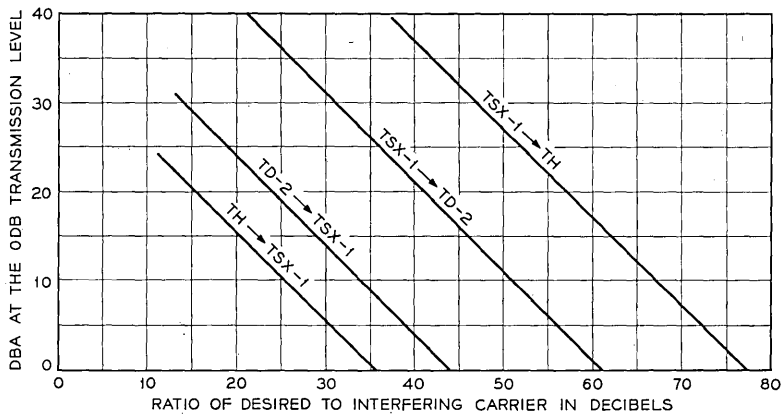


Fig. 3 — Expected message channel interference, all-message case.

mc from the satellite carrier itself. Fig. 2 shows the spectral power is 73 db below the TSX-1 unmodulated carrier or 141 db below the TH carrier. Therefore the interference on the TH carrier corresponds to phase modulation of -144 dbR/cbw. The TH signal power (see Table II) is -88.2 dbR/cbw. The signal-to-interference ratio is 56 db and the interference is $+9$ dba0.

Interference into the TSX-1 system from the TH system increases rapidly with baseband frequency. In order not to jeopardize the potential use of the satellite system above 2 mc, interference is computed at 5 mc but is referred to the signal as specified in Table I, thus giving a conservative value for the minimum allowable separation.

Finally, interference from the TSX-1 system into the TD-2 system is a maximum at the bottom baseband frequencies since the two systems are nearly co-channel. Computations are made in a manner similar to that for interference in the reverse direction. Fig. 3 shows the relationships between baseband interference to the ratio of desired carrier to interfering carrier for these four cases. These relationships are linear, and are valid for carrier ratios greater than about 10 db.

In the frequency plan shown on Fig. 1, the satellite channels are uniformly spaced between TH assignments. If a plan were used with a spacing different from 15 mc, the magnitude of the interference would, of course, be affected. Fig. 4 shows the computed baseband interference spectrum as a function of carrier spacing, and it will be noted that the increase due to reducing the separation to 10 mc is only about 2 db at the top baseband frequency.

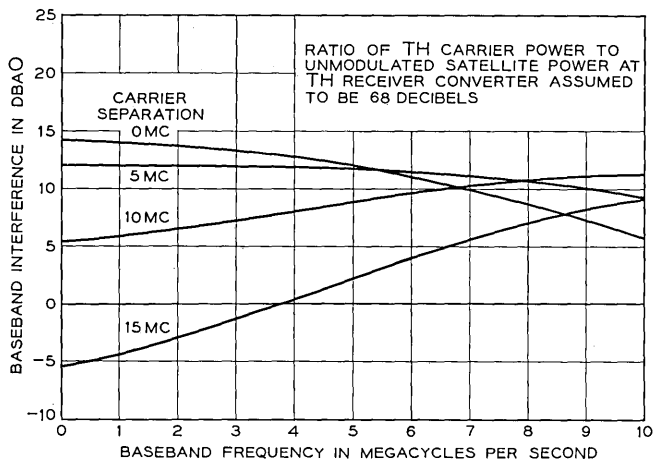


Fig. 4 — Interference — satellite transmitter into TH system.

TABLE III

System		Normal Received Carrier	Required Interference Ratio	Interfering Carrier Objectives at receiver-converter	
Interfering	Interfered- with			100%	0.01%
TH	TSX-1	-98 dbm	27 db	-125 dbm	-110 dbm
TSX-1	TH	-27	68	-95	-80
TD-2	TSX-1	-98	35	-133	-118
TSX-1	TD-2	-38	52	-90	-75

V. PERMISSIBLE INTERFERENCE

On the basis of the normal received carrier power for the TD-2 and TH systems and the minimum received satellite carrier, permissible interference carrier powers at the input to the receiver-converter of the interfered-with system can be obtained (Table III). The values tabulated in the "required interference ratio" column in Table III are from Fig. 3, and correspond to a baseband interference value of 9 dba0.

The above objectives have been based on busy-hour telephone interference. During such periods the power in the carrier of the satellite system is negligible, but under light loads it may be strong enough to make the beat note between it and the carrier of the ground common carrier system a serious source of interference. Ideally this interference would be a pure tone, but actually it will be spread over a group of telephone channels, its location depending on the frequency difference between the two carriers at any moment.

An order of magnitude estimate of this interference can readily be made as follows for the case of interference from the TD-2 system into the satellite system when the latter is very lightly loaded:

- a. Permissible carrier to interference ratio from Table III = +35 db
- b. Resulting tone interference = -38 dbR
- c. Signal (noise load) from Table II = -50.3 dbR/cbw
- d. Signal in a 3-kc band = -15.5 dbR
- e. Signal-to-interference ratio = +22.5 db
- f. If the interference is assumed to have the character of noise and spread over one channel only, it will read = +42.5 dba0.

This, it will be noted, is 4.5 db above the total 4000-mile objective. Actually it will probably be spread over a number of telephone channels,

depending upon the steadiness of the carriers, but still it is a potentially important interference source. The interference from the satellite system into the TD-2 system computed in the same manner will have the same value.

This source of interference can be decreased substantially by applying a very low frequency baseband signal to the transmitter to keep the carrier in motion even when the total speech load is low.¹¹

The above objectives have been based on telephone channel interference considerations. It appears possible to produce interference into the satellite system when the interfering carrier is weakly modulated, thus having a strong carrier spike, and when at the same time the interfered-with carrier is handling a television picture with a large gray area, thus also having a large component of energy in a fairly narrow band.

For example, assume the television signal content in the TSX-1 channel is such that there is a concentration of energy one mc removed from an assumed interfering TD-2 carrier. A sinusoidal baseband tone will result whose peak-to-peak amplitude relative to the peak-to-peak amplitude of the desired baseband signal for full deviation is given by the ratio of the carrier powers involved in db, plus the FM improvement. The assumed carrier ratio from Table III is 35 db; the FM improvement for a TSX-1 deviation of 20 mc is 26 db. Therefore, the signal-to-tone ratio is 61 db as shown later in Table IV. Bearing in mind that line-of-sight signals during fading conditions may "up-fade" as much as 5 db, this TV signal-to-tone ratio would then be reduced to about 56 db. This is of the same order of magnitude as the tolerable tone interference ratio in a television picture.

The various impairments that may be expected using the interference ratios given in Table III are summarized in Table IV. Since the ground satellite station in the TSX-1 experiment transmits in the 6-kmc band

TABLE IV

System		Telephone Interference		TV Sp-p/Tp-p	Loss of Fade Margin
Interfering	Interfered- with	Full Load	No Load		
TH	TSX-1	+9 dba0	nil	53 db	0.3 db
TSX-1	TH	9 dba0	nil	80	1.6
TD-2	TSX-1	9 dba0	42.5 dba0*	61	0.1
TSX-1	TD-2	9 dba0	42.5 dba0*	64	2.2

* This interference will appear in a few telephone channels only, and there is also considerable uncertainty in the values given.

and receives in the 4-kmc band, it is evident that TD-2 interference into TSX-1 is controlling from the standpoint of interference to television.

The loss of fading margin is computed on the basis of the increase in peak noise voltage at the converter when the interfering carrier is present. A noise peak factor of 12 db is assumed for the noise at the converter. The interference values used in computing loss of fading margin are the "100 per cent" values, so-called, on Table III since the use of the "0.01 per cent" values seemed unduly conservative.

An examination of Table IV appears to indicate that relaxing any of the RF interference objectives will result in undesirable impairment increases in one or more of the categories listed.

VI. REQUIRED PHYSICAL SEPARATION

The separation between satellite station and common carrier station should be such that the received signal does not exceed the values stated on Table III. This separation depends, of course, upon the transmitter power, path loss and antenna discrimination involved.

In this study the satellite transmitter power is assumed to be 2 kw. The transmitter power in the TH system is 5 watts. Because of atmospheric effects it is expected that the minimum useful elevation of the satellite antenna above the horizontal will be limited to about 7.5°. If so, the effective gain of the satellite antenna in the horizontal direction may be expected to be 0 db above isotropic or less. The gain of the conven-

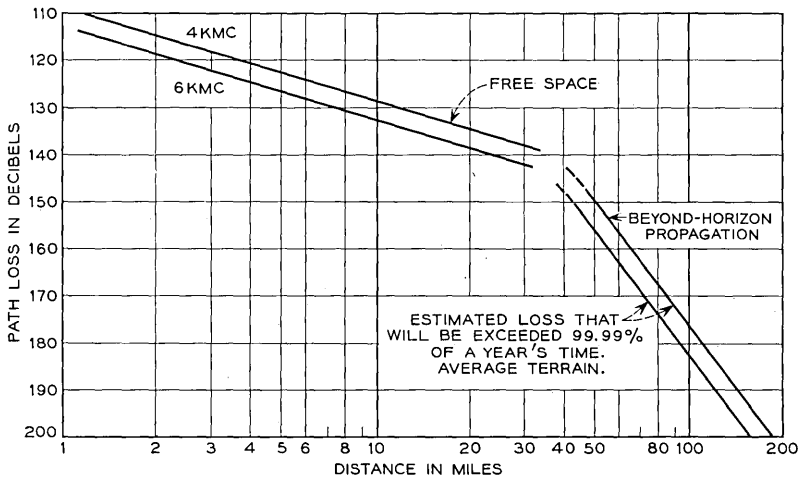


Fig. 5 — Path loss — line-of-sight and beyond-horizon.

tional horn-reflector antenna used with the TD-2 and TH microwave systems at a given angle off-beam is conveniently expressed as the on-beam gain (about 40 db above isotropic at 4 kmc) reduced by the relative directivity pattern of the antenna, a typical example of which is shown later as Fig. 6. Thus, the net response of the horn-reflector in the backward direction is +40 -70, or 30 db below isotropic.

Typical path-loss curves between isotropic antennas as a function of distance are shown on Fig. 5. For distances up to about 30 miles, free-space loss values are plotted; from about 40 miles to 140 miles, scatter propagation over average terrain is assumed. The scatter loss curve is that which will be exceeded 99.99 per cent of a year's time as estimated by K. Bullington of Bell Telephone Laboratories.¹²

The interfering carrier power in dbm is given by

$$P_R = P_T + A_T - L + A_R \quad (6)$$

where P_T = effective transmitter power

A_T = transmitter antenna gain in the direction of the interfered-with receiver

L = path loss (from Fig. 5)

A_R = receiving antenna gain in direction of the interfering transmitter.

As an example, interference from the TD-2 system into the TSX-1 system will be considered. Transmitter power and on-beam antenna gain for the TD-2 system are approximately +27 dbm and 40 db, respectively. The required separation when the antenna of the interfering transmitter is pointed at the satellite site will be such that propagation will presumably be by scatter and hence the 0.01 per cent objective of -118 dbm given in Table III will apply. The total required path loss will be then 185 db, corresponding to 123 miles.

If the satellite receiving station is off the beam of the TD-2 transmitting antenna, the energy radiated thereto is decreased by the antenna discrimination, and the required separation is therefore decreased. Fig. 6 shows the measured directivity pattern of an individual horn-reflector antenna relative to its forward gain, and it will be noted that the backward gain is 70 db below the on-beam gain. Tentatively applying the line-of-sight objective of -133 dbm from Table III, and substituting the appropriate values into (6), we have

$$-133 = +27 + (40-70) - L + 0$$

from which $L = 130$ db. Reference to Fig. 5 shows that this loss corre-

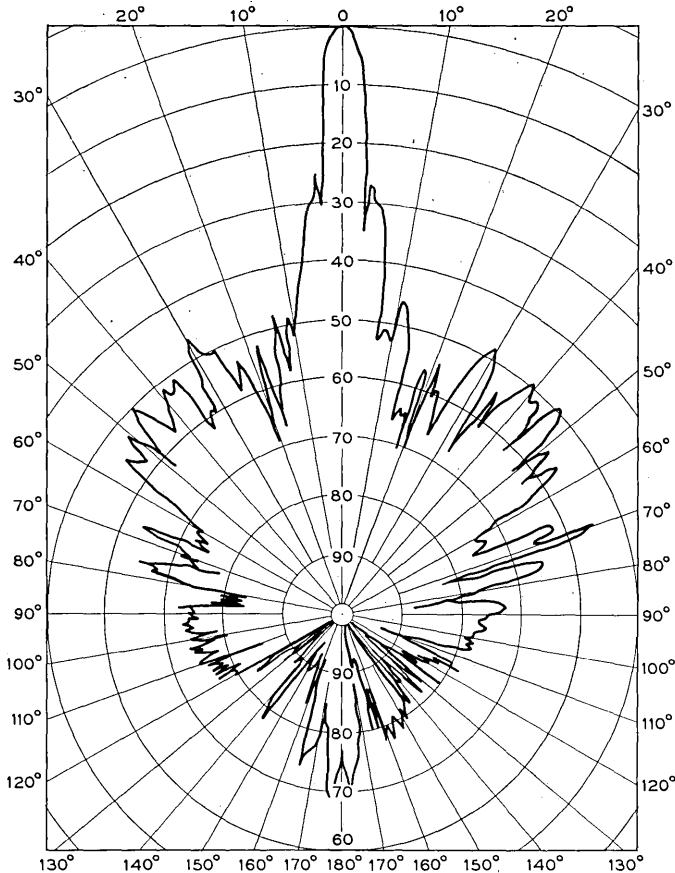


Fig. 6 — Measured pattern in db of horn-reflector antenna at 3740 mc.

sponds to a separation of about 11 miles. Thus, this distance is adequate when the horn-reflector antenna is beamed directly away from the satellite site, even with line-of-sight propagation.

The measured antenna pattern shows a large number of nulls, the position and depth of which will vary from antenna to antenna. Therefore, in developing contours of minimum permissible separation, it has seemed a reasonable and conservative approach to use not the measured patterns but instead an envelope drawn through the maximum values of the pattern as in Fig. 8 of Ref. 6. An additional element of conservatism lies in the fact that the TSX-1 antennas will use circularly polar-

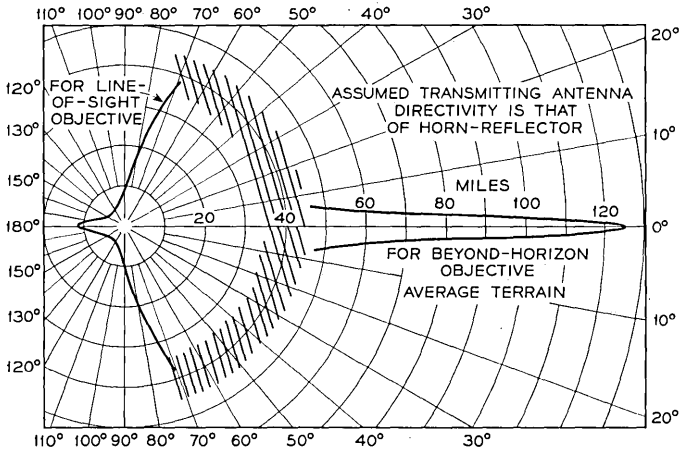


Fig. 7 — Contours of minimum permissible separation between TD-2 transmitter and satellite ground receiver.

ized waves, whereas plane polarized waves are used by the ground systems.

Fig. 7 shows a contour of minimum permissible separation between a TD-2 transmitter and a satellite receiver as a function of the angle between the bearing of the TD-2 antenna and the bearing of the satellite site, based on a smoothed envelope of the antenna discrimination pattern of Fig. 6. For angles greater than 90° , the separation can be as low as about 5 miles. For angles less than about 5° it must be 60 miles or greater. Distances in the order of about 40 miles are cross-hatched because of the uncertainties in propagation. Specific cases that fall in this range should be examined individually to ascertain whether line-of-sight or beyond-horizon objectives should be applied.

Figs. 8, 9 and 10 show similar contours of minimum permissible separation for the three other interference combinations.

It should be emphasized that these values of separation may vary substantially, depending on the local terrain. If the path is obviously line-of-sight, the "100 per cent" objective shown on Table III should be used and met on a free-space propagation basis. In the case of stations only somewhat beyond line-of-sight, it must be kept in mind that ducting may occur occasionally and the signal may become strong enough to give interference much stronger than expected under true scatter conditions. In this situation the chance of exceeding the 0.01 per cent objective given on Table III must be estimated for the specific case in-

volved, due allowance being made, of course, for the directivity of the antennas; if the estimated probability is 0.01 per cent or less, the locations may be considered safe.

The close permissible spacing indicated when the horn-reflector antenna is oriented 90° or more from the direction of the satellite station is, of course, due to its very low backward response. An 8-foot parabolic

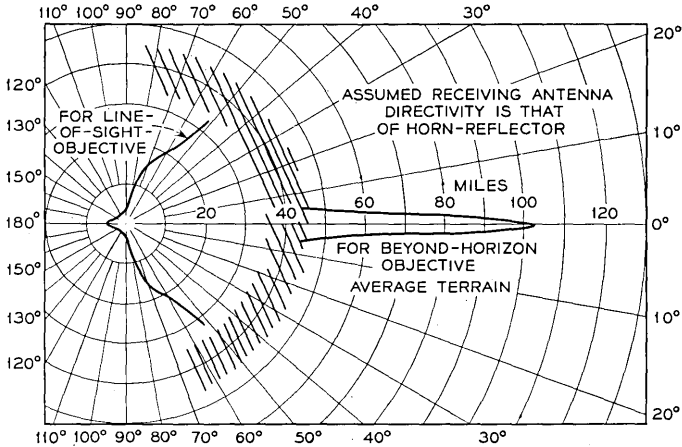


Fig. 8 — Contours of minimum permissible separation between satellite ground transmitter and TD-2 receiver.

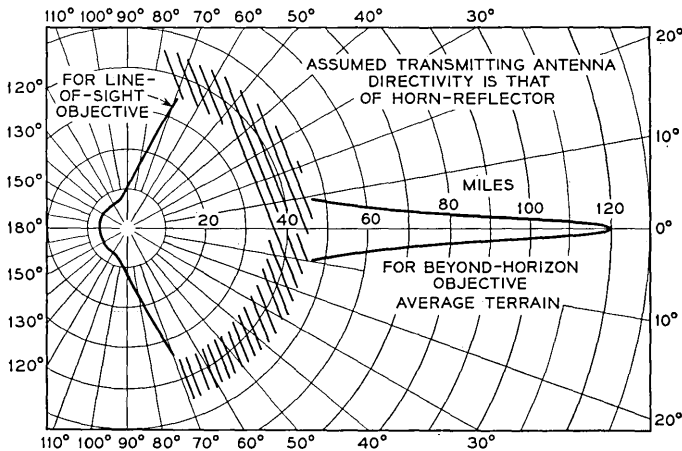


Fig. 9 — Contours of minimum permissible separation between TH transmitter and satellite ground receiver.

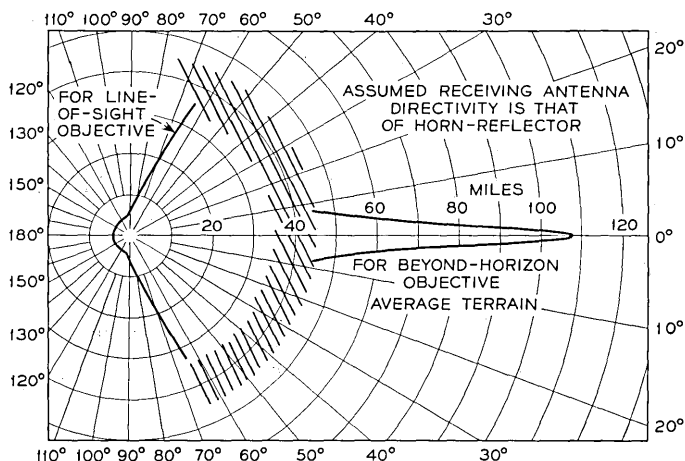


Fig. 10 — Contours of minimum permissible separation between satellite ground transmitter and TH receiver.

antenna may be expected to provide some 20 db to 30 db less discrimination in this general direction. The application of the line-of-sight objective would then result in a minimum distance so great that it would be highly improbable that line-of-sight transmission could take place. The application of the “0.01 per cent” objective indicates, however, a minimum separation of 30 to 40 miles.

In this case, it would be necessary to make certain that propagation is by scatter. The process of estimating propagation over relatively short non-line-of-sight paths involves a detailed study of the exact profile and knowledge of climatic conditions. This subject is outside the scope of the present paper.

While these contours indicate that separations of the order of 5 to 10 miles are possible under the conditions assumed, it should be emphasized that these conclusions are valid only in the absence of significant reflection paths between the interfering and interfered-with sites. Thus, reflecting objects such as houses or trees in the foreground of the horn-reflector antenna may degrade its directivity pattern in the backward direction so that the values shown on Fig. 6 are not attained. Normally it may, however, be expected that such reflections can be made adequately low by careful examination of the terrain illuminated by the ground system antenna.

Transient reflections may be produced by objects in the sky such as airplanes and birds. Also, reflections from rain clouds and precipitation

represent a possible way in which interference might reach the satellite ground receiver. However, calculations of the probable magnitude of the effects of such reflections indicate that with station separations of the order of 10 miles or more, interference of this kind from TD-2 into TSX-1 will not exceed the objective of -118 dbm for any appreciable percentage of time.

VII. INTERFERENCE FROM SPACECRAFT TO GROUND RECEIVERS

Another possibility is, of course, interference from the spacecraft transmitter into ground microwave systems. For example, assume a power of one watt and an isotropic antenna on the spacecraft, and on-beam gain of 40 db for the ground antenna. The interference at the receiver converter would have a value of -90 dbm with the spacecraft at a distance of about 370 statute miles, therefore just meeting the "100 per cent" objective for interference from the satellite system into the TD-2 system under these assumed conditions. On this same basis, the maximum power on the spacecraft could be increased to 100 watts if the distance from the spacecraft to the earth were increased ten-fold.

In the case of the horn-reflector antenna, the beamwidth is sufficiently narrow that any particular spacecraft would be in the beam of a given antenna for only brief and infrequent intervals.

APPENDIX

Using an approach similar to that developed in Ref. 13, Messrs. S. O. Rice and L. H. Enloe of Bell Telephone Laboratories have shown independently that the distributed portion, $W(f)$, of the interference spectrum is given by

$$W(f) = k^2 \int_0^{\infty} \exp[-R_w(0)](\exp[R_w(\tau)] - 1) \cdot [\cos(2\pi f - \omega_0)\tau + \cos(2\pi f + \omega_0)\tau] d\tau \quad (7)$$

where the amplitude of the stronger carrier is unity and

k = ratio of the weaker to the stronger carrier; this must be small compared to unity

f = any baseband frequency

$\omega_0/2\pi$ = frequency difference between the two carriers

$R_w(\tau)$ = sum of the autocorrelation functions of the two applied noise signals.

If the noise signals applied to the carriers have powers w_1 and w_2 mean square radians per cycle of bandwidth for the weaker and stronger

carriers, respectively, and f_1 and f_2 are the top baseband frequencies of the noise signals applied to the same two carriers, then

$$R_w(\tau) = R_{w_1}(\tau) + R_{w_2}(\tau) \quad (8)$$

$$= \frac{w_1 \sin 2\pi f_1 \tau}{2\pi\tau} + \frac{w_2 \sin 2\pi f_2 \tau}{2\pi\tau}. \quad (9)$$

Also, let

$$w_2 f_2 = D_2 = \text{mean square phase deviation of the stronger carrier}$$

and

$$w_1 f_1 = D_1 = \text{mean square phase deviation of the weaker carrier;}$$

then

$$R_w(0) = D_1 + D_2. \quad (10)$$

In addition to the distributed interference, there is a sinusoidal component of magnitude

$$k e^{-\frac{1}{2}R_w(0)} \sin \omega_0 t \text{ radians.} \quad (11)$$

The expression $W(f)$ as given by (7) is not readily evaluated except over a somewhat limited range of parameters. However, the quantity $\exp[R_w(\tau)]$ in (7) is the product of the two autocorrelation functions $\exp[R_{w_1}(\tau)]$ and $\exp[R_{w_2}(\tau)]$. It therefore follows from the convolution theorem for Fourier integrals that the power spectrum of the interference can be obtained by convoluting the power spectra of the two phase-modulated waves. Thus, baseband interference spectrum in mean square radians is given by twice the value obtained by convoluting the spectral power of the two waves, both in a resistance of one ohm. This provides the method* used in this paper for computing interference.

* While the paper was in page proof it was brought to the writer's attention that the convolution method of computing interference between two PM waves had previously been described in a paper (in Japanese) entitled "On the Interference Characteristics of the Phase Modulation Receiver for the Multiplex Transmission," Shinji Hayashi, The Journal of the Institute of Electrical Communication Engineers of Japan, **35**, pp. 522-528, November 1952.

REFERENCES

1. Fine, H., and Dixon, J. T., Suggested Criteria for Limiting Satellite Power in Channels Shared with Ground Systems, Office of the Chief Engineer, Technical Research Division, FCC, TRR Report No. 5.2.1, June 1, 1961.
2. Downing, J. J., Interference Considerations for Communications Satellites, Seventh National Communication Symposium Record, 1961, pp. 302-309.
3. Final Preliminary Views of the United States of America, as Submitted by the Federal Communications Commission, Frequency Allocations for Space

- Radiocommunication, September 7, 1961. Hearings on Space Communications and Allocation of Radio Spectrum before the Communications Subcommittee on the Committee on Commerce, United States Senate, August 1, 23 and 24, 1961 — United States Government Printing Office, Washington, D. C.
4. Roetkin, A. A., Smith, K. D., and Friis, R. W., The TD-2 Microwave Radio Relay System, *B.S.T.J.*, **30**, October 1951, p. 1041.
 5. Mc Davitt, M. B., 6000-Megacycle-per-Second Radio Relay System for Broadband Long-Haul Service in the Bell System, *AIEE Trans.*, **76**, Part 1, January 1958, p. 715.
 6. Curtis, H. E., Radio Frequency Interference Considerations in the TD-2 Radio Relay System, *B.S.T.J.*, **39**, March, 1960, p. 369.
 7. Holbrook, B. D., and Dixon, J. T., Load Rating Theory for Multi-Channel Amplifiers, *B.S.T.J.*, **18**, October 1939, p. 624.
 8. Medhurst, R. G., Hicks, Mrs. E. M., and Grossett, W. G., Distortion in Frequency-Division-Multiplex FM Systems Due to an Interfering Carrier. *Proc. I.E.E.*, **105**, Part B., No. 21, May 1958.
 9. Hamer, R., Radio-Frequency Interference in Multi-Channel Telephony FM Radio Systems, *Proc. I.E.E.*, **108**, Part B, No. 37, January 1961.
 10. Stewart, J. L., The Power Spectrum of a Carrier Frequency Modulated by Gaussian Noise, *Proc. I.R.E.* **42**, October 1954, pp. 1539-1542.
 11. U. S. Patent 2,352,254, Frequency Modulated Wave Transmission, June 27, 1944, H. E. Curtis.
 12. Response of the A. T. & T. Co. before the Federal Communications Commission, Docket No. 13522, February 27, 1961.
 13. Bennett, W. R., Curtis, H. E., and Rice, S. O., Interchannel Interference in FM and PM Systems Under Noise Load Conditions, *B.S.T.J.*, **34**, May 1955, pp. 601-636.

Long-Term Frequency Stability for a Reflex Klystron without the Use of External Cavities

By GEORGE B. GUCKER

(Manuscript received December 20, 1961)

The reflex klystron is widely used as a tunable low-power oscillator in frequency-modulated microwave radio relay systems, both as the transmitter and as the local oscillator in the receiver. Automatic frequency control is generally used to limit frequency error due to drift of the transmitter klystron when the environmental temperature changes. In the TL Radio System developed at Bell Telephone Laboratories, the transmitting klystron frequency is stabilized by minimizing the effects of environmental temperature changes with completely passive methods. The 457A klystron frequency stability is better than 50 parts per million over an ambient temperature range of 100°C when used at mid-band. In the TL Radio System, the design objective for frequency stability of 500 parts per million is attained at extreme conditions of ambient temperature, atmospheric pressure, and voltage regulation; therefore the need for automatic frequency control has been eliminated.

The operating temperature of the 457A klystron in the TL Radio System is controlled by a liquid-vapor heat exchanger, which is described. The design of the klystron mechanical tuner includes compensation to minimize the thermal coefficient of frequency. The compensation feature is described, and typical results of the program to stabilize the klystron frequency are presented.

I. INTRODUCTION

The reflex klystron is widely used as a tunable low-power oscillator in frequency-modulated microwave radio relay systems, both as the transmitter and as the receiver local oscillator. Automatic frequency control is generally applied to the transmitting klystron to maintain the output frequency within acceptable limits. The frequency control is necessary

because the klystron frequency is dependent on the operating voltages and the ambient temperature. The use of a high-Q stabilizing cavity in the output is not generally feasible since modulation is applied to the repeller of the transmitter klystron to obtain frequency modulation of the output. With well regulated power supplies, most of the frequency error is due to changes in the ambient temperature. This means that automatic frequency control mainly corrects for frequency errors due to ambient temperature changes.

Automatic frequency control requires a discriminator, coupled to the output of the system, which will develop an error signal when the frequency drifts away from the desired value. The discriminator usually consists of a high-Q stable reference cavity with associated networks to develop the proper error signal. For electronic automatic frequency control, the error signal is usually used to change the repeller voltage of the transmitting klystron in the direction and magnitude necessary to correct the frequency error. Changing the level of the repeller voltage affects the operating characteristics of the klystron, particularly the modulation linearity. Much less effect on modulation linearity is obtained when mechanical tuning of the klystron is used to correct for frequency errors due to changes in ambient temperature. In an electro-mechanical automatic frequency control, the discriminator error signal is used to actuate the tuning mechanism of the transmitting klystron to correct the frequency error. This method imposes requirements on the klystron tuning mechanism and the equipment needed for transforming the electric error signal to mechanical motion.

The TL Radio System uses the 457A klystron in conjunction with a liquid-vapor heat exchanger to eliminate completely the need for automatic frequency control of the transmitter. Stability of the output frequency over a wide range of ambient temperatures is obtained by completely passive methods. Two 457A klystrons, one as the transmitter and one as the receiver local oscillator, are used in each TL Radio unit. Essentially constant operating temperature for both klystrons is maintained by the constant boiling point of the liquid in the liquid-vapor heat exchanger. The effect of the residual temperature changes has been minimized by designing a temperature-compensating feature into the klystrons. Elimination of automatic frequency control results in lower system costs and maintenance. Further savings are realized since a blower to cool the klystrons is not required, and electronic automatic frequency control in the receiver is simplified because the receiver klystron frequency is also stabilized.

II. OBJECTIVES

The TL Radio System operates in the 10,700 to 11,700 megacycles common carrier frequency band. The design objective for frequency stability is a tolerance of ± 0.05 per cent under all operating conditions for an ambient temperature range from -40° to $+140^{\circ}\text{F}$. This objective was interpreted as an allowable error of ± 5 megacycles. The allowable error was allocated to various parts of the TL Radio System; an error of ± 2 megacycles was allotted for the 457A klystron and ± 3 megacycles for variation in electrode voltages. In the klystron design this was considered to be an objective of less than ± 1 megacycle error for thermal drift and less than ± 1 megacycle error for mechanical variations due to shock or vibration. The objectives for the klystron were met, well within the maximum allowable error, by the following methods:

1. The mechanical tuner of the 457A klystron was designed to withstand the anticipated shock and vibration levels in the TL Radio System with a frequency error of less than 0.2 megacycle.

2. The klystron design was made compatible with the TL Radio liquid-vapor heat exchanger, which is an essentially constant temperature heat sink for the tube. The klystrons and heat exchanger are enclosed in a thermally insulated chamber to minimize the effects of changes in ambient temperature. The insulated chamber and liquid-vapor heat exchanger maintain the operating temperature of the klystron within 3°F for an ambient temperature range of -40° to $+140^{\circ}\text{F}$ at constant atmospheric pressure.

3. The mechanical tuner of the 457A klystron was designed to provide compensation for the thermal effects on frequency. The 457A specifications include a requirement that the thermal coefficient of frequency shall not exceed ± 0.15 megacycle per degree Fahrenheit (klystron body temperature) when measured at the mid-band frequency of 11,200 megacycles. An adjustment of the compensation is provided so that the distribution during production may be maintained within these limits with an average value for the coefficient near zero. The specified thermal coefficient limits are equivalent to a frequency stability better than 50 parts per million over an ambient temperature range of 100°C for the operating conditions of the TL Radio System. The maximum thermal coefficient may be as high as 0.24 megacycle per degree Fahrenheit at an operating frequency of 10,700 megacycles. With extreme temperature conditions and maximum coefficient applied simultaneously, the frequency error due to thermal effects does not exceed 0.72 megacycle.

4. The klystron frequency stability is evaluated for the ambient tem-

perature range of -40° to $+140^{\circ}\text{F}$ at constant atmospheric pressure. Variations in atmospheric pressure will change the liquid boiling point and the klystron operating temperature. The extreme variation in atmospheric pressure for the worst areas of the country is not expected to exceed 0.75 pound per square inch or 40 mm Hg. This range of pressure will change the liquid boiling point slightly less than 3°F .

The effect of atmospheric pressure fluctuations is not included in evaluation of the klystron frequency stability for the following reasons:

1. Short-term variations in atmospheric pressure are generally small and are not correlated with changes in ambient temperature. The pressure effects may add to or subtract from the ambient temperature effects on the klystron. The combined effect will result in small random fluctuations.

2. Both ambient temperature and atmospheric pressure affect the klystron operating temperature in the same direction. That is, a decrease in ambient temperature or atmospheric pressure will lower the klystron operating temperature. Over extended periods of time, very low temperatures are frequently associated with high pressure conditions. The high pressure will oppose the effect of low temperature.

3. The entire ambient temperature range is used to evaluate the klystron frequency stability. This is equivalent to using the frequency at one temperature extreme as the reference. It is unlikely that any particular TL Radio location will be subjected to these extreme conditions. The additional margin obtained by adding the effect of adverse pressure conditions is not considered to be necessary.

III. LIQUID-VAPOR HEAT EXCHANGER

The liquid-vapor heat exchanger was selected as the simplest and most economical method for obtaining thermostatic control of the klystron frequency. The attributes of this method are:

1. No auxiliary power is required.
2. By attaching the liquid chamber directly to the heat source (the klystron body), an essentially constant temperature of the source can be maintained independent of ambient temperature.
3. The equipment is compact and the quantity of liquid required is small.
4. A wide range of klystron power input can be handled with negligible variation in operation temperature.
5. This method is completely passive.

The equipment used for making feasibility studies on the klystron is shown in Fig. 1. The klystron is bolted to the liquid chamber, or evapora-

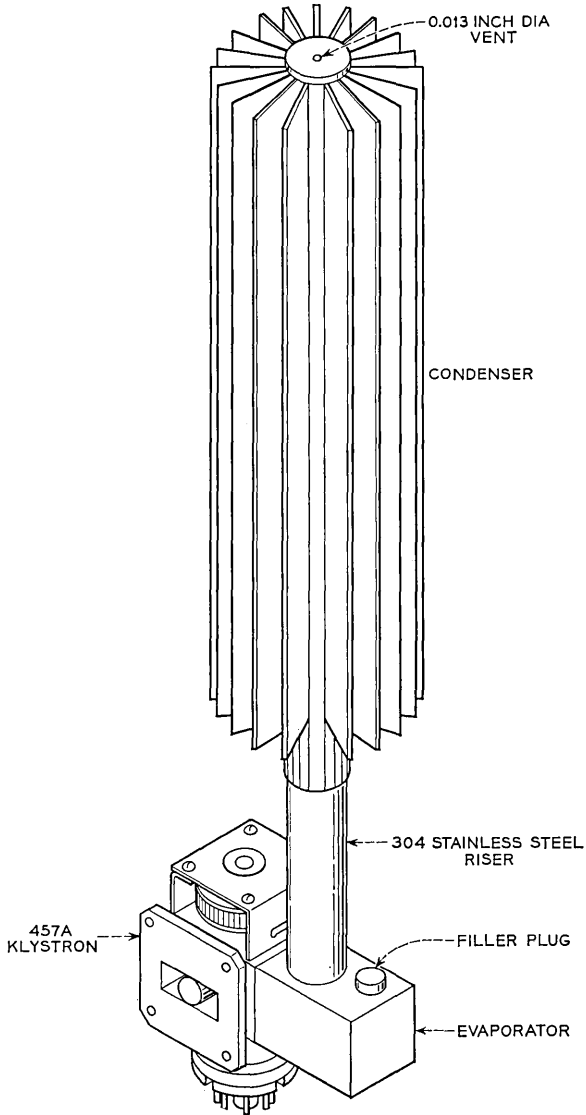


Fig. 1 — Experimental liquid-vapor system.

tor, and local boiling occurs at the evaporator wall attached to the klystron, which is the heat source. The contact surface is maintained at an essentially constant temperature slightly in excess of the liquid boiling point. In this common type of boiling, the heat energy is transmitted by

conduction and convection to the liquid in contact with the heat source surface and then by convection from the liquid to the bubbles that form on the surface. The bubbles grow in size as more liquid is vaporized until they break away from the surface due to their own buoyancy and rise in the liquid. As a bubble rises, more energy is transferred to the bubble by vaporization and the size of the bubble increases until finally it bursts at the liquid-vapor interface. In the vapor region, condensation takes place on the condenser walls and the condensate flows down the walls to the liquid region. The energy released in condensation of the vapor is conducted through the condenser walls and finally dissipated to the environment by free convection and radiation.

Preliminary calculations, which were later substantiated by experimental data, indicated that an open system at atmospheric pressure was required to obtain a satisfactory temperature regulation for the klystron. A 0.013-inch diameter vent at the top maintains atmospheric pressure in the condenser and also limits the rate of liquid loss by evaporation to a satisfactory low level. Initial data were collected with 20 cubic centimeters of water in the evaporator. The low ambient temperatures anticipated for the TL Radio System involve a risk of equipment damage from freezing if water is used. A fluorocarbon liquid, FC-75, developed by the Minnesota Mining and Manufacturing Company, was also evaluated since this liquid does not freeze at temperatures as low as -80°F . The FC-75 liquid has a boiling point approximately the same as water, and both liquids were found to be equally satisfactory for controlling the klystron temperature. A slight advantage in the liquid loss due to evaporation was found with the use of the FC-75 liquid.

The feasibility study was made for an ambient temperature range of $+35^{\circ}\text{F}$ to $+140^{\circ}\text{F}$. Typical results of this study indicated that the klystron body temperature varied by 20°F with a frequency change of 4.8 megacycles. Liquid loss measurements were also made for a 2000-hour (approximately 3 months) operating period. Typical results showed a total liquid loss of 3 cc. for FC-75 liquid and 5.5 cc. for water in the 2000-hour test period. In each case the test was started with 20 cc. of liquid in the evaporator, and temperature regulation of the klystron was not affected by the change in liquid volume. This study indicated that a considerable reduction in the effect of ambient temperature changes was possible with the liquid-vapor heat exchanger. The data collected during the study also indicated that the condenser could be much smaller and that reduction of the heat losses from the exposed surfaces of the klystron and evaporator would improve the temperature regulation of the klystron. Addition of thermal insulation around the unit yielded test results which substantiated these early conclusions.

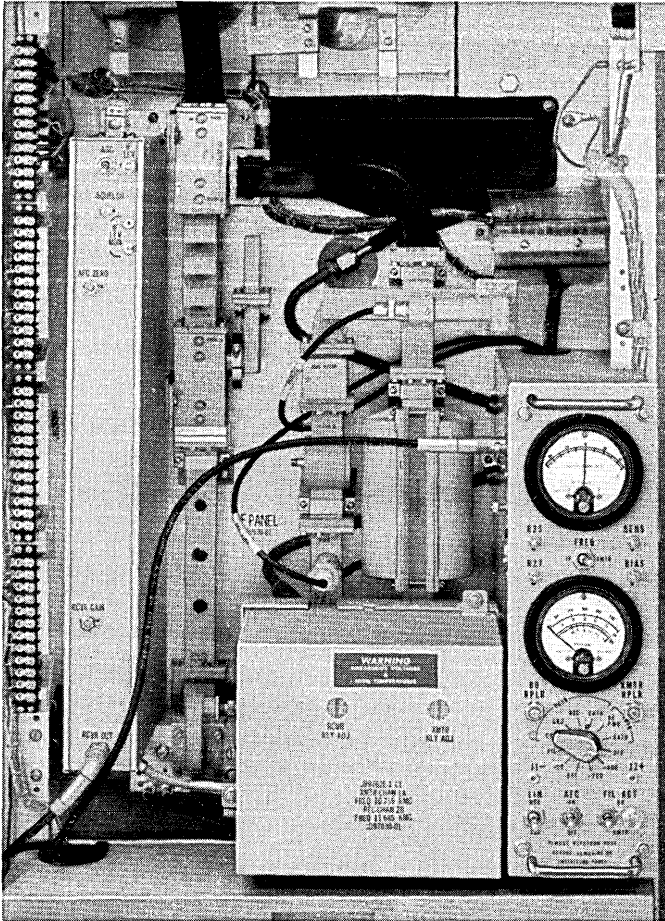


Fig. 2 — TL Radio System transmitter.

The experimental equipment shown in Fig. 1 was modified by molding an enclosure around the tube and evaporator of a very low density plastic (Eccofoam FP) having a wall thickness of approximately one inch. The insulation effectively isolated the klystron and evaporator from the ambient environment. Typical results of tests with the insulated unit indicated that the klystron body temperature varied 3°F for an ambient temperature range of -40° to $+140^{\circ}\text{F}$.

The experience gained during the feasibility studies led to considerable simplification of the final heat exchanger used in the TL Radio System. The appearance of the unit during operation is shown in Fig. 2. The

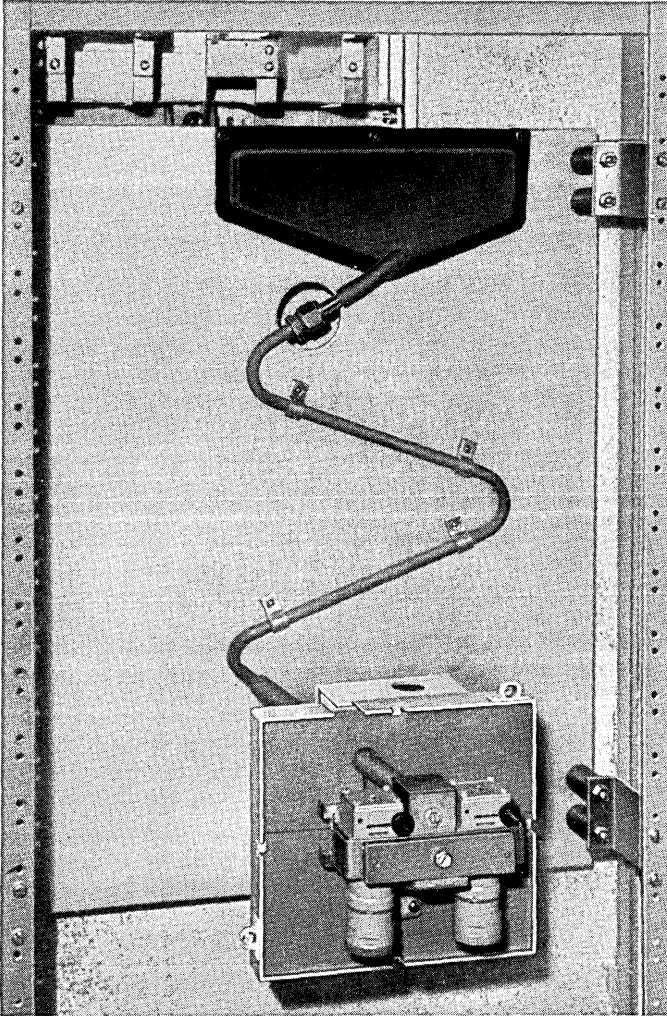


Fig. 3 — Cutaway view showing heat exchanger and condenser.

heavily insulated cover for the klystrons and the evaporator may be seen in the foreground. A cutaway view to show details is given in Fig. 3. Both the receiver and transmitter klystrons are cooled by a single evaporator. The condenser consists of a length of copper tubing clamped to an aluminum plate on which the receiver and transmitter components are mounted. The copper tubing is connected to the evaporator by stainless

steel tubing to reduce conduction losses from the evaporator. The end of the condenser tubing is closed off by a rubber expansion chamber that is soft enough to permit expansion without any significant increase in pressure. The chamber, which may be seen at the top of Fig. 3, effectively limits loss of FC-75 liquid due to evaporation. Test results on this complete unit have shown that the temperature of the two klystrons is held to within 3°F for an ambient temperature range of -40°F to $+140^{\circ}\text{F}$. These results, coupled with the thermal coefficient limits for the klystrons, mean that the thermal frequency drift of the transmitting klystron will not exceed 0.72 megacycle in the TL Radio System.

IV. 457A KLYSTRON THERMAL COEFFICIENT

The thermal coefficient of frequency for the 457A klystron is defined to be the change in frequency as a function of body temperature. The

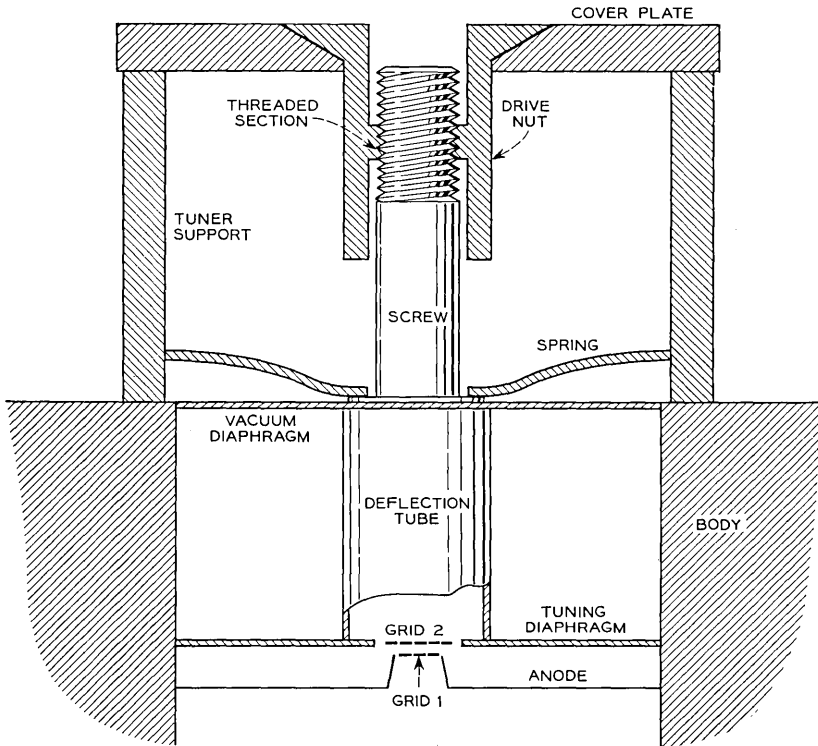


Fig. 4 — 457A tuner schematic.

limits for this coefficient are specified as ± 0.15 megacycle per degree Fahrenheit when measured at the mid-band frequency of 11,200 megacycles. The klystron design includes a method for adjusting the coefficient so that the average value for continuous production of tubes can be maintained near zero. The tuning mechanism is shown in schematic form in Fig. 4. The frequency of oscillation depends on the spacing between the two grids. An increase in spacing increases the frequency; approximately six microinches change in spacing results in a one megacycle change in frequency. The first grid is part of the anode assembly, which is a rigid member brazed to the tube body. The second grid is mounted on the flexible tuning diaphragm, which in turn is connected to the vacuum diaphragm and tuning screw by the deflection tube. The drive nut seats on the cover plate, which is rigidly supported from the body by the tuner supports. The threaded section of the drive nut engages the tuning screw threads, and rotation of the drive nut moves the screw to change the grid spacing. A wormgear and tuning shaft, not shown in Fig. 4, are coupled to the drive nut to provide a means for manual tuning to any frequency from 10,700 to 11,700 megacycles. The tuner spring loads the screw to maintain intimate contacts between the threads and between the drive nut and cover plate. When the tube has been mechanically tuned to a desired operating point, the frequency is affected by thermal expansions of the various parts due to temperature changes. The frequency depends on the dimensions of the piece parts and their relationship to each other. The thermal properties of seven piece parts largely determine the thermal coefficient. These parts are: the anode, body, tuner supports, cover plate, drive nut, screw, and the deflection tube. Rigid control of the thermal properties of each part would be an expensive and complicated process. The degree of control required may be estimated from the thermal coefficient requirements in terms of grid spacing. The thermal coefficient is dictated by the TL Radio System requirements to within ± 0.15 megacycle per degree Fahrenheit. To meet this requirement, the grid spacing change with temperature must be less than one microinch per degree Fahrenheit. A prohibitive accuracy of individual piece parts would be necessary to meet such requirements, and other factors, such as assembly variations, would further complicate the effect contributed by each part. Compensation for the gross effect of all thermal changes has been incorporated in the tube design, and adjustment of the compensation is included as a simple dimensional change on one piece part — the drive nut.

The basic principle of compensation is that at least two parts have opposite thermal effects. Adjustment of the compensation is readily made

by changing the thermal expansion of one of these parts. The application of this principle to the 457A klystron may be explained with reference to Fig. 4. Expansion of the tuner supports increases the grid spacing with a resulting increase in frequency, while expansion of the drive nut and screw decreases the grid spacing and frequency. Expansions of all other parts may simply be considered to modify the relative effects of the supports, drive nut, and screw. The net change in grid spacing due to expansions of all parts determines the thermal coefficient of frequency for the klystron. The adjustable expansion characteristic is obtained by changing the effective lengths of the drive nut and screw, which are made from materials with widely different expansion coefficients.

The thermal coefficient of the klystron is defined as the frequency change in megacycles per degree Fahrenheit when the tube body temperature is changed within the range of 235°F to 255°F. The temperature range was selected to exceed the actual operating temperature changes in the TL Radio System but is small enough to assume that all thermal expansion coefficients of the parts are constants. The klystron thermal coefficient may be considered to be the result of two compensating expansions. Let $-E$ represent the expansion of the drive nut and screw combination expressed in terms of megacycles per degree Fahrenheit, with the negative sign indicating that frequency is decreased by an increase in temperature. Also let A equal the sum of all other expansions expressed in terms of megacycles per degree Fahrenheit, with the positive sign indicating that frequency is increased by an increase in temperature. The thermal coefficient (C) for the klystron may be simply expressed as

$$C = A - E \text{ megacycles per degree Fahrenheit.}$$

For the ideal coefficient of zero, A must equal E , and perfect compensation is realized. When the compensation is upset by unavoidable causes, for example by a new supply of raw material in some piece part, an adjustment in E can be made to again equalize the effects of A and E . This adjustment has been confined to the term E , since this is related to the drive nut. The materials used for the drive nut and screw were chosen to have widely different expansion coefficients. The bronze drive nut has a coefficient which is more than three times the coefficient for the iron-nickel alloy screw. The expansion of each part depends on its coefficient and effective length. By referring to Fig. 4, it is readily seen that the effective lengths of the drive nut and screw are determined by the mating of the threads on these two parts. Changing the position of the threaded portion of the drive nut changes the effective lengths of both parts. A considerable change in the klystron thermal coefficient can

be attained by this method; the present piece parts can be modified to change the klystron thermal coefficient by as much as ± 0.2 megacycle per degree Fahrenheit. Numerical values have been derived relating the dimensional change to the thermal coefficient change for application of thermal coefficient correction as required. The dimensional change in the drive nut is obtained by a simple machine set-up.

By application of reasonable tolerances on piece parts, adjustment of the thermal coefficient for individual klystrons is not necessary for production purposes. This is important since individual adjustment would require two thermal coefficient measurements, one before and one after adjustment. This would be a time-consuming and expensive procedure. A large assortment of drive nuts would also be required for selective use. For the first year of the 457A klystron production, the thermal coefficient for all tubes will be measured to obtain data on the distribution of the coefficient. Individual limits of ± 0.15 megacycle per degree Fahrenheit will be applied to satisfy the TL Radio System requirements. On subse-

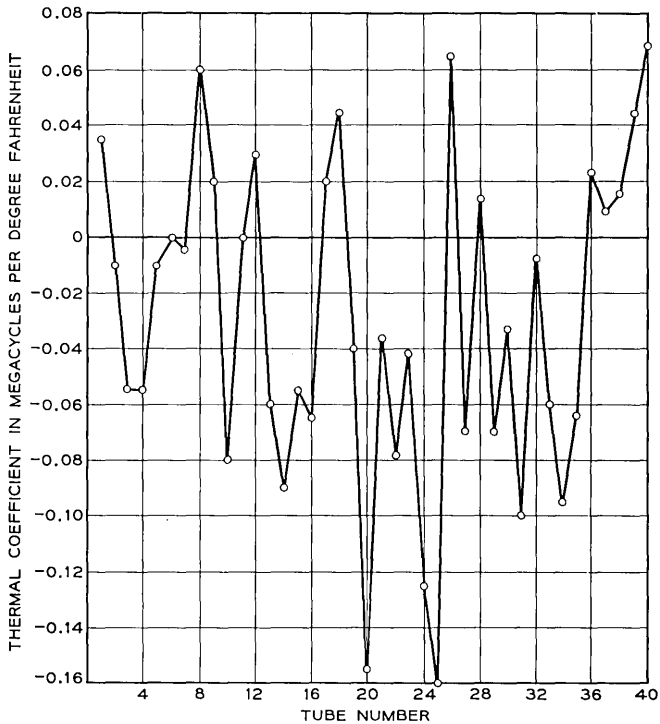


Fig. 5 — 457A thermal coefficients — 40 production tubes.

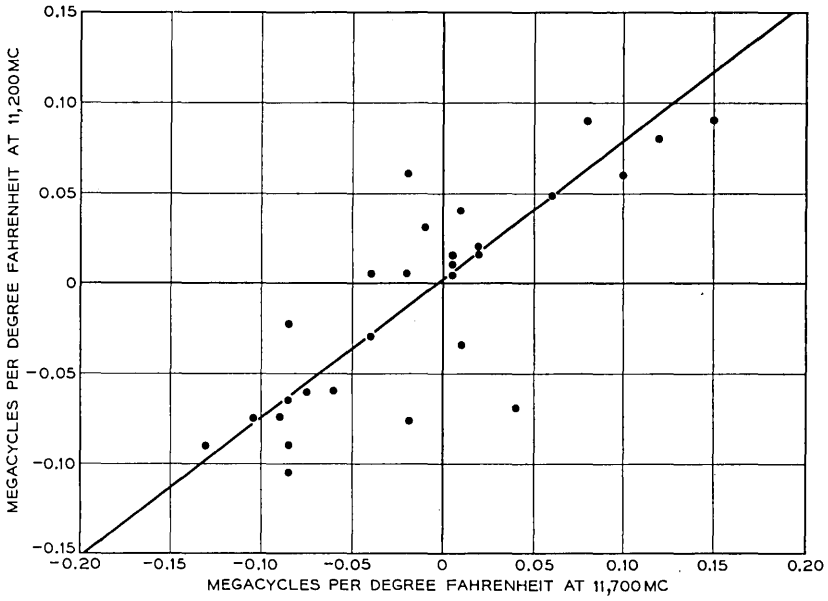


Fig. 6 — 457A klystron thermal coefficients.

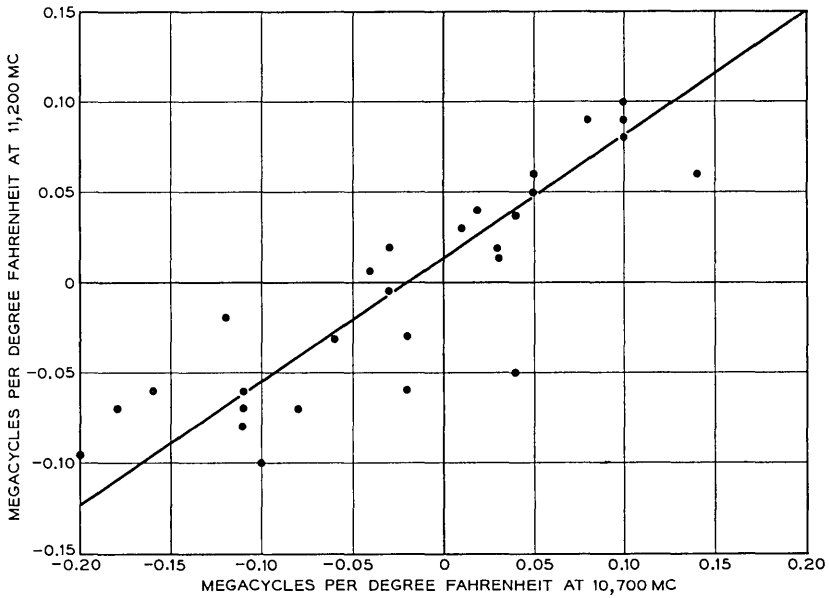


Fig. 7 — 457A klystron thermal coefficients.

quent production, the coefficient will be measured by a sampling plan to be determined by the experience of the first year of production. A dimensional change in the drive nut will be applied when necessary to maintain the average coefficient near zero.

The results of coefficient measurements for the initial 40 klystrons made under production conditions are shown on Fig. 5. These data show a distribution for the coefficient having an average of -0.029 mc per $^{\circ}\text{F}$ with a 3-sigma value of 0.165 mc per $^{\circ}\text{F}$. This distribution represents early production experience with considerable variations in the supply of piece parts. For stable production with a continuity of supply, a tighter distribution is anticipated. Two individual points outside the 0.15 limit are shown in the distribution. In both cases a manufacturing deviation accounted for the dispersion. The data shown on Fig. 5 are for the thermal coefficient measured at the mid-band frequency of $11,200$ megacycles. Correlation data for the coefficient at the extremes of the TL Radio frequency band are shown on Figs. 6 and 7. These data were collected on 28 tubes and indicate that limits of ± 0.15 mc per $^{\circ}\text{F}$ at mid-band correspond to limits of ± 0.20 mc per $^{\circ}\text{F}$ at $11,700$ mc and $+0.20$ to -0.24 mc per $^{\circ}\text{F}$ at $10,700$ mc. The extreme limits from these data will result in a frequency drift of 0.72 mc for the 457A klystron in the TL Radio System when subjected to an ambient temperature range of -40° to $+140^{\circ}$ Fahrenheit. For these extreme conditions, the 457A frequency error due to thermal effects is within the one-megacycle objective for the TL Radio System.

V. CONCLUSIONS

The frequency stability of the 457A klystron is within the TL Radio System objectives, and the need for automatic frequency control of the transmitter has been eliminated by passive methods. Elimination of the need for automatic frequency control reduces system costs and maintenance.

VI. ACKNOWLEDGEMENTS

As in any corporate development, many members of Bell Laboratories made important contributions to the designs of the 457A klystron and the heat exchanger. Particular mention should be made of the work of D. M. Sutter on heat transfer calculations, C. E. Apgar on the klystron mechanical design, and E. J. Neupauer on frequency drift measurements. The TL Radio System heat exchanger was developed at the Merrimack Valley Laboratories.

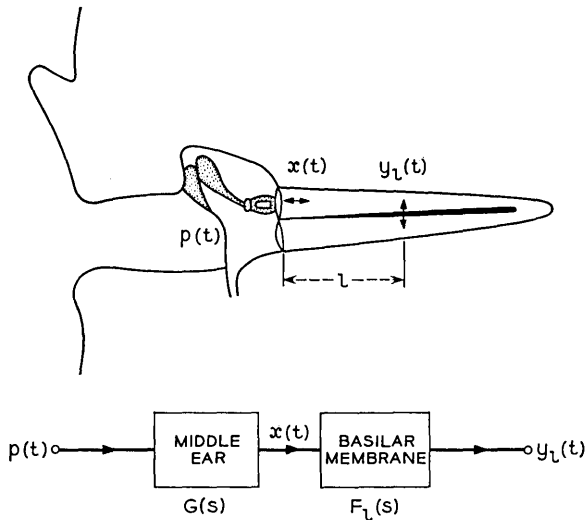
Models for Approximating Basilar Membrane Displacement—Part II. Effects of Middle-Ear Transmission and Some Relations between Subjective and Physiological Behavior

By JAMES L. FLANAGAN

(Manuscript received December 26, 1961)

This report presents the second half of results of a study on the peripheral ear. There are two objectives: (1) to derive computational models for approximating the mechanical displacement of the basilar membrane when the sound pressure at the eardrum is known, and (2) to demonstrate certain relations between subjective behavior measured experimentally and physiological behavior calculated from the models. The report describes a rational function approximation of middle-ear transmission. This result, in combination with previously derived models for the inner ear, permits an analytical approximation of basilar membrane displacements in both apical and basal regions. Because the models are rational functions, they can, if desired, be simulated by lumped-constant electrical networks. Their computational tractability also permits straightforward approximations to temporal and spatial derivatives of displacement. Relations between computed membrane displacement and subjective behavior are illustrated for several psychoacoustic phenomena, namely pitch perception, binaural lateralization, binaural time-intensity trade, threshold discrimination, and pure-tone masking. The extent to which some of these phenomena can be correlated with, identified in, and predicted by the mechanical operation of the peripheral ear appears to be substantial.

Part I of this report¹ described three analytical models for approximating the displacement of the basilar membrane when the human ear is stimulated by sound. These models were valid for points lying roughly in the apical half of the membrane, that is, for frequencies less than about



$$\frac{Y_L(s)}{P(s)} = \frac{X(s)}{P(s)} \cdot \frac{Y_L(s)}{X(s)} = G(s) \cdot F_L(s)$$

Fig. 1 — Schematic diagram of peripheral ear and functional relations between acousto-mechanical quantities.

1000 cps. Over this frequency range the elastic effects of the middle ear predominate, and the displacement of the stapes footplate is essentially proportional to, and in phase with, the sound pressure at the eardrum. At higher frequencies the mass and viscous properties of the middle ear become important, and the displacement transmission to the stapes is no longer constant with frequency. Applicability of the previously derived models to this range of frequencies depends upon being able to account for middle-ear transmission. This report describes an effort to derive a computational model for middle-ear transmission and to examine its relationship with the models for membrane displacement. Subsequent to this, an attempt is made to relate the mechanical operation of the ear, as described by the models, to several facets of subjective auditory behavior.

I. EFFECTS OF MIDDLE-EAR TRANSMISSION UPON MEMBRANE DISPLACEMENT*

The physiological functions to be considered are illustrated schematically in Fig. 1. $p(t)$ represents the sound pressure at the eardrum as a

* The material in this section was presented orally before the 60th meeting of the Acoustical Society of America, San Francisco, California, October, 1960. The abstract appears in *J. Acoust. Soc. Am.* **32**, 1960, p. 1494.

function of time; $x(t)$ is the equivalent linear displacement of the stapes footplate;* and $y_l(t)$ is the displacement of the basilar membrane (cochlea shown uncoiled) at a distance l from the stapes. In terms of frequency-domain (Laplace) transforms, the middle-ear transmission is represented by $G(s)$ and the stapes-to-membrane transmission by $F_l(s)$.

In deducing approximations to these functions, the peripheral ear is assumed both to be mechanically linear over the range of interest and to constitute a passive system. A passive system is stable by definition. It has no normal modes whose amplitudes increase indefinitely with time. The functions $G(s)$ and $F_l(s)$ can therefore be approximated by rational functions of frequency whose coefficients are real and whose poles and zeros are either real or occur in complex conjugates. The functions can have no poles with positive real parts and only simple poles with zero real parts.

The earlier paper¹ essentially treated functional approximations to $F_l(s)$ (that is, middle-ear transmission was assumed constant with frequency, or, in the present notation, $G(s) = 1$). Two of the previously derived models will be useful in the present discussion. They are the first and third which, according to the notation used earlier, were called $F_1(s)$ and $F_3(s)$. For convenience they are reproduced here and are:

$$F_1(s) = c_1 \beta_l^{4+r} \left(\frac{s + \epsilon}{s + \gamma} \right) \left[\frac{1}{(s + \alpha_l)^2 + \beta_l^2} \right]^2 e^{-s \frac{3\pi}{4\beta_l}}, \quad (1)$$

and

$$F_3(s) = c_3 \beta_l^{4+r} \frac{\left[s^2 + 2\alpha_l s + \left(\alpha_l^2 - \frac{\beta_l^2}{3} \right) \right]}{[(s + \alpha_l)^2 + \beta_l^2]^3} e^{-s \frac{3\pi}{4\beta_l}}, \quad (2)$$

where

- $s = (\sigma + j\omega)$ is the complex frequency,
- β_l is the radian frequency to which the point l distance from the stapes responds maximally,
- β_l^{4+r} is a factor which matches the physiologically measured variations in peak amplitude of displacement with resonant frequency $\beta_l \dagger$,
- $e^{-s \frac{3\pi}{4\beta_l}}$ is a delay factor of $3\pi/4\beta_l$ seconds which brings the

* In man, the stirrup does not move longitudinally as a planar piston but usually exhibits some rotational motion. $x(t)$ is taken here as the volume displacement of the footplate divided by its area.

† The present form of this factor is applicable only to the frequency range below 1000 cps. Here, as previously discussed,¹ the value of $r = 0.8$. A minor modification will be made in this factor presently to make it appropriate for higher frequencies.

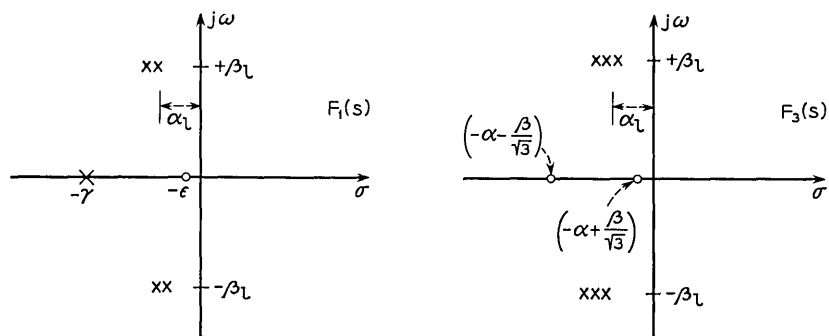


Fig. 2 — Pole zero diagrams for two functional approximations of $F(s)$.

phase delay of the model into line with the phase measured on the human ear. This factor is principally transit delay from stapes to point l on the membrane*.

The membrane characteristics are therefore approximated in terms of the poles and zeros of these two functions. Because the resonant properties of the membrane are nearly constant Q in character, the real and imaginary parts of the pole frequencies are related by a constant factor, i.e., $\beta_l = k\alpha_l$. For the present models, the best fits to the experimental data are obtained for the following choice of parameters:

$$\begin{aligned} \text{For } F_1(s): \frac{\epsilon}{\beta_l} &= 0.1 \text{ to } 0.0\dagger, & \text{For } F_3(s): \frac{\beta_l}{\alpha_l} &= 1.7 \\ \frac{\gamma}{\beta_l} &= 1.0 & & (3) \\ \frac{\beta_l}{\alpha_l} &= 2.0. & & \end{aligned}$$

Therefore, to within a multiplicative constant, the imaginary part of the pole frequency β_l completely describes the model. The pole-zero diagrams for the two models are shown in Fig. 2.

The real frequency responses of the models are evidenced by letting $s = j\omega$. If frequency is normalized in terms of $\zeta = \omega/\beta_l$, then the rela-

* At low frequencies the phase of the model departs somewhat from the experimental data. See the discussion of this point in Ref. 1 and also in J. L. Flanagan and C. M. Bird, Minimum Phase Responses for the Basilar Membrane, *J. Acoust. Soc. Am.* **34**, 1962, p. 114.

† See earlier comments about fitting phase response.

tive phase and amplitude responses of $F_1(j\zeta)$ and $F_3(j\zeta)$ are shown in Fig. 3 for the parameters stated in (3).

The inverse Laplace transforms of (1) and (2) are the displacement responses of the membrane to an impulse of displacement by the stapes. These representations will also be useful in the present discussion. If the mathematics is carried out the inverse transforms are found to be:

$$\begin{aligned}
 f_1(t) = & c_1\beta_i^{1+r} \{ [0.033 + 0.360\beta_i(t - T)] e^{-\frac{\beta_i(t-T)}{2}} \sin \beta_i(t - T) \\
 & + [0.575 - 0.320\beta_i(t - T)] e^{-\frac{\beta_i(t-T)}{2}} \cos \beta_i(t - T) - 0.575e^{-\beta_i(t-T)} \} \quad (4) \\
 & \text{for } t \geq T; \quad \epsilon/\beta_i = 0.1 \\
 & = 0; \quad \text{for } t < T,
 \end{aligned}$$

and

$$\begin{aligned}
 f_3(t) = & \frac{c_3\beta_i^{1+r}}{6} [\beta_i(t - T)]^2 e^{-\beta_i(t-T)/1.7} \sin \beta_i(t - T) \quad \text{for } t \geq T \\
 & = 0 \quad \text{for } t < T,
 \end{aligned} \quad (5)$$

where the delay $T = 3\pi/4\beta_i$, as previously stated. In the earlier paper, the simplicity of $f_3(t)$ was the main reason that $F_3(s)$ was considered as an approximation to the experimental frequency domain data. A plot

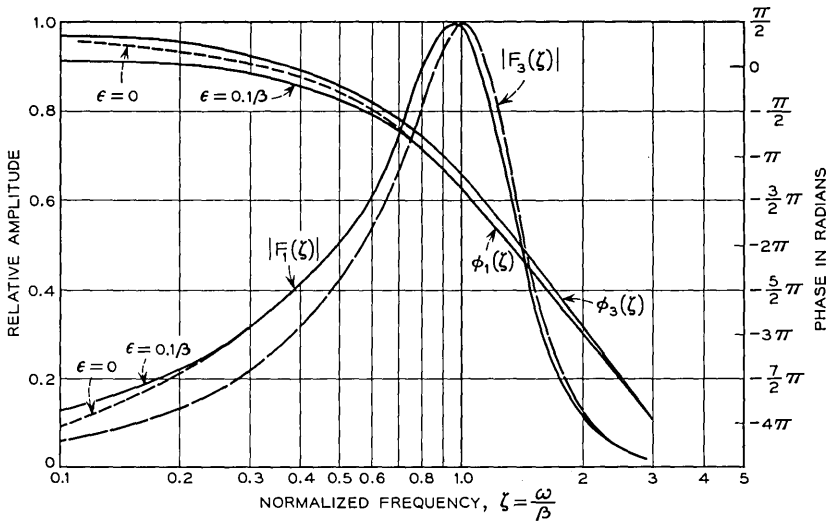


Fig. 3 — Amplitude and phase responses for two $F(s)$ models.

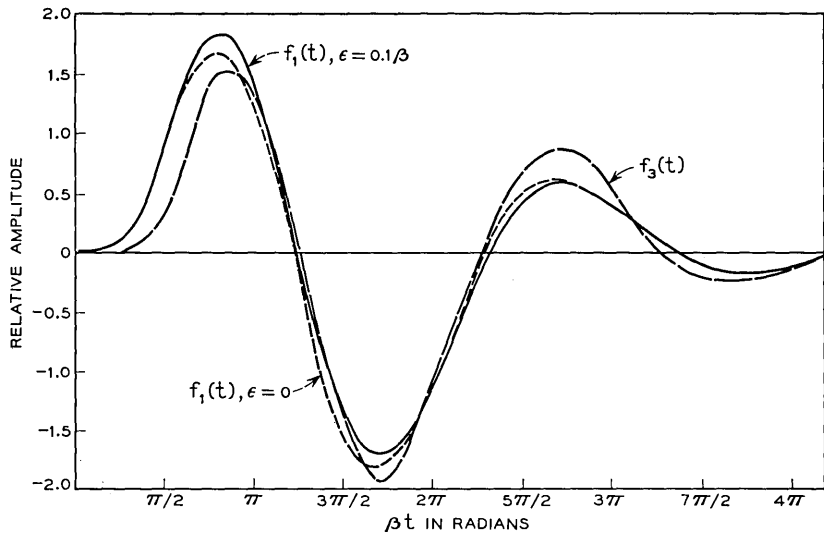


Fig. 4 — Impulse responses for the membrane models. These responses are the inverse transforms of the frequency data in Fig. 3 and represent membrane displacement caused by an impulse of stapes displacement.

of the responses (4) and (5), on a relative amplitude scale and with delay equalized, is shown in Fig. 4. The absolute time origins for the two traces are to the left of the relative origin by 1.9 and 1.5 radians respectively.

As indicated above, the factor β_i^{4+r} in (1) and (2) has a form appropriate to the frequency range below 1000 cps. If the membrane models are to be used at higher frequencies, this factor should be modified according to data given by Bekesy² on the peak membrane displacement as a function of frequency (see Fig. 4 in Ref. 1). For constant sinusoidal displacement of the stapes, Bekesy's data indicate that the peak membrane displacement increases at about 5 db/octave up to around 1000 cps, and then tends to flatten off and become roughly constant (at least up to about 2000 cps).

This amplitude variation can be accounted for by altering the multiplicative amplitude factor to $\beta_i^{4+r}(2\pi \cdot 1000/\beta_i + 2\pi \cdot 1000)^r$. The modification causes the peak response (of the curve shown in Fig. 3) to rise at about 5 db/octave below 1000 cps, and to flatten off above this frequency. At low β_i frequencies the amplitude factor is the same as before if the constant c_1 is readjusted by multiplying it by 2^r . With this minor modification, then, the functional approximations to $F_i(s)$ are appropriate for use at frequencies higher than 1000 cps.

1.1 *A Model for Middle-Ear Transmission*

To account for middle-ear transmission one would like an analytical specification of the stapes displacement produced by a given sound pressure at the eardrum for all frequencies of interest. Quantitative physio-acoustical data on the operation of the middle ear are very sparse. The data which are available are due largely to Bekesy³ and to Zwislocki.⁴ By considering the topology of the mechanical circuit and the values of elastic, mass, and viscous constants measured in physiological preparations, it is possible to deduce information about the middle-ear transmission. Zwislocki used this approach to develop an analog electrical circuit for the middle ear in which voltage is analogous to pressure, and current to volume velocity. The circuit includes ten components representing the acousto-mechanical elements of the middle ear. Seven of the elements are energy storage elements.

Using the constants suggested by Zwislocki, we measured the transfer characteristics of the middle-ear circuit when terminated in an impedance analogous to the input mechanical impedance of the cochlea. For a constant pressure at the eardrum, the amplitude and phase responses of the stapes displacement are shown by the curves in Fig. 5.*

Although the characteristic equation corresponding to Zwislocki's analog circuit is of seventh degree, the stapes displacement can be analytically approximated reasonably well by a function of third degree. (As discussed in the earlier paper, the criterion of fit is again taken as an intuitive one.) Such an approximating function is of the form:

$$G(s) = \frac{c_0}{(s + a)[(s + a)^2 + b^2]}, \quad (6)$$

where c_0 is a positive real constant. [When combined with $F_l(s)$, the multiplying constants are chosen to yield proper absolute membrane displacement. For convenience, therefore, consider $c_0 = a(a^2 + b^2)$ so

* After the present work was carried out, an excellent paper by A. Møller (Network Model of the Middle Ear, *J. Acoust. Soc. Am.* **33**, 1961, p. 168) appeared in which analogous electrical circuits for the middle ear are deduced on the basis of input impedance measurements at the drum and the middle-ear topology. For a comparison with Zwislocki's data (which we had already used), we constructed several of Møller's circuits and measured their transfer characteristics. Although their frequency responses differ in fine detail, the results of Zwislocki and Møller agree in the gross aspects of the transmission characteristics. As do Bekesy's earlier results, both sets of the latter data suggest some uncertainty and variability in the middle-ear transmission, particularly in regard to the frequency at which the transmission begins to diminish appreciably. Apparently the function differs among individuals. One of the main objectives of the present paper, however, is to demonstrate a computational technique which has been found useful in explaining certain auditory phenomena. Whenever physiological data are improved and extended, the results can easily be incorporated into the analytical technique presented here.

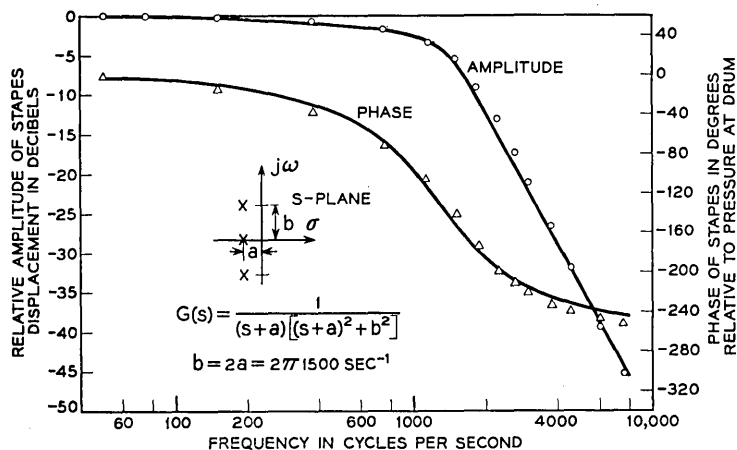


Fig. 5 — Functional approximation of middle ear transmission. The solid curves are from Zwislocki⁴ and the plotted points are amplitude and phase values of $G(s)$.

that the low-frequency transmission of $G(s)$ is unity.] When the pole frequencies of $G(s)$ are related according to

$$b = 2a = 2\pi(1500) \text{ rad/sec}, \quad (7)$$

the fit to Zwislocki's data is shown by the plotted points in Fig. 5.

The inverse transform of (6) is the displacement response of the stapes to an impulse of pressure at the eardrum. It is obtained easily and will be useful in the subsequent discussion. Let

$$G(s) = G_1(s)G_2(s),$$

where

$$G_1(s) = \frac{c_0}{s+a}; \quad G_2(s) = \frac{1}{(s+a)^2 + b^2}. \quad (8)$$

The inverses of the parts are:

$$g_1(t) = c_0 e^{-at}; \quad g_2(t) = \frac{e^{-at}}{b} \sin bt. \quad (9)$$

The inverse of $G(s)$ is then

$$g(t) = \int_r^t g_1(\tau)g_2(t-\tau) d\tau,$$

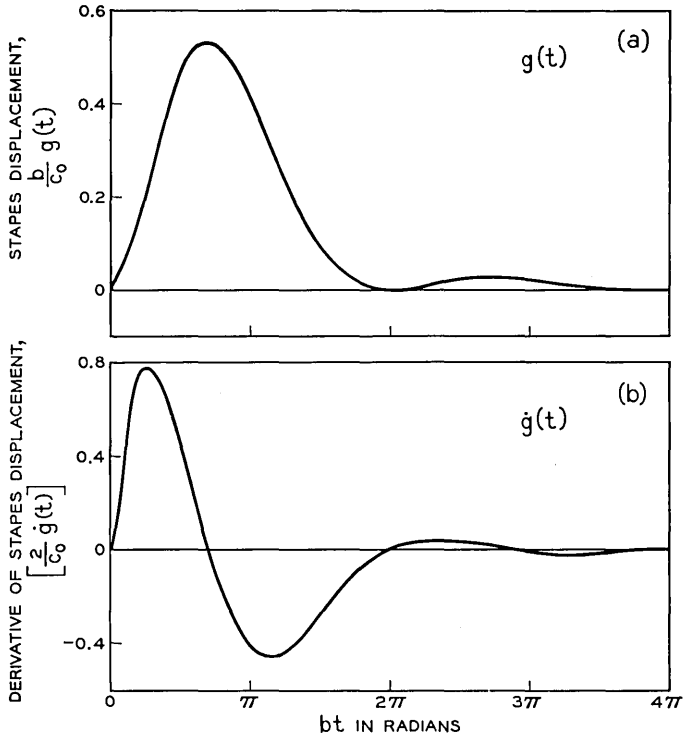


Fig. 6 — Displacement response of the stapes and its time derivative to an impulse of pressure at the eardrum.

or

$$g(t) = c_0 \frac{e^{-at}}{b} (1 - \cos bt) = \frac{c_0 e^{-bt/2}}{b} (1 - \cos bt). \quad (10)$$

Also for use at a future point in the discussion we note that the time derivative of the stapes displacement is:

$$\dot{g}(t) = \frac{c_0 e^{-bt/2}}{2} (2 \sin bt + \cos bt - 1). \quad (11)$$

Plots of $g(t)$ and $\dot{g}(t)$ are shown in Fig. 6.

1.2 Combined Response of Middle Ear and Basilar Membrane

The combined response of the models for the middle ear and basilar membrane is simply:

$$\begin{aligned} H_1(s) &= G(s)F_1(s) \\ h_1(t) &= g(t)*f_1(t). \end{aligned} \quad (12)$$

To simplify the computations and to illustrate both results, the response for model $F_1(s)$ will be computed in the frequency domain and that for $F_3(s)$ will be computed in the time domain.

1.2.1 Inverse Transform of $H_1(s)$

Disregarding for the moment the constant delay and amplitude terms, which can be resupplied at the end if needed, the problem of transforming $H_1(s) = G(s)F_1(s)$ amounts to computing the inverse of:

$$H_1'(s) = \frac{1}{s+a} \cdot \frac{1}{(s+a)^2 + b^2} \cdot \frac{s+\epsilon}{s+\gamma} \cdot \frac{1}{[(s+\alpha)^2 + \beta^2]^2}. \quad (13)$$

Expand $H_1'(s)$ as partial fractions:

$$H_1'(s) = \frac{A}{s+a} + \frac{Bs+c}{(s+a)^2 + b^2} + \frac{D}{s+\gamma} + \frac{E(s)}{[(s+\alpha)^2 + \beta^2]^2} \quad (14)$$

where

$$A = \frac{1}{b^2} \cdot \frac{\epsilon - a}{\gamma - a} \cdot \frac{1}{[(\alpha - a)^2 + \beta^2]^2}, \quad \gamma \neq a$$

$$B = 2 \operatorname{Re} B'$$

$$B' = \frac{1}{2b^2} \cdot \frac{\epsilon - a + j\bar{b}}{\gamma - a + j\bar{b}} \cdot \frac{1}{[(\alpha - a)^2 + \beta^2 - b^2 + j2b(\alpha - a)]^2}$$

$$C = [a(2 \operatorname{Re} B') - b(2 \operatorname{Im} B')]$$

$$D = \frac{1}{a - \gamma} \cdot \frac{1}{(a - \gamma)^2 + b^2} \cdot \frac{\epsilon - \gamma}{[(\alpha - \gamma)^2 + \beta^2]^2}$$

$$E(s) = (a_0 + a_1s + a_2s^2 + a_3s^3).$$

On the basis of the previous findings the problem is particularized to the conditions:

$$\begin{aligned} \beta &= 2\alpha & \text{Also let:} & \quad \eta = \beta/b \\ b &= 2a & & \quad \gamma \neq a \\ \gamma &= \beta & & \quad \beta \neq b. \\ \epsilon &= 0 & & \end{aligned} \quad (15)$$

If the arithmetic is followed through, the constants are found to be:

$$A = -\frac{1}{b^6(2\eta - 1)(1.25\eta^2 - 0.50\eta + 0.25)^2}$$

$$B' = \frac{(0.50 - j1.00)}{2b^6(\eta - 0.50 + j1.00)[(1.25\eta^2 - 0.50\eta - 0.75) + j(\eta - 1.00)]^2}$$

$$B = 2 \operatorname{Re} B'$$

$$C = b(\operatorname{Re} B' - 2 \operatorname{Im} B')$$

$$D = \frac{1}{b^6} \frac{1}{(\eta - 0.50)(\eta^2 - \eta + 1.25)(1.25)^2\eta^3},$$

and the coefficients of $E(s)$ are found to be:

$$a_0 = -\eta^3 b^2 (3.12\eta b A + 1.25\eta C + 1.56b D)$$

$$a_1 = -b^2 \left[A(3.50\eta^2 - \eta + 0.25) + B(3.50\eta^2 - 2.00\eta - 0.25) + D(2.5\eta^2) + C \left(\frac{2\eta - 1}{b} \right) \right] \tag{16}$$

$$a_2 = -[Ab(2\eta - 0.50) + Bb(2\eta - 1) + Db\eta + C]$$

$$a_3 = -(A + B + D).$$

Although somewhat involved numerically, the inverse transform of $H_1'(s)$ can now proceed termwise as indicated in (14). The basic procedure from this point has already been indicated in the appendix of the earlier paper. When the details of this instance are carried through, the result is:

$$\begin{aligned} h_1'(t) = & A e^{-bt/2} + B e^{-bt/2} (\cos bt - 0.50 \sin bt) \\ & + \frac{C}{b} (e^{-bt/2} \sin bt) + D e^{-\eta b t} + (e^{-\eta b t/2} \sin \eta b t) \\ & \cdot \left\{ \frac{1}{2\eta^3 b^3} \left[a_0 - a_1 \frac{\eta b}{2} + a_2(1.25\eta^2 b^2) - a_3(1.63\eta^3 b^3) \right] \right\} \\ & + (\eta b t e^{-\eta b t/2} \sin \eta b t) \left\{ \frac{1}{2\eta^2 b^2} [a_1 - a_2 \eta b - a_3(0.25\eta^2 b^2)] \right\} \tag{17} \\ & + a_3 e^{-\eta b t/2} \cos \eta b t + (\eta b t e^{-\eta b t/2} \cos \eta b t) \left\{ \frac{1}{2\eta^3 b^3} \left[-a_0 + a_1 \frac{\eta b}{2} \right. \right. \\ & \left. \left. + a_2(0.75\eta^2 b^2) - a_3(1.38\eta^3 b^3) \right] \right\}; \text{ for } t \geq 0. \end{aligned}$$

$h_1(t)$ is obtained from $h_1'(t)$ by resupplying the constant delay $T =$

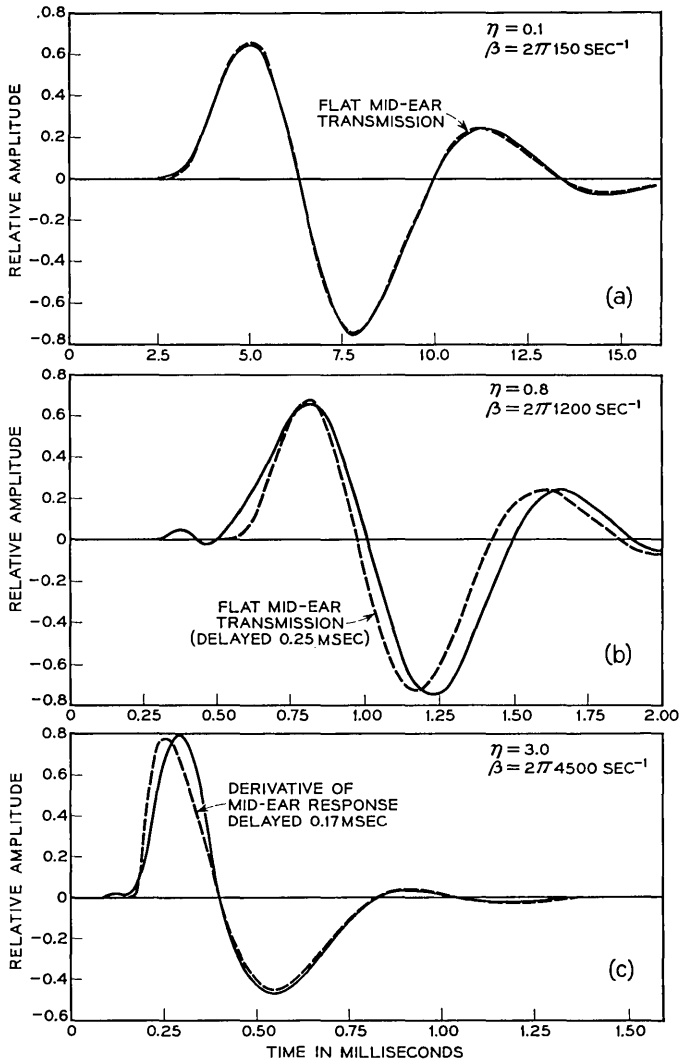


Fig. 7 — Displacement responses of apical, middle and basal points on the membrane to an impulse of pressure at the eardrum. These are computed from the inverse transform of $[G(s)F_1(s)]$.

$3\pi/4\beta$ and the multiplicative amplitude constants; that is, by letting $t = (t - T)$ and by multiplying $h'(t - T)$ by

$$[c_0 c_1 \beta_i^{4+r} (2\pi \cdot 1000/\beta_i + 2\pi \cdot 1000)^r], \text{ where } r = 0.8.$$

The form of the impulse response is thus seen to depend upon the

parameter $\eta = \beta/b$. Values of $\eta < 1.0$ refer to (apical) membrane points whose frequency of maximal response is less than the critical frequency of the middle ear. For these points the middle-ear transmission is essentially constant with frequency, and the membrane displacement is very nearly indicated by $f_1(t)$ in (4). On the other hand, values of $\eta > 1.0$ refer to (basal) points which respond maximally at frequencies greater than the critical frequency of the middle ear. For these points the middle-ear transmission is highly dependent upon frequency and would be expected to influence strongly the membrane displacement. To illustrate this point, (17) has been evaluated for $\eta = 0.1, 0.8, \text{ and } 3.0$. The result, with the delay resupplied, is shown in Fig. 7.

For an impulse of pressure delivered to the eardrum, the three solid curves represent the membrane displacements at points which respond maximally to frequencies of 150, 1200, and 4500 cps. Each of the plots also includes a dashed curve. In Figs. 7(a) and 7(b), the dashed curve is the membrane displacement computed by assuming the middle-ear transmission to be flat with zero phase. [This is simply the response $\mathcal{L}^{-1}F_1(s)$.] In Fig. 7(c) the dashed curve is the time derivative of the stapes displacement, $\dot{g}(t)$, taken from Fig. 6. The suggestion is that in the basal region, the form of the membrane displacement is very similar to the derivative of the stapes displacement. This apparently is the case, and this point will be considered again presently.

1.2.2 *Inverse Transform for $H_3(s)$*

If, as in the previous section, delay and scale constants are temporarily disregarded, the inverse transform for $[G(s)F_3(s)]$ is given by the time-domain convolution:

$$h_3'(t) = \left[(\beta t)^2 e^{\frac{\beta t}{1.7}} \sin \beta t \right] * \left[\frac{e^{\frac{bt}{2}}}{b} (1 - \cos bt) \right];$$

or

$$h_3'(t) = \int_0^t [(\beta\tau)^2 e^{\frac{\beta\tau}{1.7}} \sin \beta\tau] \left\{ \frac{e^{\frac{b(t-\tau)}{2}}}{b} [1 - \cos b(t - \tau)] \right\} d\tau, \quad (18)$$

for $t \geq 0$.

When this integration is carried through, the result is:

$$h_3'(t) = \left(\frac{\eta}{b} \right)^2 [\text{Im}(P) - \frac{1}{2} \text{Im}(Q) - \frac{1}{2} \text{Im}(R)]; \quad t \geq 0 \quad (19)$$

where

$$(P) = \frac{1}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j\eta\right]} \left[\left[e^{-\frac{\eta bt}{1.7}} \{ \cos \eta bt + j \sin \eta bt \} \right. \right. \\ \left. \left. \cdot \left\{ (bt)^2 - \frac{2bt}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j\eta\right]} + \frac{2}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j\eta\right]^2} \right\} \right. \right. \\ \left. \left. - \frac{2e^{-bt/2}}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j\eta\right]^2} \right] \right]$$

$$(Q) = \frac{1}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta - 1)\right]} \left[\left[e^{-\frac{\eta bt}{1.7}} \{ \cos \eta bt + j \sin \eta bt \} \right. \right. \\ \left. \left. \cdot \left\{ (bt)^2 - \frac{2bt}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta - 1)\right]} + \frac{2}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta - 1)\right]^2} \right\} \right. \right. \\ \left. \left. - \frac{2e^{-bt/2}(\cos bt + j \sin bt)}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta - 1)\right]^2} \right] \right]$$

$$(R) = \frac{1}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta + 1)\right]} \left[\left[e^{-\eta bt/1.7} \{ \cos \eta bt + j \sin \eta bt \} \right. \right. \\ \left. \left. \cdot \left\{ (bt)^2 - \frac{2bt}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta + 1)\right]} + \frac{2}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta + 1)\right]^2} \right\} \right. \right. \\ \left. \left. - \frac{2e^{-bt/2}(\cos bt - j \sin bt)}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta + 1)\right]^2} \right] \right]$$

As before, $h_3(t)$ is obtained by resupplying the amplitude factors and the delay T . By way of examining the form of $h_3(t)$, (19) has been evaluated for $\eta = 0.1$ and 3.0 . The resulting $h_3(t)$ is plotted in Fig. 8. Comparison with the previous response for $h_1(t)$ shows the results to be similar.

1.2.3 Combined Frequency-Domain Responses

The individual frequency-domain responses for $G(s)$ and $F_l(s)$ have been shown in Figs. 3 and 5 respectively. The combined response in the

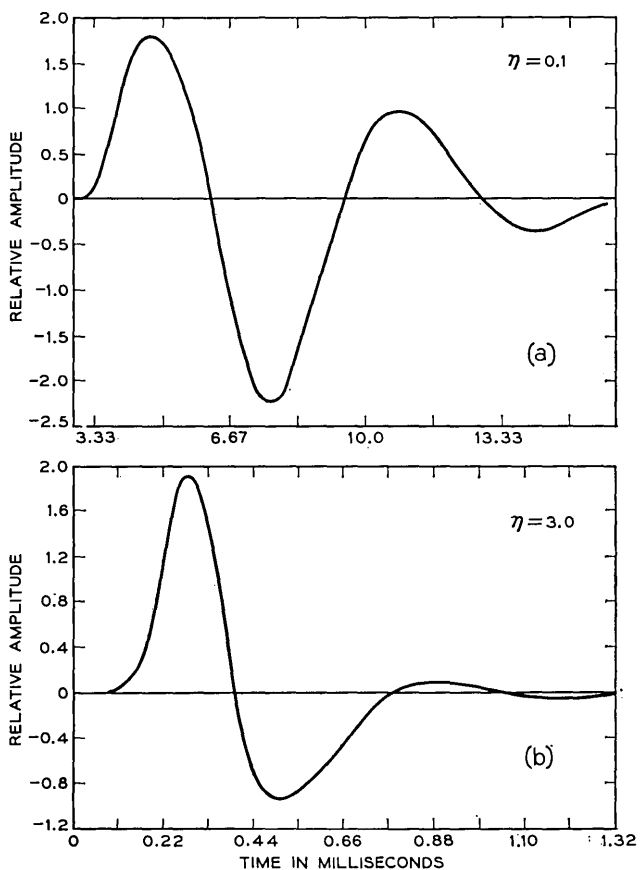


Fig. 8 — Impulse responses for apical and basal points computed from $[g(t) * f_3(t)]$.

frequency domain is simply the sum of individual amplitude (in db) and phase (in radians) responses. The combined amplitude and phase responses for the model $G(s)F_1(s)$ are shown in Figs. 9 and 10, respectively.

As already indicated by the impulse responses, one sees that the response of apical (low-frequency) points on the membrane is given essentially by $F(s)$, while for basal (high-frequency) points the response is considerably influenced by the middle-ear transmission $G(s)$. In particular, notice two things about the frequency response of the membrane model [i.e., $F(\omega)$]. One, the low-frequency skirt of the amplitude curve rises at about 6 db/octave. And two, the phase of the membrane

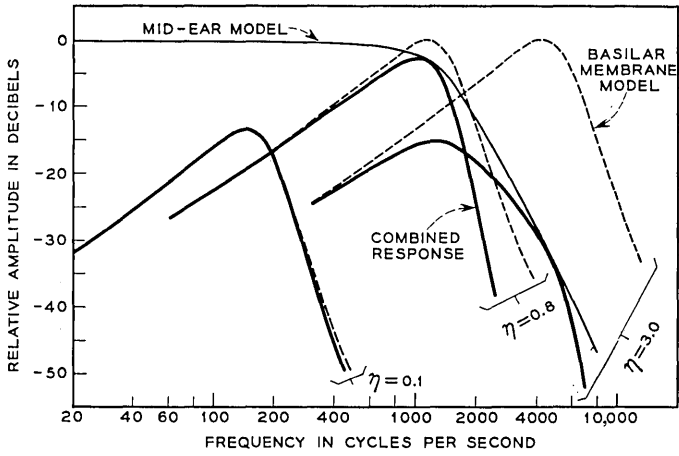


Fig. 9 — Frequency responses for the combined models $[G(\omega)F_1(\omega)]$.

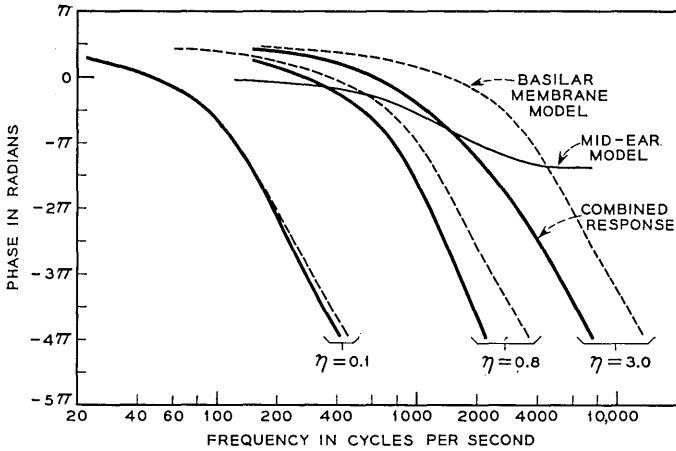


Fig. 10 — Phase responses for the combined models $[G(\omega)F_1(\omega)]$.

model [i.e., $F(\omega)$] approaches $+\pi/2$ radians at frequencies below the peak amplitude response. In other words, at frequencies appreciably less than its peak response frequency, the membrane function $F(\omega)$ behaves approximately as a differentiator.

Because the middle-ear transmission begins to diminish in amplitude at frequencies above about 1500 cps, the membrane displacement in the basal region is roughly the *time derivative of the stapes displacement*. The

waveform of the impulse response along the basal part of the membrane is therefore approximately constant in shape. Along the apical part, however, the impulse response oscillates more slowly as the apex is approached. This has already been illustrated in Fig. 7. [If the apical response is considered on a time scale normalized in terms of (βt) , then the displacement waveform is constant in shape.] This relation can, and has been, supported by psychoacoustic measurements. These results will be discussed in the second part of the paper.

Notice one other thing from Fig. 9. Because the amplitude response of the middle ear declines appreciably at high frequencies, the amplitude response of a basal point is highly asymmetrical (for example, the combined response for $\eta = 3.0$.) The result is that a given basal point, while responding with greater amplitude than any other membrane point at its characteristic frequency (i.e., at β_i), responds with greatest amplitude (but not greater than some other point) at some lower frequency.

1.3 *Some Temporal and Spatial Relations For Membrane Displacement*

Certain results from physiological research⁵ suggest that shear stresses along the basilar membrane may be as important in the mechanical-to-neural transduction as absolute displacements of the membrane. Accordingly, the spatial derivative of the displacement may be the mechanical factor of consequence. The computational tractability of the model permits a straightforward consideration of some temporal and spatial relations for the displacement.*

As a beginning, because they are easiest to talk about, consider only apical membrane points where the middle-ear transmission is essentially constant. In this case the displacement is nearly $f(t)$ [see (4) and (5)], and is only a function of t and the point parameter β . The variable β is a function of the distance along the membrane and can be so specified. (This functional relationship will be developed presently.) The impulse response is essentially a function of the product $\beta(t - T)$ and has a multiplicative factor involving β^{1+r} (i.e., $\beta^{1+r}g[\beta(t - T)]$). This fact points up a simple aspect of the dispersive nature of the basilar membrane.

If a disturbance is propagating in a nondispersive medium, the wave moves with a velocity which is the same for all frequency components, and the waveform is maintained undistorted. Let the wave for a one-dimensional situation be $p(t, x) = p(ct - x)$, where c is the velocity.

* See the further discussion of spatial derivatives (displacement gradients) in Section II.

Then,

$$\frac{\partial p}{\partial t} = c \frac{\partial p}{\partial(ct - x)}, \quad \text{and} \quad \frac{\partial p}{\partial x} = -\frac{\partial p}{\partial(ct - x)}, \quad (20)$$

and the time and space derivatives have the same waveform. The corresponding relations for the displacement responses of the membrane, however, must differ somewhat in time waveform. The model $f_3(t)$, Eq. (1.5), because of its simplicity, is particularly useful for illustrating this.

Again neglecting the amplitude constants which do not involve β or t , and which can be resupplied in the result, $f_3(t)$ reduces to:

$$f_3''(t) = \beta^{1+r} g[\beta(t - T)]; \quad t \geq T$$

where

$$g[\beta(t - T)] = [\beta(t - T)]^2 e^{-\beta(t-T)/1.7} \sin \beta(t - T), \quad (21)$$

and

$$T = 3\pi/4\beta.$$

Then,

$$\frac{\partial f_3''}{\partial \beta} = \beta^r \left[\beta \frac{\partial g}{\partial \beta} + (1 + r)g \right]. \quad (22)$$

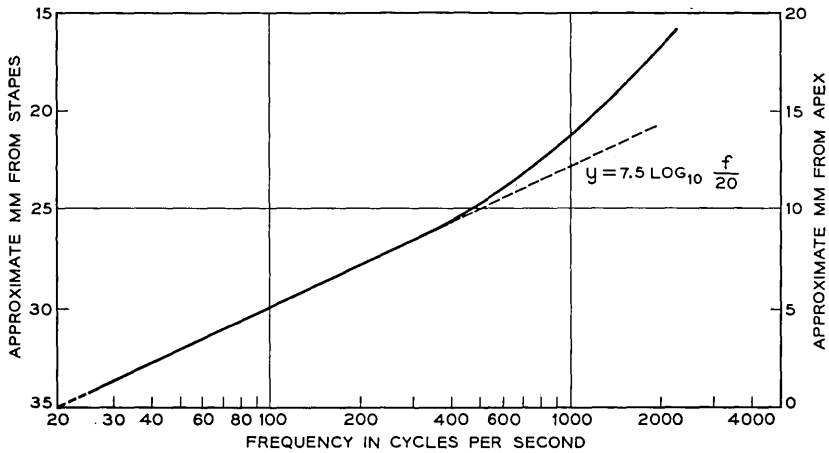


Fig. 11 — Location of peak displacement of basilar membrane as a function of frequency (after Bekesy³).

When the differentiation is carried out the result is:

$$\frac{\partial f_3''}{\partial \beta} = \beta^r [\beta(t - T)]^2 e^{-\beta(t-T)/1.7} \left[\beta t \cos \beta(t - T) + \left(1 + r + \frac{2\beta t}{\beta(t - T)} - \frac{\beta t}{1.7} \right) \sin \beta(t - T) \right]; \quad t \geq T. \tag{23}$$

As indicated earlier, the functional relation between the frequency β and the distance along the membrane is needed to put (23) into the form of a space derivative. Bekesy³ gives data on the place of maximal displacement along the membrane as a function of frequency. These data are replotted in Fig. 11. For purposes of the present discussion, the data for frequencies less than about 1000 cps are of main interest. If the basilar membrane is assumed to be 35 mm long, and if distance is now reckoned from the *apical* end, the low-frequency data are reasonably well approximated by:

$$x = 7.5 \log_{10} \frac{\beta}{40\pi}, \tag{24}$$

where x is the distance from the apex in mm. This line is drawn in Fig. 11.

It is now easy to compute

$$\frac{\partial f_3''}{\partial x} = \frac{\partial f_3''}{\partial \beta} \cdot \frac{\partial \beta}{\partial x},$$

where

$$\begin{aligned} \frac{\partial \beta}{\partial x} &= \frac{\beta}{7.5 \log_{10} e}, \\ &= 0.31\beta. \end{aligned} \tag{25}$$

Applying this result to (23) yields:

$$\begin{aligned} \frac{\partial f_3''}{\partial x} &= 0.3\beta^{1+r} [\beta(t - T)]^2 e^{-\beta(t-T)/1.7} \left[\beta t \cos \beta(t - T) + \left(1 + r + \frac{2\beta t}{\beta(t - T)} - \frac{\beta t}{1.7} \right) \sin \beta(t - T) \right]. \end{aligned} \tag{26}$$

Except for the constant amplitude factor, this is the spatial derivative of the impulse response for apical membrane points. It is plotted in Fig. 12. One notices its form is not radically different from the displacement.

The time derivative follows directly from (21), and is:

$$\frac{\partial f_3''}{\partial t} = \beta^{1+r} \frac{\partial g}{\partial t} = \beta^{2+r} \frac{\partial g}{\partial \beta t},$$

or,

$$\frac{\partial f_3}{\partial t} = \beta^{2+r}[\beta(t - T)]e^{-\beta(t-T)/1.7} \left[\beta(t - T) \cos \beta(t - T) + \left(2 - \frac{\beta(t - T)}{1.7} \right) \sin \beta(t - T) \right]; \quad t \geq T. \quad (27)$$

This function, except for amplitude factor, is the time derivative of the apical impulse response. It is plotted in Fig. 13. One notices that for a given (apical) point on the membrane, the time derivative of displacement is not greatly different in form from the spatial derivative. As men-

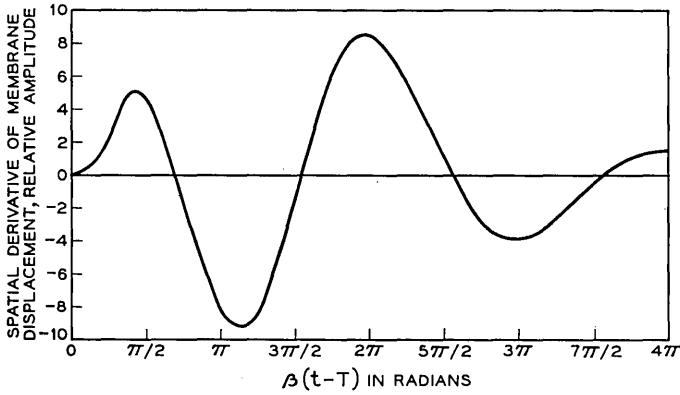


Fig. 12 — First spatial derivative of membrane displacement.

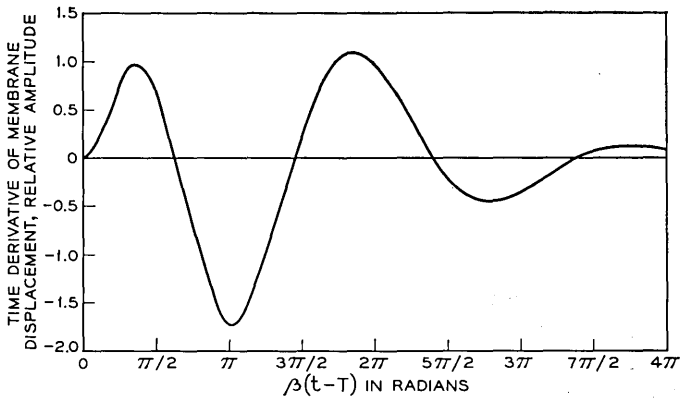


Fig. 13 — First time derivative of membrane displacement.

tioned earlier, the derivatives would have the same form if all frequency components propagated at the same velocity.

It also is of interest to consider the frequency-domain correlate of the spatial derivative. In this case it is equally easy to begin in a general way and not initially restrict the discussion to the apical region. For the model $H_1(s)$ the impulse response can be written in terms of its Laplace transform:

$$h(t) = \frac{1}{2\pi j} \int_{-j\omega}^{j\omega} H_1(s) e^{st} ds,$$

where

$$H_1(s) = G(s)F_1(s), \tag{28}$$

and where $G(s)$ and $F_1(s)$, the latter a function of the point parameter β , have been specified previously. The spatial derivative, in terms of the frequency parameter β , is therefore:

$$\frac{\partial h}{\partial \beta} = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} G(s) \frac{\partial F_1(s, \beta)}{\partial \beta} e^{st} ds. \tag{29}$$

The quantity of interest is $\partial F_1/\partial \beta$. From the previous discussion:

$$F_1(s, \beta) = c_1 \beta^{4+r} \left(\frac{2000\pi}{\beta + 2000\pi} \right)^r \left(\frac{s + \epsilon}{s + \gamma} \right) \left[\frac{1}{(s + \alpha)^2 + \beta^2} \right]^2 e^{-\frac{3\pi s}{4\beta}}. \tag{30}$$

Taking, as earlier indicated, $\beta = 2\alpha$, $\gamma = \beta$, and $\epsilon = 0$, and carrying through the differentiation gives:

$$\begin{aligned} \frac{\partial F_1}{\partial \beta} = F_1(s, \beta) & \left[\frac{4+r}{\beta} - \frac{r}{(\beta + 2000\pi)} \right. \\ & \left. + \frac{3\pi s}{4\beta^2} - \frac{2(s + 2.5\beta)}{(s + 0.5\beta)^2 + \beta^2} - \frac{1}{s + \beta} \right]. \end{aligned} \tag{31}$$

If we consider the result for real frequencies (i.e., $s = j\omega$) and normalize frequency by letting $\zeta = \omega/\beta$, then (31) becomes:

$$\begin{aligned} \frac{\partial F_1}{\partial \beta} = F_1(\zeta, \beta) & \frac{1}{\beta} \left[\frac{[4\beta + 2000\pi(4+r)]}{(\beta + 2000\pi)} \right. \\ & \left. + j\zeta \frac{3\pi}{4} - \frac{2(2.5 + j\zeta)}{(1.25 - \zeta^2 + j\zeta)} - \frac{1}{1 + j\zeta} \right]. \end{aligned} \tag{32}$$

This can be put in terms of the spatial derivative (at least for apical points) by applying (25). If this is done, then the spatial derivative be-

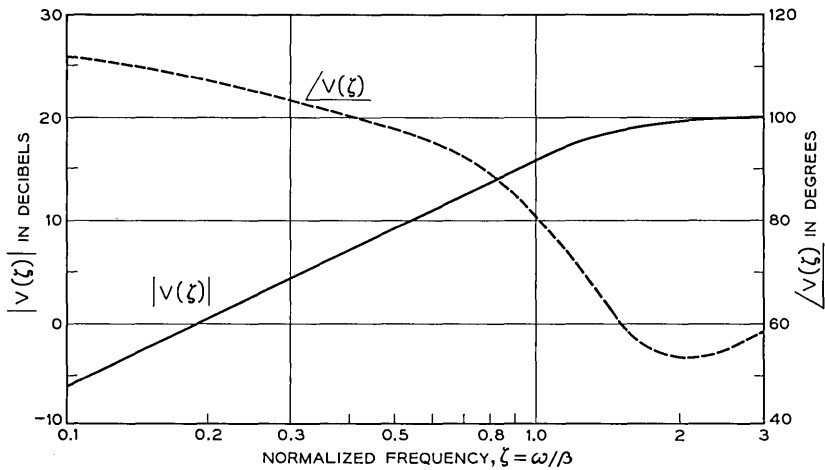


Fig. 14 — Frequency domain correlate, $V(\zeta)$, of the first spatial derivative.

comes:

$$\begin{aligned} \frac{\partial F_1}{\partial x} &= (0.3)F_1(\zeta, \beta) \left[4 + \left(\frac{2000\pi r}{\beta + 2000\pi} \right) \right. \\ &\quad \left. + j\zeta \frac{3\pi}{4} - \frac{2(2.5 + j\zeta)}{(1.25 - \zeta^2 + j\zeta)} - \left(\frac{1}{1 + j\zeta} \right) \right] \quad (33) \\ &= 0.3F_1(\zeta, \beta)V(\zeta). \end{aligned}$$

For points lying in the apical half of the membrane, therefore, (i.e., $\beta < 2000\pi$) the frequency-domain representation of the spatial derivative is simply $F_1(\zeta)$ (see Fig. 3) multiplied by the bracketed factor $V(\zeta)$ in (33). The phase and amplitude of this factor for $\beta \ll 2000\pi$ are plotted in Fig. 14.

One notices that for $\zeta < 1$ the bracketed term, to a crude approximation, is similar to a time differentiation. That is, the amplitude variation is +6 db/octave and the phase is $+\pi/2$. This indication is consonant with the previous time-domain results shown in Figs. 12 and 13.

1.4 An Electrical Circuit for Simulating Basilar Membrane Displacement*

On the basis of the relations developed in the previous sections [Eqs. (6) and (30), for example], it is possible to construct electrical circuits

* The material in this section was presented orally before the Sixty-Second Meeting of the Acoustical Society of America, Cincinnati, Ohio, November, 1961. The abstract appears in *J. Acoust. Soc. Am.*, **33**, 1961, p. 1670.

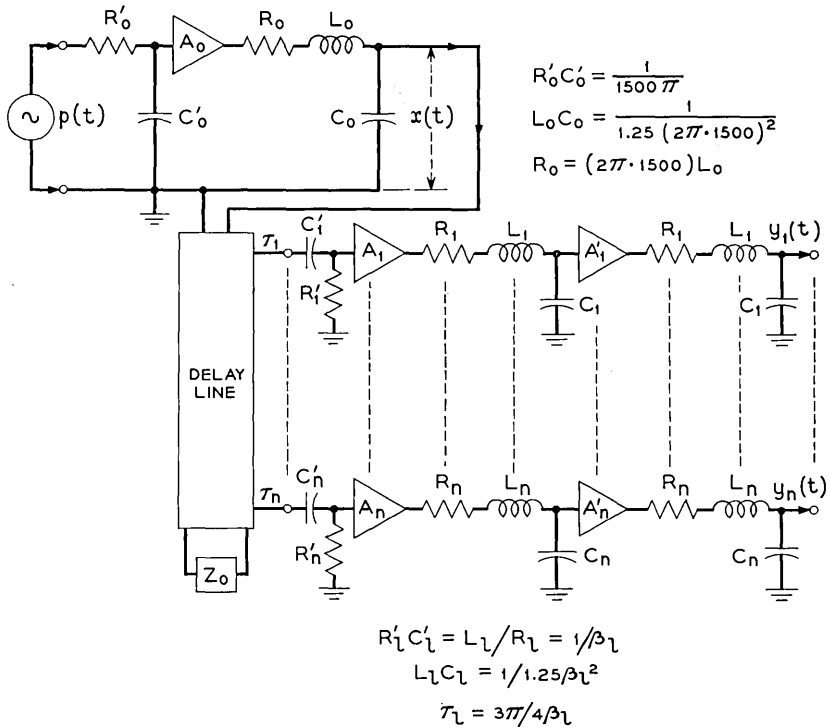


Fig. 15 — Electrical network representation of the model $[G(s)F_1(s)]$.

whose transmission properties are identical to those of the functions $G(s)$ and $F_1(s)$. This is most easily done by representing the critical frequencies in terms of simple cascaded resonant circuits. The additional phase delay can be supplied by means of an electrical delay line. A simulation of $G(s)$ as given in (6) and $F_1(s)$ as given in (30) (for $\epsilon = 0$) is shown in Fig. 15. The voltage at an individual output tap represents the membrane displacement at a specified distance from the stapes. The electrical voltages analogous to the sound pressure at the eardrum and to the stapes displacement are also indicated. The buffer amplifiers labelled A have fixed gains which take account of the proper multiplicative amplitude constants.

The circuit elements are selected according to the relations stated for $G(s)$ and $F_1(s)$. [See Eqs. (3) and (7).] For example, the procedure can be as follows. For the middle-ear simulation, choose a convenient R_o' , say 10K. Then, because $b = 2a = 2\pi \cdot 1500$, and because $a = 1/R_o' C_o'$,

$$C_o' = 0.02 \mu f.$$

Select a convenient L_0 , say 2h. Then, because $1.25b^2 = 1/L_0C_0$ and $a = R_0/2L_0$,

$$C_0 = 0.005 \mu f,$$

and

$$R_0 = 19K.$$

The components for the basilar membrane networks are chosen in the same way. In this case:

$$\beta_i = \frac{1}{R_i' C_i'}, \quad 1.25\beta_i^2 = \frac{1}{L_i C_i},$$

and

$$\alpha_i = \frac{R_i}{2L_i}.$$

Consider, for example, the membrane point which responds maximally to 4500 cps (i.e., $\beta_i = 2\pi \cdot 4500$). For convenience take R_i' and L_i as 10K and 1h, respectively. Then:

$$C_i' = 0.0035 \mu f$$

$$C_i = 0.001 \mu f$$

$$R_i = 28K.$$

For each membrane point the relative gains of the amplifiers are set to satisfy the amplitude relations implied in Fig. 9. This takes account of the constant multiplying factors in the model specification.

Some representative impulse responses of the analog circuit of Fig. 15 are shown in Fig. 16(a). One notices the degradation in time resolution as the response is viewed at points more apicalward.

As indicated earlier, the spatial derivative may figure in the conversion of mechanical to neural activity. In previous psychoacoustic work⁶ it was found useful to approximate the first spatial derivative by a finite difference. The present circuit can be used to provide such an approximation by taking the differences between the deflections of adjacent, uniformly spaced points. Fig. 16(b) shows first-difference approximations to the spatial derivative obtained from the analog circuit by taking:

$$\frac{\partial y}{\partial x} \approx \frac{y(t, x + \Delta x) - y(t, x)}{\Delta x},$$

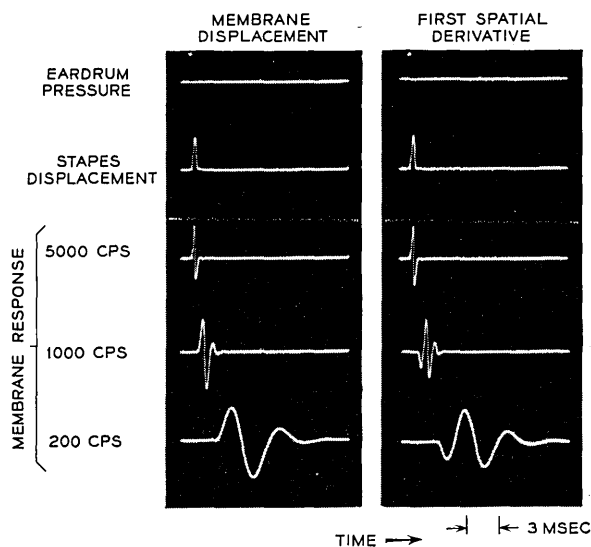


Fig. 16 — (a) Impulse responses measured on the network of Fig. 15; (b) first difference approximations to the spatial derivative measured from the network of Fig. 15.

with $\Delta x = 0.3$ mm. These responses can be compared (for apical points) with the calculated derivative in Fig. 12. Because of amplification, the polarity of the derivative traces in Fig. 16(b) is inverted from that shown in Fig. 12.

II. SOME RELATIONS BETWEEN SUBJECTIVE AND PHYSIOLOGICAL BEHAVIOR

The preceding discussion derived computationally tractable models for the operation of the middle ear and basilar membrane. Can these models be used to further our understanding of auditory subjective behavior? In particular can they help us to relate psychoacoustic phenomena to the physical operation of the peripheral ear?

The models describe only the mechanical functioning of the ear. Any comprehensive hypothesis about auditory perception must make provisions for the transduction of mechanical displacement into neural activity. The details of this process are not well understood and the assumptions that presently can be made must be of an approximate and simplified nature. Three such assumptions will be useful to us. Although

gross simplifications, they do not seem to violate known physiological facts.

The first assumption (actually a fact) is that sufficient local deformation of the basilar membrane elicits neural activity in the terminations of the auditory nerve at the organ of Corti. Such neural activity may be in the form of volleys triggered synchronously with the stimulus, or in the form of a signaling of place localization of displacement. Implicit in this is the notion that the displacement, or perhaps spatial derivatives of displacement,⁵ must exceed a certain threshold before nerve firings take place.* The number of neurons activated depends upon amplitude of membrane displacement in a monotonic fashion. Psychological and physiological evidence suggests that the intensity of the neural activity is a power-law function of the mechanical displacement. A single neuron is presumably a binary (fired or unfired) device. It is refractory for a given period after firing; hence a limit exists upon the rate at which it can fire. Large populations of neurons, all of which are not refractory at the same time, can give rise to neural volleys at rates greater than the maximum rate of a single element.

Second, neural firings occur on only one "polarity" of the displacement, or of the spatial derivative.⁷ In other words, some process like half-wave rectification operates on the displacement function, or on its spatial

* Earlier, in Section 1.3, it was suggested that the spatial derivative of displacement, as well as the displacement, may be important in the mechanical-to-neural conversion process. Further explication of this allusion and the present one is necessary.

Electrophysiological experiments on guinea pig [G. von Békésy, *J. Acoust. Soc. Am.*, **25** (1953) p. 786; H. Davis, *Ann. Oto. Rhin. Laryn.* **67** (1958) p. 789.] suggest that the outer and inner hair cells of the organ of Corti differ in their sensitivities to mechanical stimulation. The outer hair cells are sensitive to bending only in a direction transverse to the long dimension of the membrane. Further than this, only outward bending of the hairs (away from arch of Corti) produces an electrical potential in the scala media favorable for exciting the auditory nerve endings. This outward bending is produced on upward motions of the basilar membrane — that is, motions which drive it toward the tectorial membrane and produce a relative shear.

On the other hand, the inner hair cells, which reside between the arch of Corti and the axis of the cochlear spiral, are sensitive to bending in a direction parallel to the long dimension of the membrane. In this case only bending toward the apex of the cochlea produces a scala media potential favorable for stimulating the nerve. So far as a given point on the membrane is concerned, the inner hair cells are essentially sensitive to the longitudinal gradient of displacement — that is, to the spatial derivative in the long dimension. Furthermore, the inner cells fire only on that polarity of the gradient which corresponds to bending toward the apex. Threshold for firing of the inner cells apparently is about 20 db higher than that for the outer cells.

If this behavior is common to the human ear, displacement gradient, as well as displacement may be significant. As the results of Section 1.3 show, the displacement and the spatial derivative have gross features which are similar. For this reason the hypotheses and arguments to be put forward in this section generally can apply equally to displacement and gradient.

derivatives. Third, the membrane point displacing with the greatest amplitude originates the predominant neural activity. (More strictly, perhaps, this is the point experiencing the greatest transverse and longitudinal bending.) The latter may also operate to suppress or inhibit activity arising from neighboring points.

These assumptions, along with the results from the models, have in a number of instances been helpful in interpreting subjective auditory behavior. Without going into any case in great depth, let us consider several of these instances.

2.1 *Pitch Perception*

Pitch is that subjective attribute which admits of a rank ordering on a scale ranging from low to high. As such, it correlates strongly with objective measures of frequency. One important facet of auditory perception is the ability to assign pitch to sounds which exhibit time periodicity.

Consider first the pitch of pure (sinusoidal) tones. For such stimuli the basilar membrane displacement at any point is sinusoidal. The frequency responses given previously in Fig. 9 indicate the relative amplitudes of displacement versus frequency for different membrane points. At any given frequency, one point on the membrane responds with greater amplitude than all others. In accordance with the previous assumptions, the most numerous neural volleys are elicited at this maximum point. For frequencies sufficiently low (less than about 1000 cps) they are triggered once per cycle and at some fixed epoch on the displacement waveform. Subsequent processing by higher centers presumably appreciates the periodicity of the stimulus-locked volleys. For frequencies greater than about 1000 to 2000 cps, electro-physiological evidence suggests that synchrony of neural firings is not maintained.⁹ Pitch is apparently perceived through a signaling of the place of greatest membrane displacement or displacement gradient. The poorer frequency resolution of points lying in the basal part of the basilar membrane probably also contributes to the psychoacoustic fact that pitch discrimination becomes less acute at higher frequencies.¹⁰

Suppose the periodic sound stimulus is not a simple sinusoidal tone but is more complex, say repeated sharp pulses. What pitch is heard? For purpose of illustration, imagine the stimulus to be alternately positive and negative periodic impulses. Such a pulse train has a spectrum which is odd-harmonic. Pulse rate and fundamental frequency are in the ratio of two-to-one. If the pulses occur slowly enough, the membrane

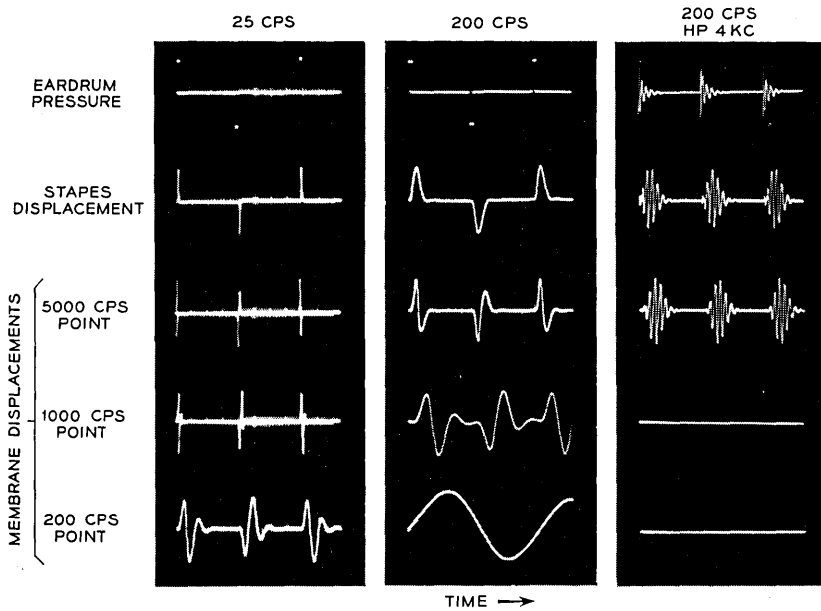


Fig. 17 — Displacement responses for alternate positive and negative pulses simulated by the network of Fig. 15.

displacements at all points along its length will resolve each pulse in time. That is, the membrane will have time to execute a complete, damped impulse response at all points for each pulse, positive or negative. If, however, the fundamental frequency of the train is sufficiently high, the fundamental component will be resolved (in frequency) at the most apically responding point. This situation is illustrated by the traces in the first and second columns of Fig. 17. These waveforms were measured on analog networks as illustrated in Fig. 15. The oscilloscope gain was adjusted for constant peak-to-peak amplitude to display the waveforms more effectively. The proper relative amplitudes are therefore not indicated in the traces.

For the low pulse-rate condition (25 cps fundamental) in the first column, one might imagine that neural firing synchronous with each pulse, regardless of polarity, would be triggered at all points along the membrane. The perceived pitch might be expected to be that of the pulse rate, and it is.⁶ For such stimulation, the models indicate that the greatest membrane displacements occur near the middle portion, in the region maximally responsive to 1000 to 2000 cps.

In the second column, the fundamental frequency is 200 cps. This is

high enough for the apical end of the membrane to resolve the fundamental component. The displacement of the 200 cps point on the membrane is the fundamental sinusoid, while the more basal points continue to resolve each pulse in time. At the apical end, neural volleys might be expected to be triggered synchronously at the fundamental frequency, while toward the basal end the displacements favor firings at the pulse rate. For this condition, the apical fundamental-correlated displacements are generally of greater amplitude and subjectively more significant than the basal, pulse-rate displacements. The fundamental-rate volleys generally predominate, and a pitch is heard corresponding to 200 sec^{-1} .

If this same stimulus is high-pass filtered at a sufficiently high frequency, only the basal displacements remain effective in producing the pitch percept. If the present arguments continue to hold, this filtering should again give rise to a pulse-rate pitch because the time resolution in the basal end separates each pulse, whether positive or negative. Psychoacoustic measurements show this in fact to be the case.¹¹ Representative membrane displacements for this condition, as given by the models, are shown in the third column of Fig. 17.

A slightly more subtle effect is obtained if the high-pass filtering is made at a low harmonic number, for example, at the second harmonic so as to remove only the fundamental component. Under certain of these conditions the significant membrane displacement can be seen to exhibit displacements that favor fundamental-rate neural activity. The pitch percept would then be expected to be the fundamental, even though the fundamental is not present in the stimulus. Again psychoacoustic measurements give this result.¹² The effect is the so-called residue pitch.

Another of the many variations of pulse stimuli, but one which is diagnostically useful in exploring pitch perception, is the periodic, unipolar impulse train in which the equispaced pulses have amplitudes (areas) alternately a and b . Such a stimulus exhibits an infinite number of complex spectral zeros, the imaginary parts of which occur at every other spectral line. The spectral envelope is cycloidal and is described by:

$$K(s) = (a + be^{-sT/2}), \quad (34)$$

where T is the fundamental period. The spectral zeros lie at

$$s = -\frac{2}{T} \ln |a/b| \pm j \frac{2(2n+1)\pi}{T},$$

and the ratio of odd-line amplitude to even-line amplitude is:

$$R = \frac{|K(s)|_{s=j2\pi/T}}{|K(s)|_{s=j4\pi/T}} = \left[\frac{a-b}{a+b} \right]. \quad (35)$$

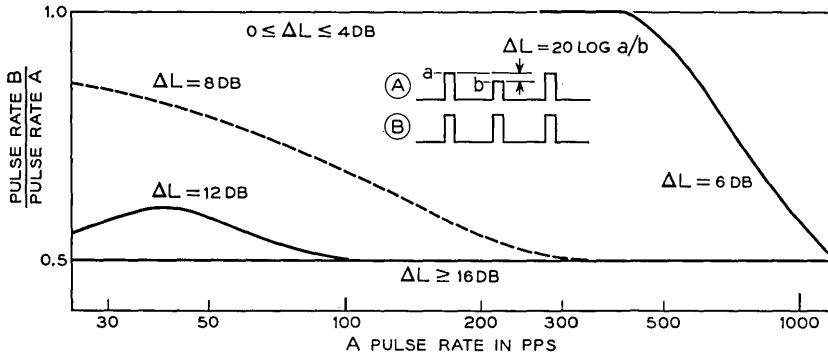


Fig. 18—Subjective pitch assigned to a periodic pulse train composed of sharp pulses with alternate amplitudes a and b . $\Delta L = 20 \log a/b$. The pitch equation is determined by matching the uniform pattern B to the test pattern A.

One psychophysical question that could be posed is "For a given sensation level, fundamental frequency and a/b ratio, what is the pitch?" When this question is answered by means of a pitch-matching experiment, the result for several values of the variables is shown in Fig. 18. These results are for a sensation level of approximately 45 db. When the a/b ratio is less than about 4 db, one never hears any pitch except the pulse rate. On the other hand, when the a/b ratio is greater than about 16 db, one never hears any pitch but one-half the pulse rate, i.e., the fundamental. Between these level differences, a transition from one pitch mode to the other takes place. The transition depends upon fundamental frequency as shown in the figure. As in the previous case, calculations and observations with the analog networks show the correlation between these modes and the displacement patterns of the basilar membrane. Unlike the situation depicted previously in Fig. 17(b), however, a pitch percept equivalent to half the pulse rate does not necessarily mean that the fundamental frequency is resolved by the membrane.

A somewhat different example of pitch stimulus is periodically interrupted random noise. Under certain conditions of interruption rate, duty factor and frequency content, chopped noise possesses a pitch.¹³ It is relevant to consider how such a signal is represented in the mechanical displacements of the basilar membrane.

Because of the ear's frequency characteristics, a broad-spectrum noise would be expected to produce the greatest displacements somewhere near the middle of the membrane, around the 1000 to 2000 cps point. Let us look at these displacements for a flat-spectrum noise which is chopped with constant duty factor of 0.2. The response waveforms for

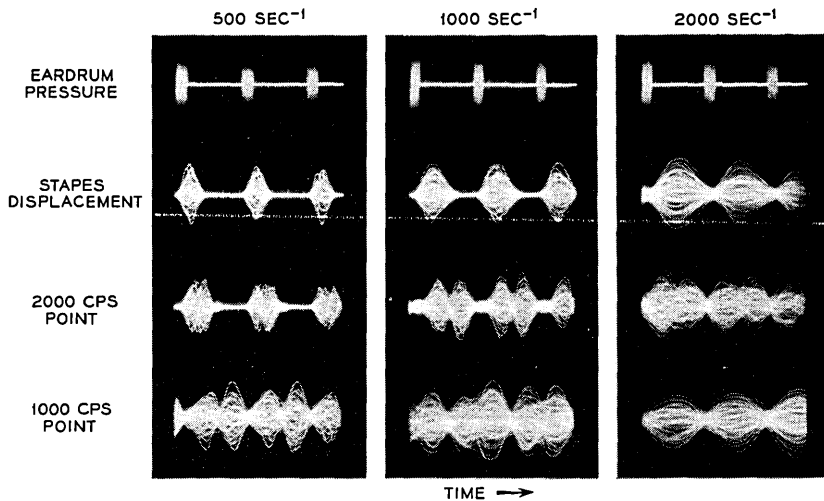


Fig. 19 — Displacement responses simulated by the network of Fig. 15 for periodically interrupted noise. Constant duty factor = 0.2.

several interruption rates are given in Fig. 19. For the slow rate, 500 sec^{-1} , the noise bursts are well resolved in time. Nerve volleys synchronous with the onset of the noise bursts might be expected for this stimulation. As the interruption rate is increased to upwards of 2000 sec^{-1} , however, neither the stapes nor membrane displacements resolve each burst separately in time. Stimulus-locked synchrony of the neural activity might be expected to be impaired or lost, even if the neural volleys could be elicited at this rate. Psychoacoustic observations bear this out. They show that it is difficult to ascribe a pitch to interrupted noise for rates greater than about 1000 sec^{-1} even under favorable conditions of low duty factor. It is not clear how much of this limit is determined by neural resolution, and how much by mechanical. Very likely both factors contribute to the resultant behavior.

2.2 Binaural Lateralization

Another aspect of perception is binaural lateralization. This is the subjective ability to locate a sound image at a particular point inside the head. The phenomenon is conventionally observed in earphone listening. If identical clicks (impulses of sound pressure) are produced simultaneously at the two ears, the average listener hears the sound image to be located in the center of his head. If one click is produced a little earlier (or with slightly greater intensity) than the other, the sound image shifts

toward the earlier (or more intense) ear. This shift continues with increasing time difference until the image moves entirely to one side and eventually breaks apart. One then hears individual clicks located at the ears.

Naively we suppose the subjective position of the image to be determined by some sort of computation of coincidence between neural volleys. The volleys originate at the periphery and travel to higher centers via synaptic connections. The volley initiated earliest progresses to a point in the neural net where a coincidence occurs with the later volley to produce a subjective image appropriately off center. To the extent that intensity differences can shift the image position, intensity probably is coded, at least partially, in terms of the volley timing. As was the case in pitch perception, there are several areas in binaural phenomena where the ear models have been helpful in suggesting explanations of, and correlations with, subjective behavior. One of these areas is the effects of phase upon the binaural lateralization of clicks.

Suppose one produces impulses of pressure at the two ears, of equal intensity but of opposite polarity (i.e., one a rarefaction and the other a condensation). How would a listener adjust the times of occurrence of such pulses so that he hears the sound image exactly in the center of head? Let us consider what the displacement waveforms and the mechanical-to-neural conversion hypotheses would predict.

An impulse of pressure rarefaction draws the eardrum and stapes initially outward and causes the membrane displacement to be initially upward. A condensation pulse, on the other hand, causes an initially inward displacement of drum and stapes and consequently an initially downward movement of the membrane. At any given point on the membrane the waveforms of displacement produced by these two stimuli differ only in sign; that is, one is the negative of the other. Typical displacements of apical and basal points caused by rarefaction and condensation pulses are shown in Fig. 20. (These traces are essentially the impulse responses calculated previously in Fig. 7.)

The top diagram in Fig. 20 is illustrative of the displacement response of points lying in the apical (low-frequency) half of the membrane. The solid curve is the displacement for a rarefaction pulse, the dashed for a condensation. The abscissa at the top is in terms of the product βt , where β is the radian frequency of maximal response for the particular apical point. The lower abscissa on the top graph shows time scales appropriate to the specific points maximally responsive to 1200 and 600 cps, respectively. The lower graph shows the displacement appropriate to points lying in the basal (high-frequency) half of the membrane. As discussed

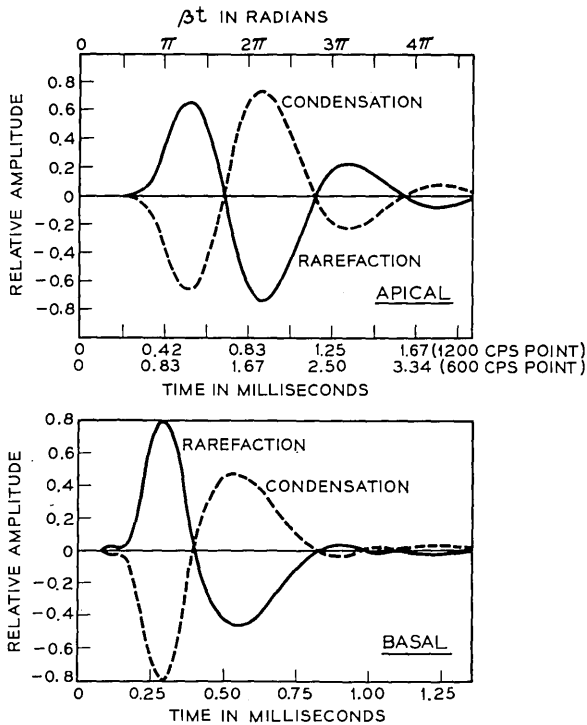


Fig. 20 — Apical and basal displacements of the basilar membrane for rarefaction and condensation pressure impulses at the eardrum. These responses from those computed in Fig. 7 for $[G(s)F_1(s)]$.

earlier, this waveform has essentially the same shape and time scale for all basal points.

Following our earlier assumptions, we suppose that neural firings (at least of the more sensitive outer hair cells) take place at some amplitude level on the upward deflections of the membrane. The curves suggest, therefore, that a time difference should exist between the firings for a rarefaction pulse and those for a simultaneous condensation pulse. The difference should be about one-half cycle on the displacement waveforms. The earlier results indicate that for broad-spectrum excitation the greatest deflections occur near the middle of the membrane, in the vicinity of the region maximally responsive to 1000 to 2000 cps. For such a place, the half-cycle intervals are of the order of 250 to 500 μ sec. The time scale for the 1200 cps point in the top graph is indicative of this magnitude.

Assuming that simultaneity of neural firings at the two ears produces a central sound image, a rarefaction pulse and a condensation pulse should produce a centered image if the condensation is advanced in time to bring its positive displacement peak approximately into coincidence with that of the rarefaction. This means advancing the condensation, or letting it lead, by about 250 to 500 μsec . Furthermore, the periodic nature of the displacements suggests that multiple fusions of the sound image might occur by virtue of neural firings triggered at secondary positive excursions. These should occur for interaural times which are full-cycle increments of the principal fusion, including lead and lag shifts. A half-cycle lead of the condensation would represent a principal fusion; a half-cycle lag would be a secondary fusion.

If cophasic pulses are delivered to the two ears, that is, rarefaction-rarefaction or condensation-condensation, the same argument says that the principal fusion should obtain for zero interaural time difference, and secondary fusions for full-cycle shifts, either lead or lag.

The preceding remarks relate to broadband, unmasked pulses, where the neural response is likely to originate near the central portion of the basilar membrane. Suppose the dominant response were elicited from some other place on the membrane. The interaural time difference for lateralization ought to change in accordance with what the curves in Fig. 20 imply. Band-filtering of the pulse stimuli is an obvious means for confining membrane activity to specific regions. This has the disadvantage, however, that the stimulus signal is contaminated with the impulse response of the filter, so that it is inconvenient, if not difficult, to analyze the membrane displacement. The objective can be achieved more conveniently by selectively masking the membrane response with filtered random noise. The significant neural information can then be originated in a normally less responsive region by obscuring the maximally responding place with noise.

The top graph in Fig. 20 suggests that if the disparity between the interaural times for lateralizing cophasic and antiphase pulses is to be increased, the significant response must originate from places more apicalward, that is, at points which respond maximally at lower frequencies. In such a case high-pass (HP) noise should be used to obscure activity in the basal part of the membrane. Low-pass (LP) noise, on the other hand, causes the coherent information to arise from the basal section. Here, because of the nature of the pulse response, the disparity between cophasic and antiphase fusions is predicted to be roughly constant with place along the membrane, and should be of the order of 250 μsec . This, too, ought to be the minimum interaural disparity that can be produced for the antiphase situation.

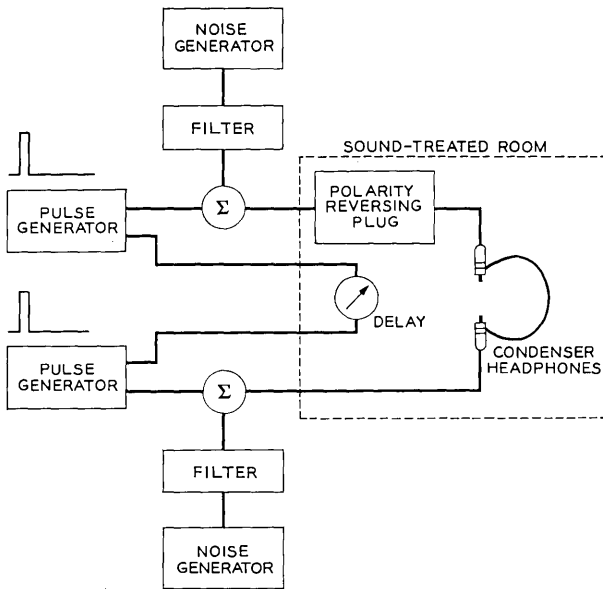


Fig. 21 — Equipment arrangement for binaural lateralization of cophasic and antiphasic clicks.

An experiment was performed to determine whether the predicted phenomena are in fact manifest.¹⁴ The arrangement to measure the effects is shown in Fig. 21. Twin pulse generators produced identical 0.1-msec rectangular pulses in separate channels at a sensation level of 40 db. The repetition rate of the pulses was 10 sec^{-1} . Random noise from two uncorrelated generators was filtered by identical filters and added to the signal channels. This noise level completely masked the selected portions of the pulse spectra. HP and LP noise cutoff frequencies of 600, 1200, and 2400 cps were used in addition to no masking. Condenser microphones fitted with ear-insert plugs were used as earphones to provide good transduction of the pulse signals. The subject was provided a delay control which permitted continuous adjustment of the time of occurrence of one pulse relative to the other over the range $\pm 5 \text{ msec}$. A switch could reverse the polarity of the pulse delivered to one earphone.

The results of this experiment for three listeners are summarized in terms of median responses in Fig. 22. For HP masking, Fig. 22(a), the interaural time for the principal antiphase lateralizations is seen to increase as the cutoff frequency of the HP noise is lowered. For these conditions the maximally responding unmasked place on the membrane

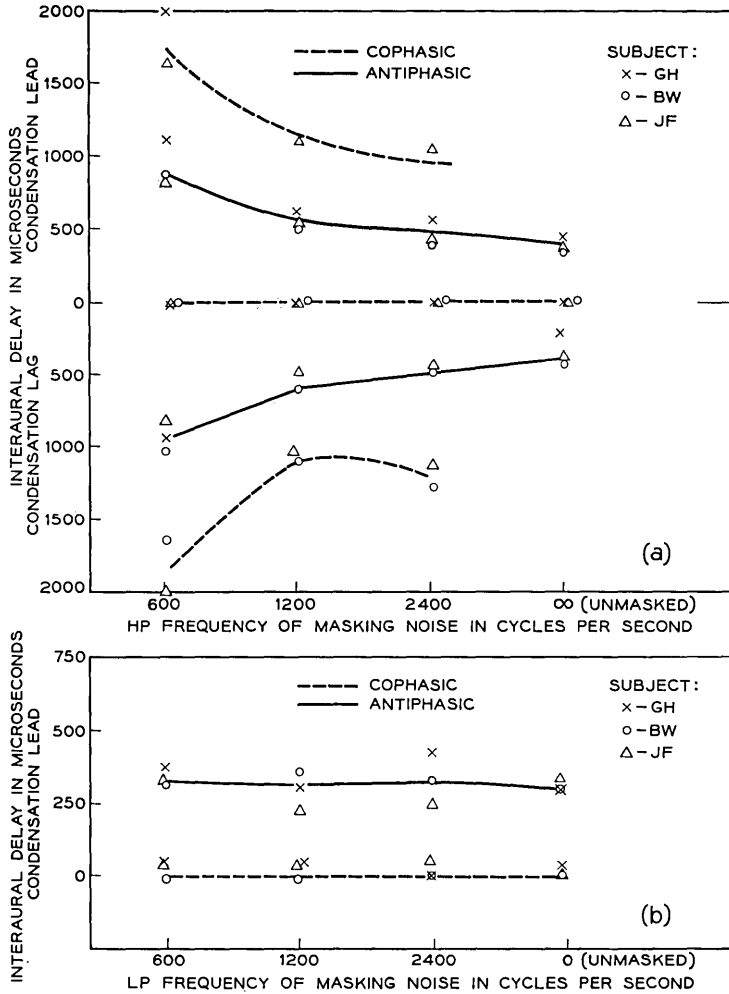


Fig. 22 — Effects of masking upon the lateralization of cophasic and antiphasic pulses.

should be that just below the cutoff frequency, f_c , of the filter. The interaural time for the antiphase fusion ought then to be about $\pm 1/2f_c$. The data follow this value reasonably well.

The secondary antiphase fusions (condensation lag) are roughly a reflection of the principal ones in the x -axis. The time separation between the principal and secondary points is approximately the predicted full-

cycle shift, or about $1/f_c$. The principal cophasic fusions fall along the axis for zero interaural times, and the secondary cophasic fusions (highest and lowest curves) fall at about the right value for the full-cycle shift.

The results for the LP masking, Fig. 22(b), indicate that the hypotheses about fusions of basal-end information are essentially sustained. The cophasic-antiphase disparity is roughly constant at about 300 μ sec. Secondary images, however, are not easily heard because the LP noise is a more potent out-of-band masker. Both masking conditions make it clear that the significant neural timing information can be made to originate from different points along the membrane. Further, the neural timing is intimately related to the individual mechanical excursions of the membrane at the significant point.

Some electrophysiological evidence also exists to support these psychoacoustic results and the assumptions made earlier. Peake¹⁵ measured the latency of the gross neural component, N_1 , in cat's ear for stimulation by rarefaction and condensation pulses. For moderately high signal levels, the difference in latencies is found to be of the order of 200 to 300 μ sec with condensation pulses giving the greater latencies. In addition, very recent data by Kiang⁷ on the activity of single, peripheral nerve units suggest that the firings are synchronized with the individual unipolar displacements of the membrane, as conjectured here.

2.3 *Time-Intensity Trade*

In other binaural experimentation it has been observed that the position of a sound image can be maintained stationary by trading relative intensity against relative time of occurrence of pulses at the two ears. That is, the movement of the sound image towards the ear receiving a leading click can be offset by an increase in intensity of the pulse at the lagging ear. To a certain extent such a trading relation is implied in the calculated membrane responses and in the simple hypotheses about conversion of mechanical to neural activity. It is worth considering the degree to which the experimentally observed trade can be explained by the membrane relations.

Consider again binaural excitation of the ears by short, unipolar pulses. The earlier assumptions about neural firings on upward displacements of the membrane, in excess of a fixed threshold, imply a time-intensity trade. For ease of illustration, consider that the form of the impulse response of the membrane for middle to apical points is essentially specified by the model $f_3(t)$ given in (5). For simplicity this can be written

without constants as

$$f_3(\theta) = (\theta)^2 e^{-\theta/1.7} \sin \theta, \quad (36)$$

where $\theta = \beta t$. The waveform of this function has already been plotted in Fig. 4.

Imagine that the amplitude of a rarefaction stimulus has been set to threshold value so that neural firings are produced exactly at the first positive crest of the displacement. Now if the intensity of the stimulating pulse is steadily increased, the threshold level will be crossed at successively earlier times on the initial quarter cycle of the wave. In terms of the model, the advance in time of the threshold crossing is a simple function of the stimulus amplitude, and we can compare it with experimentally measured figures.

For reasons that will be obvious presently, we take

$$\ln f_3(\theta) = 2 \ln \theta - \frac{\theta}{1.7} + \ln \sin \theta. \quad (37)$$

Differentiating with respect to θ ,

$$\frac{d[\ln f_3(\theta)]}{d\theta} = \frac{2}{\theta} - \frac{1}{1.7} + \cot \theta \quad (38)$$

and taking the partial with respect to time,

$$\frac{\partial[\ln f_3(\theta)]}{\partial t} = \beta \left(\frac{2}{\theta} - \frac{1}{1.7} + \cot \theta \right). \quad (39)$$

Equation (39) gives, in effect, the time-intensity trade for the wave in terms of nepers per second, as a function of the epoch θ at which threshold is crossed. In psychoacoustic tests the trade has customarily been specified in terms of msec/db—that is, the number of milliseconds by which the stimulus in one ear must be advanced to offset a relative intensity increase of one db in the other ear. Equation (39) can be put in terms of msec/db by taking its reciprocal, and multiplying by $10^{-3}/8.7$. One often sees this trade plotted as a function of intensity or sensation level of the stimulus. Let us arbitrarily take the positive maximum, $f_3(\theta_{+\max})$, as the threshold level of displacement. An increase in intensity of X db will then cause the threshold to be crossed at an epoch, $0 \leq \theta \leq \theta_{+\max}$ that satisfies: $-8.7 \ln \frac{f_3(\theta)}{f_3(\theta_{+\max})} = X$ db. We can therefore plot (39) [converted to msec/db] versus (37) [converted to db *re* $f_3(\theta_{+\max})$] for common values of the threshold crossing θ . This function, for three different apical points on the membrane, is shown in Fig. 23(a).

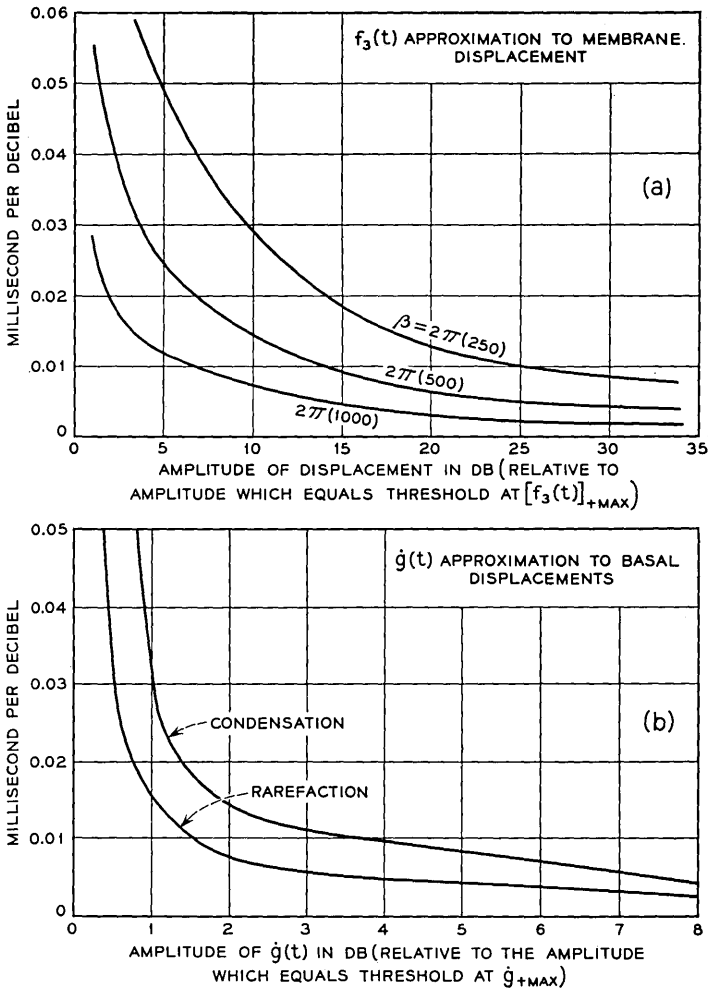


Fig. 23 — Time-intensity trade predicted for (a) apical points by membrane model $f_3(t)$; (b) for basal points by stapes derivative $\dot{g}(t)$.

The curves suggest that the trade of msec/db is greatest for low signal levels and diminishes for higher levels. It also indicates that the trade in msec/db is greater for points closer to the apex, that is, for lower β . Broadband pulse excitation of the model, as previously stated, produces greatest displacements near the middle of the membrane. The 1000-cps point is representative of this region. Low-level values of the trade for this point are on the order of 0.03 msec/db.

The earlier arguments also indicated that the impulse responses of basal points were similar to the time derivative of stapes displacement. A time-intensity trade computed from these responses ought to be suggestive of the minimum msec/db value that could be expected if the trade were based upon basal activity. We can use (11) and approximate the basal displacements by the stirrup derivative, $\dot{g}(t)$. Letting $(bt) = \Phi$,

$$\ln \dot{g}(\Phi) = -\frac{\Phi}{2} + \ln(2 \sin \Phi + \cos \Phi - 1). \quad (40)$$

And,

$$\frac{\partial t}{\partial \ln \dot{g}(\Phi)} = \frac{2(2 \sin \Phi + \cos \Phi - 1)}{b(\cos \Phi - 3 \sin \Phi + 1)}. \quad (41)$$

Again, expressing (40) in db relative to $\dot{g}(\Phi_{+\max})$ and (41) in msec/db, the two can be plotted for common values of Φ . When this is done the trading relation obtained is shown in Fig. 23(b). Because of the substantial asymmetry in $\dot{g}(\Phi)$, the function is computed for the initial quarter of the positive deflection and initial quarter of the negative deflection (that is, the initial positive deflection if the displacement phase were reversed). The former would be appropriate for rarefaction pulse excitation; the latter for condensation. These figures are, of course, susceptible of the uncertainty connected with the value b and the approximation of the basal displacements by $\dot{g}(\Phi)$. Nevertheless, the trading values thus obtained fall reasonably close to those for the

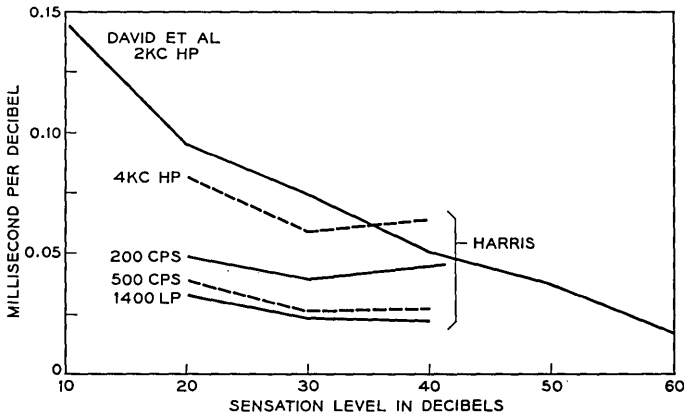


Fig. 24 — Measured time-intensity trade for several stimuli (after David et al¹⁶ and Harris¹⁷).

1000-cps point shown in Fig. 23(a). To the extent that $g(t)$ is a reasonable approximation of the basal displacements, the results indicate that the trade for condensation pulses should be slightly greater than those for rarefactions.

Some psychoacoustic data are available for comparison with these calculations. David, Guttman and van Bergeijk¹⁶ used 2-kc high-pass clicks to measure the trading function and obtained a median result shown in Fig. 24. Harris¹⁷ used both HP and LP filtered pulses and pure tones in a related investigation. Several of his results are also shown in Fig. 24. The subjective data for the HP pulses are clearly greater than the predictions from the model. The results for 1400 LP, however, are more nearly of the magnitude suggested by the computations. The previous computations also suggest that the msec/db trade should increase in magnitude as the significant neural information is elicited from more apical (low-frequency) points. Harris found, however, that for LP clicks with cutoffs between 200 to about 1000 cps, the trade was nearly constant at about 0.03 msec/db. One difficulty in comparing the computed and measured data is that we do not know how to equate values on the abscissae of Figs. 23 and 24. That is, we do not know what sensation level corresponds to the zero-db reference amplitude of the displacement wave. Only general directions and trends can therefore be legitimately compared.

Another difficulty in comparing the data is that the computed time-intensity trades assumed ideal impulse excitation of the ear. The experimental measurements, on the other hand, used pulses which were HP or LP filtered. The effect of the filter response upon that of the membrane is somewhat uncertain. The experimental determinations and the computations may not therefore be strictly comparable. To attempt to obviate this difficulty, we made some cursory measurements of the trade using the masking technique described earlier in Section 2.2. For a sensation level of 40 db, and with unmasked rarefaction clicks, one trained subject from the previous lateralization experiment made the $\Delta t - \Delta I$ swap plotted as the lower curve in Fig. 25. A binaural masking by 600-cps HP noise presumably constrains the coherent neural activity to come from a more apical point (somewhere near the 600-cps point). For such a masking the same subject made the trade indicated by the upper curve, giving values about twice as great as the unmasked trade. The slope of the unmasked function at the origin is approximately 0.03 msec/db. That for the 600 HP masking is about 0.05 msec/db.

Clearly these data for one subject are tentative and must be confirmed

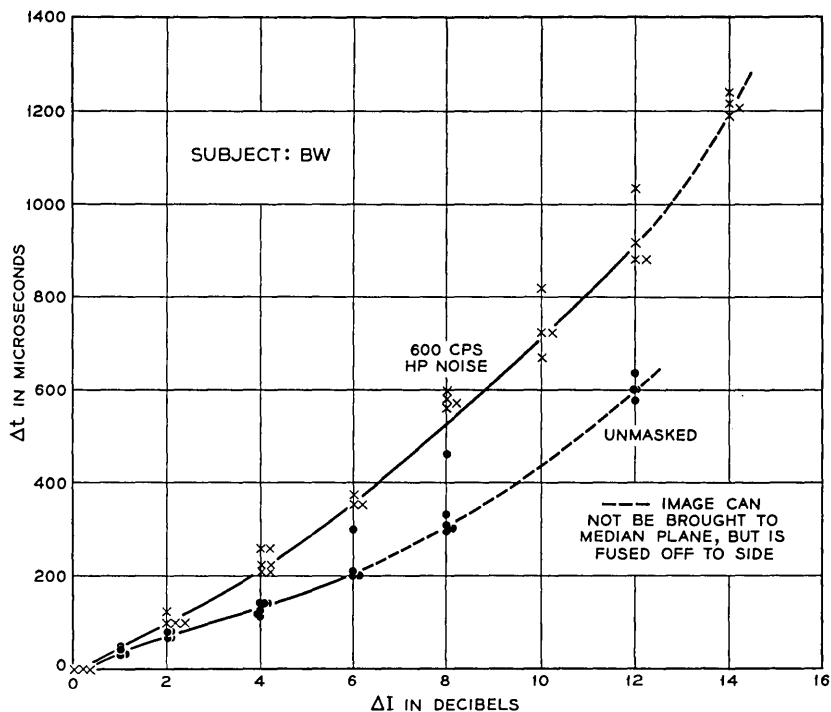


Fig. 25 — Effects of masking upon time-intensity trade for broadband, co-phasic pulses.

or refuted by additional experimentation. To the extent that they are correct, however, they support the general predictions of the model as to the frequency (or membrane place) dependence of the time-intensity trade. They do not agree well with absolute magnitudes of the computed trade, and there is still the question of how to equate abscissae. It is highly probable, too, that the time-intensity trade involves neural mechanisms not here included. Even so, the mechanical operations appear to go a long way in contributing to an explanation of the phenomenon. A time-intensity trade has also been observed at the neural level. In the cat's ear, Peake¹⁵ finds that an intensity change of about 40 db in a stimulating pressure click causes a reduction in the latency of the N_1 neural component by roughly 0.6 msec. A simple division of these figures gives 0.015 msec/db for the trade. This figure falls within the range predicted by the model for human hearing.

As a final point in this theme, the same arguments can be made for

pure tones. For such stimuli the membrane displacements are also sinusoidal (at least over a large intensity range) and have the form

$$f_{\beta}(t) = K_{\beta}(\omega) \sin \omega t, \tag{42}$$

where $K_{\beta}(\omega)$ is an amplitude versus frequency factor appropriate to the membrane point maximally responsive to radian frequency β ; $K_{\beta}(\omega)$ is largest, of course, for $\omega = \beta$. The previous argument gives

$$\ln f_{\beta}(t) = (\ln \sin \omega t + \ln K_{\beta}), \tag{43}$$

and,

$$\frac{d}{dt} (\ln f_{\beta}) = \omega(\cot \omega t). \tag{44}$$

A plot of this last result in terms of msec/db versus amplitude in db (relative to the amplitude for threshold crossing at $\omega t = \pi/2$) is shown in Fig. 26. In order of magnitude and frequency dependence, these values seem to compare reasonably well with results of Harris for 200- and 500-cps pure tones, previously shown in Fig. 24.

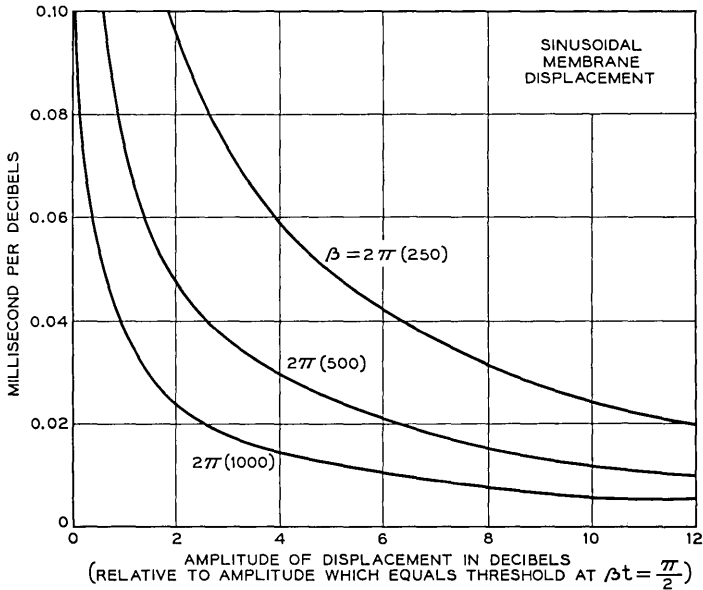


Fig. 26 — Time-intensity trade predicted for sine wave stimuli.

2.4 Threshold Sensitivity

The combined response curves in Fig. 9 indicate that the ear is more sensitive to certain frequencies than to others. This is well known to be subjectively true. To what extent, then, are the variations in the threshold of audibility accounted for by the mechanical sensitivity of the ear? We can use the model responses to examine the question.

The envelope of the peak responses in Fig. 9 can be compared with the subjectively determined minimum audible pressure for pure (sine) tones. Fig. 27 shows this comparison. The agreement is quite poor, although the gross trends are similar. The model responses here are on the basis of a 1500-cps critical frequency for the middle ear. The earlier discussion has pointed up the uncertainty of this value. The middle-ear critical frequency chosen to illustrate the computational technique was that derived from Zwislocki's data. The latter, in turn, were based upon one of Bekesy's investigations. In other investigations, Bekesy also found middle ear cutoffs higher than 1500 cps, so some uncertainty exists as to where this number should be fixed. Obviously the choice of this constant does not alter the computational method or analytical

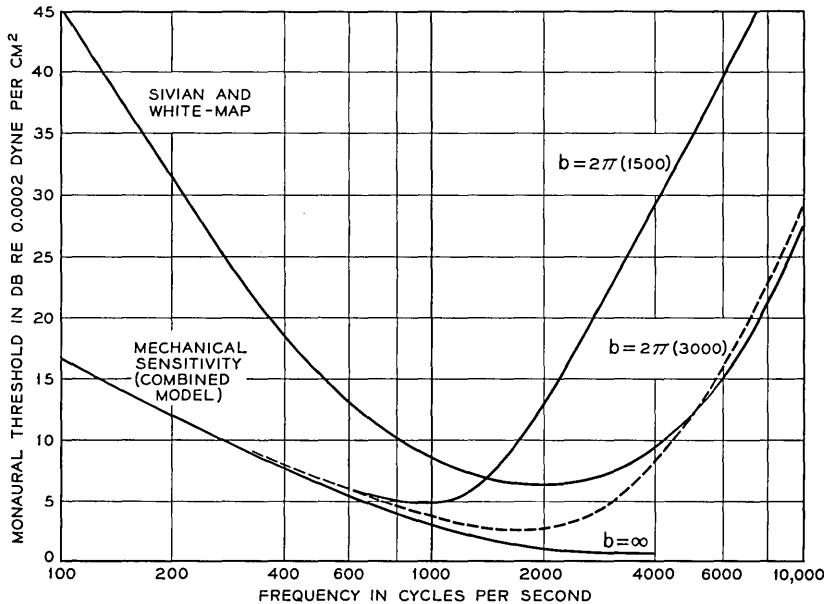


Fig. 27 — Relation of mechanical sensitivity of model to subjective monaural threshold for pure tones.

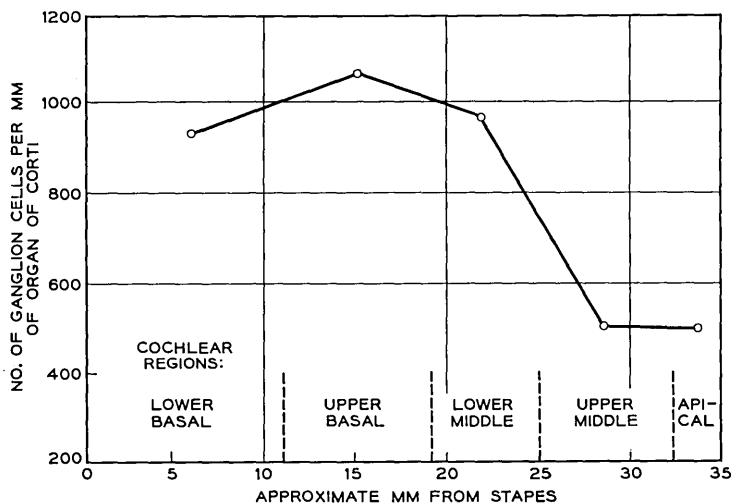


Fig. 28 — Average number of ganglion cells per mm length of organ of Corti (after Guild et al¹⁸).

technique. If we choose instead a critical frequency of 3000 cps for the middle ear, the fit to the threshold curve at high frequencies is more respectable. The match at low frequencies, however, is not improved, but we are less concerned about this for a different reason.

For the low frequencies, the disparity between mechanical and subjective sensitivity probably is a neural effect. According to our earlier assumptions, the number of neurons activated bears some monotonic relation to amplitude of membrane displacement. Perception of loudness is thought to involve possibly temporal and spatial integrations of neural activity. If a constant integrated activity were equivalent to constant loudness, the difference between mechanical and subjective sensitivities might be owing to a sparser neuron density in the apical (low-frequency) end of the cochlea. There is physiological evidence to this effect.

In histological studies Guild et al¹⁸ counted the number of ganglion cells per mm length of the organ of Corti. Their results for normal ears are summarized in Fig. 28. These data show a slight decrease in the number of cells at the basal end and a substantial decrease in the density as the apex is approached. The innervation over the middle of the membrane is roughly constant.

One can pose the same questions about threshold sensitivity for short pulses or clicks of sound. For brief pulses of sufficiently low repetition

rate, the maximal displacements of the membrane, as stated before, are near the middle. According to the model, this continues to be the case for pulse rates well in excess of several hundred per second. The resonance properties of the membrane in this region are such as to resolve in time each individual exciting pulse. If, then, the predominant displacement takes place at one point for a large range of pulse rates, polarity patterns, and pulse durations, how might the subjective threshold vary and how might it be correlated with the membrane motion. One investigation of this question has led to a model for pulse threshold loudness.¹⁹ These results can be partially summarized.

Thresholds of audibility for a variety of periodic pulse trains with various polarity patterns, pulse rates and durations are shown in Fig. 29. One notices that the thresholds are relatively independent of polarity pattern. For pulse rates less than 100 pps, the thresholds are relatively independent of rate, and dependent only upon pulse duration. Above 100 pps, the thresholds diminish with increasing pulse rate. Amplitude of membrane displacement would be expected to be a function of pulse duration and to produce a lower threshold for the longer pulses, which is the case. For rates greater than 100 sec⁻¹, however, some other non-mechanical effect apparently is of importance. The way in which audible pulse amplitude diminishes suggests a temporal integration with a time constant of the order of 10 msec.

Using the earlier assumptions about conversion of mechanical to

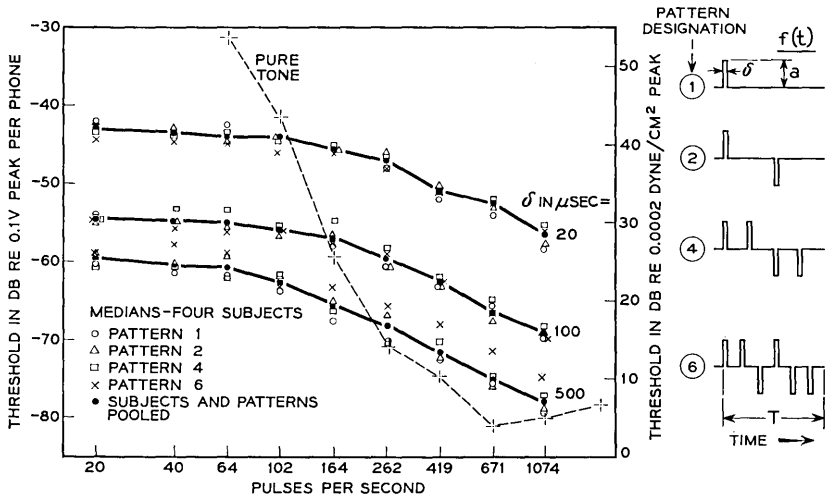


Fig. 29 — Thresholds of audibility for periodic pulses (after Flanagan¹⁹).

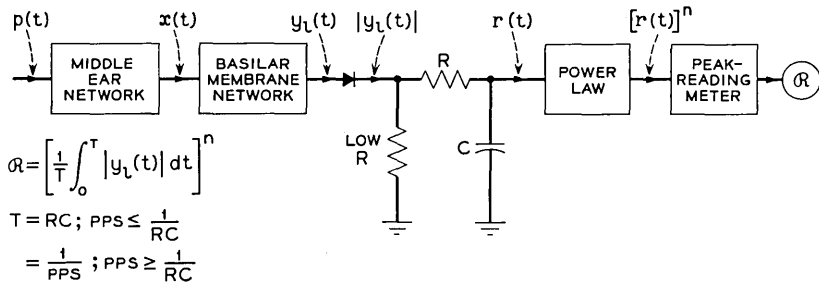


Fig. 30 — Model of the threshold for pulse data shown in Fig. 29.

neural activity, one might ask “What processing of the membrane displacement at the point of greatest amplitude would reflect the constant loudness percept at threshold?” A possible answer is suggested by the operations illustrated in Fig. 30.¹⁹ The first two blocks represent middle-ear transmission [as specified in (6)] and basilar membrane displacement [vicinity of the 1000-cps point, as specified in (31)]. The diode represents the half-wave rectification associated with neural firings on unipolar motions of the membrane. The RC integrator has a 10-msec time constant, as suggested by the threshold data. The power-law element (exponent = 0.6) represents the power-law relation found in loudness estimation.* A meter indicates the peak value of the output of the power-law device. When all stimulus conditions represented by points on the threshold curves in Fig. 29 are applied to the circuit, the output meter reads the same value: that is, threshold.

One can also notice how this model might be expected to perform for sine wave inputs. Because the integration time is 10 msec, frequencies greater than about 100 cps produce meter readings proportional to the average value of the half-wave rectified sinusoid. In other words, the meter reading is proportional to the amplitude of the sine wave into the rectifier. Two alterations in the network circuitry are then necessary. First, the basilar membrane network appropriate to the point maximally responsive to the sine frequency must be used. This may be selected from an ensemble of networks. And second, to take account of the sparser apical innervation, the signal from the rectifier must be attenuated for the low-frequency networks in accordance with the difference between the mechanical and subjective sensitivity curves in Fig. 27. The power-law device still operates to simulate the appropriate growth of loudness with sound level.

* The power-law device is not necessary for threshold indications of “audible-inaudible.” It is necessary, however, to represent the growth of loudness with sound level, and to provide indications of subjective loudness above threshold.

2.5 Pure-Tone Masking

Masking is defined as the increase in the threshold of audibility of one sound caused by the presence of another. The models with which we have been dealing describe the mechanical frequency sensitivity of the ear and hence ought also to imply something about the masking of one pure tone by another.

The significant neural information for a pure-tone stimulus is assumed to come from the membrane point which responds maximally (mechanically) to that frequency. The ability to detect activity correlated with such a tone ought likewise to be related to the amplitude of displacement produced at this same point by any interfering (masking) sound. In other words, the relative amplitudes of displacement caused at the point by the tone and masker might be expected to be related to the shift in threshold of the tone. From the model we can determine the relative amplitudes produced by tone and masker at the membrane point which responds maximally to the tone. For low frequencies, where middle-ear attenuation is not appreciable, this can be done simply from the membrane response curves such as shown in Fig. 3.

Let us take, for example, a masker of 400 cps (because there are corresponding subjective data for this condition). The relative levels of maskee and masker are shown by the dashed curve in Fig. 31. (These levels are read on the righthand ordinate.) Subjective threshold measurements for the same conditions produce the solid curve.²⁰ One sees

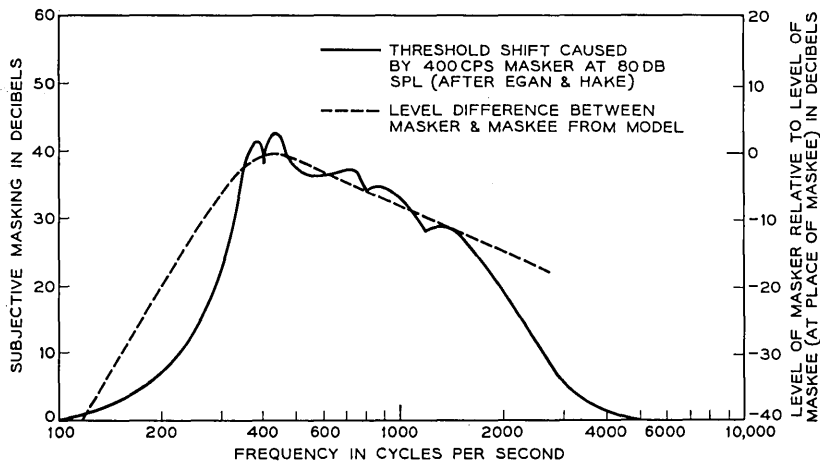


Fig. 31 — Masking of one tone by another (a) predicted by the model; and (b) measured by Egan and Hake.²⁰

that the agreement is not particularly close, although the curves have similar gross shapes. The psychoacoustic measurement indicates less masking at frequencies removed from the masker than the mechanical response implies. This might suggest at least two possibilities: one, that the upper and lower skirts of the membrane resonances are a little steeper than we think; or two, that when the maskee has a relative level as much as about 10 db or so greater than the masker (at the maskee point) some neural inhibitory mechanism functions to suppress the masker even more.

One notices irregularities in the subjective masking curve at frequencies where the tone is an integral multiple of the masker. These are produced by beats and subjectively generated harmonics. One notices, too, that when tone and masker are the same frequency, the measurement is essentially a determination of the intensity limen. For example, the masking at 400 cps is 40 db, which means that a 400-cps tone must be raised 40 db above its unmasked threshold to be just audible in the presence of another 400-cps tone at a sound-pressure level of 80 db (*re* 0.0002 dyne/cm²). The unmasked threshold (minimum audible pressure) for a 400-cps tone is approximately 10 db spl (see Fig. 27). The maskee is just detectable, therefore, when its level is about 50 db spl, or 30 db less than the masker. For an in-phase (or out of phase) condition, the maskee could maximally increase (decrease) the intensity of the masker by about 0.3 db. This is roughly the size of the intensity limen measured at this sensation level.²¹

When this same masking comparison is made for higher frequencies, the middle-ear transmission must be considered. If the middle-ear cutoff used in the model calculations is used, the agreement between subjective and mechanical results is poor at high frequencies. This again argues that the normal critical frequency for the middle ear is somewhat higher than that used to illustrate the model calculations.

The mechanical response also shows why a lower-frequency tone is a more effective masker than a higher-frequency tone. The reason is simply that the frequency response of a given point on the basilar membrane has a low-frequency skirt less steep than its high-frequency skirt. This same fact also suggests why low-frequency hearing is so difficult to impair by local injury or disease in the ear. The shallow low-frequency skirts of the response of all points along the membrane show that even basal points can respond appreciably to low-frequency stimuli. Even if the apical end of the basilar membrane were destroyed, basal locations could provide some low-frequency response.

Essentially in the same vein, these relations suggest why high-fre-

quency hearing is so susceptible of impairment. The high-frequency skirt of the frequency responses is quite sharp. Damage to any basal location leaves no other point capable of responding substantially to that frequency.

2.6 Conclusion

It seems clear that the extent to which subjective behavior can be correlated with, identified in, and predicted by the mechanical operation of the peripheral ear is rather substantial. The models developed here have been found to be useful computational tools in the analyses of a number of different psychoacoustic problems. They have, in fact, precipitated several experiments by predicting hearing phenomena which were later confirmed by the experiments. Further, electrophysiological data obtained recently link neural activity intimately with the individual mechanical excursions of the membrane. These findings also lend support to the simple assumptions about the conversion of mechanical to neural information.

The models do not, of course, account for higher-order neural functions and hence describe only a peripheral part of the hearing process. Even so, they seem in many cases to contribute substantially to physiological explanations of subjective behavior. As more knowledge is gained about mechano-neural conversion and about neural processing, analytical specification of the mechanical operation, such as developed here, may prove increasingly useful.

III. ACKNOWLEDGMENT

I wish to thank B. J. Watson of the Speech and Auditory Research Department for his capable technical assistance in the experimental phases of this work.

REFERENCES

1. Flanagan, J. L., Models for Approximating Basilar Membrane Displacement, *B.S.T.J.*, **39**, 1960, pp. 1163-1191.
2. von Bekesy, G., Über die Resonanzkurve und die Abklingzeit der verschiedenen Stellen der Schneckentrennwand, *Akust. Zeit.*, **8**, 1943, p. 66; *J. Acoust. Soc. Am.*, **21**, 1949, p. 245.
3. von Bekesy, G., Über die Schwingungen der Schneckentrennwand beim Präparat und Ohrenmodell, *Akust. Zeit.*, **7**, 1942, p. 173; *J. Acoust. Soc. Am.*, **21**, 1949, p. 233.
4. Zwislocki, J., Some Impedance Measurements on Normal and Pathological Ears, *J. Acoust. Soc. Am.*, **29**, 1957, p. 1312; Electrical Model of the Middle Ear, *J. Acoust. Soc. Am.*, **31**, 1959, p. 841 (A); also personal communication.
5. Huggins, W. H., and Licklider, J. C. R., Place Mechanisms of Auditory Frequency Analysis, *J. Acoust. Soc. Am.*, **23**, 1951, p. 290.

6. Flanagan, J. L., and Guttman, N., Pitch of Periodic Pulses, *J. Acoust. Soc. Am.*, **32**, 1960, p. 1308.
7. Kiang, N., Watanabe, T., Thomas, E., and Clark, L., *Stimulus Coding at the Periphery of the Auditory System*, book in preparation.
8. Hughes, J. R., and Rosenblith, W. A., Electrophysiological Evidence for Auditory Sensitization, *J. Acoust. Soc. Am.*, **29**, 1957, p. 275.
9. Galambos, R., Neural Mechanisms in Audition, *Laryngoscope*, **68**, 1958, p. 388.
10. Shower, E. G., and Biddulph, R., Differential Pitch Sensitivity of the Ear, *J. Acoust. Soc. Am.*, **3**, 1931, p. 275.
11. Guttman, N., and Flanagan, J. L. Pitch of High-Pass Filtered Periodic Pulses, *J. Acoust. Soc. Am.*, **33**, 1961, p. 839(A); see also last footnote in Ref. 1.
12. Flanagan, J. L., and Guttman, N., Pitch of Periodic Pulses without Fundamental Component, *J. Acoust. Soc. Am.*, **32**, 1960, p. 1319.
13. Miller, G. A., and Taylor, W. G., The Perception of Repeated Bursts of Noise, *J. Acoust. Soc. Am.*, **20**, 1948, p. 171; Harris, G. G., Pitch in Gated Random Noise, *J. Acoust. Soc. Am.*, **32**, 1960, p. 1506(A).
14. Flanagan, J. L., David, E. E., Jr., and Watson, B. J., Binaural Lateralization of Cophasic and Antiphasic Clicks, *J. Acoust. Soc. Am.*, **33**, 1961, p. 840(A); Effects of Masking Upon the Binaural Lateralization of Cophasic and Antiphasic Clicks, *J. Acoust. Soc. Am.*, **33**, 1961, p. 1670. (A).
15. Peake, W. T., An Analytic Study of Electric Responses at the Periphery of the Auditory System, Technical Report 365, Research Laboratory of Electronics, MIT, 1960.
16. David, E. E., Jr., Guttman, N., and van Bergeijk, W. A., Binaural Interaction of High-Frequency Complex Stimuli, *J. Acoust. Soc. Am.*, **31**, 1959, p. 774.
17. Harris, G. G., Binaural Interactions of Impulsive Stimuli and Pure Tones, *J. Acoust. Soc. Am.*, **32**, 1960, p. 685.
18. Guild, S. R., Crowe, S. J., Bunch, C. C., and Polvogt, L. M., Correlations of Differences in the Density of Innervation of the Organ of Corti with Differences in the Acuity of Hearing, *Acta Oto-Laryngologica*, **15**, 1931, p. 269.
19. Flanagan, J. L., Audibility of Periodic Pulses and a Model for the Threshold, *J. Acoust. Soc. Am.*, **33**, 1961, p. 1540.
20. Egan, J. P., and Hake, H. W., On the Masking Pattern of a Simple Auditory Stimulus, *J. Acoust. Soc. Am.*, **22**, 1950, p. 622.
21. Riesz, R. R., Differential Intensity Sensitivity of the Ear for Pure Tones, *Phys. Rev.*, **31**, 1928, p. 867.

Approximate Solutions for the Coupled Line Equations

By H. E. ROWE

(Manuscript received May 19, 1961)

The coupled line equations for only two modes, representing the TE₀₁ signal mode and a single spurious mode in circular waveguide, are solved in series form by the method of successive approximations. Bounds are found on the magnitudes of the terms in the series solution. These bounds decrease rapidly only for "short" waveguides; for long guides many terms of the series must be included in the solution.

The coupled line equations are transformed to a new form, in which one of the unknowns Λ is given by $\Lambda = -\ln G_0$, where G_0 is the (complex) TE₀₁ transfer function of the original coupled line equations. Thus $\text{Re } \Lambda = -\ln |G_0|$, the TE₀₁ loss in nepers, $\text{Im } \Lambda = -\angle G_0$, the TE₀₁ phase in radians. These transformed equations are again solved by successive approximations; the first term is the commonly used solution that has been obtained by physical arguments. Bounds are determined for the magnitudes of the terms in these series solutions; for a suitable restriction on the coupling coefficient that includes many cases of practical interest, these bounds decrease rapidly for long guides.

In present calculations of the TE₀₁ loss statistics in random guides, only the first term of the series expansion for Λ is considered. Unfortunately this approximation has not so far been justified.

I. INTRODUCTION

Consider the coupled line equations:

$$\begin{aligned} I_0'(z) &= -\Gamma_0 I_0(z) + jc(z) I_1(z), \\ I_1'(z) &= +jc(z) I_0(z) - \Gamma_1 I_1(z). \end{aligned} \tag{1}$$

These equations are of interest in many applications. Our particular interest in them in a companion paper¹ is that they describe the effects

of coupling between the TE_{01} signal mode, represented by the complex wave amplitude I_0 , and a single spurious mode, represented by the complex wave amplitude I_1 , caused by geometric imperfections in circular waveguide. We have, of course, assumed that only a single spurious mode has significant magnitude, so that all other spurious modes may be neglected. For example, we may consider copper waveguide with a rather general straightness deviation; the most important spurious mode under many conditions will be the forward TE_{12}^1 (both polarizations must, of course, be considered unless the straightness deviation is confined to a single plane). However, these equations apply to a variety of other problems which may be described by only two modes with varying degrees of accuracy.¹

In copper waveguide if the wall losses may be neglected the propagation constants Γ_0 and Γ_1 are pure imaginary and the coupling coefficient $c(z)$ is pure real. In helix guide, where loss is added to the spurious mode, the propagation constant Γ_1 has a significant negative real part; further, as shown by H. G. Unger,² the coupling coefficient $c(z)$ also becomes complex.

The case where the geometric imperfection (e.g., straightness deviation) and hence the coupling coefficient is a stationary random process, perhaps Gaussian, is of great interest; here it is desired to compute the statistics of the TE_{01} transmission I_0 in terms of the statistics of the coupling coefficient $c(z)$. Since exact solutions to (1) are easily found in only a few special cases, this has been done by using an approximate solution to these equations that is essentially a second-order perturbation solution, and by studying the statistics of this approximate solution.¹ The present paper will discuss this approximate solution, will give some bounds on the convergence of the approximation, and will indicate a basic gap in our knowledge concerning this problem.

Equation (1) represents a drastic idealization of the real TE_{01} transmission problem, in that it contains only one spurious mode and neglects all other spurious modes. The approximate solution includes all second-order terms; a physical interpretation of this solution states that conversion of TE_{01} to each spurious mode and subsequent reconversion to TE_{01} is considered at all pairs of elementary mode converters, but that higher-order terms involving more than one pair of elementary mode converters are neglected.¹ The exact solution of (1) includes all higher-order terms involving the single spurious mode, but neglects many more higher-order terms involving the many other spurious modes that have been neglected in (1). In view of this it may appear questionable to try

to deal with (1) in more exact terms for the TE_{01} mode conversion problem. However, a start has to be made somewhere, and it seems unlikely that the general case involving an infinite number of modes will be understood before the two-mode case of (1) is understood. Even this simple idealized case does not yet have a really satisfactory treatment. Also, (1) does apply more or less rigorously to many other situations than the TE_{01} mode conversion problem.¹

In dealing with these equations it is convenient to introduce the following change of variables:

$$\begin{aligned} I_0(z) &= e^{-\Gamma_0 z} \cdot G_0(z), \\ I_1(z) &= e^{-\Gamma_1 z} \cdot G_1(z). \end{aligned} \quad (2)$$

Then (1) becomes:

$$\begin{aligned} G_0'(z) &= jc(z) e^{\Delta\Gamma z} G_1(z), \\ G_1'(z) &= jc(z) e^{-\Delta\Gamma z} G_0(z). \end{aligned} \quad (3a)$$

$$\Delta\Gamma = \Delta\alpha + j\Delta\beta = \Gamma_0 - \Gamma_1 ;$$

$$\Delta\alpha = \alpha_0 - \alpha_1 < 0, \quad (3b)$$

$$\Delta\beta = \beta_0 - \beta_1 .$$

Note that we assume $\Delta\alpha < 0$, because in circular waveguide the TE_{01} signal mode will have lower heat loss than any of the spurious modes. We will always take as initial conditions at $z = 0$ a TE_{01} wave of unit magnitude and zero phase, and a spurious mode of zero magnitude:

$$G_0(0) = 1, \quad G_1(0) = 0. \quad (4)$$

II. SOLUTION OF THE COUPLED LINE EQUATIONS BY SUCCESSIVE APPROXIMATIONS (PICARD'S METHOD)

We summarize the solution of (3) by successive approximations. Let $G_{0(n)}(z)$ and $G_{1(n)}(z)$ be the n^{th} approximation to the solution of (3). Let the initial approximation be given simply by the initial conditions of (4):

$$G_{0(0)}(z) = 1, \quad G_{1(0)}(z) = 0. \quad (5)$$

Then following Picard's method^{3,4} we obtain the successive approximations as follows:

$$G_{0(0)}(z) = 1, \quad G_{1(0)}(z) = 0$$

$$\begin{aligned} G_{0(1)}(z) &= 1 + j \int_0^z c(s)e^{\Delta\Gamma s} G_{1(0)}(s) ds \\ &= 1 \end{aligned}$$

$$\begin{aligned} G_{1(1)}(z) &= j \int_0^z c(s)e^{-\Delta\Gamma s} G_{0(0)}(s) ds \\ &= j \int_0^z c(s)e^{-\Delta\Gamma s} ds \end{aligned}$$

$$\begin{aligned} G_{0(2)}(z) &= 1 + j \int_0^z c(s)e^{\Delta\Gamma s} G_{1(1)}(s) ds \\ &= 1 - \int_0^z c(s)e^{\Delta\Gamma s} ds \int_0^s c(t)e^{-\Delta\Gamma t} dt \end{aligned} \tag{6}$$

$$\begin{aligned} G_{1(2)}(z) &= j \int_0^z c(s)e^{-\Delta\Gamma s} G_{0(1)}(s) ds \\ &= j \int_0^z c(s)e^{-\Delta\Gamma s} ds \end{aligned}$$

⋮
⋮
⋮

$$G_{0(n)}(z) = 1 + j \int_0^z c(s)e^{\Delta\Gamma s} G_{1(n-1)}(s) ds$$

$$G_{1(n)}(z) = j \int_0^z c(s)e^{-\Delta\Gamma s} G_{0(n-1)}(s) ds.$$

The n^{th} approximation is obtained by substituting the $(n - 1)^{\text{th}}$ approximation in the right-hand side of (3) and integrating. Writing

$$\begin{aligned} G_{0(n)}(z) - G_{0(n-1)}(z) &= g_{0(n)}(z), \\ G_{1(n)}(z) - G_{1(n-1)}(z) &= g_{1(n)}(z), \end{aligned} \tag{7}$$

we have

$$G_{0(n)}(z) = 1 + \sum_{k=1}^n g_{0(k)}(z), \tag{8a}$$

$$G_{1(n)}(z) = \sum_{k=1}^n g_{1(k)}(z), \tag{8b}$$

where the g 's are given as follows:

$$g_{0(n)}(z) = j \int_0^z c(s) e^{\Delta \Gamma s} g_{1(n-1)}(s) ds, \quad n \geq 1. \tag{9a}$$

$$g_{1(n)}(z) = j \int_0^z c(s) e^{-\Delta \Gamma s} g_{0(n-1)}(s) ds, \quad n \geq 1. \tag{9b}$$

$$g_{0(0)}(z) = 1, \quad g_{1(0)}(z) = 0. \tag{9c}$$

It is readily seen that

$$\begin{aligned} g_{0(n)}(z) &= 0, & n \text{ odd} \\ g_{1(n)}(z) &= 0, & n \text{ even} \end{aligned} \tag{10}$$

so that only even terms appear in the summation of (8a) and only odd terms in the summation of (8b).

In the standard proof of Picard's method the series of (8) are shown to converge to the unique solution of the coupled line equations, (3), and bounds are given on the magnitudes of the terms in (8). Thus we may write

$$G_0(z) = \sum_{n=0}^{\infty} g_{0(n)}(z), \tag{11a}$$

$$G_1(z) = \sum_{n=0}^{\infty} g_{1(n)}(z), \tag{11b}$$

where the g 's are given in (9) and (10). However, better bounds than those given by Picard's general method may be found for the present special case. We show that

$$\begin{aligned} |g_{0(n)}(z)| &\leq \frac{\left[\int_0^z |c(s)| ds \right]^n}{n!}, & n \text{ even.} \\ &= 0, & n \text{ odd.} \end{aligned} \tag{12a}$$

$$\begin{aligned} |g_{1(n)}(z)| &\leq \frac{\left[\int_0^z |c(s)| ds \right]^n}{n!} e^{-\Delta \alpha z}, & n \text{ odd.} \\ &= 0, & n \text{ even.} \end{aligned} \tag{12b}$$

Suppose that (12a) is true for some even value of n . Then from (9b)

$$\begin{aligned}
 |g_{1(n+1)}(z)| &\cong \int_0^z |c(t)| e^{-\Delta\alpha t} \frac{\left[\int_0^t |c(s)| ds\right]^n}{n!} dt \\
 &\cong \frac{e^{-\Delta\alpha z}}{n!} \int_0^z \left[\int_0^t |c(s)| ds\right]^n d\left[\int_0^t |c(s)| ds\right] \quad (13) \\
 &= \frac{\left[\int_0^z |c(s)| ds\right]^{n+1}}{(n+1)!} e^{-\Delta\alpha z},
 \end{aligned}$$

where we recall from (3b) that $\Delta\alpha < 0$. Substituting (13) into (9a),

$$\begin{aligned}
 |g_{0(n+2)}(z)| &\cong \int_0^z |c(t)| e^{\Delta\alpha t} e^{-\Delta\alpha t} \frac{\left[\int_0^t |c(s)| ds\right]^{n+1}}{(n+1)!} dt \\
 &= \frac{1}{(n+1)!} \int_0^z \left[\int_0^t |c(s)| ds\right]^{n+1} d\left[\int_0^t |c(s)| ds\right] \quad (14) \\
 &= \frac{\left[\int_0^z |c(s)| ds\right]^{n+2}}{(n+2)!}.
 \end{aligned}$$

Noting (9c), the results of (12) hold for all n by induction.

We may ask whether the bounds of (12) are the best that can be obtained in general, or if by being sufficiently clever we can do better. It is easy to find examples whose terms are actually as large as those given in (12), so that no improvement in these results is to be expected unless suitable restrictions are placed on the problem. Thus, consider the following special case:

$$\Delta\Gamma = 0. \quad (15)$$

The coupling coefficient is non-negative but otherwise arbitrary. The general solution to (3a), subject of course to the initial conditions of (4), is¹

$$G_0(z) = \cos \left[\int_0^z c(s) ds \right]. \quad (16)$$

Expanding the cosine in power series,

$$G_0(z) = 1 - \frac{\left[\int_0^z c(s) ds\right]^2}{2!} + \frac{\left[\int_0^z c(s) ds\right]^4}{4!} - \dots \quad (17)$$

The successive terms of (17) are simply the $g_{0(n)}$ given in (9) and (10).

It is readily seen that the magnitudes of these terms are equal to the bounds given in (12a) if we require $0 \leq c(z)$ so that $|c(z)| = c(z)$.

As another similar example, let $\Delta\Gamma \neq 0$ and $c(z)$ be a single δ -function located at z_0 ,

$$c(z) = C \cdot \delta(z - z_0). \quad (18)$$

In our present case, i.e., straightness deviation, a discrete coupling of the type given in (18) corresponds to a discrete tilt located at $z = z_0$. The solution to (3a), subject again to the initial conditions of (4), is¹

$$G_0(z) = \cos C, \quad z > z_0. \quad (19)$$

Expanding the cosine,

$$G_0(z) = 1 - \frac{C^2}{2!} + \frac{C^4}{4!} - \dots \quad (20)$$

Again the terms of (20) are the $g_{0(n)}$ of (9) and (10), and their magnitudes are equal to the bounds given in (12a). Of course, this above solution, which mathematically is valid for an arbitrarily large tilt in the present idealized two-mode case, must fail for large tilts in the physical case, the error being caused by neglecting the higher-order spurious modes excited by the tilt. While this serves as a further warning against uncritical application of the results of the two-mode theory to the physical problem, it is still of interest to inquire into the mathematical properties of the solutions to (3).

It is often desirable to express the TE_{01} loss in db rather than as the magnitude of the TE_{01} normalized gain, $|G_0|$. Define the complex TE_{01} loss Λ as

$$\Lambda = -\ln G_0 = A - j\Theta. \quad (21)$$

Then

$$\begin{aligned} A &= -\ln |G_0|, \\ \Theta &= \angle G_0. \end{aligned} \quad (22)$$

A is the TE_{01} loss in nepers; the TE_{01} loss in db is simply $8.686 A$. If we have a number of sections of waveguide separated by ideal mode filters, the over-all TE_{01} gain, G_{OT} , and loss Λ_T , are given by

$$\begin{aligned} G_{OT} &= G_{0,1} G_{0,2} \dots, \\ \Lambda_T &= \Lambda_1 + \Lambda_2 + \dots, \\ A_T &= A_1 + A_2 + \dots, \\ \Theta_T &= \Theta_1 + \Theta_2 + \dots. \end{aligned} \quad (23)$$

The statistics of A and Θ for the composite guide may thus be expressed simply in terms of the statistics of A and Θ for the individual guide sections.

Suppose that the series solution for $G_0(z)$ given in (11a) converges very rapidly, so that only the first two terms need be retained. Then we have from (8-11)

$$G_0(z) \approx G_{0(2)}(z) \quad (24)$$

so that approximately

$$G_0(z) = 1 - \int_0^z c(s)e^{\Delta\Gamma s} ds \int_0^s c(t)e^{-\Delta\Gamma t} dt. \quad (25)$$

Then, assuming that the second term is small compared to 1, we have approximately:

$$|G_0(z)| = \text{Re } G_0 = 1 - \text{Re } \iint \quad (26a)$$

$$\Lambda = -\ln G_0 = \iint \quad (26b)$$

$$A = -\ln |G_0| = \text{Re } \iint \quad (26c)$$

$$\Theta = \angle G_0 = -\text{Im } \iint \quad (26d)$$

where \iint is shorthand for

$$\iint = \int_0^z c(s)e^{\Delta\Gamma s} ds \int_0^s c(t) e^{-\Delta\Gamma t} dt \quad (27a)$$

$$= \int_0^z e^{\Delta\Gamma u} du \int_0^{z-u} c(s)c(s+u) ds \quad (27b)$$

$$= \frac{1}{2} \int_0^z \int_0^z c(s)c(t)e^{\Delta\Gamma|t-s|} ds dt. \quad (27c)$$

If the coupling coefficient $c(z)$ is pure real but the differential propagation constant is complex (possibly not a physical case), (26c) becomes, using (27c),

$$A = -\ln |G_0| = \int_0^z e^{\Delta\alpha u} \cos \Delta\beta u du \int_0^{L-u} c(s)c(s+u) ds. \quad (28)$$

If $c(z)$ is complex, it turns out that for uniform waveguides its phase

angle remains constant and only its magnitude varies, so that we may write²

$$c(z) = (C_r + j C_i) \bar{c}(z), \tag{29}$$

where $\bar{c}(z)$ is real. Then (28) becomes

$$A = -\ln |G_0| = (C_r^2 - C_i^2) \int_0^z e^{\Delta\alpha u} \cos \Delta\beta u \, du$$

$$\cdot \int_0^{L-u} \bar{c}(s)\bar{c}(s+u) \, ds - 2C_r C_i \int_0^z e^{\Delta\alpha u} \sin \Delta\beta u \, du \tag{30}$$

$$\cdot \int_0^{L-u} \bar{c}(s)\bar{c}(s+u) \, ds.$$

These approximate expressions of (26-30) may be regarded as the first terms of series expansions for the various quantities. The above approximations will be valid when $\left| \iint \right| \ll 1$ and when the higher-order terms may be neglected. From the above analysis it would appear that when $\left| \iint \right| \gg 1$, all of the above approximations would fail, since, in particular, (26a) obviously fails. In spite of this fact, (26b-d), (28) and (30) may remain valid for a wide class of long guides of practical interest; roughly speaking, the required conditions are that the differential loss $|\Delta\alpha|$ be large enough and that the coupling coefficient $c(z)$ be sufficiently small and uniformly distributed in an appropriate sense. This result has been suggested by simple physical arguments;¹ a formal mathematical derivation starting with the appropriate restriction on $c(z)$ and $\Delta\alpha$ is given in the following section. These results are of importance because in a random guide the expected value of the \iint term increases linearly with distance z ;¹ while the approximations of (25) and (26a) fail, the results of (26b-d), (28) and (30) may remain valid, and so provide us with a theory for long guides.

It is apparent that further restrictions are required to obtain these additional results, by considering the example of (18-20). Thus, let the magnitude of the δ -function coupling coefficient be $\pi/2$, so that we have in (18)

$$C = \frac{\pi}{2}, \quad c(z) = \frac{\pi}{2} \cdot \delta(z - z_0). \tag{31}$$

Then from (19)

$$G_0(z) = 0, \quad z > z_0 \tag{32}$$

so that

$$\Lambda = -\ln G_0 = \infty. \quad (33)$$

However, the approximate result of (26b) yields

$$\Lambda = -\ln G_0 \approx \frac{C^2}{2} = \frac{\pi^2}{8}. \quad (34)$$

The approximation of (34) is obviously invalid; since these relations are independent of $\Delta\alpha$, this approximation remains invalid no matter how high $|\Delta\alpha|$ becomes. Cases such as this are ruled out by the additional restrictions that require the coupling coefficient to be more or less uniformly distributed with z in a certain sense, described in the following section.

III. TRANSFORMATION OF THE COUPLED LINE EQUATIONS TO LOGARITHMIC FORM, AND SOLUTION BY SUCCESSIVE APPROXIMATIONS

We repeat for convenience the coupled line equations, given in (3), together with the desired initial conditions, (4).

$$G_0'(z) = j c(z) e^{\Delta\Gamma z} G_1(z), \quad (35)$$

$$G_1'(z) = j c(z) e^{-\Delta\Gamma z} G_0(z).$$

$$G_0(0) = 1, \quad G_1(0) = 0. \quad (36)$$

Next, the following transformation of variables is made:

$$G_0(z) = e^{-\Lambda(z)}. \quad (37a)$$

$$G_1(z) = e^{-\Lambda(z)} \cdot H(z). \quad (37b)$$

The transformation of (37a) is dictated by the desire to obtain a series solution for Λ , defined in (21). That of (37b) was obtained partly by trial and error and partly by intuitive means. Substituting (37) into (35), we obtain:

$$\Lambda'(z) = -j c(z) e^{\Delta\Gamma z} H(z) \quad (38a)$$

$$H'(z) = j c(z) e^{\Delta\Gamma z} + \Lambda'(z) H(z). \quad (38b)$$

By substituting (38a) into the second term on the right-hand side of (38b), we have:

$$\Lambda'(z) = -j c(z) e^{\Delta\Gamma z} H(z) \quad (39a)$$

$$H'(z) = j c(z) e^{-\Delta\Gamma z} - j c(z) e^{\Delta\Gamma z} H^2(z). \quad (39b)$$

The initial conditions of (36) transform via (37) to

$$\Lambda(0) = 0, \quad H(0) = 0. \quad (40)$$

The method of successive approximations may now be applied to (39) (or 38), subject to the initial conditions of (40). We note that $\Lambda(z)$ is absent from (39b), so that this equation contains only a single dependent variable, $H(z)$. Thus the successive approximations to $H(z)$ may be found without reference to (39a) or to $\Lambda(z)$; the corresponding approximations for $\Lambda(z)$ are then found by a simple integration of (39a). We note further that (39b) for $H(z)$ is a Riccati equation.³

Thus, let $\Lambda_n(z)$ and $H_n(z)$ be the n^{th} approximation to the solution of (39), subject to the initial conditions of (40). Then:

$$\begin{aligned}
 H_0(z) &= 0 \\
 H_1(z) &= j \int_0^z c(s)e^{-\Delta\Gamma s} ds \\
 H_2(z) &= j \int_0^z c(s)e^{-\Delta\Gamma s} ds - j \int_0^z c(s)e^{\Delta\Gamma s} H_1^2(s) ds \\
 &= j \int_0^z c(s)e^{-\Delta\Gamma s} ds \\
 &\quad + j \int_0^z c(s)e^{\Delta\Gamma s} ds \int_0^s \int_0^s c(t)c(u)e^{-\Delta\Gamma(t+u)} dt du
 \end{aligned} \tag{41a}$$

⋮
⋮

$$H_n(z) = j \int_0^z c(s)e^{-\Delta\Gamma s} ds - j \int_0^z c(s)e^{\Delta\Gamma s} H_{n-1}^2(s) ds$$

$$\Lambda_0(z) = 0$$

$$\begin{aligned}
 \Lambda_1(z) &= -j \int_0^z c(s)e^{\Delta\Gamma s} H_1(s) ds \\
 &= \int_0^z c(s)e^{\Delta\Gamma s} ds \int_0^s c(t)e^{-\Delta\Gamma t} dt \\
 \Lambda_2(z) &= -j \int_0^z c(s)e^{\Delta\Gamma s} H_2(s) ds \\
 &= \int_0^z c(s)e^{\Delta\Gamma s} ds \int_0^s c(t)e^{-\Delta\Gamma t} dt \\
 &\quad + \int_0^z c(s)e^{\Delta\Gamma s} ds \int_0^s c(t)e^{\Delta\Gamma t} dt \int_0^t \int_0^t c(u)c(v)e^{-\Delta\Gamma(u+v)} du dv
 \end{aligned} \tag{41b}$$

⋮
⋮

$$\Lambda_n(z) = -j \int_0^z c(s)e^{\Delta\Gamma s} H_n(s) ds.$$

Note that Λ_1 is identical to the approximation of (26b). Writing as before

$$\begin{aligned}\Lambda_n(z) - \Lambda_{n-1}(z) &= \lambda_n(z), \\ H_n(z) - H_{n-1}(z) &= h_n(z),\end{aligned}\tag{42}$$

we have

$$\begin{aligned}\Lambda_n(z) &= \sum_{k=1}^n \lambda_k(z), \\ H_n(z) &= \sum_{k=1}^n h_k(z).\end{aligned}\tag{43}$$

The quantities $\lambda_n(z)$ and $h_n(z)$ are given as follows:

$$\lambda_n(z) = -j \int_0^z c(s) e^{\Delta \Gamma s} h_n(s) ds, \quad n \geq 1.\tag{44a}$$

$$\begin{aligned}h_n(z) &= -j \int_0^z c(s) e^{\Delta \Gamma s} [H_{n-1}^2(s) - H_{n-2}^2(s)] ds \\ &= -j \int_0^z c(s) e^{\Delta \Gamma s} h_{n-1}(s) [H_{n-1}(s) + H_{n-2}(s)] ds, \quad n \geq 2.\end{aligned}\tag{44b}$$

$$h_1(z) = H_1(z) = j \int_0^z c(s) e^{-\Delta \Gamma s} ds.\tag{44c}$$

Then under certain conditions described below,

$$\Lambda(z) = \sum_{n=1}^{\infty} \lambda_n(z),\tag{45a}$$

$$H(z) = \sum_{n=1}^{\infty} h_n(z).\tag{45b}$$

We next obtain bounds on $|\lambda_n(z)|$ and $|h_n(z)|$. As stated in the last section, it is first necessary to impose additional restrictions on the problem. We assume that the coupling coefficient $c(z)$ and the differential loss $\Delta\alpha$ are such that a number K exists satisfying the following relation:

$$\int_0^z |c(s)| e^{\Delta\alpha(z-s)} ds \leq K \quad \text{for every } z \geq 0.\tag{46}$$

We recall from (3) that $\Delta\alpha < 0$. It will subsequently appear that convergence of the approximate solution can be guaranteed in general only for $K \leq 0.455$; further, the smaller K the more rapid the bounds

on $|\lambda_n(z)|$ and $|h_n(z)|$ decrease as n increases. The restriction of (46) was again obtained partly by physical reasoning and partly by trial and error. Roughly speaking, for a given K it guarantees that the coupling coefficient $c(z)$ is more or less uniformly distributed along z ; we thus rule out cases where the coupling coefficient is zero over most of the guide and large over a very short section (e.g., the example of (31)-(34), where $c(z)$ is a single δ -function). Physically, such a condition says that for small K the spurious mode is dissipated much faster than it is coupled from the signal mode; the larger $c(z)$ the larger must be $|\Delta\alpha|$ in order to satisfy (46) for a given value of K . This will normally be the only case of practical interest in long random guides.

Bounds on the first few $h_n(z)$ are readily obtained. From (44c) and (46),

$$\begin{aligned}
 |h_1(z)| &\leq \int_0^z |c(s)| e^{-\Delta\alpha s} ds = e^{-\Delta\alpha z} \int_0^z |c(s)| e^{\Delta\alpha(z-s)} ds \\
 &\leq e^{-\Delta\alpha z} \cdot K.
 \end{aligned}
 \tag{47}$$

Next, from (44b), (43) and (41a),

$$h_n(z) = -j2 \int_0^z c(s) e^{\Delta\Gamma s} h_{n-1}(s) \left[\sum_{k=1}^{n-2} h_k(s) + \frac{h_{n-1}(s)}{2} \right] ds. \tag{48}$$

Thus,

$$\begin{aligned}
 |h_n(z)| &\leq 2 \int_0^z |c(s)| e^{\Delta\alpha s} |h_{n-1}(s)| \left[\sum_{k=1}^{n-2} |h_k(s)| \right. \\
 &\quad \left. + \frac{|h_{n-1}(s)|}{2} \right] ds.
 \end{aligned}
 \tag{49}$$

Equation (49) yields for the first few $h_n(z)$:

$$\begin{aligned}
 |h_2(z)| &\leq 2 \int_0^z |c(s)| e^{\Delta\alpha s} e^{-2\Delta\alpha s} \frac{K^2}{2} ds = e^{-\Delta\alpha z} K^2 \\
 &\quad \cdot \int_0^z |c(s)| e^{\Delta\alpha(z-s)} ds \leq e^{-\Delta\alpha z} \cdot K^3
 \end{aligned}
 \tag{50}$$

$$\begin{aligned}
 |h_3(z)| &\leq 2 \int_0^z |c(s)| e^{\Delta\alpha s} e^{-\Delta\alpha s} K^3 e^{-\Delta\alpha s} \left[K + \frac{K^2}{2} \right] ds \\
 &= e^{-\Delta\alpha z} 2K^4 \left[1 + \frac{K^2}{2} \right] \int_0^z |c(s)| e^{\Delta\alpha(z-s)} ds \\
 &\leq e^{-\Delta\alpha z} \cdot 2K^5 \left[1 + \frac{K^2}{2} \right].
 \end{aligned}
 \tag{51}$$

By an exactly similar process:

$$|h_4(z)| \leq e^{-\Delta\alpha z} \cdot 2^2 K^7 \left[1 + \frac{K^2}{2}\right] \left[1 + K^2 + K^4 + \frac{K^6}{2}\right] \quad (52)$$

$$|h_5(z)| \leq e^{-\Delta\alpha z} \cdot 2^3 K^9 \left[1 + \frac{K^2}{2}\right] \left[1 + K^2 + K^4 + \frac{K^6}{2}\right] \cdot \left[1 + K^2 + 2K^4 + 3K^6 + 3K^8 + 3K^{10} + 2K^{12} + \frac{K^{14}}{2}\right]. \quad (53)$$

It is difficult to continue the above process and write out explicitly the n^{th} term. However, by accepting a slightly poorer bound the analysis may be greatly simplified. We show that

$$|h_n(z)| \leq e^{-\Delta\alpha z} M^{n-2} K^{n+1}; \quad n \geq 2. \quad (54)$$

M is a constant to be determined, as a function of K . Assume that (54) is true for some value of n . Then, from (49)

$$\begin{aligned} |h_{n+1}(z)| &\leq 2 \int_0^z |c(s)| M^{n-2} K^{n+1} e^{-\Delta\alpha s} \left[K + \sum_{k=2}^{n-1} M^{k-2} K^{k+1} + \frac{M^{n-2} K^{n+1}}{2} \right] ds \\ &= e^{-\Delta\alpha z} \cdot 2M^{n-2} K^{n+2} \left[1 + K^2 \sum_{k=0}^{n-3} M^k K^k + \frac{M^{n-2} K^n}{2} \right] \cdot \int_0^z |c(s)| e^{\Delta\alpha(z-s)} ds \\ &\leq e^{-\Delta\alpha z} \cdot 2M^{n-2} K^{n+3} \left[1 + K^2 \sum_{k=0}^{n-3} M^k K^k + \frac{M^{n-2} K^n}{2} \right]. \end{aligned} \quad (55)$$

If (54) is to remain true for $n \rightarrow n + 1$, we have from (55)

$$\begin{aligned} 2M^{n-2} K^{n+3} \left[1 + K^2 \sum_{k=0}^{n-3} M^k K^k + \frac{M^{n-2} K^n}{2} \right] &\leq M^{n-1} K^{n+2}, \\ K \left[1 + K^2 \sum_{k=0}^{n-3} M^k K^k + \frac{M^{n-2} K^n}{2} \right] &\leq \frac{M}{2}. \end{aligned} \quad (56)$$

But since the left-hand side of (56) is increased by allowing $n \rightarrow \infty$ and dropping the final term inside the brackets, the inequality of (56) will be satisfied if

$$K \left[1 + \frac{K^2}{1 - MK} \right] \leq \frac{M}{2}. \quad (57)$$

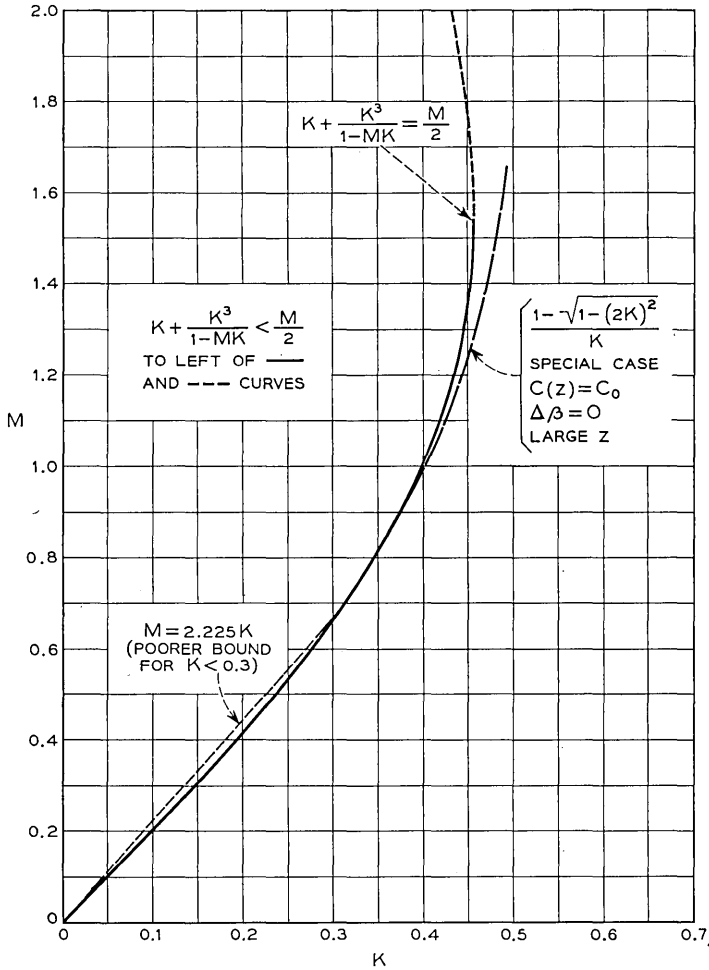


Fig. 1 — M vs K

A plot of M vs K taking (57) as an equality is shown in Fig. 1; the inequality of (57) is satisfied to the left of this curve. If for a given K we have chosen M to satisfy the inequality of (57), then since (54) holds true for $n = 2$, (50), it is valid for all n by induction. For a given K we should choose the smallest value of M satisfying (57) in order to obtain the best bound. This smallest value of M is given by the solid curve of Fig. 1 (i.e., $M < 1.554$); the other branch, indicated by the dashed curve (i.e., $M > 1.554$), thus has no significance for our problem. We

note that convergence of the series solution of (45), and hence of the successive approximations of (41), is guaranteed only for $0 \leq K \leq 0.455$; for greater values of K the present analysis cannot guarantee convergence.

Summarizing the above results:

$$\begin{aligned} |h_1(z)| &\leq e^{-\Delta\alpha z} \cdot K \\ |h_n(z)| &\leq e^{-\Delta\alpha z} \cdot M^{n-2} K^{n+1}; \quad n \geq 2. \end{aligned} \quad (58)$$

$$0 \leq K \leq 0.455.$$

M is given as a function of K by the solid curve of Fig. 1. If K is further restricted and if we are willing to degrade the bounds slightly, their form becomes simpler still. For example, if $0 \leq K \leq 0.3$ we may replace the bound of the solid curve on Fig. 1 by the slightly poorer dotted chord drawn from the origin. For this chord $M = 2.225 K$ and the results of (58) become:

$$\begin{aligned} |h_1(z)| &\leq e^{-\Delta\alpha z} \cdot K \\ |h_n(z)| &\leq e^{-\Delta\alpha z} \cdot K^3 (2.225 K^2)^{n-2}; \quad n \geq 2. \end{aligned} \quad (59)$$

$$0 \leq K \leq 0.3.$$

Finally, by (45), (57) and (58)

$$|H(z)| \leq e^{-\Delta\alpha z} \left[K + \sum_{n=2}^{\infty} M^{n-2} K^{n+1} \right] = e^{-\Delta\alpha z} \left[K + \frac{K^3}{1 - MK} \right], \quad (60)$$

$$|H(z)| \leq e^{-\Delta\alpha z} \cdot \frac{M}{2},$$

where M is again given as a function of K by the solid curve of Fig. 1.

Having found bounds on $h_n(z)$, we may now find bounds on $\lambda_n(z)$, our original objective. From (44a),

$$|\lambda_n(z)| \leq \int_0^z |c(s)| e^{\Delta\alpha s} |h_n(s)| ds; \quad n \geq 1. \quad (61)$$

From (58):

$$|\lambda_1(z)| \leq K \int_0^z |c(s)| ds.$$

$$|\lambda_n(z)| \leq M^{n-2} K^{n+1} \int_0^z |c(s)| ds; \quad n \geq 2. \quad (62)$$

$$|\Lambda(z)| \leq \frac{M}{2} \int_0^z |c(s)| ds.$$

$$0 \leq K \leq 0.455.$$

M is again given by Fig. 1. Again, if K is further restricted, simpler but slightly poorer results are obtained. For example:

$$|\lambda_1(z)| \leq K \int_0^z |c(s)| ds.$$

$$|\lambda_n(z)| \leq K^3(2.225K^2)^{n-2} \int_0^z |c(s)| ds; \quad n \geq 2. \quad (63)$$

$$0 \leq K \leq 0.3.$$

Finally, the slightly better bounds of (51-53) may be used for the smaller values of n .

We may again ask whether these bounds are the best that can be obtained. The answer is that we might be able to do a little better, but not much. Thus, consider the following special case:

$$\Delta\beta = 0,$$

$$c(z) = c_0 = \text{pure real.} \quad (64)$$

From (46) we have

$$K = \frac{c_0}{-\Delta\alpha}. \quad (65)$$

The solution to the coupled line equations, (35), subject to the initial conditions of (36), for this case may be written in the following form.¹

$$G_0(z) = -\frac{1 - \sqrt{1 - (2K)^2}}{2\sqrt{1 - (2K)^2}} \exp \frac{\Delta\alpha}{2} [1 + \sqrt{1 - (2K)^2}] z$$

$$+ \frac{1 + \sqrt{1 - (2K)^2}}{2\sqrt{1 - (2K)^2}} \exp \frac{\Delta\alpha}{2} [1 - \sqrt{1 - (2K)^2}] z \quad (66)$$

For $K < 0.5$, all of the radicals in (66) are pure real. Under these conditions the first term of (66) has a smaller coefficient and a more rapidly decaying exponential factor than the second term. Therefore, for a large enough value of z the second term dominates, and we may write

$$G_0(z) \approx \frac{1 + \sqrt{1 - (2K)^2}}{2\sqrt{1 - (2K)^2}} \exp \frac{\Delta\alpha}{2} [1 - \sqrt{1 - (2K)^2}] z; \quad (67a)$$

$$\Lambda = -\ln G_0(z) \approx -\ln \frac{1 + \sqrt{1 - (2K)^2}}{2\sqrt{1 - (2K)^2}}$$

$$- \frac{\Delta\alpha}{2} [1 - \sqrt{1 - (2K)^2}] z; \quad (67b)$$

$$K < 0.5, \quad -\Delta\alpha z \gg \frac{1}{\sqrt{1 - (2K)^2}}. \quad (67c)$$

The first term of (67b) is constant and the second increases linearly with z , so that we may write for large z

$$\Lambda \approx -\frac{\Delta\alpha}{2} [1 - \sqrt{1 - (2K)^2}] z = \frac{1 - \sqrt{1 - (2K)^2}}{2K} \cdot c_0 z, \quad (68)$$

large z .

The bound of (62) becomes simply

$$|\Lambda| \leq \frac{M}{2} c_0 z. \quad (69)$$

Therefore, a comparison between the bound of the present analysis and the exact results for the special case of (64) for large z is obtained by plotting

$$\frac{1 - \sqrt{1 - (2K)^2}}{K}$$

on Fig. 1 and comparing this quantity with M .

We see that for $K < 0.36$, the exact solution is indistinguishable from the bound of (69) on the plot of Fig. 1; consequently, little improvement may be obtained in these bounds without further restricting the problem. We also note that for $K > 0.5$, the above approximations made in the exact solution of (66) no longer apply. For $K > 0.5$ the character of the solution changes from monotonic to oscillatory; $G_0(z)$ now has periodic zeros, at which $\ln G_0(z)$ must approach infinity. Consequently, the series expansion for Λ in this case will diverge for $K > 0.5$. The present analysis guarantees convergence only for $K < 0.455$; while this might be a little smaller than necessary, the series solution may diverge for values of K not much larger.

IV. DISCUSSION

If K of (46) is very small compared to 1, $K \ll 1$, the bounds of (62-63) on $|\lambda_n(z)|$ converge very rapidly. Under these conditions it is tempting to assume that $\Lambda(z)$ is satisfactorily approximated by the first term of the summation of (45); i.e., from (41b),

$$\Lambda \approx \Lambda_1(z) = \int_0^z c(s) e^{\Delta\Gamma s} ds \int_0^s c(t) e^{-\Delta\Gamma t} dt, \quad (70)$$

or one of the alternative forms given in (27). [Alternately we might wish to make a similar statement for $G_0(z)$, as in Equation (24), when $\int_0^z |c(s)| ds \ll 1$.] This assumption has been made in all calculations of transmission statistics that have so far been made.¹

Unfortunately there is, at present, no satisfactory justification for this assumption. If $|\lambda_1(z)|$ turns out to be equal to its bound, as given in (62), and if $K \ll 1$, then, of course, we are guaranteed that the higher terms will have much smaller magnitudes than the principal term $\lambda_1(z)$. However, this situation is quite improbable, and occurs only in very specially selected cases. Thus if $|\lambda_1(z)|$ is much smaller than its bound, as will be the usual case, we have no assurance that the magnitude of the next term $|\lambda_2(z)|$ or higher terms may not be much greater than $|\lambda_1(z)|$. However, no instance is known in which $|\lambda_2(z)|$ is *not* small compared to $|\lambda_1(z)|$, for $K \ll 1$.

We do not know whether or not the perturbation solution of (70) provides a useful approximation for *all* cases of interest (i.e., for all cases where $c(z)$ satisfies (46) for some small value of K , e.g., $K = 0.1$). Even if this approximation fails in some cases, we may still hope that it holds true in most cases, so that (70) will yield the correct statistical properties of the loss and phase when the coupling coefficient is a stationary random process, perhaps Gaussian, with a sufficiently small rms value, at least for the simpler statistics of interest. Although this is believed to be true by a number of people, there is nothing in the present paper that bears on this question (and no other information known to the author). It would be most desirable to obtain further information on the way in which $\Lambda_1(z)$ of (70) approximates the true solution $\Lambda(z)$; e.g., does $\Lambda_1(z)$ approximate the fine structure of $\Lambda(z)$ as well as its average value as $\Delta\beta$ (which varies with the frequency of the applied wave) varies.

REFERENCES

1. Rowe, H. E., and Wartens, W. D., Transmission in Multimode Waveguide with Random Imperfections, B.S.T.J., this issue, pp. 1031-1170.
2. Unger, H. G., Helix Waveguide Theory and Application, B.S.T.J., **37**, November, 1958, pp. 1599-1647.
3. Ince, E. L., *Ordinary Differential Equations*, Dover, New York, N. Y., 1956.
4. Bellman, R., *Stability Theory of Differential Equations*, McGraw-Hill, New York, N. Y., 1953.

Transmission in Multimode Waveguide with Random Imperfections

By H. E. ROWE and W. D. WARTERS

(Manuscript received November 27, 1961)

The effects of random geometric imperfections on the transmission of the TE_{01} wave in circular waveguide are studied; the necessary theory of guides with known arbitrary imperfections is first developed. The TE_{01} transmission statistics are determined in terms of the statistics of the various types of geometric imperfections. Both discrete mode converters — i.e., localized imperfections such as tilts, offsets, or diameter changes at joints between pipes that are perfect right-circular cylinders — and continuous geometric imperfections — such as straightness deviation, diameter variation, ellipticity, etc., that vary smoothly with distance along the guide — are considered. The average, variance, power spectrum, and probability distribution of the TE_{01} loss-frequency curve are discussed.

Continuous straightness deviation (of the individual pipes of the guide) appears to be the most serious tolerance in present copper waveguide, and a significant factor in helix guide as well. The power spectrum of the straightness deviation is all-important in determining the TE_{01} loss due to mode conversion. Fourier components of straightness deviation having wavelengths between roughly 1.4 and 4.4 feet are the significant ones for the present 2-inch I.D. guide operated in a frequency band from 35 to 90 kmc.

TABLE OF CONTENTS

I. INTRODUCTION.....	1033
II. THEORY OF GUIDES WITH KNOWN IMPERFECTIONS.....	1034
2.1 Scattering Matrices of Discrete Mode Converters.....	1036
2.1.1 General Properties of Scattering Matrices.....	1037
2.1.2 Scattering Matrix for a Tilt.....	1038
2.1.3 Scattering Matrix for an Offset and a Diameter Change.....	1046
2.1.4 Discrete Mode Converters in Helix Guide.....	1050
2.2 The Discrete Case—Single Spurious Mode.....	1051
2.2.1 Matrix Analysis.....	1051
2.2.2 Perturbation Theory.....	1054
2.2.3 Discussion.....	1058
2.3 The Continuous Case—Single Spurious Mode.....	1059
2.3.1 Generalized Telegraphist's Equation.....	1060

2.3.2	Transformation from the Discrete to the Continuous Case..	1062
2.3.3	Transformation from the Continuous to the Discrete Case...	1067
2.3.4	Perturbation Theory for the Coupled Line Equations.....	1072
2.3.5	Transformation between Discrete and Continuous Perturbation Theory.....	1075
2.3.6	Logarithmic Form of the Coupled Lind Equations, and Improved Approximate Solution.....	1078
2.3.7	TE ₀₁ Loss in Terms of Fourier Coefficients of $c(z)$ when $\Delta\alpha = 0$	1081
2.3.8	Morgan's Coupling Coefficients for Small Cross-Sectional Deformations in Lossless Metallic Guide.....	1090
2.3.9	Relationships between Various Metallic Guide Coupling Coefficients.....	1092
2.4	Extension to Many Spurious Modes and Two Polarizations.....	1095
III.	THEORY OF GUIDES WITH RANDOM DISCRETE IMPERFECTIONS.....	1100
3.1	TE ₀₁ Loss—Summary of Previous Results.....	1101
3.2	Statistical Model of Guide.....	1103
3.2.1	Tilts and Offsets.....	1103
3.2.2	Diameter Changes.....	1104
3.3	Statistics of the TE ₀₁ Loss for a Single Section of Waveguide between Mode Filters.....	1105
3.3.1	Offsets.....	1105
3.3.2	Tilts.....	1109
3.3.3	Diameter Changes.....	1111
3.4	TE ₀₁ Loss Statistics of a Long Guide with Ideal Mode Filters.....	1115
3.5	Numerical Examples.....	1116
3.5.1	Offsets.....	1116
3.5.2	Tilts.....	1117
3.5.3	Diameter Changes.....	1118
3.6	Helix Guide.....	1118
3.7	Conclusions.....	1119
IV.	THEORY OF GUIDES WITH RANDOM CONTINUOUS IMPERFECTIONS.....	1120
4.1	TE ₀₁ Loss—Summary of Previous Results.....	1121
4.2	Statistics of Fourier Coefficients of $c(z)$	1122
4.3	TE ₀₁ Loss Statistics for a Single Section of Waveguide between Mode Filters.....	1124
4.3.1	Single Spurious Mode, Single Polarization.....	1124
4.3.2	Single Spurious Mode, Two Polarizations.....	1126
4.3.3	Many Spurious Modes.....	1128
4.3.4	Discussion.....	1130
4.4	TE ₀₁ Loss Statistics for Random Straightness Deviations.....	1130
4.4.1	Introduction.....	1130
4.4.2	Analysis.....	1133
4.4.3	Numerical Example.....	1136
4.4.4	Discussion.....	1137
4.5	TE ₀₁ Loss Statistics for Random Diameter Variations, Ellipticity, and Higher-Order Deformations.....	1138
4.5.1	Introduction.....	1138
4.5.2	Random Diameter Variations.....	1140
4.5.3	Random n -foils.....	1141
4.5.4	Discussion.....	1143
4.6	Conclusions.....	1146
	REFERENCES.....	1149

APPENDICES

A.	<i>Coupling Coefficients for Tilts, Offsets, and Diameter Changes.....</i>	1151
B.	<i>Geometry of Discrete Tilts.....</i>	1153
C.	<i>Energy Relations for Guides with Real Coupling Coefficients.....</i>	1155
D.	<i>Coupling Coefficients $\Xi_{[nm]}$ for General Continuous Deformations and Beat Wavelengths $B_{[nm]}^+$, for Metallic Guide.....</i>	1156
E.	<i>Geometry of Continuous Bends.....</i>	1160

F. <i>Rigorous Treatment of TE_{01} Loss Statistics for the Discrete Case</i>	1163
G. <i>Correlation Coefficient of TE_{01} Loss Components Due to Different Spurious Modes for the Continuous Case</i>	1168

I. INTRODUCTION

Long distance waveguide transmission via the TE_{01} mode in circular waveguide is an attractive goal because the theoretical TE_{01} heat loss decreases monotonically as the operating frequency increases. As is well known, operating frequencies far above the TE_{01} cutoff frequency are required to realize sufficiently low heat loss and delay distortion, so that the guide must operate far into the multimode region. Thus, considering a typical case, a 2-inch I.D. perfect copper circular guide operating at 55 kmc will have a theoretical TE_{01} heat loss of 1.54 db/mile; but this guide will propagate 223 additional modes, which we call spurious modes, at this frequency.¹

The TE_{01} transmission loss will approach the theoretical TE_{01} heat loss in a copper waveguide only if the waveguide is a geometrically perfect right-circular cylinder over its entire length. Any departure from this ideal geometry will couple the TE_{01} mode to some of the spurious modes^{2 to 7} the net effect of this coupling will be to increase the TE_{01} transmission loss above the theoretical heat loss, and to cause the TE_{01} transmission loss to vary with frequency.^{1,8}

Two types of geometric imperfections are of interest:

(a) Intentional deformations introduced in the guide for various reasons, e.g., to go around corners,^{3,6,7,9} to taper from a small guide to a larger one,¹⁰ etc. Mode conversion effects control the design of such devices, but we will not discuss them further.

(b) Random geometric imperfections arising during the manufacturing or the laying of the guide; these imperfections will increase the TE_{01} loss and cause it to vary randomly with frequency.⁸ The study of such effects is the purpose of the present paper.

The transmission characteristics of multimode waveguide with such random imperfections may be improved by adding heat loss to the spurious modes while keeping the TE_{01} heat loss low, i.e., close to its value for ideal copper guide. Examples are helix waveguide¹¹ and copper waveguide with a thin lossy dielectric lining.^{12,13} This additional spurious mode loss will reduce the TE_{01} loss fluctuations with frequency, but will not reduce the average TE_{01} loss. (In contrast, for large intentional bends, it is desirable to alter the phase constant of one particular copper guide mode without increasing the heat loss to either TE_{01} or any of the spurious modes.^{3,14})

The present paper is concerned primarily with determining the statistics of the TE_{01} transmission in terms of the statistics of the various geometric imperfections. The results of this analysis will indicate the required tolerances on the various types of geometric imperfections in different types of guide as a function of the allowable transmission degradation. Finally, the computed transmission statistics will be useful in determining the over-all degradation in various possible communication systems using imperfect waveguide as a transmission medium.

II. THEORY OF GUIDES WITH KNOWN IMPERFECTIONS

In this section we summarize the theory of circular waveguide with known geometric imperfections. These results yield, at least in principle, the TE_{01} transfer function for a circular waveguide with an arbitrary, known departure from perfect geometry. If the various geometric imperfections are assumed to be random processes, then the TE_{01} transfer function will also be a random process. In Sections III and IV these results for known imperfections are used to determine the TE_{01} transmission statistics in terms of the statistics of the various geometric imperfections.

We first require a solution to Maxwell's equations, in terms of the normal modes of the guide in question, for boundary conditions given by different types of geometric imperfections in various types of guide. Several people have studied these problems over the past ten years; we will give below a brief description of some of this work.

Transmission of TE_{01} through bends was first studied by M. Jouguet,⁶ and by S. O. Rice in unpublished work. S. E. Miller made use of these results to devise several methods for transmitting TE_{01} around intentional bends.⁷

S. P. Morgan first computed via perturbation theory the first-order spurious modes scattered from a unit incident TE_{01} wave by small, abrupt tilts, offsets, and diameter changes in ideal lossless metallic guide.¹⁵ These results were derived independently and published by S. Iiguchi.¹⁶ Such discontinuities will often be called discrete mode converters, because the guide possesses perfect cylindrical geometry except at isolated, discrete points along its axis.

Next, Morgan determined the first-order spurious modes scattered by an arbitrary small continuous deformation of the surface of an ideal lossless metallic guide, again via perturbation theory.⁵ These results may be used to determine the first-order spurious modes excited by small continuous diameter variation, straightness deviation, ellip-

ticity, and higher-order cross-sectional deformations of metallic guide. In addition, Morgan used these results to evaluate the expected value of the additional TE_{01} loss due to mode conversion caused by random distortion of the guide for a rather special mathematical model of the guide distortion.

Finally, Morgan applied the generalized telegraphist's equations of Schelkunoff¹ to the problem of lossless metallic waveguide with an arbitrary curvature of its axis;² if desired, the dielectric constant of the material filling the guide may be an arbitrary function of position. By this analysis, Maxwell's equations are reduced to an infinite set of coupled differential equations, the coupled line equations,¹⁷ in which the dependent variables are the complex mode amplitudes of the normal modes of the unperturbed metallic guide. In principle, the coupled line equations provide an exact description of a lossless metallic waveguide with an arbitrary straightness deviation, and are not subject to the severe restrictions of the perturbation theory which was described in the preceding paragraph. Thus, if this infinite set of differential equations could be solved for an arbitrary straightness deviation, we would have an exact solution for Maxwell's equations for the particular deformed guide. As will appear below, solutions to these equations in the general case are not available, and useful results are obtained only by applying perturbation theory of one form or another to the coupled line equations;¹⁸ however, the equations themselves are an exact description of the field problem.

Since the loss in real copper guide is low, we expect that the above coupling coefficients, which strictly speaking apply only to lossless metallic guide, will provide a good approximation for copper guide, and that the coupled line equations for lossless metallic guide need be modified only by changing the propagation constants for the various modes to take account of the small losses actually present in copper guide.

Equivalent results have been derived by H. G. Unger for various geometric imperfections in helix waveguide via the generalized telegraphist's equations;⁴ these analyses have been carried out both in terms of metallic guide modes and helix guide modes.¹¹ Unger has studied straightness deviations³ and cross-sectional deformations^{19,20} in helix, as well as winding imperfections²¹ in helix, and has given propagation constants and coupling coefficients for the various cases.²² He has similarly studied continuous diameter variations (tapers),¹⁰ which have identical behavior in both helix and metallic guide.

Thus, the study of random geometric imperfections in copper or helix waveguide has been reduced to the study of solutions to an infinite set

of differential equations with random coupling coefficients. However, this latter problem is a formidable one for which there is no really satisfactory treatment except in rather special cases. In the present paper, we use perturbation theory to approximate the solution to the coupled line equations; however, there is so far no rigorous justification for this approach. The convergence of this approximate solution is discussed elsewhere in this issue for the idealized case in which there are only two modes (rather than an infinite number of modes).¹⁸ Even in this simple case, we do not know how good an approximation the perturbation solution provides.

It is obvious that the various results of S. P. Morgan for metallic guide must be related to each other, even though they may appear somewhat dissimilar. In this section, after first developing the necessary theory for a long guide with many discrete mode converters, we show how the coupled line equations and Morgan's results for discrete mode converters may be derived from each other, and how perturbation theory derived from either the discrete case or from the coupled line equations yields equivalent results to perturbation theory applied directly to the field equations.⁵ This discussion is intended to provide a better physical understanding of the coupled line equations themselves, as well as of the approximate solution that we use, than would be obtained by merely a formal treatment based entirely on the coupled line equations. We will often simplify the problem by including only a few of the spurious modes (sometimes only a single spurious mode), in addition to the TE_{01} signal mode. While this procedure is useful in studying some of the basic problems, it of course does not provide a rigorous treatment for the real problem, which involves an infinite number of modes. However, experimental results for copper guides show that often only one or two spurious modes are present with significant magnitude,⁸ and thus provide some additional justification for the study of the idealized problem.

2.1 *Scattering Matrices of Discrete Mode Converters*

While S. P. Morgan has performed a field analysis for discrete mode converters in lossless metallic guide, many of the general properties of the scattering matrices for discrete mode converters may be derived from conservation of energy, reciprocity, and the symmetry properties of the different mode converters. We give such a discussion in the present section, making use of Morgan's results^{15,16} where necessary. Most of the discussion will be confined to cylindrical guides of infinite conduc-

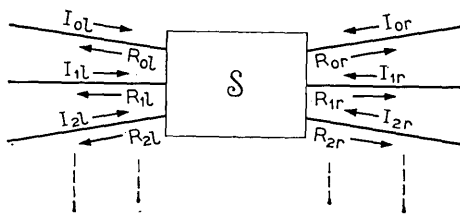


Fig. 1 — Generalized mode converter.

tivity. We choose the particular case of the discrete tilt to illustrate the general approach; briefer discussions are given for offsets and diameter changes.

2.1.1 *General Properties of Scattering Matrices*

Consider the general mode converter illustrated in Fig. 1. On the left of the obstacle we have the modes $0l, 1l, 2l, \dots$, and on the right the corresponding modes $0r, 1r, 2r, 3r, \dots$, where the letters l and r stand for “left” and “right” respectively. The subscript zero will denote the TE_{0l} mode and the other subscripts will denote the spurious modes. This convention will be used throughout this paper. We assume that there are the same number of propagating modes on each side of the obstacle, thus ruling out cases in which one of the spurious modes is below cutoff on one side of the obstacle, above cutoff on the other side. Cutoff modes are neglected throughout this analysis.

Denoting the normalized complex amplitudes of the modes incident on the obstacle from the right and left as I_{nr} and I_{nl} respectively, and of those leaving the obstacle as R_{nr} and R_{nl} respectively, the matrix equations²³ of the obstacle may be written:

$$\mathcal{R} = \mathcal{S}\mathcal{I}$$

$$\mathcal{R} = \begin{bmatrix} R_l \\ \vdots \\ R_r \end{bmatrix} \quad \mathcal{I} = \begin{bmatrix} I_l \\ \vdots \\ I_r \end{bmatrix}$$

$$R_l = \begin{bmatrix} R_{0l} \\ R_{1l} \\ R_{2l} \\ \vdots \end{bmatrix} \quad R_r = \begin{bmatrix} R_{0r} \\ R_{1r} \\ R_{2r} \\ \vdots \end{bmatrix} \quad (1)$$

$$I_l = \begin{bmatrix} I_{0l} \\ I_{1l} \\ I_{2l} \\ \vdots \end{bmatrix} \quad I_r = \begin{bmatrix} I_{0r} \\ I_{1r} \\ I_{2r} \\ \vdots \end{bmatrix}.$$

Morgan's results show that except very close to the cutoff of a spurious mode, the power scattered from TE_{01} to the forward modes greatly exceeds the power scattered to the backward modes for "small" mode converters. Consequently, in the following treatment we shall neglect all reflected waves. Using this assumption, and the fact that S must be symmetric (reciprocity),

$$s = \left[\begin{array}{c|c} 0 & \tilde{S} \\ \hline S & 0 \end{array} \right], \quad S = \begin{bmatrix} s_{00} & s_{01} & s_{02} & \cdots \\ s_{10} & s_{11} & s_{12} & \cdots \\ s_{20} & s_{21} & s_{22} & \cdots \\ \dots & \dots & \dots & \dots \end{bmatrix}, \quad (2)$$

and from (1),

$$\begin{bmatrix} R_l \\ R_r \end{bmatrix} = \begin{bmatrix} 0 & \tilde{S} \\ \hline S & 0 \end{bmatrix} \begin{bmatrix} I_l \\ I_r \end{bmatrix}, \quad (3)$$

where \tilde{S} denotes the transpose of the submatrix S . Thus,

$$R_l = \tilde{S}I_r, \quad (4a)$$

$$R_r = SI_l. \quad (4b)$$

If the obstacle is assumed lossless

$$\tilde{S}S^* = [1], \quad (5)$$

where the $*$ denotes the complex conjugate and $[1]$ denotes the unit matrix. From (2) we thus have

$$\tilde{S}S^* = [1] \quad \text{or equivalently} \quad S\tilde{S}^* = [1]. \quad (6)$$

2.1.2 Scattering Matrix for a Tilt

Consider the tilt of Fig. 2(a). Neglecting reflected waves, this mode converter may be characterized by the matrices of (3) or (4). Since the tilt is symmetric about the plane $A-A'$ of Fig. 2(a), its input and output terminals may be interchanged without altering its behavior. From this fact and (4), we have

$$S = \tilde{S}; \quad (7)$$

the matrix S is symmetric for a tilt. If we consider a tilt in perfectly conducting guide, so that energy is conserved, we have from (6) and (7)

$$SS^* = S^*S = [1]. \quad (8)$$

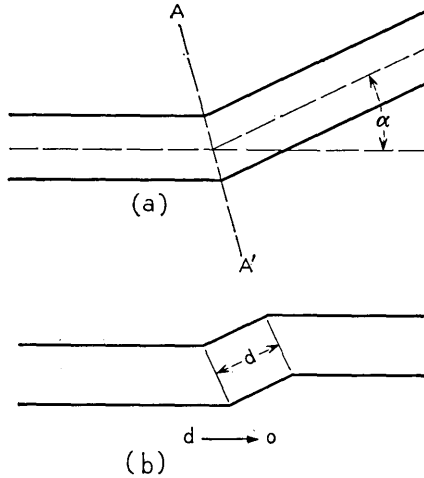


Fig. 2 — Waveguide with tilt.

Next, consider a cascade connection of the tilt of Fig. 2(a) and an identical tilt rotated 180° , as shown in Fig. 2(b); we may consider that the first tilt has an angle $+\alpha$, the second an angle $-\alpha$. Let the scattering matrix for the first tilt be S , for the second (rotated) tilt be S' . Then as the distance d between the two tilts approaches zero, the over-all scattering matrix of the two tilts becomes $S'S$ (neglecting reflected waves). But it is clear from physical considerations that this cascade connection of two opposite tilts of equal magnitude must be equivalent to a straight piece of guide (of zero length); thus we must have

$$S'S = SS' = [1]. \tag{9}$$

From (8) and (9) we have in the lossless case

$$S^* = S'. \tag{10}$$

By utilizing the symmetry properties of the various modes involved, further restrictions on the elements of S are easily found. For the sake of definiteness, consider the case where the only modes considered are the signal mode, TE_{01} , and the first-order forward spurious modes scattered by a discrete tilt from TE_{01} , i.e., one polarization of the forward TM_{11} and TE_{1m} modes, denoted by TM_{11}^+ and TE_{1m}^+ , as shown by Morgan.^{15,16} It is obvious by symmetry that only the (linear) polarization having an asymmetric transverse field distribution with respect to the plane of the tilt will be excited by an incident TE_{01} wave.²³ While

the backward TE_{1m} modes, denoted by TE_{1m}^- , are excited to first order by a tilt, their magnitudes are much smaller than the magnitudes of the corresponding forward spurious modes, as stated earlier, and consequently we neglect the backward modes for the present. We recall the convention of (1), i.e. the top elements in the R and I column vectors of (4), having the subscript 0, denote the TE_{01} signal mode; the other elements denote the various spurious modes. Let the elements of the S and S' matrices be s_{ij} and s_{ij}' respectively. Thus, for example, if a unit TE_{01} wave is incident on the tilt of Fig. 2(a) from the left, a TE_{01} wave of (complex) amplitude s_{00} will emerge on the right. We now observe that rotation of the tilt by 180° leaves the TE_{01} mode unaffected, but reverses the sign of the field components of all of the spurious modes, since their field components vary as $\cos \varphi$ or $\sin \varphi$. Consequently, the matrix components s_{ij}' for the rotated tilt are related to the matrix components s_{ij} as follows:

$$s_{ij}' = \begin{array}{ll} -s_{ij}; & i = 0, \quad j \neq 0. \\ & i \neq 0, \quad j = 0. \\ +s_{ij}; & i = 0, \quad j = 0. \\ & i \neq 0, \quad j \neq 0. \end{array} \quad (11)$$

From (10) we have

$$s_{ij}^* = s_{ij}'. \quad (12)$$

Equations (11) and (12) thus yield

$$s_{ij} = \begin{array}{ll} \text{pure imaginary;} & i = 0, \quad j \neq 0. \\ & i \neq 0, \quad j = 0. \\ \text{pure real;} & i = 0, \quad j = 0. \\ & i \neq 0, \quad j \neq 0. \end{array} \quad (13)$$

The coupling coefficients between TE_{01} and the spurious modes are pure imaginary, while all other matrix components are pure real.

Summarizing the above results, the scattering matrix S of a discrete tilt in lossless metallic guide must satisfy the following relations, if we include only TE_{01} and the propagating first-order forward spurious modes:

$$S = \tilde{S}, \quad s_{ij} = s_{ji}. \quad (14)$$

$$SS^* = [1]. \quad (15)$$

$$s_{ij} = \begin{cases} \text{pure imaginary;} & i = 0, & j \neq 0. \\ & i \neq 0, & j = 0. \\ \text{pure real;} & i = 0, & j = 0. \\ & i \neq 0, & j \neq 0. \end{cases} \quad (16)$$

As an example, let us determine the form of S for the case where only a single spurious mode is considered, in addition to the TE_{01} signal mode. (For a guide with a large intentional bend, the most significant spurious mode might be one polarization of forward TM_{11} . For a guide with a small random straightness deviation confined to a single plane, the most significant spurious mode might be one polarization of forward TE_{12} , as discussed in Section 2.3 below.) Then (14) to (16) yield for the S -matrix

$$S = \begin{bmatrix} \sqrt{1 - c^2} & jc \\ jc & \sqrt{1 - c^2} \end{bmatrix}. \quad (17)$$

For small tilts in lossless metallic guide, Morgan^{15,16} has given the coupling coefficients s_{0j} [or jc of (17)] in terms of the tilt angle α of Fig. 2(a) to first order in α , as follows;

$$s_{0j} = jC_{t(j)}^+ \cdot \alpha + \dots, \quad j \neq 0, \quad (18)$$

where the first j on the right-hand side of (18) represents $\sqrt{-1}$, the subscripts j indicate the spurious mode. $C_{t(j)}^+$ is a constant depending on the (forward) spurious mode; formulas and numerical values at a frequency of 55 kmc in 2-inch diameter guide for the C_t^+ 's are given in Appendix A. In addition, the coupling coefficients C_t^- to the corresponding backward spurious modes are also given in Appendix A; as indicated above, these are much smaller than the forward mode coupling coefficients.

Consider a tilt of angle α_1 followed by a tilt of angle α_2 ; as the distance between the two tilts approaches zero, it is obvious that the structure approaches simply a single tilt of angle $\alpha_1 + \alpha_2$. If $S(\alpha)$ is the matrix for a tilt of angle α , then the S -matrix must satisfy the further requirement that

$$S(\alpha_1) \cdot S(\alpha_2) = S(\alpha_1 + \alpha_2), \quad (19)$$

for every value of α_1 and α_2 , where we again neglect reflected modes. Consider again the idealized case where only a single spurious mode is allowed, in addition to the TE_{01} signal mode. We show in Section 2.3.3 that the "exact" matrix for a large tilt for this idealized two-mode case

is given by

$$S(\alpha) = \begin{bmatrix} \cos C_t \alpha & j \sin C_t \alpha \\ j \sin C_t \alpha & \cos C_t \alpha \end{bmatrix}, \quad (20)$$

where C_t is Morgan's coupling coefficient for the spurious mode in question. The matrix of (20) can readily be seen to satisfy the consistency condition of (19), and to approach the results of (17) and (18) for $C_t \alpha \ll 1$. However, the reader should be warned again that (20) will *not* be valid for large tilts in the physical case, because such large tilts will excite many spurious modes with significant magnitude, and hence can *not* be described in terms of only two modes.

The above results for tilts in metallic waveguide include only one polarization of each of the forward spurious modes, i.e., TE_{1m}^+ and TM_{11}^+ . We wish to extend these results to include both polarizations of each of these spurious modes so that we will be able to treat a long line containing arbitrary tilts with arbitrary angular orientation, i.e., *not* confined to a single plane.

We first write the general results of (4b) as follows, dropping as unnecessary the subscripts l and r since we will always assume that all modes travel in the forward direction, from left to right.

$$R] = [S] \cdot I]. \quad (21)$$

$$R] = \begin{bmatrix} R_0 \\ R_x \\ R_y \end{bmatrix}, \quad I] = \begin{bmatrix} I_0 \\ I_x \\ I_y \end{bmatrix}; \quad (22a)$$

$$R_x] = \begin{bmatrix} R_{1x} \\ R_{2x} \\ \vdots \end{bmatrix}, \quad R_y] = \begin{bmatrix} R_{1y} \\ R_{2y} \\ \vdots \end{bmatrix}, \quad I_x] = \begin{bmatrix} I_{1x} \\ I_{2x} \\ \vdots \end{bmatrix}, \quad I_y] = \begin{bmatrix} I_{1y} \\ I_{2y} \\ \vdots \end{bmatrix}. \quad (22b)$$

In these and following matrix relations, we adopt the convenient notation that column vectors ($n \times 1$ matrices) are denoted by the symbol], row vectors ($1 \times n$ matrices) by symbol $_$, and square matrices by the symbol [] where it is not obvious from the context that something else is intended (e.g., the column vectors on the right-hand sides of (22a) and (22b)). The top elements in (22a), R_0 and I_0 , represent the transmitted and incident TE_{01} wave. $R_x]$, $I_x]$ and $R_y]$, $I_y]$ are column vectors whose elements represent the two orthogonal (linear) polarizations of each of the spurious modes.

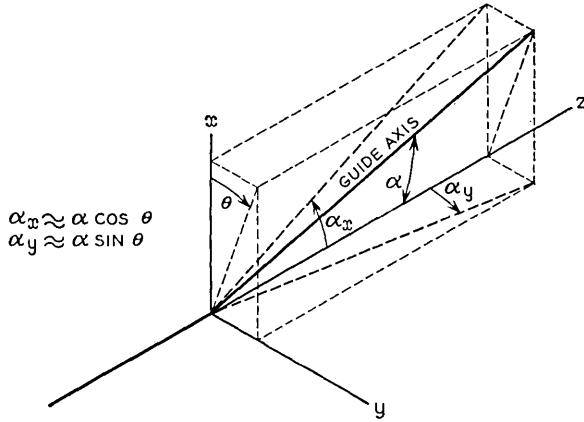


Fig. 3 — Geometry of a tilt in waveguide.

Consider the tilt of Fig. 3, in which the two polarizations of each of the TE_{1m}^+ and TM_{11}^+ spurious modes are defined along the x and y axes. To use the previous treatment including only a single polarization of each spurious mode, we must set $\theta = 0$ in Fig. 3, so that the plane of the tilt lies in one of the planes defining the spurious mode polarization. Then, using the notation of (22), we may write the results for the x -polarization as follows:

$$\begin{bmatrix} R_0 \\ R_x \end{bmatrix} = \begin{bmatrix} s_{00} & j\tilde{C} \\ jC & [D] \end{bmatrix} \cdot \begin{bmatrix} I_0 \\ I_x \end{bmatrix}, \tag{23}$$

$$C] = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix}, \quad jc_j \equiv s_{0j}. \tag{24}$$

The components of the column vector of (24) are given to first order by (18). In (23), we have partitioned the S -matrix to conform to the partitioning of the R and I column vectors. The restrictions of (14) to (16) become respectively, in terms of the submatrices of (23):

$$[D] = [\tilde{D}]. \tag{25}$$

$$s_{00}^2 + \sum_j c_j^2 = 1,$$

$$(s_{00} - [D]) \cdot C] = 0], \tag{26}$$

$$C] \cdot \tilde{C} + [D]^2 = [1].$$

$$s_{00}, \quad C], \quad [D] \quad \text{pure real.} \quad (27)$$

In (26), $[1]$ denotes the unit matrix.

Now by symmetry an incident TE_{01} wave in this particular tilt ($\theta = 0$) will not excite the y -polarization of any of the spurious modes as stated earlier, and by reciprocity the y -polarization of any of the spurious modes will not excite TE_{01} . Further, the x - and y -polarizations of all of the spurious modes are uncoupled from each other by symmetry. Therefore, we may expand the matrix relation of (23) to include the y -polarization of the spurious modes as follows:

$$\begin{bmatrix} R_0 \\ R_x \\ R_y \end{bmatrix} = \begin{bmatrix} s_{00} & j\tilde{C} & 0 \\ jC & [D] & [0] \\ 0 & [0] & [1] \end{bmatrix} \cdot \begin{bmatrix} I_0 \\ I_x \\ I_y \end{bmatrix}, \quad (28)$$

where (18) and (24) to (27) still apply.

Now let us rotate the axes by an angle $-\theta$ with respect to the plane of the tilt, as shown in Fig. 3, and write the field in the guide in terms of modes referred to these new axes. The geometry of the rotated tilt may be specified precisely in the following way. Imagine that before the guide is tilted, lines $\varphi = \text{constant}$ are drawn on the surface of the guide parallel to the guide axis. Then the tilt, of orientation θ , is constructed at a specified point on the axis by tilting the guide in the plane defined by the axis and the $\varphi = \theta$ line, by an angle α . If

$$\begin{aligned} \rho &= \text{distance from the (tilted) guide axis in a plane perpendicular to the guide axis,} \\ s &= \text{distance measured along the (tilted) guide axis,} \end{aligned} \quad (29)$$

the three coordinates ρ, φ, s constitute "bent cylindrical coordinates," as used by Morgan for continuous bends confined to a single plane.² In subsequent analysis for a guide with many tilts of arbitrary orientation, we will adopt the convention that $\alpha > 0$ while θ is unrestricted; in contrast, for a guide with tilts confined to a single plane, we will set $\theta = 0$ and allow α to be unrestricted.

Now let $R_0]$ and $I_0]$ denote the fields with respect to the old axes and $R_\theta]$ and $I_\theta]$ the fields with respect to the rotated axes. First, we note that TE_{01} is the same in both sets of coordinates. Next, we note that the field components of all of the spurious modes under consideration, i.e., TE_{1m}^+ and TM_{11}^+ , vary as $\cos \varphi$ or $\sin \varphi$. If we call φ_0 the "old" and φ_θ the "rotated" angular coordinates, the transformation is simply

$$\varphi_\theta = \varphi_0 + \theta. \quad (30)$$

For each spurious mode this yields in terms of wave amplitudes

$$\begin{bmatrix} R_{x_0} \\ R_{y_0} \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} R_{x_\theta} \\ R_{y_\theta} \end{bmatrix}, \tag{31}$$

with a similar result for the I 's. Note that (31) applies to a single spurious mode only (two polarizations); the index denoting the particular mode has been omitted for convenience. Thus, the R 's (or I 's) in (31) are single (complex) numbers, and not column vectors. We may write the corresponding general rotation matrix including TE_{01} and all spurious modes in the following convenient form:

$$R_0 = [M] \cdot R_\theta, \quad I_0 = [M] \cdot I_\theta. \tag{32a}$$

$$[M] = \begin{bmatrix} 1 & \underline{0} & \underline{0} \\ 0 & \cos \theta \cdot [1] & \sin \theta \cdot [1] \\ 0 & -\sin \theta \cdot [1] & \cos \theta \cdot [1] \end{bmatrix}. \tag{32b}$$

$$\tilde{M} = M^{-1}. \tag{32c}$$

R and I in (32a) are as given in (22); the $[1]$'s in (32b) represent the unit matrix. We note from (32c) that M is an orthogonal matrix.

Now rewriting (28) as

$$R_0 = [S_0] \cdot I_c, \tag{33}$$

we substitute the relations of (32a) into (33) to obtain

$$\begin{aligned} [M] \cdot R_\theta &= [S_0] \cdot [M] \cdot I_\theta \\ R_\theta &= [M]^{-1} \cdot [S_0] \cdot [M] \cdot I_\theta \\ &= [\tilde{M}] \cdot [S_0] \cdot [M] \cdot I_\theta. \end{aligned} \tag{34}$$

Therefore,

$$R_\theta = [S_\theta] \cdot I_\theta, \tag{35a}$$

$$[S_\theta] = [\tilde{M}] \cdot [S_0] \cdot [M]. \tag{35b}$$

Substituting $[S_0]$ from (28) and $[M]$ from (32b), we have

$$[S_\theta] = \begin{bmatrix} s_{00} & j \cos \theta \cdot \underline{\tilde{C}} & j \sin \theta \cdot \underline{\tilde{C}} \\ j \cos \theta \cdot C & \cos^2 \theta \cdot [D] + \sin^2 \theta \cdot [1] & \sin \theta \cos \theta ([D] - [1]) \\ j \sin \theta \cdot C & \sin \theta \cos \theta ([D] - [1]) & \sin^2 \theta \cdot [D] + \cos^2 \theta \cdot [1] \end{bmatrix} \tag{36}$$

as the scattering matrix for a tilt in lossless metallic guide, having an orientation of θ radians with respect to the axes defining the polarization of the spurious modes. C] remains as defined in (24), with components given to first order by (18), and the restrictions of (25) to (27) still apply.

The above transformation may readily be seen to yield the correct results in a few simple cases:

1. $\theta = \pi$. $[S_x]$ is readily seen to be identical to the matrix $[S']$ described in connection with (11).

2. $\theta = \pi/2$. Here we see that $R_x]_{\theta=0} = R_y]_{\theta=\pi/2}$, $R_y]_{\theta=0} = -R_x]_{\theta=\pi/2}$, and similarly for the I 's, as is obvious from geometric considerations.

The coupling coefficients between TE_{01} and the x - and y -polarizations of the j^{th} spurious mode are from (36), (24) and (18)

$$s_{0j,x} = jC_{i(j)}^+ \cdot (\alpha \cos \theta), \quad (37a)$$

$$s_{0j,y} = jC_{i(j)}^+ \cdot (\alpha \sin \theta). \quad (37b)$$

These general results simplify so that they may be simply expressed in terms of fixed x , y , z coordinates in the following special case:

1. The angular deviation of the guide axis from the z -axis is small.

2. The unit vector perpendicular to the guide axis and lying in the plane defined by the guide axis and the $\varphi = 0$ line is almost parallel to the x -axis. These conditions insure that the $\varphi = 0$ line drawn on the guide will remain almost parallel to the xz plane. Subject to these conditions, we have approximately

$$\alpha \cos \theta \approx \alpha_x, \quad (38a)$$

$$\alpha \sin \theta \approx \alpha_y, \quad (38b)$$

where α_x and α_y are the angles made by the projections of the guide axes, adjacent to the tilt, in the xz and yz planes respectively. These results are readily verified for the case shown in Fig. 3. They are derived in Appendix B. Substitution of (38) into (37) [and thence into the matrix of (36)] will greatly simplify certain later calculations.

2.1.3 Scattering Matrix for an Offset and a Diameter Change

The scattering matrix for an offset may be found by similar methods as used in Section 2.1.2 above for a tilt. Consider the offset of Fig. 4(a), with scattering matrix S ; as above we assume the offset is small, so that backward modes may be neglected and only the propagating first-order forward spurious modes (i.e., TE_{1m}^+)^{15,16} need be considered. In Fig.

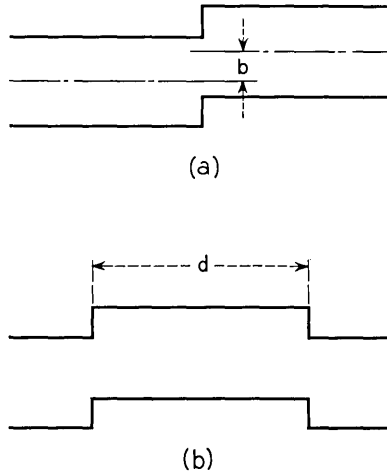


Fig. 4 — Waveguide with offset.

4(b), this offset is cascaded with an identical offset, which may be derived from the original offset in two ways:

1. Reverse input and output terminals.
2. Rotate the first offset 180°.

From 1, the matrix for the second offset is \tilde{S} , the transpose of S . From 2, the matrix for the second offset is S' , as defined in (11) above, by an argument similar to that used for the tilt. From these facts, as the distance d separating the offsets approaches zero,

$$\tilde{S}S = S\tilde{S} = [1], \tag{39a}$$

$$S'S = SS' = [1]. \tag{39b}$$

Assuming lossless metallic guide, from (6)

$$\tilde{S}S^* = S\tilde{S}^* = [1]. \tag{40}$$

From (39) and (40) we have for an offset in lossless metallic guide:

$$\tilde{S}S = [1]. \tag{41}$$

$$s_{ij} = \text{pure real} = \begin{matrix} -s_{ji} & i = 0, & j \neq 0. \\ & i \neq 0, & j = 0. \\ +s_{ji} & i = 0, & j = 0. \\ & i \neq 0, & j \neq 0. \end{matrix} \tag{42}$$

For a single spurious mode the matrix S for an offset becomes

$$S = \begin{bmatrix} \sqrt{1 - c^2} & c \\ -c & \sqrt{1 - c^2} \end{bmatrix}, \quad (43)$$

corresponding to (17) for a tilt. For small offsets in lossless metallic guide, Morgan^{15,16} has given the coupling coefficients s_{0j} [or c of (43)] in terms of the offset b of Fig. 4(a) to first order in b , as follows:

$$s_{0j} = C_{o(j)}^+ \cdot b + \dots, \quad j \neq 0. \quad (44)$$

$C_{o(j)}^+$ is a constant depending on the spurious mode; formulas and numerical values for the C_o^\pm 's at a frequency of 55 kmc in 2-inch diameter guide are given in Appendix A.

Analogous results to those of (19) to (38) for tilts are readily found for offsets, but will not be discussed in detail here. In particular, the scattering matrix for an offset with an arbitrary angular orientation is found in the same way as given in (21) to (37), making use of the rotation operator of (32). The geometry for offsets of arbitrary orientation is much simpler than for tilts. As before, imagine that lines $\varphi = \text{constant}$ are drawn on the surface of the initially perfect guide, parallel to the guide axis. Then the offset of Fig. 5 is constructed at a specified point on the guide axis by translating the guide a distance b in the $\varphi = \theta$ plane without rotating the guide, so that corresponding φ -lines on the two sides of the offset are separated by a distance b . For a guide with many offsets of arbitrary orientation we take $b > 0$ with θ unrestricted, while for a guide with offsets confined to a single plane we set $\theta = 0$ and allow b to be unrestricted, as in the case of tilts. Then using the notation of Section 2.1.2, the scattering matrix S_θ for an offset of orientation θ is given as follows:

$$S_\theta = \begin{bmatrix} s_{00} & \cos \theta \cdot \underline{\tilde{C}} & \sin \theta \cdot \underline{\tilde{C}} \\ -\cos \theta \cdot C & \cos^2 \theta \cdot [D] + \sin^2 \theta \cdot [1] & \sin \theta \cos \theta ([D] - [1]) \\ -\sin \theta \cdot C & \sin \theta \cos \theta ([D] - [1]) & \sin^2 \theta \cdot [D] + \cos^2 \theta \cdot [1] \end{bmatrix} \quad (45)$$

where

$$C = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix}, \quad c_j \equiv s_{0j}. \quad (46)$$

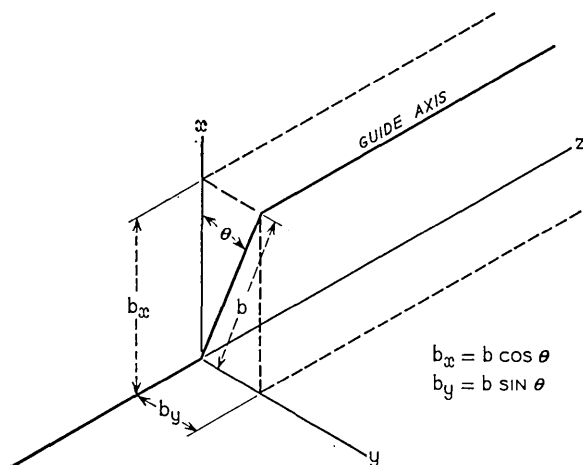


Fig. 5 — Geometry of an offset in waveguide.

The components of $C]$ are given to first order by (44). Equation (45) (for offsets) has been written in such a way that the restrictions of (41) and (42) become identical to those of (25) to (27) (for tilts).

The coupling coefficients between TE_{01} and the x - and y -polarizations of the j^{th} spurious mode are from (44) to (46)

$$s_{0j,x} = C_{o(j)}^+ \cdot (b \cos \theta), \tag{47a}$$

$$s_{0j,y} = C_{o(j)}^+ \cdot (b \sin \theta). \tag{47b}$$

From Fig. 5 it is readily seen that $b_x = b \cos \theta$ and $b_y = b \sin \theta$ are exactly the x - and y -components of the offset, in analogy to the approximate results of (38) for tilts.

A similar treatment may be applied to a discrete diameter change. Here the spurious modes are the higher order TE_{0m} ; again for small offsets only the forward modes need be considered. This case differs in one fundamental respect from that of the offset. For an offset the TE_{01} signal mode and the TE_{1m} spurious modes have a different angular dependence; for a diameter change the TE_{01} signal mode and the higher order TE_{0m} spurious modes have the same (i.e., no) angular dependence. Thus, for a diameter change the signal and all spurious modes are coupled to each other to first order.

First by conservation of energy (true in helix as well as copper guide, because the TE_{0m} modes are the same in both) (6) yields

$$\tilde{S}S^* = S\tilde{S}^* = [1]. \tag{48}$$

Next, consider the cascade connection of two identical diameter changes connected back-to-back. As in the case of an offset, this yields

$$\tilde{S}S = S\tilde{S} = [1]. \quad (49)$$

From (48) and (49) we find that

$$S = S^* = \text{pure real}. \quad (50)$$

However, there is no operation corresponding to the 180° rotation, used for tilts and offsets, because the signal and spurious modes have the same symmetry in the present case. While we have obtained no further information from general considerations than contained in (49) and (50), for the case of a single spurious mode the scattering matrix for a diameter change is identical to that of (43) for an offset.

For small diameter changes, Morgan^{15,16} gives the coupling coefficients s_{0j} in terms of the change in radius Δr as follows:

$$s_{0j} = C_{a(j)}^\dagger \cdot \Delta r + \dots, \quad j > 0. \quad (51)$$

$C_{a(j)}^\pm$ is given in Appendix A.

There is of course only a single polarization of each spurious mode in this case, and consequently, there is no analysis in the present case corresponding to those for tilts and offsets with two polarizations.

2.1.4 Discrete Mode Converters in Helix Guide

While the above results for diameter changes apply equally well to both helix and copper guide, those for tilts and offsets apply to only copper guide (strictly speaking, ideal lossless metallic guide). This is so because (6) no longer holds true in helix; energy is not conserved in helix, and in addition the various normal modes of helix are not even orthogonal with respect to power.²² The coupling coefficients for discrete mode converters in helix have been obtained by Unger.¹⁹ However, one useful result is readily obtained from general considerations without performing a detailed field analysis.

We show that at a discrete tilt or offset, the TE_{01} transfer coefficient s_{00} (the upper left-hand element in the scattering matrices for a tilt or offset given in Sections 2.1.2 and 2.1.3 above) is identical in both copper and helix guides, if we neglect backward modes and include all forward modes. This fact will permit the average TE_{01} loss in helix guide with discrete random tilts and offsets to be readily calculated in terms of similar results for copper guide.

The above statement is proven as follows. First consider an offset or tilt as a joint between a helix and a copper guide, with a unit TE_{01}

wave incident from the helix. If backward waves are neglected, the forward waves in the copper guide will be identical to those in a copper-copper joint because the incident TE_{01} is the same in helix as it would be in copper. Therefore, the TE_{01} transfer coefficient s_{00} is the same for this helix-copper joint as for a copper-copper joint. Now reverse input and output terminals, so that the TE_{01} is incident from the copper guide, and TE_{01} and spurious modes travel away from the joint in helix. The spurious modes are now quite different than before, since they must be normal modes of helix guide; however, by reciprocity s_{00} must remain the same. Finally, we may replace the copper guide containing the incident TE_{01} with helix without further altering the fields in any way. Thus if backward modes are neglected, s_{00} is identical in helix and copper guides with identical tilts or offsets.

This conclusion has been verified experimentally by the authors in 2-inch diameter helix guide at a frequency of 55 kmc. It has also been verified by a field analysis by H. G. Unger.²⁰

2.2 *The Discrete Case — Single Spurious Mode*

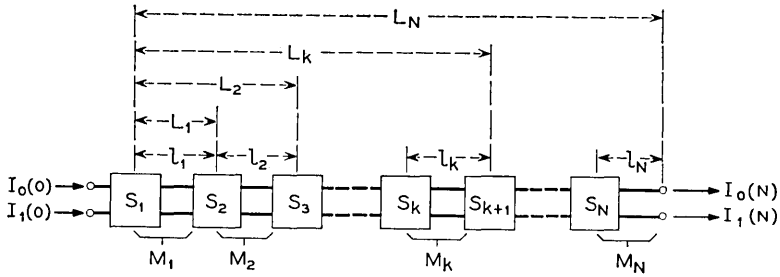
We next apply the results of Section 2.1 to the study of long guides with many discrete mode converters separated by guide sections that are ideal, i.e., geometrically perfect right-circular cylinders. We restrict our attention to the case of a single spurious mode, in addition to the TE_{01} signal mode. (If the spurious mode is polarized, such as TE_{1m}^+ , we consider only one of its linear polarizations.)

The over-all transmission matrix for such a guide with N discrete mode converters consists of a product of $2N$ matrices, one matrix for each mode converter (as given in the preceding section) and one diagonal matrix for each section of ideal guide. These matrix results are then used to derive a perturbation theory, valid when the mode converters are sufficiently small, that greatly simplifies further calculations.

2.2.1 *Matrix Analysis*

Consider the guide illustrated schematically in Fig. 6. This guide consists of N discrete mode converters separated by N sections of ideal copper guide, of length l_k . Since we consider only a single spurious mode, if the discrete mode converters are tilts or offsets they must lie in a single plane (taken to be the $\varphi = 0$ plane in the notation of Section 2.1). We seek the response of such a guide to a unit input TE_{01} wave.

The scattering matrix S_k for the k^{th} mode converter is given in Section 2.1 for the case of a single spurious mode, i.e., (17) and (18) for tilts,



$$L_k = \sum_{m=1}^k l_m$$

S_k = SCATTERING MATRIX FOR k^{TH} MODE CONVERTER

M_k = SCATTERING MATRIX FOR k^{TH} MODE CONVERTER PLUS FOLLOWING SECTION OF GUIDE

I_0 = SIGNAL MODE (TE_{01}) NORMALIZED AMPLITUDE

I_1 = SPURIOUS MODE NORMALIZED AMPLITUDE

Fig. 6 — Waveguide line with discrete mode converters separated by ideal guide sections.

(43) and (44) for offsets, (43) and (51) for diameter changes. In the ideal guide sections connecting the discrete mode converters, the signal and spurious modes propagate independently with their respective propagation constants; the scattering matrix W_k corresponding to the k^{th} section of ideal guide, of length l_k as shown in Fig. 6, is given by

$$W_k = \begin{bmatrix} e^{-\Gamma_0 l_k} & 0 \\ 0 & e^{-\Gamma_1 l_k} \end{bmatrix}, \tag{52}$$

where Γ_0 is the propagation constant for the TE_{01} signal mode, Γ_1 the propagation constant for the spurious mode. Thus, the scattering matrix M_k for the k^{th} mode converter plus the (following) k^{th} section of guide, as shown in Fig. 6, is

$$M_k = W_k \cdot S_k. \tag{53}$$

For tilts M_k is given as

$$M_k = \begin{bmatrix} e^{-\Gamma_0 l_k} \cdot \sqrt{1 - c_k^2} & je^{-\Gamma_0 l_k} \cdot c_k \\ je^{-\Gamma_1 l_k} \cdot c_k & e^{-\Gamma_1 l_k} \cdot \sqrt{1 - c_k^2} \end{bmatrix}, \tag{54}$$

with c_k given by (18). For offsets and diameter changes

$$M_k = \begin{bmatrix} e^{-\Gamma_0 l_k} \cdot \sqrt{1 - c_k^2} & e^{-\Gamma_0 l_k} \cdot c_k \\ -e^{-\Gamma_1 l_k} \cdot c_k & e^{-\Gamma_1 l_k} \cdot \sqrt{1 - c_k^2} \end{bmatrix}, \tag{55}$$

with c_k given by (44) and (51) respectively.

The over-all transmission matrix T for the entire guide, N sections long, will be

$$T = \begin{bmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{bmatrix} = \prod_{k=N}^1 M_k = M_N \cdot M_{N-1} \cdots M_2 \cdot M_1, \quad (56)$$

and the output mode amplitudes will be given in terms of the input mode amplitudes by

$$\begin{bmatrix} I_0(N) \\ I_1(N) \end{bmatrix} = T \begin{bmatrix} I_0(0) \\ I_1(0) \end{bmatrix}. \quad (57)$$

We assume the guide is excited by a pure TE₀₁ wave of unit amplitude and zero phase; the initial conditions on (57) become

$$I_0(0) = 1, \quad I_1(0) = 0. \quad (58)$$

Then,

$$I_0(N) = T_{00}I_0(0) = T_{00}, \quad \text{TE}_{01} \text{ gain.} \quad (59)$$

$$I_1(N) = T_{10}I_0(0) = T_{10}, \quad \text{TE}_{01}\text{-spurious mode transfer coefficient.} \quad (60)$$

It will subsequently be convenient to normalize these quantities as follows:

$$T_{00} = e^{-\Gamma_0 L_N} \cdot G_0; \quad G_0 = \text{TE}_{01} \text{ normalized gain.} \quad (61)$$

$$T_{10} = e^{-\Gamma_1 L_N} \cdot G_1; \quad G_1 = \text{TE}_{01}\text{-spurious mode normalized transfer coefficient.} \quad (62)$$

In each case, the propagation factor of the corresponding mode has been removed; as shown in Fig. 6, L_N is the total length of the N sections of guide being considered. Since TE₀₁ has a lower heat loss (i.e., $\alpha_0 = \text{Re } \Gamma_0$) than any other mode, in a physical guide $G_0 \leq 1$.

The exact solution above is of limited value both because of its complexity and also because, as discussed in Section 2.1, the available expressions for the coupling coefficients of discrete mode converters are valid only to first order. Consequently, we seek approximate expressions for T_{00} and T_{10} , or equivalently G_0 and G_1 , valid when the coupling coefficients are sufficiently small that the guide departs only slightly from ideal. Under these conditions, it will be shown that G_1 is of first order and G_0 departs from unity only to second order.

It is convenient to write

$$G_0 = 1 - \rho, \quad (63)$$

where ρ , the (complex) departure of G_0 from unity, will be of second order. We further define the TE₀₁ normalized magnitude g , complex loss Λ , loss (in nepers) A , and phase Θ as follows:

$$\begin{aligned} G_0 &= ge^{j\Theta} = e^{-\Lambda} = e^{-A}e^{j\Theta} = 1 - \rho; \\ \Lambda &= A - j\Theta. \end{aligned} \quad (64)$$

Then if the coupling coefficients are sufficiently small so that $|\rho| \ll 1$, we have to second order:

$$g \approx 1 - \text{Re } \rho. \quad (65a)$$

$$\Lambda \approx \rho. \quad (65b)$$

$$A \approx \text{Re } \rho. \quad (65c)$$

$$\Theta \approx -\text{Im } \rho. \quad (65d)$$

2.2.2 Perturbation Theory

Consider the transmission matrix T of (56), with the M_k given by (54) or (55). Let us expand the square root in the diagonal elements of (54) or (55) in a power series as follows;

$$\sqrt{1 - c_k^2} = 1 - \frac{1}{2}c_k^2 - \frac{1}{8}c_k^4 + \dots \quad (66)$$

It is apparent from the rules of matrix multiplication that the components of T may be expressed as power series in the c_k 's. Since each c_k may be expressed as a power series in the appropriate geometric parameter, with the first term given by (18), (44) or (51), we can thus obtain expressions for G_1 and G_0 as power series in the geometric parameters. The first-order results of (18), (44) and (51) are sufficient to give G_1 to first order, G_0 to second order; if the mode converters are sufficiently small we may hope that these results will give a valid approximation for the TE₀₁ gain and the TE₀₁-spurious mode transfer coefficient.

We first determine the first few terms of expansions for T_{10} and T_{00} . For convenience, we write for M_k in (56)

$$M_k = \begin{bmatrix} A_k & \epsilon_k \\ \delta_k & B_k \end{bmatrix}, \quad (67)$$

where A_k , ϵ_k , δ_k and B_k are determined by comparison with (54) or

(55). For small coupling, A_k and B_k will have magnitudes a little less than 1, while ϵ_k and δ_k will have magnitudes much smaller than 1. In writing out T_{10} and T_{00} , we group terms according to the number of "small" quantities (i.e., ϵ 's and δ 's) they contain. Thus, we have:

$$T_{10} = \sum_{i=1}^N A_1 \cdots A_{i-1} \delta_i B_{i+1} \cdots B_N + \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^N A_1 \cdots A_{i-1} \delta_i B_{i+1} \cdots B_{j-1} \epsilon_j A_{j+1} \cdots A_{k-1} \delta_k B_{k+1} \cdots B_N + \dots \tag{68}$$

$$T_{00} = A_1 \cdots A_N + \sum_{i=1}^{N-1} \sum_{j=i+1}^N A_1 \cdots A_{i-1} \delta_i B_{i+1} \cdots B_{j-1} \epsilon_j A_{j+1} \cdots A_N + \sum_{i=1}^{N-3} \sum_{j=i+1}^{N-2} \sum_{k=j+1}^{N-1} \sum_{l=k+1}^N A_1 \cdots A_{i-1} \delta_i B_{i+1} \cdots B_{j-1} \epsilon_j A_{j+1} \cdots A_{k-1} \delta_k B_{k+1} \cdots B_{l-1} \epsilon_l A_{l+1} \cdots A_N + \dots \tag{69}$$

Consider first (68) for T_{10} . Referring to (67), (66), (54) and (55), the successive terms of (68) (of which we have written down the first two) contain components of the following orders in the c_k :

<i>term</i>	<i>order of components</i>
1	c, c^3, c^5, \dots
2	c^3, c^5, c^7, \dots

Equation (18), (44) or (51) gives the c_k to first order in terms of the tilts α_k , offsets b_k , or diameter changes Δr_k . Therefore, we may obtain T_{10} to first order in the geometric parameters by retaining only the components of order c , which occur in only the first term of (68) (the single summation), and using Morgan's first-order coupling coefficients. Indeed, since we do not have the coupling coefficients to more than first order, it would be totally unjustified to retain any additional components in (68), and in particular, to seek an exact result for the matrix multiplication of (56) via numerical techniques or otherwise, without first obtaining the coupling coefficients c_k to higher order.

Similarly, the successive terms of (69) for T_{00} contain components of the following orders in the c_k :

<i>term</i>	<i>order of components</i>
1	c^0, c^2, c^4, \dots
2	c^2, c^4, c^6, \dots

We obtain T_{00} to second order in the geometric parameters by retaining only the components of order c^0 and c^2 in the first term and the components of order c^2 in the second term, and using Morgan's first-order coupling coefficients. As above, any additional terms would be unjustified. We note that there are no first-order terms present in T_{00} .

We thus obtain the following approximate results for G_0 and G_1 , the normalized transmission parameters defined in (61) to (64), from (66) to (69), (54), (55) and (18), (44) and (51). In these results we use the differential propagation constant $\Delta\Gamma$, defined as

$$\Delta\Gamma = \Delta\alpha + j\Delta\beta = \Gamma_0 - \Gamma_1. \quad (70a)$$

$$\Delta\alpha = \alpha_0 - \alpha_1, \quad \Delta\beta = \beta_0 - \beta_1. \quad (70b)$$

The real and imaginary parts of $\Delta\Gamma$, $\Delta\alpha$ and $\Delta\beta$ respectively, are called the differential attenuation constant and the differential phase constant respectively. Since the TE_{01} signal mode has lower heat loss than any other mode, $\Delta\alpha < 0$ throughout the present paper. The geometry of the guide is shown in Fig. 6; L_{k-1} is the length of guide up to the k^{th} mode converter, l_k the distance between the k^{th} and the $(k+1)^{\text{th}}$ mode converters. x_i denotes Morgan's first-order approximation for the coupling coefficient of the i^{th} mode converter.

For tilts and offsets the diameter of the different guide sections is of course identical, so that $\Delta\Gamma$ is strictly constant. This is not true for diameter changes, so that strictly speaking we should include in the analysis the fact that $\Delta\Gamma$ changes from section to section. However, we assume that the total range of the guide diameter is very small, centered about its nominal value, and neglect the small changes in $\Delta\Gamma$ in all following analysis, both for the discrete case and for the continuous case. This approximation is not necessary; it would be possible to include the variation of $\Delta\Gamma$ in the present analysis without great difficulty. However, we choose to ignore this effect without careful study in the interests of simplicity.

i. G_0 — Tilts, Offsets and Diameter Changes

$$G_0 = 1 - \rho. \quad (71a)$$

$$\begin{aligned} \rho &\approx \frac{1}{2} \sum_{i=1}^N x_i^2 + \sum_{i=1}^{N-1} \sum_{j=i+1}^N x_i x_j e^{\Delta\Gamma(L_{j-1} - L_{i-1})} \\ &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N x_i x_j e^{\Delta\Gamma|L_{j-1} - L_{i-1}|}. \end{aligned} \quad (71b)$$

Special case — equally spaced mode converters

$$l_k = l_0, \quad L_k = kl_0. \tag{71c}$$

$$\begin{aligned} \rho &\approx \frac{1}{2} \sum_{i=1}^N x_i^2 + \sum_{i=1}^{N-1} \sum_{j=i+1}^N x_i x_j e^{\Delta\Gamma l_0(j-i)} \\ &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N x_i x_j e^{\Delta\Gamma l_0|j-i|} \end{aligned} \tag{71d}$$

$$= \frac{1}{2} \sum_{i=1}^N x_i^2 + \sum_{k=1}^{N-1} e^{\Delta\Gamma l_0 \cdot k} \sum_{i=1}^{N-k} x_i x_{i+k}.$$

$C_t \cdot \alpha_i$, tilts.

$$x_i = C_o \cdot b_i, \quad \text{offsets.} \tag{71e}$$

$C_d \cdot \Delta r_i$, diameter changes.

ii. G_1 — Tilts

$$G_1 \approx j \sum_{i=1}^N x_i e^{-\Delta\Gamma L_{i-1}}. \tag{72a}$$

Special case — equally spaced mode converters

$$l_k = l_0, \quad L_k = kl_0. \tag{72b}$$

$$G_1 \approx j \sum_{i=1}^N x_i e^{-\Delta\Gamma l_0(i-1)}. \tag{72c}$$

$$x_i = C_t \cdot \alpha_i. \tag{72d}$$

iii. G_1 — Offsets and Diameter Changes

$$G_1 \approx - \sum_{i=1}^N x_i e^{-\Delta\Gamma L_{i-1}}. \tag{73a}$$

Special case — equally spaced mode converters

$$l_k = l_0, \quad L_k = kl_0. \tag{73b}$$

$$G_1 \approx - \sum_{i=1}^N x_i e^{-\Delta\Gamma l_0(i-1)}. \tag{73c}$$

$$\begin{aligned} x_i &= C_o \cdot b_i, \quad \text{offsets.} \\ &= C_d \cdot \Delta r_i, \quad \text{diameter changes.} \end{aligned} \tag{73d}$$

2.2.3 Discussion

The above results for tilts and offsets apply strictly to only lossless metallic guide, for which $\Delta\alpha = 0$. However, we expect on intuitive grounds that they provide a satisfactory approximation for real copper guide. The results for diameter changes apply equally well to both copper and helix; for tilts and offsets in helix we must use the helix guide coupling coefficients, but otherwise the analysis is the same.

The results of (71) to (73) yield the TE_{01} signal mode to second order (there are no first-order terms), the spurious mode to first order, in the appropriate geometric parameters. In each case we have determined the first correction term to the solution for a geometrically perfect guide with all $c_k = 0$, i.e., $G_0(z) = 1$, $G_1(z) = 0$ where z is distance measured along the guide axis.

A rough physical interpretation may be given for the approximate results of (71) to (73). The spurious mode at any point in the guide is regarded as a sum of waves arising at each mode converter. Each of these waves is computed by assuming an incident TE_{01} wave at each mode converter identical to the TE_{01} wave that would be present in a perfect guide, with all $c_k = 0$, and further assuming that the converted spurious mode wave is unaffected by subsequent mode converters, i.e., propagates as it would in perfect guide. The TE_{01} signal mode is regarded as the sum of three components:

1. The TE_{01} wave that would be present in a perfect guide.
2. The signal lost from TE_{01} at each mode converter, assuming an incident TE_{01} wave identical to that in perfect guide and no incident spurious mode wave.
3. A sum of waves reconverted to TE_{01} from the spurious mode at each mode converter. Each of these waves is computed by taking the approximate spurious mode as computed above, and assuming that the reconverted TE_{01} wave is unaffected by subsequent mode converters. It is easy to see that this component may be expressed as a sum over all pairs of mode converters.

The approximate solutions of (71) to (73) may be regarded as the initial terms of power series expansions in the x_i 's (or equivalently the α_i 's, b_i 's, or Δr_i 's). It is reasonable to assume that if the x_i 's are sufficiently small these power series will converge sufficiently rapidly so that their first terms will provide valid approximations; under these conditions we would expect that $|\rho| \ll 1$, $|G_1| \ll 1$. However, to use these results with confidence we must have bounds on the errors introduced by these approximations that will give quantitative information on the way in which the solutions of (71) to (73) approximate the true solution. We

postpone consideration of these questions to Section 2.3 below, where we study the continuous case, since we will show there that the discrete and continuous cases are closely related to each other. It turns out that although we can give bounds on the higher terms of the series expansions that are in a sense the best possible, we still lack sufficiently precise information to make any useful statement on the way in which the perturbation solutions approximate the true solutions. Consequently, confidence in the accuracy of practical calculations using the results of (71) to (73) can at present be justified only on intuitive grounds.

In the region where we hope the perturbation solution is valid, $|\rho| \ll 1$. Under these conditions, the TE_{01} normalized magnitude g , loss A , and phase Θ are given approximately by (65), where ρ is given by (71). If $|\rho|$ is not much smaller than 1, it would seem that the approximation of (65) must be invalid. However, it will be shown in Section 2.3 on the continuous case that (65b), (65c), and (65d) together with (71) remain plausible for a wide class of interesting cases where the magnitude of the right-hand side of (71b) or (71d) becomes much greater than 1 and (65a) fails, although here again the justification is no more rigorous than that for (71) to (73) when $|\rho| \ll 1$, $|G_1| \ll 1$. This extension is important, for otherwise we should be limited to considering only cases where the total loss $|\Lambda|$, and hence, A and $|\Theta|$, are small.

2.3 *The Continuous Case — Single Spurious Mode*

In this section, we study mode conversion caused by distributed geometric imperfections, such as continuous curvature of the guide axis or continuous variation of the guide diameter. We discuss briefly the general telegraphist's equations or coupled line equations for the general case, but again restrict the detailed treatment to the case of a single spurious mode. (If the spurious mode is not TE_{0m} , we consider only one of its linear polarizations.)

It is obvious that a close relationship must exist between corresponding discrete and continuous cases. A guide with an arbitrary curved axis may be regarded as the limit of a guide with many small discrete tilts; similarly, a guide with a continuous diameter variation may be regarded as the limit of a guide with many small discrete diameter changes. We will show how the matrix equations for the discrete case and the differential equations (coupled line equations) for the continuous case may be obtained from each other by suitable limiting processes.

We next discuss the perturbation theory for the coupled line equations, and show its relationship to the discrete perturbation theory of Section

2.2. An improved perturbation theory for the TE_{01} complex loss $\Lambda = A - j\Theta$ [see (64)] is given that permits the treatment of many cases of practical interest.

For moderate lengths of copper guide the differential loss is small compared to 1, $|\Delta\alpha|L \ll 1$, where L is the guide length. We obtain an approximate treatment for this case by setting $\Delta\alpha = 0$, as in lossless metallic guide. The TE_{01} loss A for this case may be expressed in terms of the Fourier coefficients of the coupling coefficient $c(z)$; this treatment makes evident the relationship between the power spectrum of the coupling coefficient and the corresponding transmission statistics of the TE_{01} loss.

Finally, we discuss Morgan's coupling coefficients for general continuous cross-sectional deformations of copper guide, and the relationship between the various copper guide coupling coefficients.

2.3.1 Generalized Telegraphist's Equations

By means of the generalized telegraphist's equations,⁴ Maxwell's equations for the fields in a deformed guide may be expressed in terms of the normal modes of the undeformed guide;

$$\frac{dI_m}{dz} = \sum_n \kappa_{mn} I_n. \quad (74)$$

I_m represents the normalized complex amplitude of the m^{th} mode and the summation is extended over all modes. κ_{mm} represents the propagation constant of the m^{th} mode; κ_{mn} , with $m \neq n$, represents the coupling coefficient between the n^{th} and m^{th} modes. The κ_{mn} are functions of the geometry of the guide imperfection. Of course, (74) is in rather general symbolic form; in particular problems a more specific notation is often used to denote the various modes. For example, each subscript in (74) often becomes a double subscript, to conform to the usual ways of indexing waveguide modes. Square brackets enclosing a subscript pair are often used to indicate a TE mode, round brackets a TM mode; some sort of notation, such as \pm superscripts or different symbols, is used to differentiate between forward and backward modes.²

The κ_{mn} have been evaluated for a variety of different types of imperfections in a variety of guides, as discussed in the introduction to Section II. Morgan has given the κ_{mn} for curved metallic waveguide filled with an inhomogeneous dielectric.² For a homogeneous lossless dielectric, the propagation constants κ_{mm} are simply equal to the corresponding propagation constants in undeformed metallic guide;²

$$\kappa_{mm} = \pm j\beta_m. \quad (75)$$

Morgan finds that curvature couples the TE_{01}^+ to only TM_{11}^+ and TE_{1m}^\pm . (This should not be taken to mean that no other spurious modes will be excited in a curved guide with a homogeneous dielectric; such other modes will indeed arise, e.g., because of the coupling between TE_{1m} and TE_{2m} . However, in good guides these other modes will normally be negligible compared to the first-order spurious modes.) Using his results the κ 's of (74) involving the TE_{01}^+ mode may be expressed in terms of the C_i 's of (18) and Appendix A as follows, for a curved circular guide with a homogeneous dielectric:

$$\kappa_{[01]^+[1m]^\pm} = \kappa_{[1m]^\pm[01]^+} = \frac{jC_{t[1m]^\pm}}{R(z)}, \tag{76a}$$

$$\kappa_{[01]^+[11]^+} = \kappa_{[11]^+[01]^+} = \frac{jC_{t[11]^+}}{R(z)}. \tag{76b}$$

$R(z)$ in (76a) and (76b) is the radius of curvature of the axis of the curved guide, with z measured along the curved axis; the curvature of the axis is confined to a single plane. Similarly, Unger¹⁰ has evaluated the κ_{mn} for a tapered metallic waveguide as follows:†

$$\kappa_{[0m]^+[0n]^\pm} = \frac{1}{r} \frac{k_{0m}k_{0n}}{k_{0n}^2 - k_{0m}^2} \cdot \frac{\beta_{0n} \pm \beta_{0m}}{\sqrt{\beta_{0m}\beta_{0n}}} \frac{dr(z)}{dz}, \tag{77}$$

$m \neq n$ for the upper signs.

In (77), $r(z)$ is the radius of the guide, the k 's are the Bessel roots of the modes in question (see Appendix A). When the incident mode is the TE_{01} , (77) becomes

$$\kappa_{[01]^+[0m]^\pm} = C_{d[m]^\pm} \cdot \frac{dr(z)}{dz}, \quad m \neq 1, \tag{78}$$

in terms of C_d of (51) and Appendix A. Finally, we note from (77) that if both modes are forward modes,

$$\kappa_{[0m]^+[0n]^+} = -\kappa_{[0n]^+[0m]^+}, \tag{79}$$

which, as we shall subsequently see, must be true by (43).

Various geometric imperfections in helix guide have been similarly treated by Unger.^{3,19,20,21}

The generalized telegraphist's equations of (74) contain an infinite number of modes. To simplify the problem, we wish to approximate the true situation with a finite number of modes; indeed, in much of the present study we include only two modes, the TE_{01} signal mode and the

† The sign of the second term inside the summation in (40) of Reference 10 appears to be in error.

most important spurious mode. It is obvious that this approximation must distort the results in some respects; however, we hope that if the significant spurious modes are included, any errors will be small. Thus, in (74) we will set all I 's equal to zero except those corresponding to the TE_{01}^+ signal mode and one or a few of the forward spurious modes. Other ways of approximating the true situation with a finite number of modes might of course be possible, but we choose this one as the simplest.

In (76) and (78) above we observe that the coupling coefficients for continuous bends and diameter changes may be expressed in terms of the coupling coefficients for discrete tilts and diameter changes. This, plus the obvious fact that a continuous deformation may be regarded as the limit of many small, closely spaced discrete imperfections, suggests a closer study of the relation between the discrete and continuous cases. In Sections 2.3.2 and 2.3.3 below we study the correspondence between the coupled line equations for the continuous case and the matrix equations of Section 2.2.1 above for the discrete case.

2.3.2 Transformation from the Discrete to the Continuous Case

Consider a lossless metallic guide containing many small discrete mode converters spaced an equal distance Δz apart. Fig. 7 illustrates three cases of interest — tilts, offsets, and diameter changes — with of course a greatly exaggerated vertical scale. c_k is in each case the conversion coefficient in the corresponding matrix, (17) or (43) for tilts or offsets and diameter changes respectively; c_k is given by (18), (44), and (51) for tilts, offsets, and diameter changes respectively. Then it is clear that by the proper limiting process, in which $\Delta z \rightarrow 0$ and the discontinuities become smaller and more closely spaced, we may approach the continuous deformations illustrated in Fig. 7 by the dotted lines.

The limiting continuous deformation of Fig. 7(a) corresponds to a guide with a continuously varying curvature of its axis. The continuous deformation of Fig. 7(b) corresponds to a waveguide made of very thin circular punchings which may slide with respect to each other like a stack of cards. Finally, the continuous deformation of Fig. 7(c) corresponds to a continuous taper. The case of Fig. 7(a) will be important in determining the effect of random straightness deviation on the TE_{01} transmission. For the continuous bend and offset of Figs. 7(a) and 7(b) the guide axis must lie in a single plane if the analysis is to be restricted to a single polarization of the most important spurious mode (TE_{12}^+).

For purposes of illustration we consider the transition to the continuous bend, illustrated in Fig. 7(a). A larger-scale drawing for this case

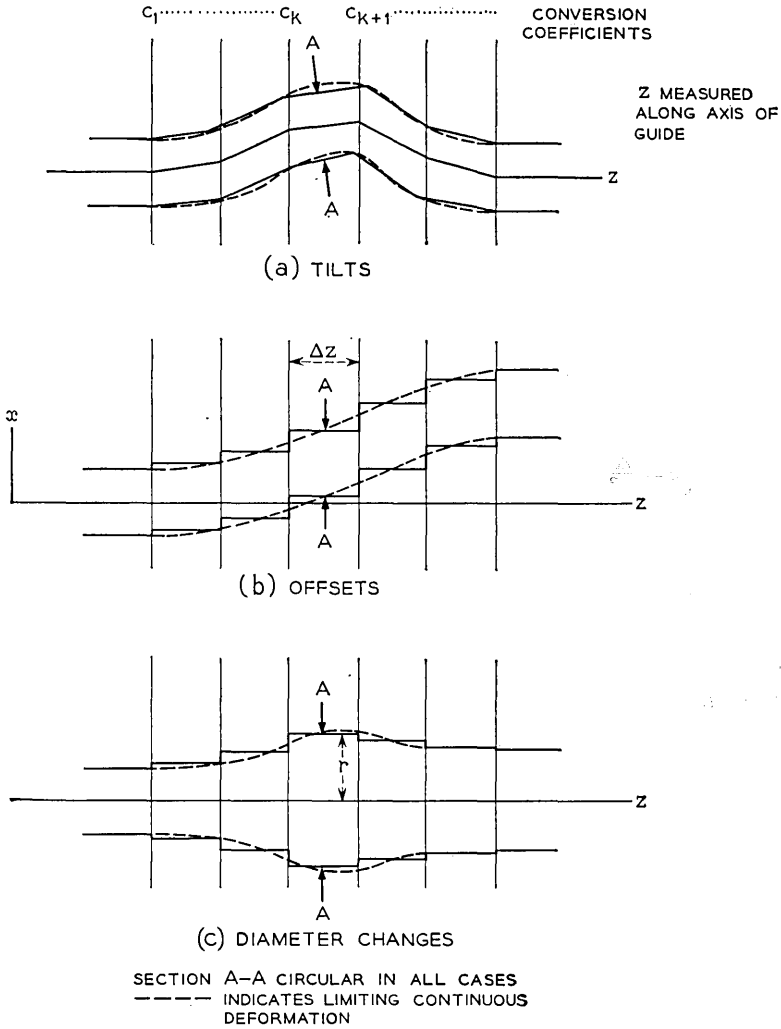


Fig. 7 — Waveguides with tilts, offsets, and diameter changes.

is shown in Fig. 8; only the center line of the continuously bent guide and of the tilt approximation to it are shown, for the two adjacent sections lying on either side of the k^{th} tilt.

Referring to Fig. 6, we set the distance between discrete tilts l_k equal to Δz . Let $I_0(z)$ and $I_1(z)$ denote the TE_{01} and spurious mode complex amplitudes just to the left of the k^{th} tilt; then $I_0(z + \Delta z)$ and $I_1(z + \Delta z)$

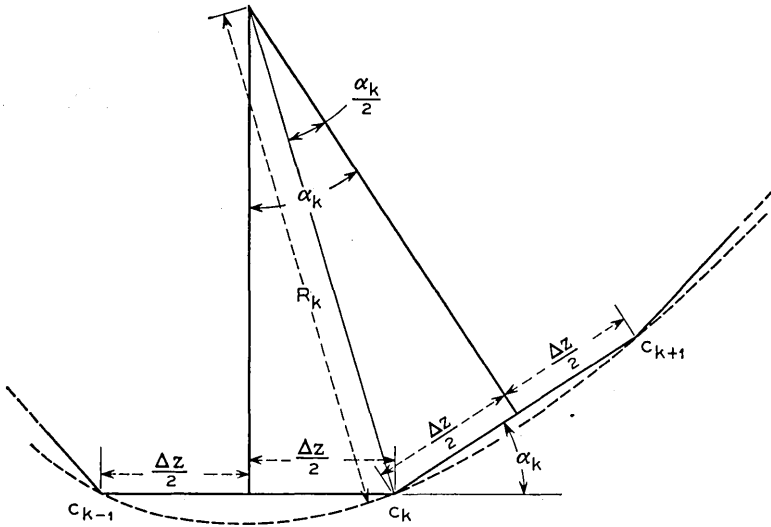


Fig. 8 — Fig. 7(a) in larger scale.

denote the corresponding complex mode amplitudes just to the left of the $(k + 1)^{\text{th}}$ tilt. Then taking (17) for the scattering matrix of the tilt, and setting $l_k = \Delta z$ in (52), (53), and (54), we have

$$\begin{bmatrix} I_0(z + \Delta z) \\ I_1(z + \Delta z) \end{bmatrix} = \begin{bmatrix} e^{-\Gamma_0 \Delta z} \sqrt{1 - c_k^2} & j e^{-\Gamma_0 \Delta z} c_k \\ j e^{-\Gamma_1 \Delta z} c_k & e^{-\Gamma_1 \Delta z} \sqrt{1 - c_k^2} \end{bmatrix} \begin{bmatrix} I_0(z) \\ I_1(z) \end{bmatrix}, \quad (80)$$

where c_k is the conversion coefficient for the k^{th} tilt. From (18)

$$c_k = C_t \alpha_k + \dots, \quad (81)$$

where C_t is Morgan's coupling coefficient for tilts for the spurious mode in question, given in Appendix A. From Fig. 8

$$\alpha_k = \frac{\Delta z}{R_k} + \dots \quad (82)$$

Thus,

$$c_k = \frac{C_t}{R_k} \Delta z + \dots, \quad (83)$$

where the dots in (82) and (83) represent terms of higher than first order in Δz .

Expanding the exponentials and the square roots in the matrix of (80) in power series [see (66)], making use of (83), and writing the right-hand sides of the resulting equations as power series in Δz , we have

$$I_0(z + \Delta z) = I_0(z) + \left[-\Gamma_0 I_0(z) + j \frac{C_t}{R_k} I_1(z) \right] \Delta z + \dots \tag{84}$$

$$I_1(z + \Delta z) = I_1(z) + \left[-\Gamma_1 I_1(z) + j \frac{C_t}{R_k} I_0(z) \right] \Delta z + \dots$$

We transfer the terms $I_0(z)$ and $I_1(z)$ to the left-hand sides of these equations and divide by Δz . Then as $\Delta z \rightarrow 0$, $R_k \rightarrow R(z)$, where $R(z)$ is the radius of curvature of the guide axis, and (84) becomes

$$I_0'(z) = -\Gamma_0 I_0(z) + jc(z) I_1(z), \tag{85}$$

$$I_1'(z) = jc(z) I_0(z) - \Gamma_1 I_1(z),$$

$$c(z) = \frac{C_t}{R(z)}, \quad \text{coupling coefficient,} \tag{86}$$

where the primes denote differentiation with respect to z , and $R(z)$ is the radius of curvature of the guide axis.

We now compare (85) and (86) with (74) to (76); in the case of lossless metallic guide, where $\Gamma_0 = j\beta_0$, $\Gamma_1 = j\beta_1$, the two sets of equations are identical if we retain only two modes in the results of (74) to (76). Thus by taking the proper limiting form of Morgan's results for discrete tilts we arrive at Morgan's results for continuous bends obtained via the generalized telegraphist's equations, in the two-mode case. It is easy to see that this method extends readily to additional first-order spurious modes. It should in principle be possible to include as many modes as desired in this type of discussion, but such calculations have not been actually carried out. We will be content in the present paper to take the two-mode model as suggestive of these more general results.

As in the discrete case, we assume the guide is excited by a pure TE_{01} wave of unit amplitude. Thus, the initial conditions on the differential equations (85) become

$$I_0(0) = 1, \quad I_1(0) = 0, \tag{87}$$

corresponding to (58) for the discrete case. Then $I_0(z)$ will be the TE_{01} gain, $I_1(z)$ the TE_{01} -spurious mode transfer coefficient, corresponding to (59) and (60).

It proves convenient to follow the normalization used in the discrete case [see (61) and (62)], removing the propagation factor of each mode in ideal guide;

$$I_0(z) = e^{-\Gamma_0 z} \cdot G_0(z) \quad (88)$$

$$I_1(z) = e^{-\Gamma_1 z} \cdot G_1(z). \quad (89)$$

$G_0(z)$ and $G_1(z)$ are again the TE₀₁ normalized gain and the TE₀₁-spurious mode normalized transfer coefficient. Substituting the transformation of (88) and (89) into (85), we have:

$$\begin{aligned} G_0'(z) &= jc(z)e^{\Delta\Gamma z}G_1(z) \\ G_1'(z) &= jc(z)e^{-\Delta\Gamma z}G_0(z) \end{aligned} \quad (90)$$

$\Delta\Gamma = \Gamma_0 - \Gamma_1$, differential propagation constant.

$$c(z) = \frac{C_t}{R(z)}, \quad \text{coupling coefficient.} \quad (91)$$

The initial conditions of (87) become

$$G_0(0) = 1, \quad G_1(0) = 0. \quad (92)$$

A similar treatment may be given for the limiting case of many discrete offsets or diameter changes, illustrated in Figs. 7(b) and 7(c) respectively. For continuous offsets we find:

$$I_0'(z) = -\Gamma_0 I_0(z) + c(z)I_1(z), \quad (93)$$

$$I_1'(z) = -c(z)I_0(z) - \Gamma_1 I_1(z).$$

$$c(z) = C_o \cdot x'(z), \quad \text{coupling coefficient.} \quad (94)$$

C_o is Morgan's offset coupling coefficient for the spurious mode in question, given in Appendix A. $x(z)$ is the transverse displacement of the guide axis, as illustrated in Fig. 7(b); primes of course denote differentiation with respect to z . Equations (93) and (94) must agree with the generalized telegraphist's equations for a guide with a continuous offset, when restricted to two modes. Using the transformation of (88) and (89), (93) and (94) become:

$$G_0'(z) = c(z)e^{\Delta\Gamma z}G_1(z), \quad (95)$$

$$G_1'(z) = -c(z)e^{-\Delta\Gamma z}G_0(z),$$

$\Delta\Gamma = \Gamma_0 - \Gamma_1$; differential propagation constant. (96)

$$c(z) = C_o \cdot x'(z), \quad \text{coupling coefficient.}$$

The continuous diameter change may be similarly treated by replacing $x'(z)$ by $r'(z)$, C_o by C_a in (93) to (96), where $r(z)$ is the guide radius.

2.3.3 Transformation from the Continuous to the Discrete Case

We next determine the scattering matrix for a discrete mode converter by considering the limiting form of the generalized telegraphist's equations as the continuous coupling coefficient approaches a δ -function. For purposes of illustration we consider the case of a discrete tilt, regarded as the limiting form of a continuous bend. We again restrict our treatment to only two modes, the TE_{01} signal mode and a single spurious mode, for simplicity.

Fig. 9 shows a tilt of angle α obtained as the limit of a continuous bend. From (86) or (91) the continuous coupling coefficient is given by

$$\begin{aligned}
 &0, \quad z < 0; \\
 c(z) &= \frac{C_t \alpha}{\Delta}, \quad 0 < z < \Delta; \\
 &0, \quad \Delta < z.
 \end{aligned} \tag{97a}$$

$$\lim_{\Delta \rightarrow 0} c(z) = C_t \alpha \cdot \delta(z). \tag{97b}$$

Thus the coupling coefficient is zero for $z < 0$ and $z > \Delta$, constant in the region $0 < z < \Delta$, z being measured along the guide axis as usual. Now exact solutions for (85) or (90) are known for only a few special $c(z)$; indeed this is why we are forced to use approximate solutions of various types in the present paper. One important case in which exact solutions are available is the present case $c(z) = \text{constant}$;¹⁷ this is obvious from the fact that the equations (85) then become simply two simultaneous differential equations with constant coefficients, whose solutions are well known.

The solution for (85) with $c(z) = c_0$ is given as follows:

$$\begin{bmatrix} I_0(z) \\ I_1(z) \end{bmatrix} = T(z) \begin{bmatrix} I_0(0) \\ I_1(0) \end{bmatrix} \tag{98a}$$

$$\begin{aligned}
 T(z) &= \frac{\exp\left(-\frac{\Gamma_0 + \Gamma_1}{2} z\right)}{K_+ - K_-} \\
 &\cdot \begin{bmatrix} -K_- e^{(\Delta\Gamma/2)z\sqrt{}} + K_+ e^{(-\Delta\Gamma/2)z\sqrt{}} & e^{(\Delta\Gamma/2)z\sqrt{}} - e^{(-\Delta\Gamma/2)z\sqrt{}} \\ e^{(\Delta\Gamma/2)z\sqrt{}} - e^{(-\Delta\Gamma/2)z\sqrt{}} & K_+ e^{(\Delta\Gamma/2)z\sqrt{}} - K_- e^{(-\Delta\Gamma/2)z\sqrt{}} \end{bmatrix} \tag{98b}
 \end{aligned}$$

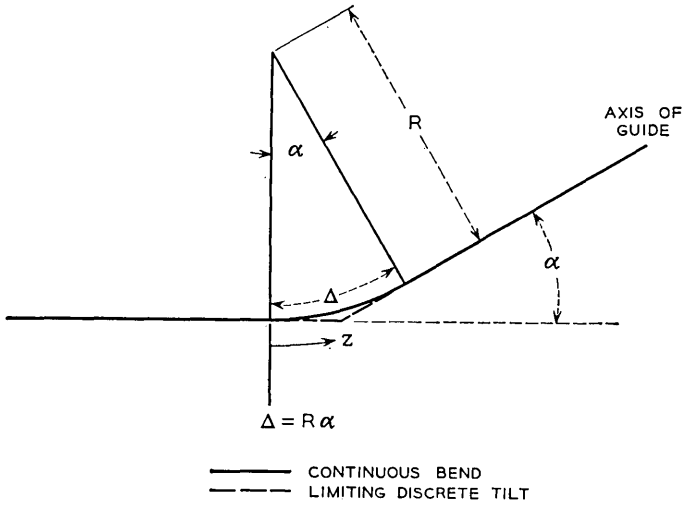


Fig. 9 — Tilt of angle α as limit of continuous bend.

$$K_{\pm} = -j \frac{1 \pm \sqrt{}}{2 \frac{c_0}{\Delta\Gamma}}; \quad K_+ K_- = -1 \tag{98c}$$

$$\frac{1}{K_+ - K_-} = j \frac{c_0/\Delta\Gamma}{\sqrt{}} \tag{98d}$$

$$\sqrt{} = \sqrt{1 - \left(\frac{2c_0}{\Delta\Gamma}\right)^2}. \tag{98e}$$

These results are applied to the situation of Fig. 9 and (97) by setting

$$\begin{aligned} z &\rightarrow \Delta, \\ c_0 &\rightarrow \frac{C_t \alpha}{\Delta}. \end{aligned} \tag{99}$$

Thus

$$\begin{bmatrix} I_0(\Delta) \\ I_1(\Delta) \end{bmatrix} = T(\Delta) \begin{bmatrix} I_0(0) \\ I_1(0) \end{bmatrix}. \tag{100}$$

Now in (97) and Fig. 9 let us keep α fixed and let $\Delta \rightarrow 0, R \rightarrow 0$; then the curved guide of Fig. 9 will approach a discrete tilt of angle α , and

the matrix T should approach the scattering matrix of this discrete tilt. From (98) and (99) we have:

$$\begin{aligned} \lim_{\Delta \rightarrow 0} \frac{\Delta \Gamma}{2} \sqrt{1 - \left(\frac{2C_t \alpha}{\Delta \Gamma \cdot \Delta}\right)^2} \cdot \Delta &= jC_t \alpha \\ \lim_{\Delta \rightarrow 0} K_{\pm} &= \pm 1 \\ \lim_{\Delta \rightarrow 0} \frac{1}{K_+ - K_-} &= \frac{1}{2} \\ \lim_{\Delta \rightarrow 0} \exp\left(-\frac{\Gamma_0 + \Gamma_1}{2} \Delta\right) &= 1 \end{aligned} \tag{101}$$

Then from (98) to (101) we have

$$\lim_{\Delta \rightarrow 0} T(\Delta) \equiv T(0) = \begin{bmatrix} \cos C_t \alpha & j \sin C_t \alpha \\ j \sin C_t \alpha & \cos C_t \alpha \end{bmatrix} \tag{102a}$$

for

$$\lim_{\Delta \rightarrow 0} c(z) = C_t \alpha \cdot \delta(z) \tag{102b}$$

as the scattering matrix of a discrete tilt of angle α . This is the result given in (20) as the “exact” two-mode scattering matrix for a finite tilt.

We have thus found the scattering matrix for a two-mode, finite tilt starting from the scattering matrix for an infinitesimal tilt by first passing to the limit of continuous mode conversion and then transforming back to the discrete case by allowing the continuous coupling coefficient to approach a δ -function. Alternately, we have found the scattering matrix that satisfies the following requirements (for a single forward spurious mode):

1. Conservation of energy and symmetry.
2. Agrees with Morgan’s small tilt theory [see (17) and (18)].
3. Satisfies an additional requirement for finite tilts, so that the matrix for the sum of two tilts equals the product of the matrices of the individual tilts [see (19)]. In practice the above matrix for a finite tilt will of course not provide an exact description for large tilts because of the presence of additional spurious modes.

An offset and a diameter change may be similarly treated. For a finite offset of magnitude b we have corresponding to (102) as the “exact” two-mode scattering matrix

$$\lim_{\Delta \rightarrow 0} T(\Delta) \equiv T(0) = \begin{bmatrix} \cos C_o b & \sin C_o b \\ -\sin C_o b & \cos C_o b \end{bmatrix}. \quad (103)$$

Diameter changes may be similarly treated if they are small enough so that the variation of $\Delta\Gamma$ may be neglected.

The present analysis readily extends to include additional first-order spurious modes. It should in principle be possible to generalize it further to include as many modes as desired, but such calculations have not actually been carried out.

The astute reader may have noted a potential difficulty with the analysis above. This potential paradox may be stated as follows. The above discussion shows that the continuous bend and the discrete tilt of Fig. 9 are approximately equivalent to each other so long as the following conditions are satisfied:

$$\frac{2C_t \alpha}{|\Delta\Gamma| \cdot \Delta} \gg 1. \quad (104a)$$

$$\frac{|\Delta\Gamma|}{2} \cdot \Delta \ll 1. \quad (104b)$$

But if (104a) is *not* satisfied it is no longer obvious that the discrete tilt and continuous bend remain approximately equivalent. Stated physically, consider a short continuous bend, with length Δ fixed and small enough so that $\frac{|\Delta\Gamma|}{2} \cdot \Delta \ll 1$. Then we should expect on physical grounds that this continuous bend is approximately equivalent to its corresponding discrete tilt for all angles α . However the above analysis seems to guarantee this equivalence only for large α , and not in an obvious way for small α .

While it is not obvious, it is true that this equivalence remains valid for small α , or more precisely when (104a) is violated but (104b) remains true. To study this matter let us consider the case of a guide with zero differential attenuation constant, $\Delta\alpha = 0$, so that

$$\Delta\Gamma = j\Delta\beta. \quad (105)$$

For brevity we restrict our attention to the TE_{01} -spurious mode conversion coefficient, which for convenience we call T_{10} (e.g., see the labeling of matrix components in (56)). For the discrete tilt

$$T_{10 \text{ discrete}} = j \sin C_t \alpha \cdot \exp\left(-\frac{\Gamma_0 + \Gamma_1}{2} \Delta\right). \quad (106)$$

For the continuous bend the results of (98) may be written [using (105)] as follows:

$$T_{10 \text{ continuous}} = j \frac{\sin \left[f \left(\frac{2C_t \alpha}{|\Delta\beta \cdot \Delta|} \right) \cdot C_t \alpha \right]}{f \left(\frac{2C_t \alpha}{|\Delta\beta \cdot \Delta|} \right)} \cdot \exp \left(- \frac{\Gamma_0 + \Gamma_1}{2} \Delta \right); \quad (107)$$

$$f(x) = \frac{\sqrt{1+x^2}}{x}.$$

Now when $x \gg 1$, $f(x) \approx 1$; thus when $\frac{2C_t \alpha}{|\Delta\beta \cdot \Delta|} \gg 1$ [see (104a)] then from (106) and (107) we have $T_{10 \text{ discrete}} \approx T_{10 \text{ continuous}}$, as stated above.

Next consider the case where $\frac{2C_t \alpha}{|\Delta\beta \cdot \Delta|}$ is not large compared to 1. From (106) and (107) we see that $T_{10 \text{ discrete}} \approx T_{10 \text{ continuous}}$ if

$$f \left(\frac{2C_t \alpha}{|\Delta\beta \cdot \Delta|} \right) \cdot C_t \alpha \ll 1. \quad (108)$$

If $\frac{2C_t \alpha}{|\Delta\beta \cdot \Delta|} \ll 1$, then noting that $f(x) \approx \frac{1}{x}$ for $x \ll 1$, (108) becomes

$$\frac{|\Delta\beta|}{2} \cdot \Delta \ll 1, \quad (109)$$

which is simply the condition of (104b). If $\frac{2C_t \alpha}{|\Delta\beta \cdot \Delta|} = 1$, noting that $f(1) = \sqrt{2}$, (108) becomes

$$\sqrt{2} C_t \alpha \ll 1, \quad (110)$$

i.e., the conversion coefficient must be moderately small. However a more detailed calculation for this case shows that

$$0.95 T_{10 \text{ discrete}} \leq T_{10 \text{ continuous}} \leq T_{10 \text{ discrete}} \quad (111a)$$

for

$$C_t \alpha \leq 0.55 \text{ radian}, \quad \sin C_t \alpha \leq 0.523. \quad (111b)$$

The condition of (111b) is far less restrictive than that of (110); (111) states that for $\frac{2C_t \alpha}{|\Delta\beta \cdot \Delta|} = 1$, the continuous bend and its corresponding discrete tilt will have conversion coefficients that differ by less than 5

per cent provided that the coupling coefficient to the spurious mode is less than 0.523, which is quite large.

These considerations render it plausible that the corresponding discrete and continuous cases approximate each other well if (104b) is satisfied, for both large and small tilts.

2.3.4 Perturbation Theory for the Coupled Line Equations

Exact solutions to the coupled line equations, (85) and (86) or (90) and (91) [or (93) and (94), (95) and (96)], are known in only a very few special cases, i.e., for a few special $c(z)$. In subsequent work we will require a solution for an arbitrary $c(z)$ so that we may eventually treat the case of random $c(z)$. Therefore it is necessary to consider approximate solutions to the coupled line equations, as we have done in the discrete case.

These or similar equations arise in the design of a variety of devices in which the coupling between two modes is of interest, such as directional couplers, tapers or impedance transformers, and bends. In dominant mode transmission systems the signal mode is the forward TEM wave, the spurious mode the reflected TEM wave.^{24,25,26,27} Directional couplers in a variety of waveguide systems have been similarly studied.^{17,28} Tapers and bends in a variety of waveguides have been studied; several recent papers deal with this type of problem when the principal wave is the TE₀₁ mode in circular guide^{3,9,10} as noted earlier. In much of this work an approximate solution is used which is similar or identical to the one employed here. This approximate solution has also been used to study randomly distributed nonuniformities in ordinary (dominant mode) transmission lines.^{29,30}

The approximate solution that we use is most readily found via Picard's method of successive approximations.^{31,32} This is discussed in detail in a companion paper,¹⁸ where it is shown that the solution to (90) and (91) may be written as follows:

$$G_0(z) = \sum_{n=0}^{\infty} g_{0(n)}(z), \quad (112a)$$

$$G_1(z) = \sum_{n=0}^{\infty} g_{1(n)}(z). \quad (112b)$$

$$g_{0(n)}(z) = j \int_0^z c(s) e^{\Delta\Gamma s} g_{1(n-1)}(s) ds, \quad n \geq 1. \quad (113a)$$

$$g_{1(n)}(z) = j \int_0^z c(s) e^{-\Delta\Gamma s} g_{0(n-1)}(s) ds, \quad n \geq 1. \quad (113b)$$

$$g_{0(0)}(z) = 1, \quad g_{1(0)}(z) = 0. \tag{113c}$$

It is readily seen that

$$g_{0(n)}(z) = 0, \quad n \text{ odd} \tag{114a}$$

$$g_{1(n)}(z) = 0, \quad n \text{ even.} \tag{114b}$$

Bounds on the terms of the series solution of (112) are given as follows:¹⁸

$$|g_{0(n)}(z)| \leq \frac{\left[\int_0^z |c(s)| ds \right]^n}{n!}, \quad n \text{ even.} \tag{115a}$$

$$= 0, \quad n \text{ odd.}$$

$$= 0, \quad n \text{ even.} \tag{115b}$$

$$|g_{1(n)}(z)| \leq \frac{\left[\int_0^z |c(s)| ds \right]^n}{n!} e^{-\Delta \alpha z}, \quad n \text{ odd.}$$

These bounds are in a sense the best possible, since cases are known where the equalities in (115a) and (115b) are satisfied.¹⁸

Now suppose that the series solutions for $G_0(z)$ and $G_1(z)$ given in (112) and (113) converge so rapidly that only the first nonzero terms that depend on the coupling coefficient $c(z)$ need be retained. From (114) we see that the $n = 1$ term of (112a) for $G_0(z)$ is identically zero; therefore we retain in this equation only the $n = 2$ term in addition to the $n = 0$ term, which is simply 1. The first nonzero term in (112b) for $G_1(z)$ is the $n = 1$ term. We again use the notation of (63) and (64):

$$G_0 = 1 - \rho = ge^{j\Theta} = e^{-\Lambda} = e^{-A}e^{j\Theta}; \tag{116}$$

$$\Lambda = A - j\Theta.$$

Then we have:

$$\rho(z) \approx \int_0^z c(s) e^{\Delta \Gamma s} ds \int_0^s c(t) e^{-\Delta \Gamma t} dt. \tag{117}$$

$$G_1(z) \approx j \int_0^z c(s) e^{-\Delta \Gamma s} ds. \tag{118}$$

Since we have assumed that $c(z)$ is sufficiently small so that $|\rho| \ll 1$, we have to second order from (65) and (117):

$$g \approx 1 - \operatorname{Re} \iint f f \quad (119a)$$

$$\Lambda \approx -\ln G_0 = \iint f f \quad (119b)$$

$$A \approx -\ln |G_0| = \operatorname{Re} \iint f f \quad (119c)$$

$$\Theta \approx \angle G_0 = -\operatorname{Im} \iint f f \quad (119d)$$

where $\iint f f$ is shorthand for

$$\iint f f = \int_0^z c(s) e^{\Delta \Gamma s} ds \int_0^s c(t) e^{-\Delta \Gamma t} dt \quad (120a)$$

$$= \int_0^z e^{\Delta \Gamma u} du \int_0^{z-u} c(s)c(s+u) ds \quad (120b)$$

$$= \frac{1}{2} \int_0^z \int_0^z c(s)c(t) e^{\Delta \Gamma |t-s|} ds dt. \quad (120c)$$

These results for the continuous bend are analogous to those of (63) to (65) and (71) and (72) for the case of discrete tilts. The continuous offset and diameter change are of course readily handled in the same way, and the above relations hold with only minor modifications. In particular, for the corresponding solutions to (95) and (96) for the continuous offset and diameter change the j in (113a) is dropped, the j in (113b) replaced by -1 , and the j in (118) replaced by -1 (compare (73) for the discrete case); the remainder of the equations in the present section are unaltered. (For the continuous diameter change we have neglected the variation of $\Delta \Gamma$.)

These perturbation results for the continuous case may be regarded, as in the discrete case, as the first terms of power series expansions in the geometric parameter characterizing the deformed guide. Again, it is reasonable to assume that if the deformation is sufficiently small, these power series will converge so rapidly that their first terms will provide valid approximations. Equation (115) gives bounds on the magnitudes of the higher terms; for guides short enough so that $\int_0^z |c(s)| ds \ll 1$, these bounds converge rapidly. But the fact that the bounds converge rapidly does not guarantee that the terms themselves converge rapidly. For example, it might be possible for the second term of (112a) to be much smaller than its bound, while at the same time the fourth term was close to its bound, so that the fourth term would be comparable to the second term, even though the bound on the fourth term was much

smaller than the bound on the second term. (On the other hand this might not be possible; we just do not know.) Consequently we have no precise information on the way in which these perturbation solutions approximate the true solutions; the accuracy of these approximate solutions, which are used throughout the following statistical analysis of random guides, must be accepted largely on faith at present.

In Section 2.3.5 we show how the discrete and continuous perturbation theories may be obtained as limiting cases of each other. Section 2.3.6 discusses a modified perturbation theory, which makes it appear plausible that the results of (119b), (119c), and (119d) for the TE_{01} complex loss Λ , loss A , and phase Θ hold true in a wide range of important cases where $|\iint| \gg 1$, so that the other approximations of (117), (118), and (119a) fail. This extension is important in permitting the analysis of long guides.

2.3.5 Transformation between Discrete and Continuous Perturbation Theory

In this section we consider the relationship between the discrete and continuous perturbation theories; we select the case of discrete tilts and continuous bends for purposes of illustration.

Let us first consider the transformation from the discrete to the continuous case, as illustrated in Fig. 7(a); but instead of considering a single differential section of line with a single discrete mode converter, as in Section 2.3.2, we consider the entire line. Let the tilts be equally spaced, with spacing Δz , and further let the position of the k^{th} tilt be z_k ; thus

$$z_k = k\Delta z. \tag{121}$$

Let the angle of the k^{th} tilt be α_k . From Fig. 8 we have, as in (82),

$$\alpha_k = \frac{\Delta z}{R_k} + \dots, \tag{122}$$

where the dots indicate terms of higher order in Δz . R_k is defined in Fig. 8; in the limit as $\Delta z \rightarrow 0$, R_k approaches the radius of curvature of the guide axis at z_k .

Consider first the spurious mode normalized transfer coefficient, G_1 . From (72) we have, setting $l_0 = \Delta z$ and neglecting terms of higher order than Δz , and substituting (122) for α_k ,

$$G_1 \approx j \sum_{i=1}^N \frac{C_i}{R_i} \Delta z \cdot e^{-\Delta\Gamma(i-1)\Delta z}. \tag{123}$$

Setting

$$z_i = i\Delta z = s_i, \quad \Delta z = \Delta s, \quad R_i = R(s_i), \quad (124)$$

(123) becomes

$$G_1 \approx j \sum_{i=1}^N \frac{C_i}{R(s_i)} e^{-\Delta\Gamma s_i} e^{\Delta\Gamma \cdot \Delta s} \Delta s. \quad (125)$$

We now let $\Delta s \rightarrow 0$ and $N \rightarrow \infty$ in such a way that $N\Delta s$ remains constant,

$$N\Delta s = L, \quad (126)$$

where L is the length of line under consideration. Then $R(s_i) \rightarrow R(s)$, the radius of curvature of the guide axis, $e^{\Delta\Gamma \cdot \Delta s} \rightarrow 1$, and the summation of (125) becomes the following integral.

$$G_1 \approx j \int_0^L \frac{C_t}{R(s)} e^{-\Delta\Gamma s} ds. \quad (127)$$

Finally, noting (86) or (91), (127) becomes

$$G_1 \approx j \int_0^L c(s) e^{-\Delta\Gamma s} ds, \quad (128)$$

which is identical to the Picard approximate solution for the coupled line equations for the continuous case, given in (118).

Similarly the approximate solution for ρ given in (71) for the discrete case can be shown to approach the Picard approximation for the continuous case, given in (117). Setting $l_0 = \Delta z$ in (71) and substituting (122), we have on neglecting higher-order terms in Δz

$$\rho \approx \frac{1}{2} \sum_{i=1}^N \left(\frac{C_i}{R_i} \Delta z_i \right)^2 + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{C_i^2 \Delta z_i \Delta z_j}{R_i R_j} e^{\Delta\Gamma(j-i)\Delta z}. \quad (129)$$

Setting

$$\begin{aligned} z_i = i\Delta z = s_i, \quad \Delta z_i = \Delta s, \quad R_i = R(s_i), \\ z_j = j\Delta z = t_j, \quad \Delta z_j = \Delta t, \quad R_j = R(t_j), \end{aligned} \quad (130)$$

(129) becomes

$$\rho \approx \frac{1}{2} \sum_{i=1}^N \left(\frac{C_i^2}{R^2(s_i)} \Delta s \right) \Delta s + \sum_{i=1}^{N-1} \frac{C_i}{R(s_i)} e^{-\Delta\Gamma s_i \Delta s} \sum_{j=i+1}^N \frac{C_j}{R(t_j)} e^{\Delta\Gamma t_j \Delta t}. \quad (131)$$

We now let $\Delta s \rightarrow 0$ and $\Delta t \rightarrow 0$ in such a way that $N\Delta s$ and $N\Delta t$ remain constant,

$$N\Delta s = L, \quad N\Delta t = L, \quad (132)$$

where L is again the total length of line. Then $R(s_i) \rightarrow R(s)$, $R(t_i) \rightarrow R(t)$, where $R(s)$ and $R(t)$ are the radius of curvature of the guide axis.

The first term of (131) (the single summation) approaches zero; the second, however, approaches a double integral to yield

$$\rho \approx \int_0^L \frac{C_t}{R(s)} e^{-\Delta\Gamma s} ds \int_s^L \frac{C_t}{R(t)} e^{\Delta\Gamma t} dt. \tag{133}$$

Using (86) or (91), and further interchanging the order of integration and the integration variables,

$$\rho \approx \int_0^L c(s) e^{\Delta\Gamma s} ds \int_0^s c(t) e^{-\Delta\Gamma t} dt, \tag{134}$$

which is identical to the Picard approximation for the solution to the coupled line equations, given in (117). From (120) we see that the two other forms of (71d) also have their equivalents in the continuous case.

Next let us reverse the above process, and go from the continuous approximate solutions back to the discrete ones. Consider a line of length Nl_0 with N equally spaced tilts a distance l_0 apart. The k^{th} tilt is located at z_k , where

$$z_k = (k - 1)l_0, \tag{135}$$

and has an angle α_k . (The form of (135) was chosen to be consistent with the notation of Fig. 6 and Section 2.2.) From (97b) or (102b) we may write the continuous coupling coefficient for this case as

$$c(z) = C_t \sum_{k=1}^N \alpha_k \cdot \delta[z - (k - 1)l_0]. \tag{136}$$

We now substitute (136) into the approximate solution for the continuous case, and derive the approximate solution for the discrete case. Substituting (136) into (118), we find

$$G_1(Nl_0) \approx j \sum_{i=1}^N C_t \alpha_i e^{-\Delta\Gamma(i-1)l_0}, \tag{137}$$

in agreement with (72c) and (72d). Substituting (136) into (117), we have

$$\begin{aligned} \rho(Nl_0) &\approx \int_0^{Nl_0} C_t \sum_{i=1}^N \alpha_i \cdot \delta[s - (i - 1)l_0] e^{\Delta\Gamma s} ds \\ &\quad \cdot \int_0^s C_t \sum_{j=1}^N \alpha_j \cdot \delta[t - (j - 1)l_0] e^{-\Delta\Gamma t} dt \\ &= \int_0^{Nl_0} C_t \sum_{i=1}^N \alpha_i \cdot \delta[s - (i - 1)l_0] e^{\Delta\Gamma s} ds \\ &\quad \cdot \sum_{j=1}^N C_t \alpha_j \cdot u[s - (j - 1)l_0] e^{-\Delta\Gamma(j-1)l_0}, \end{aligned} \tag{138}$$

where $u(x)$ is the unit step function,

$$u(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0. \end{cases} \quad (139)$$

Continuing,

$$\rho(Nl_0) \approx \sum_{i=2}^N C_i \alpha_i e^{\Delta\Gamma(i-1)l_0} \sum_{j=1}^{i-1} C_j \alpha_j e^{-\Delta\Gamma(j-1)l_0} + \frac{1}{2} \sum_{i=1}^N (C_i \alpha_i)^2. \quad (140)$$

The second term of (140) (the single summation) was obtained via the relation

$$\int_a^b f(x)u(x)\delta(x) dx = \frac{1}{2}f(0), \quad a < 0 < b. \quad (141)$$

Alternately, the result of (140) may be obtained by regarding the δ -functions as the limit of some continuous functions [e.g., (97a)], taking the limit after the integrations have been performed. Finally, interchanging the order of summation and the summation indices, (140) becomes

$$\rho(Nl_0) = \frac{1}{2} \sum_{i=1}^N (C_i \alpha_i)^2 + \sum_{i=1}^{N-1} \sum_{j=i+1}^N (C_i \alpha_i)(C_j \alpha_j) e^{\Delta\Gamma l_0(j-i)}, \quad (142)$$

which agrees with (71d) for the discrete case. We see that the single summation, which "disappeared" in going from the discrete to the continuous case, has satisfactorily "reappeared." The alternate forms for ρ given in (120) may similarly be transformed to their discrete equivalents in (71d) via the substitution of (136) and appropriate manipulation of δ -functions.

The corresponding analysis for offsets and diameter changes is readily performed.

One consequence of the results of the present section is that bounds derived for the approximate solution in the continuous case may be directly applied to the approximate solutions given in Section II for the discrete case.

2.3.6 Logarithmic Form of the Coupled Line Equations, and Improved Approximate Solution

The perturbation results of Section 2.3.4, given in (116) to (120), were expected on intuitive grounds to be valid for short lines, whose additional loss due to mode conversion is small. In particular, the results of (117) and (119) depend on having $|\rho| \ll 1$, $|\mathcal{J}\mathcal{J}| \ll 1$. These relations may

again be regarded as the first terms in power series expansions for the various quantities.

As the length of guide increases we expect the mode conversion loss to increase; for example in Sections III and IV below it is shown that for all random guides that we deal with, $\langle \text{Re } \beta \beta \rangle$ is proportional to the total length of the guide. It is clear that as the length of guide increases, eventually $|\beta \beta| \gg 1$. Under these conditions we no longer obtain a valid approximation from the first two terms of (112a); many terms become significant, and (117) and (118) are no longer valid. It would seem that the results of (119) are also invalid for long guides. However while (119a) is certainly invalid, we will see that it is plausible that (119b), (119c), and (119d) will remain good approximations for long guides with large mode conversion loss if the differential attenuation constant is large enough in a certain sense compared to the coupling coefficient. The detailed mathematical analysis for this problem is given in a companion paper;¹⁸ however, it is not difficult to see on physical grounds that something of this sort is to be expected.

Consider a long guide of length L with a large enough differential loss so that the total differential loss $|\Delta\alpha|z$ in a short section of length z is large;

$$|\Delta\alpha|z \gg 1, \quad z \ll L. \tag{143}$$

Now let this guide be divided up into M equal sections of length z by ideal mode filters, so that

$$Mz = L. \tag{144}$$

An ideal mode filter by definition transmits TE_{01} without loss or phase shift, and suppresses all other modes completely. (Practical mode filters may consist of a section of helix guide that has low loss to TE_{01} , high loss to all other spurious modes other than the higher order TE_{0m} .) In addition to the requirement of (143), further assume that each section z is short enough so that the perturbation results of (116) to (120) do apply to the individual sections; for example, we might require from (115) that

$$\int_{(k-1)z}^{kz} |c(s)| ds \ll 1. \tag{145}$$

It is more or less obvious on physical grounds that under these conditions the insertion of ideal mode filters will not alter the over-all TE_{01} loss significantly, because the spurious mode level is not likely to build up to a significant magnitude.

Denote quantities applying to the k^{th} section by the symbol k in the following equation; quantities applying to the over-all guide will be written without subscripts. Then the over-all TE_{01} transmission parameters for the guide with ideal mode filters are given in terms of the TE_{01} parameters for individual sections as follows:

$$\begin{aligned} G_0 &= \prod_{k=1}^M k G_0, \\ g &= \prod_{k=1}^M k g, \\ \Lambda &= \sum_{k=1}^M k \Lambda, \\ A &= \sum_{k=1}^M k A, \\ \Theta &= \sum_{k=1}^M k \Theta. \end{aligned} \tag{146}$$

Then, for example, Λ may be written from (119), (120) and (146) as follows:

$$\Lambda \approx \sum_{k=1}^M \int_{(k-1)z}^{kz} ds \int_{(k-1)z}^s dt \cdot c(s)c(t) e^{\Delta\Gamma(s-t)}. \tag{147}$$

In (147) the double integral has been taken from (120a); the other two versions of this integral could of course be used equally well. Now it is more or less obvious by inspection of the integrand and the limits and from physical considerations that (147) is approximately equivalent to

$$\Lambda \approx \int_0^{L=Mz} ds \int_0^s dt \cdot c(s)c(t) e^{\Delta\Gamma(s-t)}. \tag{148}$$

In other words, the results of (119b), (119c), and (119d) should remain valid for the whole line, as stated above. Since the requirement in the above crude argument is that the mode conversion loss must be small in a section for which the total differential loss is high, we would intuitively expect that for (119b), (119c), and (119d) to hold for a long line, the ratio of the coupling coefficient $|c(z)|$ to the differential attenuation constant $|\Delta\alpha|$ must be small in some sense which we have not yet attempted to define.

While the above argument may be physically appealing, it is obviously desirable to put these statements on firmer ground; this is done elsewhere in this issue.¹⁸ We summarize here briefly some of the results of this investigation.

Under certain conditions we may write $\Lambda(z)$ as a series expansion;

$$\Lambda(z) = \int_0^z c(s) e^{\Delta\Gamma s} ds \int_0^s c(t) e^{-\Delta\Gamma t} dt + \sum_{n=2}^{\infty} \lambda_n(z). \quad (149)$$

In (149) we have written out only the first term explicitly; we see that this is identical to the previous approximation of (119b) and (120a). Bounds on the higher terms are given in detail in Ref. 18; we give here some slightly poorer but simpler bounds.

$$\begin{aligned} |\lambda_1(z)| &= \left| \int_0^z c(s) e^{\Delta\Gamma s} ds \int_0^s c(t) e^{-\Delta\Gamma t} dt \right| \leq K \int_0^z |c(s)| ds; \\ |\lambda_n(z)| &\leq K^3 (2.225K^2)^{n-2} \int_0^z |c(s)| ds; \quad n \geq 2, \\ &0 \leq K \leq 0.3; \end{aligned} \quad (150)$$

where K is defined by

$$\int_0^z |c(s)| e^{\Delta\alpha(z-s)} ds \leq K \quad \text{for every } z \geq 0. \quad (151)$$

Finally, convergence is guaranteed only for $K < 0.455$; a case is known where the series of (149) diverges for $K > 0.5$. Equation (151) requires that $|c(z)|$ be uniformly small in a certain sense, with respect to $|\Delta\alpha|$.

Once again we take as an approximate solution the first term of (149). The bounds in Ref. 18 are again almost the best possible in the same sense as in the case of Section 2.3.4; i.e., cases are known where the higher-order terms are almost as large as their bounds. In cases of interest $K \ll 1$, and the bounds converge rapidly. However, as in the case of Section 2.3.4, this is not sufficient to guarantee that the terms themselves converge rapidly; and here again we lack precise information on the way in which the first term (perturbation solution) approximates the true solution.

Because of the relationship between discrete and continuous cases, similar statements can be made regarding the results of (65) and (71) for the discrete case.

2.3.7 TE_{01} Loss in Terms of Fourier Coefficients of $c(z)$ when $\Delta\alpha = 0$

If the differential attenuation constant is equal to zero,

$$\Delta\alpha = 0, \quad (152)$$

the above perturbation results for the loss A [see (119c) and (120)] may be further simplified. This case is of interest as an approximation

to the situation present in real copper guides. In practical copper guides mode filters, consisting of a length of helix guide, must be inserted at regular intervals, perhaps a few hundred feet apart, for reasons that will be further discussed in Sections III and IV below. If the mode filters may be assumed ideal, we consider each section separately and simply add their individual TE_{01} loss and phase, i.e., Λ , A , and Θ , as in (146). If for each copper guide section of length L the total differential loss $|\Delta\alpha|L$ is small,

$$|\Delta\alpha|L \ll 1, \quad (153)$$

on intuitive grounds we approximate the true solution by setting $\Delta\alpha = 0$ in the various approximate solutions of (116) to (120).

Thus, let us set $\Delta\alpha = 0$ in (119c) and (120c) to obtain

$$A \approx \frac{1}{2} \operatorname{Re} \int_0^L \int_0^L c(s)c(t) e^{j\Delta\beta|t-s|} ds dt \quad (154)$$

for a guide of length L . Recalling that $c(s)$ and $c(t)$ are real in ideal metallic guide, (154) yields

$$\begin{aligned} A &\approx \frac{1}{2} \operatorname{Re} \int_0^L \int_0^L c(s)c(t) e^{j\Delta\beta(t-s)} ds dt \\ &= \frac{1}{2} \int_0^L \int_0^L c(s)c(t) e^{j\Delta\beta(t-s)} ds dt \\ &= \frac{1}{2} \int_0^L c(s) e^{-j\Delta\beta s} ds \int_0^L c(t) e^{+j\Delta\beta t} dt \\ &= \frac{1}{2} \left| \int_0^L c(s) e^{-j\Delta\beta s} ds \right|^2. \end{aligned} \quad (155)$$

Summarizing,

$$A \approx \frac{1}{2} \left| \int_0^L c(s) e^{-j\Delta\beta s} ds \right|^2, \quad \Delta\alpha = 0. \quad (156)$$

We note further from (119a) and (116) that

$$|G_0| = g \approx 1 - \frac{1}{2} \left| \int_0^L c(s) e^{-j\Delta\beta s} ds \right|^2. \quad (157)$$

From (157) we have to second order

$$|G_0|^2 \approx 1 - \left| \int_0^L c(s) e^{-j\Delta\beta s} ds \right|^2. \quad (158)$$

But from (118), for $\Delta\alpha = 0$

$$|G_1|^2 \approx \left| \int_0^L c(s) e^{-j\Delta\beta s} ds \right|^2. \tag{159}$$

Therefore from (158) and (159) we have to second order

$$|G_0|^2 + |G_1|^2 \approx 1, \quad \Delta\alpha = 0. \tag{160}$$

In Appendix C it is shown that (160) must hold exactly for $\Delta\alpha = 0$ [see (C-13)]. This is something like conservation of energy; in fact if $\alpha_0 = \alpha_1 = 0$, (C-13) is precisely conservation of energy.

We note in passing that a similar result to that of (156) is readily found for the discrete case of Section II. Proceeding in an analogous way from (65c) and the second form of (71d), we find for the discrete case

$$A \approx \frac{1}{2} \left| \sum_{i=1}^N x_i e^{-j\Delta\beta l_0 i} \right|^2, \quad \Delta\alpha = 0. \tag{161}$$

Similarly the result of (160) is readily seen to hold true to second order, and by Appendix C must also be true exactly.

Equation (156) states that the loss in nepers A is one half the square of the magnitude of the Fourier transform of the coupling coefficient, with transform variable $\Delta\beta$, the differential propagation constant. Since we deal with the case $\Delta\alpha = 0$, the logarithmic perturbation theory of Section 2.3.6 does not indicate that (156) is valid for long lines; consequently, this approximate result will remain valid only for short line sections, perhaps subject to a condition such as $\int_0^L |c(s)| ds \ll 1$.

It will prove convenient to rewrite (156) in a slightly modified form. Define

$$I = e^{j\pi(\Delta\beta L/2\pi)} \int_0^L c(s) e^{-j\Delta\beta s} ds. \tag{162}$$

We have

$$|I| = \left| \int_0^L c(s) e^{-j\Delta\beta s} ds \right|. \tag{163}$$

I is closely related to the approximate expression for the spurious mode transfer coefficient, given in (118). From (162) and (163), (156) may be written as

$$A \approx \frac{1}{2} |I|^2. \tag{164}$$

It turns out to be useful to expand the coupling coefficient $c(z)$ in a Fourier series as follows:†

$$c(z) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n z/L}; \quad c_{-n} = c_n^*. \quad (165)$$

Substituting (165) into (162), we have after some algebraic simplification:

$$I = L \sum_{n=-\infty}^{\infty} c_n (-1)^n \frac{\sin \pi \left(\frac{\Delta\beta L}{2\pi} - n \right)}{\pi \left(\frac{\Delta\beta L}{2\pi} - n \right)}. \quad (166)$$

Now, (166) has a rather striking form. Assume for a moment that the only variation of I (and hence A) with the frequency f of the radiation in the guide occurs through $\Delta\beta$. If we take the independent variable in (166) to be proportional to $\Delta\beta$, then (166) is simply the sampling theorem representation of a complex band-limited function,^{33,34} i.e., a complex function whose real and imaginary parts are each band-limited. Taking the dimensionless quantity $\frac{\Delta\beta L}{2\pi}$ as the independent variable, $I\left(\frac{\Delta\beta L}{2\pi}\right)$ will contain no frequencies $|\nu|$ greater than $\frac{1}{2}$.‡ By (164) the loss A is proportional to the square of the magnitude of I , or alternately to the sum of the squares of the magnitudes of the real and imaginary parts of I ; therefore the TE_{01} loss A , regarded as a function of the normalized independent variable $\frac{\Delta\beta L}{2\pi}$, will contain no frequencies greater than 1. If for the time being we neglect any variation of the c_n 's with $\Delta\beta$ in (166), then $I\left(\frac{\Delta\beta L}{2\pi}\right)$ is determined by its values at the sample points. At the n^{th} sample point

† Here c_n is the n^{th} complex Fourier coefficient of the continuous coupling coefficient $c(z)$. In contrast, in Section 2.2 above we have used c_k to represent the coupling coefficient of the k^{th} discrete mode converter. In the following work the meaning will always be clear from the context.

‡ Here, and often in the sequel, we use the word frequency to denote the independent variable ν of the Fourier transform of some quantity of interest. In the present case we consider the Fourier transform of $I\left(\frac{\Delta\beta L}{2\pi}\right)$ with respect to the independent variable $\frac{\Delta\beta L}{2\pi}$; since $\frac{\Delta\beta L}{2\pi}$ is dimensionless, the corresponding Fourier transform variable ν is also dimensionless. Thus, if $\mathcal{G}(\nu)$ is the Fourier transform of $I\left(\frac{\Delta\beta L}{2\pi}\right)$, $\mathcal{G}(\nu) = \int_{-\infty}^{\infty} I(\xi) e^{-j2\pi\nu\xi} d\xi$.

$$\frac{\Delta\beta L}{2\pi} = n, \quad B \equiv \frac{2\pi}{\Delta\beta} = \frac{L}{n}. \quad (167)$$

The n^{th} sample point occurs at the free-space wavelength or frequency (of the radiation in the guide) where the beat wavelength B is equal to the mechanical wavelength $\frac{L}{n}$ associated with the n^{th} Fourier coefficient of the coupling coefficient. The value of I at the n^{th} sample point is

$$I(n) = Lc_n(-1)^n. \quad (168)$$

By (162) and (118) the spurious mode transfer coefficient at the n^{th} sample point for the continuous bend becomes

$$G_1(n) \approx jLc_n, \quad (169)$$

with an analogous result for the continuous offset and diameter change. $I(n)$ and $G_1(n)$ are determined only by the n^{th} Fourier coefficient. At intermediate values of $\frac{\Delta\beta L}{2\pi}$, $I\left(\frac{\Delta\beta L}{2\pi}\right)$ is determined by interpolating between the sample values with $\frac{\sin x}{x}$ functions, as shown in (166).

Now the object of the present calculations is to determine the loss A as a function of the frequency f , and later to determine the statistics of the loss-frequency curves for guides with random imperfections. However, $\Delta\beta$ is approximately proportional to the free-space wavelength λ of the radiation in the guide, and thus *inversely* proportional to the frequency f , if the two modes involved (the TE_{01} signal mode and the spurious mode) are both far from cutoff. Therefore for analytical purposes it is more convenient to choose an independent variable proportional to the free-space wavelength λ rather than to the frequency f , and this is what is done in the present paper. If a single spurious mode is under consideration, it is most convenient to choose the dimensionless parameter $\frac{\Delta\beta L}{2\pi}$, which is approximately proportional to λ , as the independent variable. If more than one spurious mode is being considered, we see in Section 2.4 below that the loss A is given by a sum of terms of the form given in (156), with of course different c 's and $\Delta\beta$'s for each spurious mode. While the different $\Delta\beta$'s are all approximately proportional to λ , they have different constants of proportionality, and hence it is perhaps most convenient to take λ itself as the independent variable.

In practice we will consider only fairly narrow percentage bandwidths (although the absolute bandwidths will be enormous compared with conventional communication channels) at any one time. Therefore the

fact that we take λ as the independent variable instead of f should cause no serious inconvenience.

The sampling theorem interpretation of (166) requires that the c_n , defined by (165), be independent of λ (or frequency) and hence $\Delta\beta$; this may or may not be true. Since $c(z)$ is given by the product of one of Morgan's coupling coefficients and a geometric parameter [see (86) or (94)], the λ -dependence of the c_n 's is identical to the λ -dependence of C_t , C_o , or C_d . From the equations for these three coefficients given in Appendix A we see that far from cutoff, C_o and C_d are approximately independent of λ , as desired, but that C_t varies approximately inversely with λ . In Section 2.3.8 below we shall introduce additional coupling coefficients of Morgan, $\Xi_{[nm]}$, which permit a similar treatment for general continuous cross-sectional deformations of copper guide; the $\Xi_{[nm]}$ vary approximately directly with λ . The geometric parameters associated with C_o and C_d are the derivatives with respect to distance z of offset x and radius a respectively; hence these geometric parameters are dimensionless. The geometric parameter associated with C_t is curvature, which has the dimension of length⁻¹. The geometric parameters associated with the $\Xi_{[nm]}$ all have the dimensions of length. As a convenient device for recalling these facts, the exponent of λ in stating the λ -dependence of the coupling coefficient is the same as the exponent of length in stating the dimensions of its associated geometric parameter.

We wish to retain the sampling theorem interpretation of (166) even in those cases where the c_n 's are not independent of λ and hence of $\Delta\beta$. Now in those cases where the coupling coefficients are not approximately constant (i.e., C_t , $\Xi_{[nm]}$) the variation with λ is quite slow. Since as mentioned above we need consider only narrow percentage bandwidths, the principal variation with λ in (166) occurs through $\Delta\beta$, and not through the coupling coefficient. Consequently, over the moderate bandwidths of interest we may evaluate Morgan's coefficients at the middle of the narrow band under consideration.

There are several more elegant methods for deriving this approximation that state in effect that c_n is to be evaluated at the wavelength λ_n (or frequency) corresponding to the n^{th} sample point, given in (167), rather than at the operating wavelength, as implied in (166). This is certainly true at the sample points, by (168), and appears plausible in general. In Section 2.3.9 below we show, for example, that under certain reasonable conditions a guide with a given straightness deviation may be described equally well as either a continuous bend, for which the c_n 's vary approximately inversely with λ , or as a continuous offset, for which the c_n 's are approximately independent of λ . By a suitable transforma-

tion it is always possible to change the geometric parameter to dimensionless form (in general, to the derivative with respect to z of a parameter with the dimensions of length, such as radius, offset, ellipticity, etc.) so that the corresponding coupling coefficient is approximately independent of λ (and hence $\Delta\beta$) far from cutoff.

Alternately, this result may be derived by fairly simple manipulation of the sampling theorem.

The final result, however, is that over the fairly narrow percentage bands of interest we are justified in neglecting the λ -(or frequency) variation of the coupling coefficient (i.e., C_t or $\Xi_{(nm)}$).

The TE₀₁ loss A in terms of the Fourier coefficients c_n is, from (164) and (166),

$$A \approx \frac{L^2}{2} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_m c_n^* (-1)^{(m-n)} \frac{\sin \pi \left(\frac{\Delta\beta L}{2\pi} - m \right) \sin \pi \left(\frac{\Delta\beta L}{2\pi} - n \right)}{\pi \left(\frac{\Delta\beta L}{2\pi} - m \right) \pi \left(\frac{\Delta\beta L}{2\pi} - n \right)}. \quad (170)$$

At the sample points of I , given in (167), the TE₀₁ loss becomes simply

$$A(n) \approx \frac{L^2}{2} |c_n|^2. \quad (171)$$

As mentioned earlier the bandwidth of A is twice the bandwidth of I ; thus the Fourier transform of A with respect to the independent variable $\frac{\Delta\beta L}{2\pi}$ will contain no frequencies greater than 1.

Therefore while A is also a band-limited function, it has twice as many sample points as I . Consequently A is *not* completely determined by its values given in (171) at the sample points of I . Simple examples are readily found of two different guides in which the TE₀₁ loss is the same at the sample points of I , given in (167), but differs greatly between these sample points. Thus consider the following two cases, in which all but two adjacent Fourier components of the coupling coefficient are identically zero:

$$\begin{aligned} \text{(a)} \quad c(z) &= 2 \cos 2\pi k \frac{z}{L} + 2 \cos 2\pi(k+1) \frac{z}{L}. \\ \text{(b)} \quad c(z) &= 2 \cos 2\pi k \frac{z}{L} - 2 \cos 2\pi(k+1) \frac{z}{L}. \end{aligned} \quad (172)$$

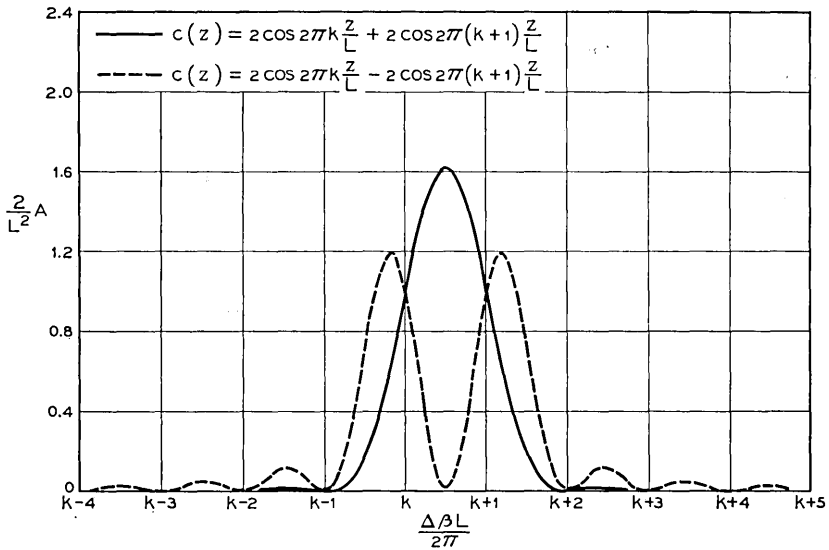


Fig. 10 — TE₀₁ loss for coupling coefficient with two sinusoidal components.

The TE₀₁ loss A is shown for these two cases in Fig. 10. While the losses are the same at the sample points of I , they differ markedly in between.

While A is not completely determined by its values at the sample points of I , it is clear from (170) that the principal contribution to A in the region near the k^{th} sample point arises from values of m and n close to k . This relation between the TE₀₁ loss and the coupling coefficient is of great importance, as it means that the additional TE₀₁ loss due to some particular spurious mode in a given frequency range is caused primarily by the components of the coupling coefficient which lie in a corresponding (mechanical) spectral region. Thus, for example, loss due to TE₀₁ — TE₁₂ coupling in 2-inch guide between 50 and 60 kmc is caused primarily by straightness deviations with periods between 2 and 2.4 feet, the range of beat wavelength B corresponding to this frequency band.

Equation (170) is also useful in considering situations where only isolated Fourier coefficients of the coupling coefficient are nonzero; the TE₀₁ loss A then becomes a series of isolated narrow peaks of shape $\left(\frac{\sin x}{x}\right)^2$. There are at least two such cases of practical interest:

i. A periodically supported copper guide that deforms elastically under its own weight (the "serpentine bend"),³⁵ with support period short com-

pared to the total length of guide between mode filters. The Fourier components occur only at the fundamental and harmonics of the support period; the amplitude of the Fourier coefficients decreases rapidly as the order of the harmonic increases. Thus the TE₀₁ loss vs wavelength curve will have a series of equally spaced narrow loss peaks of rapidly decreasing magnitude.

ii. Shuttle pulse measurements in a copper waveguide without mode filters. Here, because of the absence of mode filters, the pulse traverses an iterated structure. Since the Fourier series expansion of the coupling coefficient for any number of round trips must be identical to the expansion for a single round trip, for N round trips only every N^{th} Fourier coefficient will be nonzero. Again the TE₀₁ transmission will consist of equally spaced narrow loss peaks separated by wide pass bands of low loss.⁸

These two cases will not be discussed further in the present paper.

The spacing between sample points of I is an important parameter in the analysis and in the interpretation of experimental transmission data. Since $\Delta\beta$ is proportional to the free-space wavelength λ ,

$$\Delta\beta = D\lambda, \tag{173}$$

where the constant D depends on the spurious mode. From (167) the wavelength corresponding to the n^{th} sample point, λ_n , is given by

$$\frac{D\lambda_n L}{2\pi} = n, \quad \lambda_n = \frac{2\pi n}{DL}. \tag{174}$$

Therefore the sample point spacing in wavelength is

$$\Delta\lambda_n = \lambda_{n+1} - \lambda_n = \frac{2\pi}{DL}. \tag{175}$$

Thus from (173) and (175), noting the definition of beat wavelength B in (167),

$$\frac{\Delta\lambda}{\lambda} = \frac{B}{L}, \tag{176}$$

where we have dropped the subscript n since the result of (175) is independent of n . Since

$$\frac{\Delta\lambda}{\lambda} \approx \frac{\Delta f}{f}, \tag{177}$$

we have finally

$$\frac{\Delta f}{f} \approx \frac{\Delta \lambda}{\lambda} = \frac{B}{L} \quad (178)$$

for the sample point spacing, in either frequency or wavelength. We see from (175) that the sample points are equally spaced in wavelength, and consequently unequally spaced in frequency. However over a sufficiently narrow band the variation in sample point spacing in frequency will be small. The beat wavelength B between TE_{01} and all other TE modes is tabulated in Appendix D for the frequencies 50, 55, and 60 kmc in 2-inch diameter guide.

2.3.8 Morgan's Coupling Coefficients for Small Cross-Sectional Deformations in Lossless Metallic Guide

Morgan has determined to first order the spurious modes excited by an incident TE_{01} wave at a small arbitrary deformation of the cross section of a cylindrical metallic guide.⁵ This analysis must include the continuous offset and diameter change studied above when those deformations are small; it includes in addition higher-order deformations of the cross section, such as ellipticity, tri-foil, etc.

Let the surface of the slightly deformed guide be given in cylindrical co-ordinates by

$$r = a + \rho(\varphi, z), \quad (179)$$

$$\rho(\varphi, z) = \sum_{n=0}^{\infty} [a_n(z) \cos n\varphi + b_n(z) \sin n\varphi], \quad b_0(z) = 0, \quad (180)$$

$$\rho(\varphi, z) = 0, \quad a_n(z) = 0, \quad b_n(z) = 0; \quad \begin{array}{l} z \leq 0 \\ z \geq L \end{array} \quad (181)$$

where a is the radius of the perfect guide. $\rho(\varphi, z)$ must be suitably small (we have omitted Morgan's small dimensionless parameter ϵ). Then the $n = 0$ term corresponds to a continuous diameter variation, the $n = 1$ term to a continuous offset, the $n = 2$ term to ellipticity, the $n = 3$ term to what has been called tri-foil, etc. By (181) it is assumed that the guide is distorted only in the interval $0 < z < L$.

Morgan has shown that to first order a TE_{01} wave incident on the deformation of (179) to (181) excites the forward and backward TE_{nm} modes. We denote the various modes as follows:

$$\begin{aligned} I_{[0m]}^+ & \text{--- forward } TE_{0m} \\ I_{[0m]}^- & \text{--- backward } TE_{0m} \\ I_{[nm]}^{\parallel+}, I_{[nm]}^{\perp+} & \text{--- two polarizations of forward } TE_{nm} \\ I_{[nm]}^{\parallel-}, I_{[nm]}^{\perp-} & \text{--- two polarizations of backward } TE_{nm}. \end{aligned} \quad (182)$$

We denote the two orthogonal polarizations of each mode (except the TE_{0m}) by the symbols \parallel and \perp , rather than the subscripts x and y used in Section 2.1 above. The geometric imperfections of Section 2.1 — offsets and tilts — could naturally be resolved into components along the x - and y -axes; with the more general deformations of (179) to (181) this is of course not possible, and the \parallel and \perp notation appears more natural. In the special case $n = 1$, using the geometry of Fig. 5, we may if we wish identify $a_1(z)$ with $x(z)$, $b_1(z)$ with $y(z)$, and the \parallel and \perp polarizations of the resulting TE_{1m} spurious modes with the subscripts x and y respectively of Section 2.1.

We normalize the complex mode amplitudes in the usual way [see (88) and (89)]:

$$\begin{aligned} I_{[0m]}^{\pm}(z) &= e^{\mp\Gamma_{[0m]}z} G_{[0m]}^{\pm}(z) \\ I_{[nm]}^{\parallel\pm}(z) &= e^{\mp\Gamma_{[nm]}z} G_{[nm]}^{\parallel\pm}(z) \\ I_{[nm]}^{\perp\pm}(z) &= e^{\mp\Gamma_{[nm]}z} G_{[nm]}^{\perp\pm}(z). \end{aligned} \tag{183}$$

In lossless metallic guide the propagation constants $\Gamma_{[nm]}$ are of course pure imaginary,

$$\Gamma_{[nm]} = j\beta_{[nm]}. \tag{184}$$

Assuming a unit incident TE_{01} wave,

$$G_{[01]}^{+}(0) = 1, \tag{185}$$

the $G_{[nm]}$ become the normalized spurious mode complex transfer coefficients. They are found to first order from Morgan's analysis, and given in a slightly modified notation as follows:

$$\begin{bmatrix} G_{[nm]}^{\parallel\pm}(L) \\ G_{[nm]}^{\perp\pm}(L) \end{bmatrix} \approx -j\Xi_{[nm]} \int_0^L \begin{bmatrix} a_n(z) \\ b_n(z) \end{bmatrix} e^{-j(\beta_{01} \mp \beta_{nm})z} dz \tag{186}$$

where the coupling coefficient $\Xi_{[nm]}$ is given by

$$\Xi_{[nm]} = \frac{1}{\sqrt{\epsilon_n} a^3} \frac{k_{01}k_{nm}^2}{\sqrt{k_{nm}^2 - n^2}} \frac{1}{\sqrt{\beta_{01}\beta_{nm}}} \tag{187}$$

$$\epsilon_n = \begin{matrix} 1, & n = 0 \\ 2, & n \geq 1 \end{matrix}; \quad J_n'(k_{nm}) = 0. \tag{188}$$

It is worth noting that the angular index n of the spurious mode is identical to the subscript of the pertinent component of mechanical deformation, a_n or b_n ; therefore, to first order the modes excited by TE_{01} at a given deformation have the same angular symmetry as the defor-

mation. $\Xi_{[nm]}$ is the same for forward and backward modes, and for both polarizations; in contrast C_o , C_t , and C_d of Appendix A are different for forward and backward modes. This contrast in behavior will be discussed in Section 2.3.9. The $\Xi_{[nm]}$ are given in Appendix D for the frequencies 50, 55, and 60 kmc, in 2-inch diameter guide. We note again that these results hold true only when there is no distortion of the guide at the ends, $z = 0$ and $z = L$, by (181).

Special comment is necessary for the $G_{[0m]}$ terms in (186), and in particular for the $G_{[01]}^+$ term. The $n = 0$ terms correspond to the TE_{0m} modes, which have only a single polarization. Since $b_0(z) = 0$ in (181), the \parallel terms of (186) are the significant ones; and since we have only one polarization, the symbol \parallel may be conveniently dropped. The $G_{[01]}^+$ term, corresponding to the forward TE_{01} , requires special interpretation. It is to be considered a first-order correction to be added to the unperturbed solution, i.e., $G_{[01]}^+ = 1$. It is readily seen that $G_{[01]}^+$ in (186) represents a phase shift of the unperturbed TE_{01} wave, caused by a change in the average diameter of the guide.

By analogy with the previous analysis, we may use the results of (186) to (188) in any of the results of Section 2.3 above for the continuous case—e.g., (116) to (120)—by substituting $-\Xi_{[nm]}a_n(z)$ or $-\Xi_{[nm]}b_n(z)$ for the continuous coupling coefficient $c(z)$. While these coupling coefficients were derived by Morgan for lossless metallic guide, we expect as before that these results provide a satisfactory approximation for real copper guide by modifying the various propagation constants to take account of the small losses present in copper guide. The relationship of these results to previous ones for the continuous case is discussed in Section 2.3.9.

2.3.9 Relationships between Various Metallic Guide Coupling Coefficients

A sufficiently small straightness deviation of the guide axis may obviously be described equally well as a continuous bend, a continuous offset, or simply as a continuous displacement. For these three cases we have given above three different coupling coefficients— $C_{t[m]}$, $C_{o[m]}$, and $\Xi_{[1m]}$ —that yield, among other things, the first-order spurious mode transfer coefficient of a deformed guide. While the results of Section 2.3.4—in particular (118) for continuous bends and its analog for continuous offsets and diameter changes—were derived for only a single forward spurious mode, the result for the first-order spurious mode transfer coefficient holds true in general, with of course the proper coupling coefficient for each of the spurious modes, as discussed in Section 2.4 below. Since a given deformation that is suitably small may be described in these three different ways, and since we must get the same

answer for the spurious mode output in each calculation, the three coupling coefficients must be related to each other. Similar statements obviously apply to the two coupling coefficients $C_{d[m]}$ and $\Xi_{[0m]}$ for diameter variations.

For a guide with a continuous bend, confined to the x - z plane, the TE_{1m}^\pm transfer coefficient is, from (118) and (91), to first order

$$G_{[1m]}^{x\pm} \approx j \int_0^L \frac{C_{t[1m]}^\pm}{R_x(z)} e^{-j(\beta_{01} \mp \beta_{1m})z} dz, \tag{189}$$

where $R_x(z)$ is the radius of curvature of the guide axis, and the superscript x indicates that the x -polarization is under consideration. If the slope of the guide axis is small compared to unity, the distance along the perturbed guide axis is approximately equal to the distance along the unperturbed guide axis, so that we may regard z in (189) as measured along a fixed rectangular co-ordinate axis. Further we may approximate the curvature as

$$\frac{1}{R_x(z)} \approx x''(z). \tag{190}$$

Then (189) may be written

$$G_{[1m]}^{x\pm} \approx j C_{t[1m]}^\pm \int_0^L x''(z) e^{-j(\beta_{01} \mp \beta_{1m})z} dz. \tag{191}$$

Similarly for a guide with a continuous offset in the x - z plane the spurious mode output is, from (96) and (118) with the j replaced by -1 ,

$$G_{[1m]}^{x\pm} \approx -C_{o[1m]}^\pm \int_0^L x'(z) e^{-j(\beta_{01} \mp \beta_{1m})z} dz. \tag{192}$$

If the ends of the guide are parallel, so that

$$x'(0) = x'(L) = 0, \tag{193}$$

and if the perturbation is small, the spurious mode output must be identical in the two cases and the right-hand sides of (191) and (192) must be equal. Integrating (191) by parts, we find subject to (193)

$$G_{[1m]}^{x\pm} \approx -(\beta_{01} \mp \beta_{1m}) C_{t[1m]}^\pm \int_0^L x'(z) e^{-j(\beta_{01} \mp \beta_{1m})z} dz. \tag{194}$$

For (192) and (194) to be identical, we must have

$$\frac{C_{o[1m]}^\pm}{C_{t[1m]}^\pm} = \beta_{01} \mp \beta_{1m}. \tag{195}$$

Using (A-1) and (A-2) of Appendix A, (195) is easily shown to be an identity.

Similarly a line with a continuous offset, but with no net offset between its two ends, is described in terms of displacement of the guide axis by the results of (186). Thus setting

$$a_1(z) = x(z), \quad (196)$$

we have for the spurious mode output

$$G_{[1m]}^{x\pm} \approx -j\Xi_{[1m]} \int_0^L x(z) e^{-j(\beta_{01} \mp \beta_{1m})z} dz, \quad (197)$$

where we have replaced the superscript \parallel in (186) by x . This result is valid only when the distortion vanishes at the ends of the guide; in this case this means that there must be no offset between the two ends of the guide,

$$x(0) = x(L) = 0. \quad (198)$$

Subject to (198), (197) must agree with the previous two results. Integrating (192) by parts, and using the condition of (198),

$$G_{[1m]}^{x\pm} \approx -j(\beta_{01} \mp \beta_{1m}) C_{o[1m]} \int_0^L x(z) e^{-j(\beta_{01} \mp \beta_{1m})z} dz. \quad (199)$$

For (197) and (199) to be identical, we must have

$$\frac{\Xi_{[1m]}}{C_{o[1m]}^{\pm}} = \beta_{01} \mp \beta_{1m}. \quad (200)$$

By (A-1) of Appendix A and (187), (200) is easily seen to be an identity.

An analogous study can be made for a continuous diameter change. From Sections 2.3.2 and 2.3.4

$$G_{[0m]}^{\pm} = -C_{d[1m]}^{\pm} \int_0^L r'(z) e^{-j(\beta_{01} \mp \beta_{0m})z} dz. \quad (201)$$

From (186) we have

$$G_{[0m]}^{\pm} = -j\Xi_{[0m]} \int_0^L a_0(z) e^{-j(\beta_{01} \mp \beta_{0m})z} dz, \quad (202)$$

valid when

$$a_0(0) = a_0(L) = 0. \quad (203)$$

Subject to (203), (201) and (202) must be identical. From (179) we have $r'(z) = a_0'(z)$; substituting this into (201), integrating by parts, and using (203),

$$G_{[0m]}^{\pm} = -j(\beta_{01} \mp \beta_{0m})C_{d[m]}^{\pm} \int_0^L a_0(z)e^{-j(\beta_{01} \mp \beta_{0m})z} dz. \quad (204)$$

For (202) and (204) to be identical, we must have

$$\frac{\Xi_{[0m]}^{\pm}}{C_{d[m]}^{\pm}} = \beta_{01} \mp \beta_{0m}. \quad (205)$$

Equation (A-4) of Appendix A and (187) show that (205) is an identity.

It is interesting that coupling coefficients found in such diverse ways are so simply related, and pleasing that these different calculations are all consistent with each other.

A simple physical interpretation may be given for the fact that the mode conversion caused by a given type of deformation of the guide may be described by different coupling coefficients (with of course corresponding different geometric parameters). These different descriptions correspond to using different sets of normal modes to describe the fields within the deformed guide.

For purposes of illustration we take the case of deviation from straightness of the guide axis. The coupling coefficient $\Xi_{[1m]}$ and (197) correspond to the normal modes of the original undeformed guide. The coupling coefficient $C_{o[m]}$ and (192) correspond to the normal modes of the deformed guide with cross section perpendicular to the axis of the undeformed guide. Finally, the coupling coefficient $C_{[1m]}$ and (191) correspond to the normal modes of the deformed guide with cross section perpendicular to the axis of the deformed guide. This process may be carried one step further, using the normal modes of a curved circular guide.⁹

A similar discussion can of course be given for the two representations of mode conversion caused by changes in diameter [$C_{d[m]}$ in (201) and $\Xi_{[0m]}$ in (202)].

2.4 Extension to Many Spurious Modes and Two Polarizations

Most of the above results are readily extended to include additional spurious modes, coupled to the TE_{01} signal mode to first order, and two polarizations of all such spurious modes (except TE_{0m}). First, we may proceed via the scattering matrices for discrete mode converters including many spurious modes, and thence to the continuous case via limiting processes, as above. Alternately, we may take the generalized telegraphist's equations, including all modes, and find perturbation solutions as above, and thence treat the discrete case by limiting processes. The conclusions of such a study may be summarized as follows:

1. The first-order spurious mode transfer coefficients for each of the spurious modes are given as above [e.g., (72), (73), (118) or (186)], with of course the appropriate coupling coefficient and differential propagation constant for each of the modes.

2. For the second-order TE_{01} transmission parameters ($G_0, g, \Lambda, A, \Theta$) we must replace each term involving the coupling coefficient by a summation of similar terms, each corresponding to a particular spurious mode, with of course the appropriate coupling coefficient and differential propagation constant for each mode. For this purpose we regard the two orthogonal polarizations of a spurious mode as two distinct (but degenerate) spurious modes. Thus for example we write

$$\Lambda \approx \sum_m \Lambda_m, \quad A \approx \sum_m A_m, \quad \Theta \approx \sum_m \Theta_m. \quad (206)$$

The quantities for the m^{th} mode are given for example by (65) and (71) or (161) for the discrete case, and (116) and (117), (119) and (120), or (156), (170) for the continuous case, with in each of these equations the subscript m attached to the coupling coefficients and differential propagation constants.

Much of this analysis is a straightforward extension of the two-mode analysis above, and will not be discussed here. The study of convergence of the approximate solutions, analogous to that discussed in Sections 2.3.4 and 2.3.6 and treated in Ref. 18, appears more difficult in the general multimode case, and little work has been done. Of course even in the simpler two-mode case we lack precise information on the validity of the approximate solutions, as discussed above; the same is certainly true here.

From the results of Section 2.3.7 and (206), the contribution to the TE_{01} loss A of each of the spurious modes arises from a different portion of the (mechanical) spectrum of the geometric imperfection. For example, consider a guide whose only imperfection is straightness deviation; the most important spurious modes are the forward TE_{1m} . Consider the frequency band from 50–60 kmc in 2-inch I.D. waveguide. The beat wavelength range, which is equal to the range of mechanical wavelengths of straightness deviation that contribute significantly to the TE_{01} loss component, is shown in Table I for each of these modes (see Appendix D).

The coupling coefficients for each polarization of each spurious mode for discrete tilts and offsets are obtained from (37), (38) and (47) respectively; for discrete diameter changes the coupling coefficients are obtained from (51). Thus in (71e) each of the C 's should have a subscript denoting the spurious mode, and α_i and b_i should have subscripts

TABLE I

m	Mechanical Wavelength Range for 50-60 kmc Band in Feet
1	2.45264-2.94898
2	1.99720-2.40812
3	0.46389-0.56232
4	0.21503-0.26289
5	0.12293-0.15228
6	0.07762-0.09815
7	0.05134-0.06723
8	0.03377-0.04763

denoting the x or y components of tilt or offset, depending on the polarization of the spurious mode.

For the cross-sectional deformations of Section 2.3.8, the continuous coupling coefficients for the two polarizations of the TE_{nm} mode are

$$c_{[nm]}^{\parallel\pm}(z) = -\bar{\Xi}_{[nm]} a_n(z), \tag{207a}$$

$$c_{[nm]}^{\perp\pm}(z) = -\bar{\Xi}_{[nm]} b_n(z), \tag{207b}$$

in terms of the geometric parameters of (180). The corresponding differential propagation constants are for both cases

$$\Delta\Gamma_{[nm]}^{\pm} = \Gamma_{[01]} \mp \Gamma_{[nm]}. \tag{208}$$

For the coefficients $C_{o[m]}$ for continuous offsets, we find from (47), by taking the limiting form of the discrete case, that the coupling coefficients for the two polarizations of the TE_{1m} mode are

$$c_{[1m]}^{x\pm}(z) = C_{o[m]}^{\pm} \cdot x'(z), \tag{209a}$$

$$c_{[1m]}^{y\pm}(z) = C_{o[m]}^{\pm} \cdot y'(z), \tag{209b}$$

where as usual $x(z)$ and $y(z)$ denote the two rectangular components of the displacement of the guide axis from the z -axis (Fig. 7b) and the prime denotes differentiation with respect to z . The differential propagation constants are again given by (208).

Similarly for a continuous diameter variation

$$c_{[om]}^{\pm}(z) = C_{d[m]}^{\pm} \cdot r'(z), \tag{210}$$

where $r'(z)$ is the derivative of the guide radius, and we have again (208) for the differential propagation constants.

The geometry for general continuous bends (not confined to a single plane) is considerably more complex, as might be anticipated from the latter part of Section 2.1.2, and requires special discussion. In general the guide axis may be a quite arbitrary 3-dimensional space curve; e.g.,

it may be pretzel-shaped. In such extreme cases we require the notation of differential geometry to describe the guide axis. We will not attempt to discuss the subject of differential geometry itself,³⁶ but merely use some of its simpler results.

Let us treat the general continuous case as the limiting form of the discrete case, as in Section 2.3.2 where the curved axis of the guide was confined to a single plane. In the present more general case we use the bent cylindrical co-ordinates ρ, φ, s described in conjunction with (29). Note that distance along the (bent) guide axis will be called s in the present section (it has previously been called z), so that z can refer to distance along a fixed rectangular co-ordinate axis. The other co-ordinates ρ and φ are as defined in Section 2.1.2. We particularly require the definition of the φ co-ordinate. Briefly, lines $\varphi = \text{constant}$ are drawn on the surface of the guide before it is bent, parallel to the (straight) axis of the guide. After bending (perhaps into a pretzel shape) these now deformed φ -lines furnish the φ -co-ordinates.

Now consider briefly the differential geometry of twisted space curves.³⁶ Let \mathbf{r} be the vector from a fixed origin to the curve. Three unit vectors characterize the geometric properties of a general space curve:

\mathbf{t} — tangent vector.

\mathbf{p} — principal normal vector.

\mathbf{b} — binormal vector.

Then the following relations hold true:

$$\frac{d\mathbf{r}}{ds} = \mathbf{t}. \quad (211a)$$

$$\frac{d\mathbf{t}}{ds} = \frac{1}{\rho} \mathbf{p}. \quad (211b)$$

$$\mathbf{b} = \mathbf{t} \times \mathbf{p}. \quad (211c)$$

$$\frac{d\mathbf{b}}{ds} = -\frac{1}{\tau} \mathbf{p}. \quad (211d)$$

$$\frac{d\mathbf{p}}{ds} = \frac{1}{\tau} \mathbf{b} - \frac{1}{\rho} \mathbf{t}. \quad (211e)$$

In (211) ρ is the radius of curvature, $\frac{1}{\rho}$ the curvature; similarly τ is called the radius of torsion, $\frac{1}{\tau}$ the torsion. The principal normal \mathbf{p} lies

in the plane of the circle that best approximates the twisted curve in the neighborhood of a given point; the radius of this circle is of course the radius of curvature ρ .

Now let us consider the limiting process in which a guide with a continuous three-dimensional bend is considered as the limit of a guide with closely spaced discrete tilts. In the following let $\theta(s)$ represent the orientation of the elementary discrete tilt located at distance s along the (bent) guide axis, measured in the bent cylindrical co-ordinate system described above in the present section and in Section 2.1.2. Then in the limit as the discrete tilts become a continuous three-dimensional space curve, $\theta(s)$ is given by

$$\frac{d\theta(s)}{ds} = -\frac{1}{\tau}. \tag{212}$$

From (37) the continuous coupling coefficients to the two polarizations of the TE_{1m}^\pm spurious mode are:

$$c_{[1m]^{x\pm}}(s) = C_{t[1m]^\pm} \cdot \frac{\cos \theta(s)}{\rho(s)} \tag{213a}$$

$$c_{[1m]^{y\pm}}(s) = C_{t[1m]^\pm} \cdot \frac{\sin \theta(s)}{\rho(s)}. \tag{213b}$$

Similarly for the TM_{11}^+ spurious mode:

$$c_{(11)^{x+}}(s) = C_{t(11)^+} \cdot \frac{\cos \theta(s)}{\rho(s)} \tag{214a}$$

$$c_{(11)^{y+}}(s) = C_{t(11)^+} \cdot \frac{\sin \theta(s)}{\rho(s)}. \tag{214b}$$

$\rho(s)$ is the radius of curvature, determined from (211b). $\theta(s)$ is found from (212) as the negative of the integral of the torsion $\frac{1}{\tau}$, determined from (211d).

The results of (213) and (214) are sufficiently general to include such things as pretzel-shaped waveguides and other unusual deformations. However since we seldom expect such things in practice, we seek to simplify these results. We are guided by the simplification in the discrete case that is discussed following (37). We assume that the angular deviation of the guide axis from the z -axis of an x, y, z rectangular co-ordinate system remains small;

$$\begin{aligned} |x'(z)| &\ll 1, \\ |y'(z)| &\ll 1. \end{aligned} \tag{215}$$

Further we assume that the $\varphi = 0$ co-ordinate on the surface of the guide initially lies in the x - z plane, and that it remains close to this plane. Then from (38), by the usual transformation from the discrete to the continuous case, we have

$$s \approx z, \quad (216a)$$

$$\frac{\cos \theta(s)}{\rho(s)} \approx x''(z), \quad (216b)$$

$$\frac{\sin \theta(s)}{\rho(s)} \approx y''(z), \quad (216c)$$

to be substituted into (213) and (214). The superscripts x and y of (213) and (214) in this special case correspond to linear polarizations defined with respect to the x and y axes of the fixed rectangular coordinate system; in general the superscripts of (213) and (214) correspond to the $\varphi = 0$ and $\varphi = \pi/2$ planes in the bent cylindrical coordinate system. The approximations of (216) are found directly for the continuous case in Appendix E.

III. THEORY OF GUIDES WITH RANDOM DISCRETE IMPERFECTIONS

We now apply the results of Section II to the study of guides with random discrete mode converters. The following cases of practical interest will be discussed:

i. Guides made of individual pipes that are perfect right circular cylinders of identical diameter and length l_0 but that have randomly imperfect joints, with either tilts or offsets. The first-order spurious modes in this case will be the TE_{1m} family, with the most significant being the forward TE_{11} and TE_{12} , and for tilts the forward TM_{11} . Both polarizations of each spurious mode must of course be considered.

ii. Guides made of individual pipes that are perfect right circular cylinders of identical length l_0 with perfect joints (no tilts or offsets) but of slightly different random diameters. The spurious modes will be the TE_{0m} family, with forward TE_{02} the most important. Each mode now has only a single polarization.

The statistics of the TE_{01} loss will be determined in terms of the statistics of the guide imperfections. Only the case of individual pipes of identical length will be considered.

The necessary results from Section II are summarized below in Section 3.1. Section 3.2 states the statistical models adopted for guides with random tilts, offsets, and diameter changes. In Section 3.3 the expected value of the average TE_{01} loss, and the power spectrum and total

power (expected value of the mean square value) of the TE_{01} loss variation about its average value, are determined in terms of the statistics of the random offsets, tilts, or diameter variations of the guide. In Section 3.4 these results are extended to cover long guides with mode filters. Representative numerical examples are given in Section 3.5 for a 20-mile total guide length. Section 3.6 considers the application of certain of these results for tilts and offsets to helix guide.

Most of the work of Section III will be confined to copper guide. As before, we assume that the coupling coefficients for ideal metallic guide provide a good approximation to those for copper guide, but modify the (pure imaginary) propagation constants for ideal metallic guide to account for the loss present in copper guide. Of course the analysis for diameter changes applies equally well to copper or helix guide.

3.1 TE_{01} Loss — Summary of Previous Results

We give below the TE_{01} normalized loss (in nepers) A , written in a form suitable for the purposes of the present section. These relations are readily obtained from the results of Section II, and in particular from Section 2.4 and (37), (38), (47), (51), (61) to (65), (71d), (71e), (161), and (206). First, from (206) the total TE_{01} loss A is given as the sum of terms due to the individual first-order spurious modes;

$$A \approx \sum_m A_m. \quad (217)$$

The term A_m due to the m^{th} spurious mode is given below in (218) and (219). In these and similar relations involving only a single spurious mode we omit the subscripts denoting the spurious mode on the coupling coefficients and differential propagation constants, in order to avoid unnecessarily cluttering up the equation. Therefore for each spurious mode

$$A \approx \frac{1}{2}A_0 + \sum_{k=1}^{N-1} A_k \cos k\Delta\beta l_0 \quad (218a)$$

where

$$A_k = e^{k\Delta\alpha l_0} \sum_{i=1}^{N-k} (x_i^{\parallel} x_{i+k}^{\parallel} + x_i^{\perp} x_{i+k}^{\perp}), \quad 0 \leq k \leq N-1. \quad (218b)$$

Tilts:

$$\begin{aligned} x_i^{\parallel} &= C_t \alpha_i \cos \theta_i \approx C_t \alpha_i^{\parallel} \\ x_i^{\perp} &= C_t \alpha_i \sin \theta_i \approx C_t \alpha_i^{\perp}. \end{aligned} \quad (218c)$$

Offsets:

$$\begin{aligned} x_i^{\parallel} &= C_o b_i \cos \theta_i = C_o b_i^{\parallel} \\ x_i^{\perp} &= C_o b_i \sin \theta_i = C_o b_i^{\perp}. \end{aligned} \quad (218d)$$

Diameter change:

$$\begin{aligned} x_i^{\parallel} &= C_d \Delta r_i \\ x_i^{\perp} &= 0. \end{aligned} \quad (218e)$$

The distinction between A_k of (218a) and (218b), the k^{th} Fourier coefficient of the TE_{01} loss due to a single spurious mode, and A_m of (217), denoting the TE_{01} loss due to the m^{th} spurious mode, will always be clear from the context. In particular cases the subscripts indicating the spurious mode will always be enclosed in [] or () to denote TE or TM spurious modes respectively.

For the case of zero differential attenuation constant, an alternate form is sometimes useful. From (161) for each spurious mode we have (with mode subscripts again omitted)

$$A \approx \frac{1}{2} \left| \sum_{i=1}^N x_i^{\parallel} e^{-j\Delta\beta l_0 i} \right|^2 + \frac{1}{2} \left| \sum_{i=1}^N x_i^{\perp} e^{-j\Delta\beta l_0 i} \right|^2; \quad \Delta\alpha = 0. \quad (219)$$

The x_i 's are again given by (218c) to (218e).

In the above results we have chosen to group the two polarizations of each spurious mode (TE_{1m} and TM_{11}^+) for tilts and offsets in each term of (217). We use the symbols \parallel and \perp to distinguish the two polarizations, rather than x and y as in Section II. The length of the individual pipes is l_0 , the total length of guide is $L_N = Nl_0$.

In these results the significant frequency dependence, at least over moderate bandwidths, occurs through the $\Delta\beta$, the differential phase constants between TE_{01} and the spurious modes. Far from cutoff (the case of greatest practical interest) the $\Delta\beta$ are approximately proportional to the free-space wavelength λ . A great simplification in the subsequent analysis is obtained by neglecting the frequency dependence of $\Delta\alpha$ and the x_i 's, setting these quantities equal to their values at the middle of the band of interest. Then A_m , the contribution to the TE_{01} loss of conversion to the m^{th} spurious mode, becomes a Fourier series, periodic in λ , with random coefficients A_k , as given in (218). A , the total TE_{01} loss due to mode conversion, is given by a sum of periodic components of incommensurable periods, according to (217).

Over moderate bandwidths, the effect of the frequency dependence of $\Delta\alpha$ is small far from cutoff. For offsets and diameter changes, the fre-

quency dependence of the coupling coefficients and hence the x_i is also small far from cutoff; however, for tilts the x_i vary approximately inversely with λ . Thus, the above approximation may appear questionable for tilts, but is used even here because of the resulting simplification in the analysis. [Compare the discussion in Section 2.3.7 following (169).]

We thus regard the loss A as a function of the free-space wavelength λ , and write

$$A = \bar{A} + \delta A \quad (220)$$

where \bar{A} is the average value over free-space wavelength λ of the loss over some suitable band, and δA is the deviation of the loss from its average value. The expected value of the average loss $\langle \bar{A} \rangle$, and the power spectrum of δA and its total power or the expected value of its average mean square value $\langle \overline{(\delta A)^2} \rangle$, are easily found in terms of the statistics of the random Fourier coefficients A_k ; the bar again indicates an average over the free-space wavelength λ .

A more exact treatment of the loss statistics, one that avoids the above approximations and includes the frequency dependence of all quantities, is straightforward but lengthy. A brief discussion, given in Appendix F, of the statistics of each of the two terms of (218a) as functions of λ , verifies the approximate analysis.

3.2 Statistical Model of Guide

3.2.1 Tilts and Offsets

We assume that the parallel and perpendicular components of tilt or offset ($\alpha^{\parallel}, \alpha^{\perp}$ or b^{\parallel}, b^{\perp}) are independent Gaussian random variables with zero mean and equal variance; tilts or offsets at different joints are assumed to be independent, and to have the same distribution. Then the magnitude of tilt or offset (α or b) will have a Rayleigh distribution and the orientation (θ) will be uniformly distributed between 0 and 2π , and these two quantities will also be independent.

From (37), (38), (47), and (218c) and (218d) we have

$$x^{\parallel} = \frac{C_t \alpha^{\parallel}}{C_o b^{\parallel}}, \quad x^{\perp} = \frac{C_t \alpha^{\perp}}{C_o b^{\perp}}; \quad x = \frac{C_t \alpha}{C_o b}. \quad (221)$$

Thus identical statements to those above may be made about $x^{\parallel}, x^{\perp}, x$ and θ .

For convenience we state the following results in terms of x 's; any of these equations obviously holds true if x is replaced by α or b throughout. The rms value of x is denoted by \hat{x} ; thus

$$\hat{x}^2 \equiv \langle x^2 \rangle = \frac{C_i^2 \langle \alpha^2 \rangle}{C_o^2 \langle b^2 \rangle}, \quad (222)$$

where the symbols $\langle \ \rangle$ denote the expected value of a random variable. Then the various probability distributions at any joint may be written as follows:

$$p(x^{\parallel}) = \frac{1}{\sqrt{\pi} \hat{x}} \exp - \left(\frac{x^{\parallel}}{\hat{x}} \right)^2; \quad p(x^{\perp}) = \frac{1}{\sqrt{\pi} \hat{x}} \exp - \left(\frac{x^{\perp}}{\hat{x}} \right)^2 \quad (223)$$

$$p(x^{\parallel}, x^{\perp}) = p(x^{\parallel})p(x^{\perp})$$

$$p(x) = \frac{2x}{\hat{x}^2} \exp - \left(\frac{x}{\hat{x}} \right)^2, \quad x > 0; \quad p(\theta) = \frac{1}{2\pi}, \quad (224)$$

$$0 < \theta < 2\pi$$

$$p(x, \theta) = p(x)p(\theta).$$

We have:

$$\langle x^{\parallel} \rangle = \langle x^{\perp} \rangle = 0, \quad \langle x^{\parallel} x^{\perp} \rangle = 0. \quad (225)$$

$$\langle x^{\parallel 2} \rangle = \langle x^{\perp 2} \rangle = \frac{\hat{x}^2}{2}, \quad \langle x^{\parallel 4} \rangle = \langle x^{\perp 4} \rangle = \frac{3\hat{x}^4}{4}.$$

$$\langle x \rangle = \frac{\sqrt{\pi}}{2} \hat{x}, \quad \langle x^2 \rangle = \hat{x}^2, \quad \langle x^4 \rangle = 2\hat{x}^4. \quad (226)$$

Subscripts indicating the joint or mode converter have been omitted in the above relations. Since different joints are independent, for two different joints i and j we have

$$\langle x_i^{\parallel} x_j^{\parallel} \rangle = \langle x_i^{\parallel} x_j^{\perp} \rangle = \langle x_i^{\perp} x_j^{\perp} \rangle = 0, \quad i \neq j. \quad (227)$$

Subscripts denoting the spurious mode have been omitted in the above equations.

3.2.2 Diameter Changes

The radius of the i^{th} pipe r_i is given by

$$r_i = a + \epsilon_i \quad (228)$$

where a is the average radius and ϵ_i a small random variation about the average. The ϵ_i are assumed to be independent Gaussian random variables with zero mean and variance $\hat{\epsilon}^2$.

$$p(\epsilon_i) = \frac{1}{\sqrt{2\pi}\hat{\epsilon}} \exp - \frac{\epsilon_i^2}{2\hat{\epsilon}^2} \quad (229)$$

$$\langle \epsilon_i \rangle = 0; \quad \langle \epsilon_i^2 \rangle = \hat{\epsilon}^2; \quad \langle \epsilon_i \epsilon_j \rangle = 0, \quad i \neq j. \quad (230)$$

Then Δr_i , the change in radius at the i^{th} joint, is given by

$$\Delta r_i = \epsilon_i - \epsilon_{i-1}, \quad (231)$$

and has a Gaussian distribution with zero mean and variance $(\hat{\Delta r})^2$

$$p(\Delta r_i) = \frac{1}{\sqrt{2\pi\hat{\Delta r}}} \exp -\frac{(\Delta r_i)^2}{2(\hat{\Delta r})^2}; \quad \Delta r = \sqrt{2}\hat{\epsilon}. \quad (232)$$

However adjacent joints are no longer independent; from (231) Δr_i and Δr_j are independent only if $|j - i| > 1$. Therefore:

$$\begin{aligned} \langle \Delta r_i \rangle &= 0. \\ \langle (\Delta r_i)^2 \rangle &= (\Delta r)^2 = 2\hat{\epsilon}^2. \\ \langle \Delta r_i \Delta r_{i\pm 1} \rangle &= -\frac{(\hat{\Delta r})^2}{2} = -\hat{\epsilon}^2. \\ \langle \Delta r_i \Delta r_j \rangle &= 0, \quad |j - i| > 1. \end{aligned} \quad (233)$$

Equations (232) and (233) apply also to the corresponding x_i , given by (218c) for a diameter change, for any spurious mode, by making the substitution $\Delta r_i \rightarrow x_i$ and $\hat{\Delta r} \rightarrow \hat{x}$, where \hat{x} is given by

$$\hat{x} = C_d \hat{\Delta r}. \quad (234)$$

Subscripts denoting the spurious mode have again been omitted in the above equations.

3.3 Statistics of the TE_{01} Loss for a Single Section of Waveguide between Mode Filters

3.3.1 Offsets

For a copper waveguide section whose only defects are independent random offsets at the joints between pipes, the most important spurious modes will be the forward TE_{12} and TE_{11} , with the other forward TE_{1m} modes contributing only a small amount to the total TE_{01} loss. Neglecting the frequency dependence of all quantities except the $\Delta\beta$, we have from (218a) and (220) for each spurious mode:

$$\begin{aligned} A &= \bar{A} + \delta A, \\ \bar{A} &= \frac{1}{2}A_0, \\ \delta A &= \sum_{k=1}^{N-1} A_k \cos k\Delta\beta l_0, \end{aligned} \quad (235)$$

where the random Fourier coefficients A_k are given in (218b), together with (218d). The subscript indicating the spurious mode has again been omitted.

The following moments of A_k are easily determined from the results of Section 3.2.1:

$$\langle A_0 \rangle = \hat{x}^2 N, \quad (236a)$$

$$\langle A_0^2 \rangle = \hat{x}^4 N(N + 1).$$

$$\langle A_k \rangle = 0, \quad \langle A_k^2 \rangle = \frac{\hat{x}^4}{2} (N - k) e^{k2\Delta\alpha l_0}, \quad 1 \leq k \leq N - 1. \quad (236b)$$

$$\langle A_k A_l \rangle = 0, \quad k \neq l. \quad (236c)$$

From (235) and (236a) the expected value of the average TE₀₁ normalized loss due to each spurious mode is

$$\langle \bar{A} \rangle = \frac{\hat{x}^2}{2} N. \quad (237)$$

From (236b) the (discrete) power spectrum of δA is

$$P_k = \frac{1}{2} \langle A_k^2 \rangle = \frac{\hat{x}^4}{4} (N - k) e^{k2\Delta\alpha l_0}, \quad 1 \leq k \leq N - 1. \quad (238)$$

From (236c) the different Fourier components of δA are uncorrelated. The total power, or the expected value of the mean square value, of δA is found by summing over the discrete power spectrum P_k . From (238):

$$\begin{aligned} \langle \overline{(\delta A)^2} \rangle &= \sum_{k=1}^{N-1} P_k = \frac{\hat{x}^4}{4} \sum_{k=1}^{N-1} (N - k) e^{k2\Delta\alpha l_0} \\ &= \frac{\hat{x}^4}{4} \frac{e^{2\Delta\alpha l_0}}{1 - e^{2\Delta\alpha l_0}} \left[N - \frac{1 - e^{N2\Delta\alpha l_0}}{1 - e^{2\Delta\alpha l_0}} \right]. \end{aligned} \quad (239)$$

Strictly speaking, the average over free-space wavelength λ indicated by the bar in (237) and (239) must be taken either over a single fundamental period of δA (such that $\Delta\beta l_0$ changes by 2π) or over a band large compared to a single period. Let $\Delta\lambda$ and Δf be the interval in wavelength and frequency, respectively, corresponding to a fundamental period. From (167) and (173) we find

$$\frac{\Delta f}{f} \approx \frac{\Delta\lambda}{\lambda} = \frac{B}{l_0} \quad (240)$$

where B is the beat wavelength, tabulated in Appendix D, and l_0 is the distance between joints in the guide. (This result is similar to that of

(178) for the sample point spacing; in (178), however, the length L is the total length of guide.)

Equation (239) simplifies in two special cases of interest:

1. Small differential loss over total length $L_N = Nl_0$.

$$\langle \overline{(\delta A)^2} \rangle = \frac{\hat{x}^4}{8} N(N-1); \quad -N2\Delta\alpha l_0 \ll 1. \quad (241)$$

2. Large differential loss over total length L_N , small differential loss over pipe length l_0 .

$$\langle \overline{(\delta A)^2} \rangle = \frac{\hat{x}^4}{8} \frac{N}{-\Delta\alpha l_0}; \quad \begin{array}{l} -N2\Delta\alpha l_0 \gg 1 \\ -2\Delta\alpha l_0 \ll 1. \end{array} \quad (242)$$

We recall again that $\Delta\alpha$ is negative throughout the present treatment, in which the TE_{01} signal mode has lower heat loss than any of the spurious modes.

Referring to Section 3.2.1, $\langle A \rangle$ and $\sqrt{\langle \overline{(\delta A)^2} \rangle}$ are both proportional to the square of the rms offset at the joints between pipes. $\langle \bar{A} \rangle$ is proportional to L_N , the length of the waveguide section. $\sqrt{\langle \overline{(\delta A)^2} \rangle}$ is initially proportional to L_N , when L_N is small enough so that the differential loss may be neglected; for large L_N it becomes proportional to $\sqrt{L_N}$.

The power spectrum P_k of (238) has its maximum value for $k = 1$, and decreases monotonically as k increases to $N - 1$. For small differential loss P_k is triangular; for large differential loss it is exponential. The "3-db bandwidth" of the power spectrum P_k , the value of k for which P_k is equal to half its maximum value, is related to the rate of variation of the TE_{01} loss component due to a particular spurious mode. We have:

$$k_{3db} = \frac{N}{2}; \quad -N2\Delta\alpha l_0 \ll 1 \quad (243a)$$

$$k_{3db} = \frac{0.692}{-2\Delta\alpha l_0}; \quad -N2\Delta\alpha l_0 \gg 1. \quad (243b)$$

Making use of (167) and (173), we find

$$\frac{\Delta f_{3db}}{f} \approx \frac{\Delta \lambda_{3db}}{\lambda} = \frac{B}{k_{3db} l_0} \quad (244)$$

for the interval in free-space wavelength or frequency (of the radiation in the guide) corresponding to the 3-db bandwidth of the TE_{01} loss variation.

The statistics of the total TE_{01} loss due to mode conversion are given simply by summing over the contributions of the individual spurious modes. From (217) we have

$$\langle \bar{A} \rangle = \sum_{[m]} \langle \bar{A}_{[m]} \rangle, \quad (245)$$

$$\langle \overline{(\delta A)^2} \rangle = \sum_{[m]} \langle \overline{(\delta A_{[m]})^2} \rangle, \quad \dagger \quad (246)$$

where $[m]$ indexes the TE_{1m}^+ spurious modes. The individual terms in (245) and (246) for each spurious mode are given by (237) and (239) (or (241) and (242) in special cases), in which the subscript $[m]$ has been omitted for convenience. The power spectra of the $\delta A_{[m]}$ are given by (238). From (222) and (237),

$$\langle \bar{A} \rangle = \frac{\langle b^2 \rangle N}{2} \sum_{[m]} C_{o[m]}^2. \quad (247)$$

A practical waveguide system will contain mode filters for the TE_{1m} modes at a close enough spacing so that the differential losses for the important spurious modes are small in each section. For this special case we have from (241)

$$\langle \overline{(\delta A)^2} \rangle = \frac{\langle b^2 \rangle^2 N(N-1)}{8} \sum_{[m]} C_{o[m]}^4; \quad -N2\Delta\alpha_{[m]}l_0 \ll 1. \quad (248)$$

In (247) and (248) $\langle b^2 \rangle$ is the mean square magnitude of offset, $C_{o[m]}$ is Morgan's coupling coefficient between TE_{01} and TE_{1m} for offsets, given in Appendix A, and $\Delta\alpha_{[m]} = \alpha_{[01]} - \alpha_{[1m]}$, the difference in attenuation constants of the TE_{01} and TE_{1m}^+ modes.

Formulas and numerical values for the various coupling coefficients and beat wavelengths are given in Appendix A and D. For a frequency of 55 kmc and a 1-inch guide radius, (247) and (248) become, summing over the nine propagating TE_{1m}^+ modes:

$$\begin{aligned} \langle \bar{A} \rangle &= \frac{\langle b^2 \rangle N}{2} (1.107 + 4.581 + 0.641 + 0.271 + \dots) \\ &= \langle b^2 \rangle N(3.519); \quad \langle b^2 \rangle \text{ in inch}^2. \end{aligned} \quad (249)$$

$$\begin{aligned} \langle \overline{(\delta A)^2} \rangle &= \frac{\langle b^2 \rangle^2 N(N-1)}{8} (1.226 + 20.984 \\ &\quad + 0.411 + 0.074 + \dots) \end{aligned}$$

† No cross terms appear in the summation of (246) because, subject to the approximations of the present section that neglect the frequency variation of the coupling coefficients (C_t , C_o , and C_d) and the $\Delta\alpha$'s, the Fourier components of the different $\delta A_{[m]}$ have incommensurable periods, and hence their total powers or mean square values may be simply added. The cross terms are treated exactly in Appendix F.

$$\begin{aligned}
 &= \langle b^2 \rangle^2 N(N-1)(2.842) \\
 \sqrt{\langle (\delta A)^2 \rangle} &= \langle b^2 \rangle \sqrt{N(N-1)}(1.686); \quad \langle b^2 \rangle \text{ in inch}^2, \\
 &\quad -N2\Delta\alpha_{[m]}l_0 \ll 1.
 \end{aligned} \tag{250}$$

The most important terms are those due to TE_{11}^+ and TE_{12}^+ .

3.3.2 Tilts

We next consider a copper guide whose only defects are random tilts at the joints between pipes. The spurious modes are the forward TE_{1m} with TE_{12}^+ and TE_{11}^+ the most important, as in the offset case above, and in addition the forward TM_{11} . The effects of the TE_{1m}^+ modes on the TE_{01} transmission are given by the results of Section 3.3.1 above [see (235) to (246)], using of course the appropriate coupling coefficients for tilts [see (218c)]. However, TM_{11}^+ requires special consideration.

Equations (235) and (236) apply to TM_{11}^+ as well as to the TE_{1m} modes. For TM_{11}^+ , $\Delta\beta = \Delta\alpha$ in copper guide. Thus the beat wavelength for TM_{11}^+ is very long—3195 feet in 2-inch I.D. pipe at 55 kmc—compared to the beat wavelengths of the TE_{1m} modes, and long even compared to the length of guide sections between mode filters in a practical waveguide system.† Thus, the bandwidths we will consider (e.g., 50–60 kmc) are only a small portion of the fundamental period of $A_{(11)}^+$, the TM_{11}^+ component of the TE_{01} loss, as given in (240), and so the summation of powers of Fourier components given in (239) is no longer valid in determining the mean square loss variation. In fact $A_{(11)}^+$ will be almost independent of frequency (except for the slow variation of coupling coefficient, which is inversely proportional to free-space wavelength λ , neglected over moderate bandwidths in the present analysis).

Thus, consider a section of copper guide short compared with the TM_{11}^+ beat wavelength (3195 feet). Then both $\Delta\alpha$ and $\Delta\beta$ may be set equal to zero. From (219) the TM_{11}^+ component of the TE_{01} loss is

$$A_{(11)}^+ = \frac{1}{2} \left[\sum_{i=1}^N x_i^{\parallel} \right]^2 + \frac{1}{2} \left[\sum_{i=1}^N x_i^{\perp} \right]^2. \tag{251}$$

This is simply equal to the TE_{01} loss due to TM_{11}^+ at a single tilt equal to the net tilt between the input and output ends of the guide. This result is obvious from the fact that we have neglected both attenuation and phase shift in the relatively short sections of guide under consideration. The loss is therefore independent of the lengths of guide between

† This is true only for unmodified copper guide. A thin dielectric coating will reduce the TE_{01} - TM_{11}^+ beat wavelength to much smaller values;^{12,13,14} the present treatment of TM_{11}^+ will obviously not apply in such a case.

discrete mode converters, and these lengths may thus be set equal to zero, yielding only a single tilt whose magnitude and orientation are equal to the net tilt between the ends of the guide. This result holds true for TM_{11}^+ for any arbitrary continuous bend of the guide axis, which may be considered as the limit of a series of discrete tilts (Section 2.3).

To this approximation $A_{(11)}^+$ is thus independent of frequency (except for the slow variation of coupling coefficient with λ). TM_{11}^+ thus contributes to the average loss \bar{A} but not to $\overline{(\delta A)^2}$. Thus (245) contains an extra term, while (246) remains unaltered.

$$\langle \bar{A} \rangle = \langle \bar{A} \rangle_{TE} + \langle A_{(11)}^+ \rangle \quad (252)$$

$$\langle \bar{A} \rangle_{TE} = \sum_{[m]} \langle \bar{A}_{[m]} \rangle \quad (253)$$

$$\langle \overline{(\delta A)^2} \rangle = \sum_{[m]} \langle \overline{(\delta A_{[m]})^2} \rangle \quad (254)$$

The index $[m]$ again indexes the TE_{1m}^+ modes. $\langle \bar{A}_{[m]} \rangle$ and $\langle \overline{(\delta A_{[m]})^2} \rangle$ are again given by (237) and (239) (or (241) and (242) in special cases) and $\langle A_{(11)}^+ \rangle$ is also given by (237), with the appropriate coupling coefficients for tilts. Equations (238), (243), and (244) remain true for the TE_{1m}^+ components of the TE_{01} loss. The results analogous to (247) to (250) for tilts are summarized below.

$$\langle \bar{A} \rangle_{TE} = \frac{\langle \alpha^2 \rangle N}{2} \sum_{[m]} C_{t[m]}^2 \quad (255)$$

$$\langle A_{(11)}^+ \rangle = \frac{\langle \alpha^2 \rangle N}{2} C_{t(11)}^2 \quad (TM_{(11)}^+) \quad (256)$$

For small differential loss (because of mode filters),

$$\langle \overline{(\delta A)^2} \rangle = \frac{\langle \alpha^2 \rangle^2 N(N-1)}{8} \sum_{[m]} C_{t[m]}^4; \quad -N2\Delta\alpha_{[m]}l_0 \ll 1. \quad (257)$$

$\langle \alpha^2 \rangle$ is the mean square magnitude of tilt. Substituting numerical values for the C_t 's from Appendix A, for a frequency of 55 kmc and a 1-inch guide radius, (255)–(257) become, summing over the nine propagating TE_{1m}^+ modes:

$$\begin{aligned} \langle \bar{A} \rangle_{TE} &= \frac{\langle \alpha^2 \rangle N}{2} (29.465 + 81.085 + 0.616 + 0.057 + \dots) \\ &= \langle \alpha^2 \rangle N(55.619); \quad \langle \alpha^2 \rangle \text{ in radian}^2. \end{aligned} \quad (258)$$

$$\langle A_{(11)}^+ \rangle = \langle \alpha^2 \rangle N(14.598); \quad \langle \alpha^2 \rangle \text{ in radian}^2. \quad (259)$$

$$\begin{aligned}
 \langle (\delta A)^2 \rangle &= \frac{\langle \alpha^2 \rangle^2 N(N-1)}{8} (868.18 + 6574.74 + 0.38 + \dots) \\
 &= \langle \alpha^2 \rangle^2 N(N-1)(930.41) \\
 \sqrt{\langle (\delta A)^2 \rangle} &= \langle \alpha^2 \rangle \sqrt{N(N-1)} (30.503); \quad \langle \alpha^2 \rangle \text{ in radian}^2, \\
 &\quad -N2\Delta\alpha_{[m]l_0} \ll 1. \tag{260}
 \end{aligned}$$

Again the most important terms are those due to TE_{11}^+ and TE_{12}^+ , and for the average loss, to TM_{11}^+ in addition.

For a given (copper) guide $A_{(11)}^+$, the TM_{11}^+ contribution to the (average) TE_{01} loss is determined simply by the net tilt between the input and output of the guide, as discussed above. The present model, which assumes that the only imperfections are tilts at the joints between perfect pipes, seems grossly unrealistic as far as the effects of TM_{11}^+ are concerned for any practical guide, for at least two reasons. First, practical pipes will have long bows or gradual curvature of the guide axis; this factor will probably be much more important in determining the net tilt between the guide input and output than tilts of the very small angles of interest here. Second, practical guides may be subject to mechanical constraints of various types which will also introduce slow variations in curvature of the guide axis.

In contrast, gradual curvature of the guide axis will have little effect on the TE_{1m} components of the TE_{01} loss for reasons indicated in Section 2.3.7 and to be discussed in detail in Section IV; this is so because only straightness deviations whose wavelengths are approximately equal to the beat wavelengths of the important spurious modes contribute to the TE_{01} loss in copper guide.

Consequently the effect of TM_{11}^+ on the average loss has been stated separately for the particular model discussed here; it is given in (256) and (259) for whatever tutorial value it may have. As stated above, TM_{11}^+ will have no significant effect on the variation of TE_{01} loss about its average value for the relatively short mode filter spacings which must be used in a practical waveguide system.

3.3.3 Diameter Changes

Finally, consider a copper or helix guide whose only defects are random diameters of the individual pipes, which are perfect right circular cylinders and have no tilts or offsets at their joints. Again from (218a)

and (220) for each spurious (TE_{0m}) mode we have, as in (235) for offsets or tilts:

$$\begin{aligned} A &= \bar{A} + \delta A \\ \bar{A} &= \frac{1}{2}A_0 \\ \delta A &= \sum_{k=1}^{N-1} A_k \cos k\Delta\beta l_0. \end{aligned} \quad (261)$$

The random Fourier coefficients are again given by (218b), together with (218e).

$$A_k = e^{k\Delta\alpha l_0} \sum_{i=1}^{N-k} x_i x_{i+k}. \quad (262)$$

The subscript indicating the spurious mode is again omitted.

The moments of the A_k must be slightly modified from those given in (236) for independent offsets and tilts, because of the correlation between adjacent diameter changes imposed by the present mathematical model and because each spurious mode now has only one polarization rather than two. From (228) to (234):

$$\langle A_0 \rangle = \hat{x}^2 N, \quad \langle A_0^2 \rangle = \hat{x}^4 (N^2 + 3N - 1). \quad (263a)$$

$$\langle A_1 \rangle = -\frac{\hat{x}^2}{2} (N - 1) e^{\Delta\alpha l_0}, \quad (263b)$$

$$\langle A_1^2 \rangle = \frac{\hat{x}^4}{4} (N^2 + 5N - 8) e^{2\Delta\alpha l_0}$$

$$\langle A_k \rangle = 0, \quad \langle A_k^2 \rangle = \frac{\hat{x}^4}{2} [3(N - k) - 1] e^{k2\Delta\alpha l_0}; \quad (263c)$$

$$2 \leq k \leq N - 1.$$

$$\langle A_k A_l \rangle = \hat{x}^4 e^{(k+l)\Delta\alpha l_0} \cdot \left\{ \begin{array}{l} -(N - l) \\ + \frac{1}{4}(N - l) \\ 0 \end{array} \right\}; \left[\begin{array}{l} l - k = 1 \\ l - k = 2 \\ l - k \geq 3 \end{array} \right], \quad 0 < k < l, \quad (263d)$$

$$\langle A_0 A_l \rangle = \hat{x}^4 e^{l\Delta\alpha l_0} \cdot \left\{ \begin{array}{l} -\frac{1}{2}(N - 1)(N + 4) \\ + \frac{1}{2}(N - 2) \\ 0 \end{array} \right\}; \left[\begin{array}{l} l = 1 \\ l = 2 \\ l \geq 3 \end{array} \right]. \quad (263e)$$

From (236a) and (261) the expected value of the average loss due to each spurious mode is again

$$\langle \bar{A} \rangle = \frac{\hat{x}^2}{2} N, \quad (264)$$

as in (237) for offsets or tilts. However, the statistics of the A_k 's, as given in (263), differ from those of (236) for tilts and offsets. Because $\langle A_1 \rangle$ is no longer equal to zero, the expected value of the TE_{01} loss will have a fundamental periodic component, in addition to a dc component. Consequently it is convenient to rewrite the first relation of (261) as follows:

$$A = \bar{A} + \langle A_1 \rangle \cos \Delta\beta l_0 + \delta A'. \quad (265)$$

Thus,

$$\delta A' = (A_1 - \langle A_1 \rangle) \cos \Delta\beta l_0 + \sum_{k=2}^{N-1} A_k \cos k\Delta\beta l_0. \quad (266)$$

$\langle A_1 \rangle$ is given in (263b). \bar{A} is the average loss as before, $\langle A_1 \rangle \cos \Delta\beta l_0$ is a slowly varying sinusoidal component of loss whose period equals the fundamental period of the TE_{01} loss (see (240)), and $\delta A'$ includes the remaining random loss variations. The (discrete) power spectrum P_k' of $\delta A'$ is given by

$$P_k' = \begin{cases} \frac{1}{2}(\langle A_1^2 \rangle - \langle A_1 \rangle^2); & k = 1. \\ \frac{1}{2}\langle A_k^2 \rangle & ; 2 \leq k \leq N - 1. \end{cases} \quad (267)$$

From (263d) the different Fourier components of $\delta A'$ are now uncorrelated only if their indices differ by three or more. This does not affect the calculation of the total power of $\delta A'$, which remains simply the sum of P_k' .

$$\langle (\overline{\delta A'})^2 \rangle = \sum_{k=1}^{N-1} P_k' = \frac{\hat{x}^4}{4} \left\{ 3 \frac{e^{2\Delta\alpha l_0}}{1 - e^{2\Delta\alpha l_0}} \left[N - \frac{1 - e^{N2\Delta\alpha l_0}}{1 - e^{2\Delta\alpha l_0}} \right] - \frac{e^{2\Delta\alpha l_0}}{1 - e^{2\Delta\alpha l_0}} \left[1 - e^{(N-1)2\Delta\alpha l_0} \right] + \frac{(N-1)}{2} e^{2\Delta\alpha l_0} \right\}. \quad (268)$$

Again for small or large differential loss, (268) simplifies:

1. Small differential loss over total length $L_N = Nl_0$.

$$\langle (\overline{\delta A'})^2 \rangle = \frac{\hat{x}^4}{8} (3N - 1)(N - 1); \quad -N2\Delta\alpha l_0 \ll 1. \quad (269)$$

2. Large differential loss over total length L_N , small differential loss over pipe length l_0 .

$$\langle \overline{(\delta A')^2} \rangle = \frac{\epsilon^4}{8_j} \frac{3N}{-\Delta\alpha_0}; \quad \begin{array}{l} -N2\Delta\alpha_0 \gg 1 \\ -2\Delta\alpha_0 \ll 1 \end{array} \quad (270)$$

$\langle \bar{A} \rangle$ and $\langle \overline{(\delta A')^2} \rangle$ have the same general functional form as for tilts and offsets. The "3-db bandwidth" of P_k' is approximately the same as given in (243)–(244) for tilts and offsets.

The statistics of the total TE_{01} loss due to mode conversion are now given by summing over the TE_{0m}^+ spurious modes.

$$\langle \bar{A} \rangle = \sum_{[m]} \langle \bar{A}_{[m]} \rangle \quad (271)$$

$$\langle \overline{(\delta A')^2} \rangle = \sum_{[m]} \langle \overline{(\delta A_{[m]})^2} \rangle. \quad (272)$$

The individual terms in (271) and (272) are given by (264) and (268) (or (269) and (270) in special cases). In addition, each spurious mode will contribute a single sinusoidal component to the TE_{01} loss, of magnitude given by the middle term of (265). From Section 3.2.2 and (264),

$$\langle \bar{A} \rangle = \langle \epsilon^2 \rangle N \sum_{[m]} C_{d[m]}^2 \quad (273)$$

where $\langle \epsilon^2 \rangle$ is the mean square variation of pipe radius.

In discussing the effects of the TE_{1m}^+ modes in the first two parts of this section, it was assumed that the line contained ideal mode filters at a close enough spacing so that the differential loss in each section could be neglected. However, a practical mode filter in 2-inch guide presently consists of a section of helix guide, which has a low loss for the TE_{0m} spurious modes (although it effectively suppresses all other spurious modes). In the present case it is therefore assumed that there will be no mode filters in the entire length of line between repeaters. Thus, the total differential loss will be large, and the approximate result of (270) yields

$$\langle \overline{(\delta A')^2} \rangle = \frac{3 \langle \epsilon^2 \rangle^2 N}{2l_0} \sum_{[m]} \frac{C_{d[m]}^4}{-\Delta\alpha_{[m]}}; \quad \begin{array}{l} -N2\Delta\alpha_{[m]}l_0 \gg 1, \\ -2\Delta\alpha l_0 \ll 1. \end{array} \quad (274)$$

Substituting numerical values for the coupling coefficients and the

$\Delta\alpha$'s, for a frequency of 55 kmc and a 1-inch guide radius, (273) and (274) become:

$$\begin{aligned}\langle \bar{A} \rangle &= \langle \epsilon^2 \rangle N (2.424 + 0.771 + 0.394 + 0.244 + \dots) \\ &= \langle \epsilon^2 \rangle N (4.414); \quad \langle \epsilon^2 \rangle \text{ in inch}^2. \quad (275) \\ \langle \overline{(\delta A')^2} \rangle &= \frac{3 \langle \epsilon^2 \rangle^2 N}{2l_0} \times 10^4 (7.197 + 0.274 + 0.037 + 0.008 + \dots) \\ &= \frac{\langle \epsilon^2 \rangle^2 N}{l_0} (11.28 \times 10^4) \\ \sqrt{\langle \overline{(\delta A')^2} \rangle} &= \langle \epsilon^2 \rangle \sqrt{\frac{N}{l_0}} (335.9); \quad \langle \epsilon^2 \rangle \text{ in inch}^2 \\ &\quad l_0 \text{ in feet.} \\ -N2\Delta\alpha_{[m]}l_0 &\gg 1, \quad -2\Delta\alpha_{[m]}l_0 \ll 1. \quad (276)\end{aligned}$$

In these results the summation has been extended over all of the propagating TE_{0m}^+ spurious modes, $TE_{02}^+ - TE_{09}^+$. The most important modes are the first few TE_{0m}^+ ; for $\langle \overline{(\delta A')^2} \rangle$ only TE_{02}^+ and TE_{03}^+ are significant.

Finally, the sinusoidal component of the TE_{01} loss contributed by the TE_{0m}^+ spurious mode is, from (265) and (263b),

$$\begin{aligned}\langle A_{1[m]} \rangle \cos \Delta\beta_{[m]}l_0 &= -\langle \epsilon^2 \rangle (N - 1) C_{d[m]}^2 \cos \Delta\beta_{[m]}l_0; \\ &\quad -\Delta\alpha_{[m]}l_0 \ll 1. \quad (277)\end{aligned}$$

3.4 TE_{01} Loss Statistics of a Long Guide with Ideal Mode Filters

Consider a long guide made up of M sections of imperfect guide of equal length and the same statistical parameters, separated by ideal mode filters. We must evaluate the over-all transmission statistics of the guide in terms of the transmission statistics of each section, given in Section 3.3. The transmission parameters of such a guide with ideal mode filters are given in (146). Since for the present we are concerned with only the over-all TE_{01} loss A , we have

$$A = \sum_{k=1}^M {}_kA. \quad (278)$$

${}_kA$ is the total TE_{01} loss, due to all spurious modes, of the k^{th} section of guide. From (220) we write

$${}_kA = {}_k\bar{A} + \delta {}_kA. \quad (279)$$

Further, for the loss of the entire guide (with mode filters)

$$A = \bar{A} + \delta A \quad (280)$$

where

$$\bar{A} = \sum_{k=1}^M {}_k\bar{A}, \quad \delta A = \sum_{k=1}^M \delta_k A. \quad (281)$$

We assume that each section of guide between mode filters is statistically independent of all other sections. Then from (281)

$$\langle \bar{A} \rangle = \sum_{k=1}^M \langle {}_k\bar{A} \rangle = M \langle {}_k\bar{A} \rangle, \quad (282)$$

$$\langle \overline{(\delta A)^2} \rangle = \sum_{k=1}^M \langle \overline{(\delta_k A)^2} \rangle = M \langle \overline{(\delta_k A)^2} \rangle, \quad (283)$$

where M is the number of sections of guide separated by ideal mode filters.

Finally, from (278) and the central limit theorem,³⁷ if M is large the over-all loss A , regarded as a function of free-space wavelength λ , will be a Gaussian random process; this random process in general will not be stationary, although over the relatively narrow bands of interest it may often be assumed stationary.

3.5 Numerical Examples

In the present section several numerical examples are presented to provide concrete illustrations of the above results. A 20-mile total guide length, made of 2-inch I.D. pipes 10 feet long, with equally spaced ideal mode filters, is considered in all cases; the operating frequency is taken to be 55 kmc. It is assumed that the mode filters have infinite loss for the TE_{1m}^+ and TM_{11}^+ spurious modes, zero loss for the TE_{01} signal mode and the TE_{0m}^+ spurious modes. The results for tilts and offsets apply to copper guide; the results for diameter variations apply equally well to either copper or helix.

3.5.1 Offsets

Assume an rms offset such that the additional average loss due to mode conversion to the forward TE_{1m} modes is 1 db/mile (compared to the theoretical TE_{01} heat loss at 55 kmc of 1.54 db/mile). Two cases are considered (see Table II): (1) mode filter spacing such that the rms loss variation for the 20-mile line is 1 db, and (2) a 200-foot mode filter spacing. The formulas for zero differential loss are used for simplicity, since the differential loss for TE_{12}^+ , the most important spurious mode, will remain small.

TABLE II

	Case 1	Case 2
Average loss	1 db/mile	1 db/mile
RMS total loss fluctuation for 20-mile line	1 db	0.407 db
RMS offset	7.87 mils	7.87 mils
Mode filter spacing	1160 feet	200 feet
$\Delta f_{3 \text{ db}}$ for TE_{12}^+	209 mc	1211 mc

3.5.2 *Tilts*

Assume an rms tilt such that the additional average loss due to mode conversion to the forward TE_{1m} modes is again 1 db/mile. The additional average loss due to conversion to TM_{11}^+ is stated separately because the present model is unrealistic as far as TM_{11}^+ is concerned, for reasons stated in Section 3.3.2. Two cases are again considered (see Table III): (1) mode filter spacing such that the rms loss variation for the 20-mile line is 1 db, and (2) a 200-foot mode filter spacing. The formulas for zero differential loss are again used for simplicity.

TABLE III

	Case 1	Case 2
Average loss TE_{1m}^+ modes	1 db/mile	1 db/mile
Average loss TM_{11}^+	0.262 db/mile	0.262 db/mile
RMS total loss fluctuation for 20-mile line	1 db	0.465 db
RMS tilt	0.114°	0.114°
Equivalent crack on one side of joint	3.96 mils	3.96 mils
Mode filter spacing	890 feet	200 feet
$\Delta f_{3 \text{ db}}$ for TE_{12}^+	272 mc	1211 mc

The TE_{01} average loss due to TM_{11}^+ conversion will depend only on the net angle between the input and output of a waveguide section between mode filters for a mode filter spacing short compared to 3195 feet, the TM_{11}^+ beat wavelength in copper guide. As discussed in Section 3.3.2, this angle (Table IV) will depend principally on long bows in the pipes and on the way in which the guide is laid.

TABLE IV

Mode filter spacing.....	200 feet		1000 feet	
	1 db/ mile	0.1 db/ mile	1 db/ mile	0.1 db/ mile
Average loss, TM_{11}^+				
RMS net angle between input and out- put of waveguide sections.....	1.00°	0.316°	2.23°	0.706°

3.5.3 Diameter Changes

For diameter variations there are no mode filters for the spurious TE_{0m}^+ modes. Therefore, the mechanical tolerance required to yield an rms loss fluctuation $\langle(\delta\bar{A}')^2\rangle$ of 1 db for the 20-mile line is determined (see Table V); the additional average loss will now be very small.

TABLE V

Average loss	0.214 db/mile
$\langle(\delta\bar{A}')^2\rangle$: RMS total loss fluctuation for 20-mile line	1 db
RMS diameter variation	6.50 mils
$\Delta f_{3\text{ db}}$ for TE_{02}^+	11.3 mc

Sinusoidal Components of TE_{01} Loss for 20-Mile Line

Mode	Peak-to-Peak Amplitude	Period
TE_{02}^+	4.690 db	4791 mc
TE_{03}^+	1.492 db	1832 mc
TE_{04}^+	0.762 db	974 mc
TE_{05}^+	0.472 db	600 mc

3.6 Helix Guide

While the above results for diameter changes apply to both copper and helix guide, the results for tilts and offsets apply only to copper guide. Equivalent results for helix would require the coupling coefficients for the normal modes of the helix at tilts and offsets. However, a very simple argument shows that \bar{A} , the average loss, will be identical in a helix and a copper guide which have identical tilts or offsets; $\bar{A}_{(11)}$, the TM_{11}^+ component of the TE_{01} loss, must now be included in the copper pipe average loss in the case of tilts, as shown below. In addition, the spurious modes have such a high loss in helix that the TE_{01} loss fluctuations will be very small.

From (235), (218) and Section 2.2 we see that \bar{A} is simply the sum of the TE_{01} signal losses at each individual discrete mode converter (tilt or offset), where by signal loss we mean $-\ln s_{00}$, where s_{00} is the TE_{01} transfer coefficient of the discrete tilt or offset. From Section 2.1.4 s_{00} is identical in copper and helix guide with equal tilts or offsets.

Therefore the above results for the expected value of the average TE_{01} loss for tilts and offsets hold equally well for helix waveguide. The TE_{01} loss fluctuations in helix will be very small for these cases.

3.7 Conclusions

Experimental copper waveguides have been built whose tolerances are far better than those of the numerical examples in Section 3.5. Since the average loss and the rms loss variation are proportional to the square of the rms tolerance, it is clear that tilts and offsets at joints and uniform diameter variations of the individual pipes will not contribute significantly to the observed TE_{01} loss in these waveguides. Consequently the additional TE_{01} loss observed in present experimental waveguides must be due principally to continuous mode conversion, and in particular to continuous random deviations from straightness of the individual copper pipes themselves.⁸ The continuous case will be treated in Section IV.

The added TE_{01} average loss due to TM_{11}^+ conversion in copper waveguide (unmodified by a dielectric lining or anything else) is a function only of the net angle between the input and output of each waveguide section between mode filters, for a reasonable mode filter spacing. The tolerance on this angle must be held to a few tenths of a degree to keep this loss component down to 0.1 db/mile.

The present analysis has been restricted to equally spaced mode converters, i.e., individual pipes of the same length. If the pipe lengths are allowed to become random, instead of starting from (71d) we must start with (71a) and (71b). The TE_{01} loss due to a single spurious mode will still have a discrete power spectrum, but the discrete components will no longer be equally spaced, and consequently the TE_{01} loss will no longer be periodic. The frequencies as well as the amplitudes of the discrete components must now be treated as random variables. Aside from these minor differences, the analysis should be similar and lead to similar results.

We refer again to the treatment of Appendix F, where more exact expressions for the TE_{01} loss statistics in the discrete case are derived without neglecting the frequency (or λ) dependence of the coupling coefficients and the differential attenuation constants, as in the above treatment. It is found that these approximations are valid for our present purposes.

Finally we note that by means of the Kronecker product, it is possible to compute certain of the TE_{01} transmission statistics exactly—i.e., without using perturbation theory—for the case of statistically independent discrete mode converters.³⁸ This treatment requires that the individual conversion coefficients be known exactly; unfortunately only in the idealized two-mode case is the exact form of the coupling coeffi-

icients known. Such calculations may be used to check certain of the above approximate results in the two-mode case.

IV. THEORY OF GUIDES WITH RANDOM CONTINUOUS IMPERFECTIONS

This section applies the results of Section II to the study of multimode waveguides with random continuous mode conversion. Continuous mode conversion arises from gradual continuous changes in the geometric properties of the guide, such as curvature of the guide axis, variation of the guide diameter, or changes in the cross section of the guide such as ellipticity, etc., as opposed to the discrete case, studied in Section III. The statistics of the TE_{01} loss-frequency curve are determined in terms of the statistics of the different guide imperfections. In particular, the average TE_{01} loss and the rms value and the power spectrum of the TE_{01} loss variations are calculated.

The most important practical application of these results to date has been to study the effects of random straightness deviations of the guide axis. Here we consider only small unintentional straightness deviations, either arising in the manufacturing process of the individual pipes themselves or resulting from the way in which the guide is laid. We exclude from consideration the case of large intentional bends (to go around corners), which couple TE_{01} to the degenerate forward TM_{11} mode. The spurious modes of interest here are thus the TE_{1m} family, with the forward TE_{11} and TE_{12} the most important. The present analysis indicates that very small random straightness deviations in a certain spectral region (i.e., having mechanical wavelengths lying in a certain range), having an rms value of a fraction of a mil, are primarily responsible for the observed departure of the TE_{01} transmission from its theoretical value in present copper guide,^{8,39} causing an increased average loss and random fluctuations about this average. In addition, the analysis indicates that random straightness deviations will be equally important in helix or dielectric coated waveguide in increasing the average TE_{01} loss; however, the high spurious mode loss in helix will effectively remove the TE_{01} loss fluctuations.^{8,39}

While random straightness deviations are the most important manufacturing tolerance at present, the same methods are easily applied to study the effects of other tolerances of the guide. The present section will also consider random diameter changes, which produce the TE_{0m} modes, random ellipticity, which gives rise to the TE_{2m} modes, and higher-order deformations of the cross section, which produce TE_{nm} modes of higher angular index.

In order to specify the statistics of the guide, we assume that each

type of imperfection (e.g., deviation of the guide axis from straightness, diameter variation, ellipticity, etc.), regarded as a function of distance z along the guide axis, is a stationary Gaussian random process of known spectrum. The various continuous coupling coefficients are of course proportional to the geometric imperfections, and thus are also Gaussian random processes.

The analysis of the continuous case is greatly simplified if the differential loss between the TE_{01} signal mode and the various spurious modes may be neglected over the lengths of interest. This approximation will be made throughout the present section. As discussed above, a practical system using copper guide will contain regularly spaced mode filters that have a high loss for all spurious modes except the TE_{0m} family. The mode filter spacing will be sufficiently small so that the differential losses may be neglected for all important spurious modes except the TE_{0m} family.

For the TE_{0m} spurious modes the effective line length will be the total distance between repeaters; obviously the total differential loss is no longer negligible. However, the results for zero differential loss will be stated for this case to get at least a rough upper bound on the importance of diameter variations, for both copper and helix guide.

4.1 TE_{01} Loss—Summary of Previous Results

We give in the present section the normalized TE_{01} loss (in nepers) A , for the case in which the total differential loss (for each section of guide between mode filters) may be neglected, so that we may set $\Delta\alpha = 0$. From (206) the total normalized TE_{01} loss A is given by a sum of terms due to the individual first-order spurious modes;

$$A \approx \sum_m A_m. \quad (284)$$

The A_m are given by (285) to (288) below. These results are obtained from Section 2.3.7. We again omit subscripts denoting the spurious mode where no confusion will arise.

$$A \approx \frac{1}{2} |I|^2, \quad (285a)$$

$$I = L \sum_{n=-\infty}^{\infty} c_n (-1)^n \frac{\sin \pi \left(\frac{\Delta\beta L}{2\pi} - n \right)}{\pi \left(\frac{\Delta\beta L}{2\pi} - n \right)}. \quad (285b)$$

The c_n are the Fourier coefficients of the continuous coupling coefficient

$c(z)$, defined by

$$c(z) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n z/L}; \quad c_{-n} = c_n^* . \quad (286)$$

At the n^{th} sample point, defined by

$$\frac{\Delta\beta L}{2\pi} = n \quad \text{or} \quad B \equiv \frac{2\pi}{\Delta\beta} = \frac{L}{n}, \quad (287)$$

we have

$$I(n) = Lc_n(-1)^n, \quad (288a)$$

$$A(n) \approx \frac{L^2}{2} |c_n|^2. \quad (288b)$$

The coupling coefficient $c(z)$ is given in terms of the various geometric parameters in Section 2.3.

As in the discrete case (Section 3.1), the principal frequency dependence in these results occurs through the $\Delta\beta$'s, which far from cutoff are approximately proportional to the free-space wavelength λ . Over the moderate fractional bandwidths of interest, any frequency dependence of the coupling coefficients will be slow and may be neglected. From the discussion of Section 2.3.7, any frequency dependence of the coupling coefficient may be taken into account in (285) by calculating c_n at the frequency corresponding to the n^{th} sample point, rather than at the operating frequency.

We regard the loss A as a function of the free-space wavelength λ , and write A as follows;

$$A = \langle A \rangle + \delta A. \quad (289)$$

We determine the expected value of the loss $\langle A \rangle$, and the power spectrum of δA and its total power or mean square value $\langle (\delta A)^2 \rangle$, in terms of the power spectrum of the random coupling coefficient $c(z)$ (and consequently of the random geometric imperfection of the guide).

4.2 Statistics of Fourier Coefficients of $c(z)$ ^{37,40}

We assume that the geometric imperfection of the guide (e.g., deviation of the guide axis from straightness) is a stationary Gaussian random process with a known power spectrum. The continuous coupling coefficient $c(z)$ to the particular polarization of one of the spurious modes will be a similar random process, since the coupling coefficients are

simply proportional to the corresponding geometric imperfection. Therefore $c(z)$ is a stationary Gaussian random process with a power spectrum $S(\zeta)$. Thus, if $R(\tau)$ is the covariance of $c(z)$,

$$R(\tau) = \langle c(z)c(z + \tau) \rangle, \quad (290)$$

then

$$S(\zeta) = \int_{-\infty}^{\infty} R(\tau) e^{-j2\pi\zeta\tau} d\tau. \quad (291)$$

Consider the Fourier series expansion of $c(z)$ over the interval $0 < z < L$, given in (165) and (286).

$$c(z) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n z / L}; \quad c_{-n} = c_n^*. \quad (292)$$

$$c_n = a_n + j b_n = |c_n| e^{j\varphi_n}$$

The c_n 's will be complex Gaussian random variables; i.e., a_n and b_n , the real and imaginary parts of c_n , will be Gaussian random variables with zero mean. If L is sufficiently long, the a_n 's and b_n 's become almost independent, and hence the c_n 's become almost independent complex Gaussian random variables. Thus, the magnitude and phase of each c_n are independent and have a Rayleigh and a uniform distribution respectively. The mean square value of the n^{th} Fourier coefficient $\langle |c_n|^2 \rangle$ is then simply related to the power spectrum $S(\zeta)$ of $c(z)$. We have *approximately* for large L :

$$\langle |c_n|^2 \rangle \equiv \hat{c}_n^2 = \frac{1}{L} \cdot S\left(\frac{n}{L}\right). \quad (293)$$

$$\langle c_n c_m^* \rangle = 0, \quad n \neq m. \quad (294)$$

The quantity \hat{c}_n defined in (293) is the rms magnitude of the n^{th} Fourier coefficient. The various probability distributions for the real and imaginary parts or for the magnitude and phase of the Fourier coefficients may be written approximately as follows:

$$p(a_n) = \frac{1}{\sqrt{\pi} \hat{c}_n} \exp -\left(\frac{a_n}{\hat{c}_n}\right)^2; \quad (295a)$$

$$p(b_n) = \frac{1}{\sqrt{\pi} \hat{c}_n} \exp -\left(\frac{b_n}{\hat{c}_n}\right)^2.$$

$$p(a_n, b_n) = p(a_n)p(b_n). \quad (295b)$$

$$p(|c_n|) = \frac{2|c_n|}{\hat{c}_n^2} \exp -\left(\frac{|c_n|}{\hat{c}_n}\right)^2. \quad (296a)$$

$$p(\varphi_n) = \frac{1}{2\pi}, \quad 0 < \varphi < 2\pi.$$

$$p(|c_n|, \varphi_n) = p(|c_n|)p(\varphi_n). \quad (296b)$$

Since the different Fourier coefficients are approximately independent,

$$p(c_n, c_m) = p(c_n)p(c_m), \quad |n| \neq |m|. \quad (297)$$

Finally, the first few moments of the $|c_n|$ are of interest.

$$\langle |c_n| \rangle = \frac{\sqrt{\pi}}{2} \hat{c}_n; \quad \langle |c_n|^2 \rangle = \hat{c}_n^2; \quad \langle |c_n|^4 \rangle = 2\hat{c}_n^4. \quad (298)$$

The results of this section provide a good approximation for the practical cases of interest in which L , the line length, is of the order of a few hundred feet and the power spectrum $S(\zeta)$ varies slowly in the range of interest, which includes mechanical wavelengths from a few feet to a few inches, depending on the spurious mode. These results become exact if the coupling coefficient $c(z)$ has a white power spectrum ($S(\zeta) = \text{constant}$, or equivalently $R(\tau) \propto \delta(\tau)$, the unit impulse).

4.3 TE_{01} Loss Statistics for a Single Section of Waveguide between Mode Filters

4.3.1 Single Spurious Mode, Single Polarization

From the relations of Section 4.2, $I\left(\frac{\Delta\beta L}{2\pi}\right)$, given in (285b), will be a complex band-limited Gaussian random process; the real and imaginary parts of I are independent Gaussian random processes with flat power spectra over the range $|\nu| < \frac{1}{2}$.^{†33,34} Since by (285a) A is proportional to the square of the magnitude of I and is thus proportional to the sum of the squares of the real and imaginary parts of I , the power spectrum of A may be determined from the well-known analysis for the response of a square law device to Gaussian noise.^{37,40} The square of a Gaussian random process has, in addition to a dc component, a random component whose power spectrum is twice the convolution of the input power spectrum with itself. Since the real and imaginary parts of I have flat band-limited power spectra over the range $|\nu| < \frac{1}{2}$, the random com-

[†] ν again indicates the independent variable of the Fourier transform of I , or some other quantity of interest, with respect to the normalized independent variable $\frac{\Delta\beta L}{2\pi}$. See the footnote on p. 1084.

ponent of $A\left(\frac{\Delta\beta L}{2\pi}\right)$ will therefore have a triangular power spectrum over the range $|\nu| < 1$.

We first separate the TE_{01} loss A as before [see (289)].

$$A = \langle A \rangle + \delta A, \quad (299)$$

where as usual we omit subscripts denoting the spurious mode. Then from the results of Section 4.1 and 4.2, the expected value of the TE_{01} loss is given by

$$\langle A(\lambda_n) \rangle = \frac{L^2}{2} \langle |c_n|^2 \rangle = \frac{L}{2} \cdot S\left(\frac{n}{L}\right), \quad (300)$$

where from (174)

$$\lambda_n = \frac{2\pi n}{D} \frac{n}{L} \quad (301)$$

is the free-space wavelength corresponding to the n^{th} sample point and D is the constant relating the differential propagation constant $\Delta\beta$ to the free-space wavelength λ [see (173)]. Substituting (287) and (301) into (300), and using the result to interpolate between the sample points,

$$\langle A(\lambda) \rangle = \frac{L}{2} \cdot S\left(\frac{D}{2\pi} \lambda\right) = \frac{L}{2} \cdot S\left(\frac{1}{B}\right), \quad (302)$$

relating the expected value of the TE_{01} loss due to a single spurious mode (single polarization) to the power spectrum of the coupling coefficient between TE_{01} and the spurious mode. B is the beat wavelength between TE_{01} and the spurious mode.

The (continuous) power spectrum of δA in the region close to λ is given by

$$\begin{aligned} P(\nu) &= \frac{L^2}{4} \cdot S^2\left(\frac{D}{2\pi} \lambda\right) (1 - |\nu|) \\ &= \frac{L^2}{4} \cdot S^2\left(\frac{1}{B}\right) (1 - |\nu|), \quad |\nu| < 1. \end{aligned} \quad (303)$$

In deriving (303) we have tacitly assumed that $S(\xi)$, the power spectrum of the coupling coefficient $c(z)$, varies only slowly in the region of interest, so that I and A are approximately stationary over moderate bandwidths; however, this is not a serious restriction. Equation (303) may be obtained either from the known results on the square of a Gaussian noise^{37,40} or directly from Sections 4.1 and 4.2. The total power, or the expected value of the mean square value, of δA may be found by

integrating the power spectrum $P(\nu)$. From (303)

$$\langle(\delta A)^2\rangle = \frac{L^2}{4} \cdot S^2\left(\frac{D}{2\pi} \lambda\right) = \frac{L^2}{4} \cdot S^2\left(\frac{1}{B}\right), \quad (304a)$$

$$\sqrt{\langle(\delta A)^2\rangle} = \frac{L}{2} \cdot S\left(\frac{D}{2\pi} \lambda\right) = \frac{L}{2} \cdot S\left(\frac{1}{B}\right) = \langle A \rangle, \quad (304b)$$

for a single polarization of a single spurious mode. Alternately, the results of (304) may be obtained directly (at the sample points) from (288b), (293) and (298); it is apparent that (302) and (304) hold for quite general power spectra $S(f)$.

The power spectrum $P(\nu)$ of δA is triangular; from (303) the 3-db bandwidth is

$$\nu_{3\text{db}} = \frac{1}{2}. \quad (305)$$

The interval in free-space wavelength or frequency corresponding to the 3-db bandwidth of δA is thus

$$\frac{\Delta f_{3\text{db}}}{f} \approx \frac{\Delta \lambda_{3\text{db}}}{\lambda} = \frac{2B}{L}, \quad (306)$$

and is thus simply twice the sample point spacing (for I), given in (178). [Compare (306) with (244) and (243a).]

Finally, we consider the probability distribution for A , considering for the present only a single polarization of a single spurious mode. As discussed at the beginning of the present section, A is the sum of the squares of two independent Gaussian random variables. Alternately, A may be regarded as the square of a Rayleigh-distributed random variable. Consequently, for a single polarization of a single spurious mode, A has an exponential probability density.

$$p(A) = \frac{1}{\langle A \rangle} \exp -\frac{A}{\langle A \rangle}, \quad (307)$$

where the average loss $\langle A \rangle$ is given in (302). We recall that this result (and all others of the present section) is based on the assumption of zero differential loss, $\Delta\alpha = 0$. Equation (307) holds equally well for the corresponding discrete case of Section III.

4.3.2 *Single Spurious Mode, Two Polarizations*

The above results are easily extended to two polarizations of the spurious mode. We assume that the two orthogonal components of the geometric imperfection giving rise to mode conversion are independent

Gaussian random processes with the same statistics. For example, consider random straightness deviations, which couple TE_{01} to the TE_{1m} family but principally to the forward TE_{12} and TE_{11} . The position of the guide axis is specified by $x(z)$ and $y(z)$, the coordinates of the guide axis in the transverse plane as functions of distance along the axis z . We will assume that $x(z)$ and $y(z)$ are independent Gaussian random processes with identical power spectra.

The coupling coefficients $c_{[m]}^{\parallel}(z)$ and $c_{[m]}^{\perp}(z)$ between TE_{01} and the two polarizations of the m^{th} spurious mode will also be independent Gaussian random processes, since the coupling coefficients are proportional to the corresponding geometric imperfections. Thus, for straightness deviations we have from Section 2.4 for small deviations

$$c_{[m]}^{\parallel}(z) = C_{t[m]} \cdot x''(z), \quad c_{[m]}^{\perp}(z) = C_{t[m]} \cdot y''(z), \quad (308)$$

where the symbols \parallel and \perp distinguish the two polarizations of the spurious mode, rather than x and y as in Section II. Since $x(z)$ and $y(z)$ are independent Gaussian random processes, $x''(z)$ and $y''(z)$ and consequently $c_{[m]}^{\parallel}(z)$ and $c_{[m]}^{\perp}(z)$ will also be independent Gaussian random processes.

The TE_{01} loss $A_{[m]}$ for both polarizations of the m^{th} spurious mode is given by

$$A_{[m]} = A_{[m]}^{\parallel} + A_{[m]}^{\perp}, \quad (309)$$

where $A_{[m]}^{\parallel}$ and $A_{[m]}^{\perp}$ are independent random processes with statistics given by the results of Section 4.3.1. Writing

$$A_{[m]} = \langle A_{[m]} \rangle + \delta A_{[m]}, \quad (310)$$

we have for the expected value of the TE_{01} loss

$$\langle A_{[m]} \rangle = \langle A_{[m]}^{\parallel} \rangle + \langle A_{[m]}^{\perp} \rangle. \quad (311)$$

Since $\delta A_{[m]}^{\parallel}$ and $\delta A_{[m]}^{\perp}$, the two ac loss components, are independent, their power spectra and total powers add. Denoting the (continuous) power spectra of $\delta A_{[m]}$, $\delta A_{[m]}^{\parallel}$ and $\delta A_{[m]}^{\perp}$ by $P_{[m]}(\nu)$, $P_{[m]}^{\parallel}(\nu)$, and $P_{[m]}^{\perp}(\nu)$ respectively, we have

$$P_{[m]}(\nu) = P_{[m]}^{\parallel}(\nu) + P_{[m]}^{\perp}(\nu), \quad (312)$$

$$\langle (\delta A_{[m]})^2 \rangle = \langle (\delta A_{[m]}^{\parallel})^2 \rangle + \langle (\delta A_{[m]}^{\perp})^2 \rangle. \quad (313)$$

Since both polarizations are assumed to have identical statistics, we have from Section 4.3.1:

$$\langle A_{[m]}(\lambda) \rangle = L \cdot S_{[m]} \left(\frac{D_{[m]}}{2\pi} \lambda \right) = L \cdot S_{[m]} \left(\frac{1}{B_{[m]}} \right). \quad (314)$$

$$P_{[m]}(\nu) = \frac{L^2}{2} \cdot S_{[m]}^2 \left(\frac{D_{[m]}}{2\pi} \lambda \right) \cdot (1 - |\nu|) \quad (315)$$

$$= \frac{L^2}{2} \cdot S_{[m]}^2 \left(\frac{1}{B_{[m]}} \right) \cdot (1 - |\nu|), \quad |\nu| < 1.$$

$$\langle (\delta A_{[m]})^2 \rangle = \frac{L^2}{2} \cdot S_{[m]}^2 \left(\frac{D_{[m]}}{2\pi} \lambda \right) = \frac{L^2}{2} \cdot S_{[m]}^2 \left(\frac{1}{B_{[m]}} \right). \quad (316a)$$

$$\sqrt{\langle (\delta A_{[m]})^2 \rangle} = \frac{L}{\sqrt{2}} \cdot S_{[m]} \left(\frac{D_{[m]}}{2\pi} \lambda \right) = \frac{L}{\sqrt{2}} \cdot S_{[m]} \left(\frac{1}{B_{[m]}} \right) = \frac{\langle A_{[m]} \rangle}{\sqrt{2}}. \quad (316b)$$

In these and all subsequent results $S_{[m]}(\zeta)$ is the power spectrum of each of the orthogonal components of the coupling coefficient.

The power spectrum $P_{[m]}(\nu)$ of course remains triangular, and the 3-db bandwidth and the corresponding interval in free-space wavelength or frequency remains as given in (305) and (306) for a single polarization of the spurious mode. These latter quantities are the same as those for the corresponding discrete case, given in (243a) and (244).

Finally, since $A_{[m]}^{\parallel}$ and $A_{[m]}^{\perp}$ are independent random variables with the same probability distribution [see (307)], the probability distribution for $A_{[m]}$ is simply the convolution of (307) with itself.

$$p(A_{[m]}) = \frac{4A_{[m]}}{\langle A_{[m]} \rangle^2} \exp -\frac{2A_{[m]}}{\langle A_{[m]} \rangle}. \quad (317)$$

This result holds true for discrete tilts and offsets for zero differential loss.

4.3.3 Many Spurious Modes

For many spurious modes, the total TE_{01} loss A is given as a sum over loss components $A_{[m]}$ due to the different spurious modes. From (284)

$$A = \sum_{[m]} A_{[m]}. \quad (318)$$

Writing the total TE_{01} loss as before,

$$A = \langle A \rangle + \delta A. \quad (319)$$

Then we have

$$\langle A \rangle = \sum_{[m]} \langle A_{[m]} \rangle, \quad (320a)$$

$$\delta A = \sum_{[m]} \delta A_{[m]}. \quad (320b)$$

The average TE_{01} loss is simply the sum of the contributions of each of the spurious modes.

From (320b) we have for the variance of the total TE_{01} loss

$$\langle (\delta A)^2 \rangle = \sum_{[m]} \sum_{[n]} \langle \delta A_{[m]} \delta A_{[n]} \rangle. \quad (321)$$

The terms $\langle (\delta A_{[m]})^2 \rangle$ are given by (316a) or (304a). It would be most convenient if the different ac components $\delta A_{[m]}$ were independent, so that the cross terms $\langle \delta A_{[m]} \delta A_{[n]} \rangle$ could be neglected.

For the spurious modes produced by a geometric imperfection of a given angular symmetry (e.g., the TE_{1m} , produced by straightness deviations) the different $\delta A_{[m]}$ are not independent. The TE_{01} loss component due to the m^{th} spurious mode at one frequency is proportional to the TE_{01} loss component due to the n^{th} spurious mode at a widely separated frequency, as in the discrete case (Appendix F). Thus, knowledge of one of the $\delta A_{[m]}$ over a sufficiently wide frequency band is sufficient to determine all of the others. However, at the same frequency $\delta A_{[m]}$ and $\delta A_{[n]}$ are almost uncorrelated, so that the cross terms in (321) may be neglected.

$\delta A_{[m]}$ and $\delta A_{[n]}$ are easily seen to be almost independent in a simple way. Consider a frequency which corresponds to a sample point of the m^{th} spurious mode. Under special conditions this frequency may also correspond to a different sample point of the n^{th} spurious mode (in general, this will not be so). From (288b), $A_{[m]}$ will depend on only a single Fourier coefficient (say the k^{th}) of the geometric imperfection. Similarly, $A_{[n]}$ will depend on only a single Fourier coefficient, but on a different one (say the l^{th}), since different spurious modes have different beat wavelengths. Since the different Fourier coefficients of the geometric imperfection are almost independent, $\delta A_{[m]}$ and $\delta A_{[n]}$ will thus also be almost independent at this frequency. Since in general the sample points corresponding to different spurious modes do not precisely coincide, the correlation coefficient between $\delta A_{[m]}$ and $\delta A_{[n]}$ at a single frequency will not be identically zero, but should be small.

The correlation coefficient between the ac components of the TE_{01} loss due to two different spurious modes generated by the same type of geometric imperfection is derived in Appendix G for the special case in which the geometric imperfection and hence the coupling coefficients have white power spectra. Numerical results are given for the important practical case of TE_{12}^+ and TE_{11}^+ generated by random straightness deviations, in which the second derivatives of the rectangular co-ordinates of the guide axis are independent Gaussian random processes with white spectra. The normalized correlation coefficient for reasonable guide lengths is very small indeed.

The cross terms in (321) will consequently make only a negligible contribution to the variance of the total TE_{01} loss. Therefore, $\langle (\delta A)^2 \rangle$ is

given simply by the sum of the contributions of the individual spurious modes;

$$\langle(\delta A)^2\rangle = \sum_{[m]} \langle(\delta A_{[m]})^2\rangle, \quad (322)$$

where the $\langle(\delta A_{[m]})^2\rangle$ are given by (316a) (or by (304a) for a spurious mode with a single polarization, e.g., TE_{0m}).

4.3.4 Discussion

The TE_{01} loss in a given frequency band, resulting from a given spurious mode, depends only on the Fourier components of the corresponding geometric imperfection for a narrow range of mechanical wavelengths. This band of mechanical wavelengths corresponds to the range of beat wavelengths between TE_{01} and the spurious mode over the frequency band of interest. The statistics of the TE_{01} loss are strongly dependent on the power spectrum of the geometric imperfection.

The present results are strictly valid only for zero differential loss, $\Delta\alpha = 0$, although they will remain approximately true so long as the differential loss over the length of guide remains small, $|\Delta\alpha|L \ll 1$. However, further study shows that moderate values of differential loss $\Delta\alpha$ will change the average TE_{01} loss very little, but will smooth out the fluctuations of the TE_{01} loss, for the present case in which the coupling coefficient power spectrum is essentially flat in the range of interest.⁴¹

The following sections will present specific numerical examples for the various types of geometric imperfections.

4.4 TE_{01} Loss Statistics for Random Straightness Deviations

4.4.1 Introduction

In the present section we apply the results of Section 4.3 to the case of random deviations from straightness of the guide axis. In Section 4.5 other types of continuous geometric imperfections are similarly treated. There are two reasons for treating straightness deviations separately:

1. Straightness deviations introduced by the manufacturing process are almost entirely responsible for the additional TE_{01} loss of present 2-inch I.D. copper guide.

2. Experimental TE_{01} transmission measurements have provided a fair idea of the shape of the power spectrum of straightness deviations for different types of guide, at least over a limited range. This power spectrum differs in some respects from the power spectra that might be assumed for other types of geometric imperfections, in that under certain

conditions it contains an infinite low-frequency (i.e., long mechanical wavelength) content; some additional discussion of this particular case seems appropriate.

We must first specify the statistical properties of the coupling coefficient; since the coupling coefficient $c(z)$ is assumed to be a stationary Gaussian random process, its statistics are completely determined by its power spectrum $S(\zeta)$. The continuous case is inherently more complicated to discuss than the discrete case of Section III. For the discrete case only the rms offset, tilt, or diameter change must be specified. In the continuous case, however, the TE_{01} loss statistics are no longer determined only by the mean square value of straightness deviation or other geometric imperfection; the shape of the power spectrum of the mechanical imperfection strongly influences the resulting TE_{01} loss. We must therefore know the power spectrum of the imperfection before we can predict the TE_{01} loss statistics of a guide. Conversely, knowledge of the TE_{01} loss statistics enables us to estimate the power spectrum of the imperfections. Up to now there have been no existing mechanical methods for measuring the straightness deviation to the required accuracy so that its covariance and power spectrum can be determined;† TE_{01} transmission measurements have provided the only means of determining the significant Fourier components of the straightness deviation.

Present experimental measurements of the TE_{01} loss over a band extending from 33 kmc to 90 kmc, made by A. P. King and G. D. Mandeville, indicate that for one type of 2-inch I.D. copper guide, the radius of curvature of the straightness deviation has an approximately flat power spectrum over the range of interest.³⁹ Thus, if $x(z)$ and $y(z)$ are the rectangular components of straightness deviation with power spectra $X(\zeta)$ and $Y(\zeta)$, we have‡

$$X(\zeta) = Y(\zeta) \propto \frac{1}{\zeta^4} \quad (323)$$

for mechanical wavelengths lying in the beat wavelength range for TE_{11} and TE_{12} (the most important spurious modes), 1.4 to 4.4 feet for the 35–90 kmc band.

The power spectrum of (323) for $x(z)$ and $y(z)$, which corresponds to a white power spectrum for the radius of curvature of the guide axis or

† Methods of making these mechanical measurements are currently under development by K. J. Dahms, W. G. Nutt, and R. B. Ramsey and their associates at Bell Telephone Laboratories.

‡ ζ is the "mechanical frequency," having the dimension $\frac{1}{\text{length}}$; the corresponding mechanical wavelength is $\frac{1}{\zeta}$. [See for example (291).]

equivalently for the second derivatives $x''(z)$ and $y''(z)$, appears plausible under certain conditions, considering one way in which the guide has been made. If we imagine a guide made by drawing a copper pipe with more or less random hardness or wall thickness variations through a die, it is not hard to see that the radius of curvature of the guide axis might be a random process with a very short correlation distance, or equivalently with a very wide power spectrum. The power spectrum for $x''(z)$ and $y''(z)$ must, of course, fall off for sufficiently high mechanical frequencies (or sufficiently short mechanical wavelengths).

The coupling coefficients and hence $x''(z)$ and $y''(z)$, the second derivatives of the displacement of the guide axis, have been assumed to be stationary random processes. The displacements themselves, $x(z)$ and $y(z)$, will *not* in general be stationary random processes, unless the power spectrum of $x''(z)$ and $y''(z)$ (and the corresponding coupling coefficients) has special properties. However, this situation seems to be in accord with the physical facts. As a simple example we may consider a guide made of pipes with random uniform bows, screwed together at random; we might further assume that the first pipe of the guide starts out with zero displacement and zero slope, $x(0) = y(0) = x'(0) = y'(0) = 0$. Then it is obvious that while the second derivatives $x''(z)$ and $y''(z)$ and hence the coupling coefficients are stationary random processes, the displacements $x(z)$ and $y(z)$ are not. The variances of the displacements, $\langle x^2(z) \rangle$ and $\langle y^2(z) \rangle$, grow with distance z ; the guide tends to wander more and more from the axis as z increases, unless additional mechanical constraints are imposed in laying the guide.

Since our knowledge of the power spectrum $X(\zeta)$ is limited, any example chosen to illustrate the order of magnitude of straightness tolerance that will have a significant effect on the TE_{01} transmission must be arbitrary to a considerable extent. For the present numerical example we assume that $X(\zeta)$ and $Y(\zeta)$ are as given by (323). For this power spectrum the displacement of the guide axis $x(z)$ or $y(z)$ is not a stationary random process, and the integral of (323) or the "total power" is infinite. However, the principal spurious modes are TE_{12}^+ and TE_{11}^+ , with beat wavelengths of about 2.2 and 2.7 feet at a frequency of 55 kmc. In order to get a rough measure of the short-wavelength straightness deviations that are responsible for the additional TE_{01} loss, we shall quite arbitrarily calculate a "total" mean square straightness deviation $\langle x^2(z) \rangle + \langle y^2(z) \rangle$ by including only those components having mechanical wavelengths less than 5 feet, $\zeta > \frac{1}{5}$. While this is a rather arbitrary choice, it makes some physical sense. The significant components for TE_{11}^+ and TE_{12}^+ lie between 1.4 and 4.4 feet for a band from 35–90 kmc. The

long-wavelength components (greater than 5 feet) do not affect the TE_{01} transmission in this band; in any case, these long-wavelength components will depend strongly on the random errors made in laying the guide, and perhaps very little on the straightness deviations introduced by the manufacturing process. The mean square value of the components of wavelength less than 5 feet will give us a rough idea of the order of magnitude of the tolerance on the short-wavelength "manufacturing" straightness deviations.

For the numerical example presented below we assume a 20-mile total guide length of 2-inch I.D. copper guide with equally spaced mode filters, spaced either 200 or 1000 feet apart. These mode filters are assumed to have zero loss to the TE_{01} signal mode, infinite loss to the spurious TE_{1m}^+ modes. TM_{11}^+ is neglected, since it has been adequately treated in Section III. The differential loss is neglected even though it is not small in the distance between mode filters for all spurious modes, particularly for the 1000-foot mode filter spacing. The differential loss will not greatly affect the average TE_{01} loss, but will reduce the TE_{01} loss fluctuations below the values computed for zero differential loss. The treatment of a long line with ideal mode filters is given in Section 3.4.

The x and y components of the straightness deviation of the guide axis are assumed to have power spectra given in (323), discussed above. The magnitude of these power spectra is chosen to yield an additional average TE_{01} loss (due to all the propagating TE_{1m}^+ spurious modes) of 1 db per mile, at a frequency of 55 kmc. The total rms straightness deviation for components having wavelengths less than 5 feet is stated, for reasons discussed above. In addition, the rms straightness deviation for components having wavelengths between 2 and 3 feet, corresponding to the TE_{12}^+ and TE_{11}^+ beat wavelengths for the 50–60 kmc band, is also given. The contributions of each of the spurious modes to the average TE_{01} loss and to the TE_{01} loss fluctuations are stated separately.

4.4.2 Analysis

Let $X(\zeta)$ and $Y(\zeta)$ be the power spectra of the rectangular components of the straightness deviation of the guide axis, $x(z)$ and $y(z)$. Then

$$\begin{aligned} X(\zeta) &= \int_{-\infty}^{\infty} \langle x(z)x(z + \tau) \rangle e^{-j2\pi\zeta\tau} d\tau \\ Y(\zeta) &= \int_{-\infty}^{\infty} \langle y(z)y(z + \tau) \rangle e^{-j2\pi\zeta\tau} d\tau. \end{aligned} \tag{324}$$

We will always assume identical spectra for $x(z)$ and $y(z)$, i.e.,

$$X(\zeta) = Y(\zeta). \quad (325)$$

Then the power spectra of $x''(z)$ and $y''(z)$, the second derivatives of the x and y components of the displacement of the guide axis, are given by

$$\begin{aligned} (2\pi\zeta)^4 X(\zeta) &= \int_{-\infty}^{\infty} \langle x''(z)x''(z + \tau) \rangle e^{-j2\pi\zeta\tau} d\tau \\ (2\pi\zeta)^4 Y(\zeta) &= \int_{-\infty}^{\infty} \langle y''(z)y''(z + \tau) \rangle e^{-j2\pi\zeta\tau} d\tau. \end{aligned} \quad (326)$$

Noting (308), the power spectra $S_{[m]}(\zeta)$ of the coupling coefficients $c_{[m]}(z)$ [see (290) and (291)] are given by

$$\begin{aligned} S_{[m]}^{\parallel}(\zeta) &= C_{t[m]}^2 \cdot (2\pi\zeta)^4 X(\zeta), \\ S_{[m]}^{\perp}(\zeta) &= C_{t[m]}^2 \cdot (2\pi\zeta)^4 Y(\zeta), \end{aligned} \quad (327)$$

for the parallel and perpendicular polarizations of the TE_{1m} spurious mode respectively.

Far from cutoff $C_{t[m]}$ is approximately inversely proportional to the free-space wavelength λ . It is thus sometimes convenient to write $C_{t[m]}$ as

$$C_{t[m]} = \frac{\mathcal{C}_{t[m]}}{\lambda}, \quad (328)$$

where $\mathcal{C}_{t[m]}$ is now approximately independent of λ .

From (314) the average TE_{01} loss due to the TE_{1m} spurious mode (with two polarizations) is

$$\langle A_{[m]}(\lambda) \rangle = L \mathcal{C}_{t[m]}^2 D_{[m]}^4 \lambda^2 \cdot X\left(\frac{D_{[m]}}{2\pi} \lambda\right). \quad (329)$$

The constant $D_{[m]}$, defined in (173), is related to the beat wavelength $B_{[m]}$ by

$$\frac{D_{[m]}}{2\pi} \lambda = \frac{1}{B_{[m]}} = \frac{\beta_0 - \beta_{[m]}}{2\pi}. \quad (330)$$

In (329), $X(\zeta)$ is the power spectrum of *each* of the rectangular components of straightness deviation [see (325)]; the only frequency- (or wavelength-) dependent terms are the factors λ^2 and the mechanical power spectrum $X\left(\frac{D_{[m]}}{2\pi} \lambda\right)$. The rms fluctuation of the TE_{01} loss component due to TE_{1m} about its average is given by (316b);

$$\sqrt{\langle (\delta A_{[m]})^2 \rangle} = \frac{\langle A_{[m]} \rangle}{\sqrt{2}}. \quad (331)$$

As discussed above, for the present example we assume that over the range of interest (mechanical wavelengths less than 5 feet) $X(\zeta)$, the power spectrum for each component of the straightness deviation, has the shape given by (323). Therefore we take

$$X(\zeta) = Y(\zeta) = \frac{X_0}{(2\pi\zeta)^4}, \quad (332)$$

where X_0 is a scaling parameter determining the magnitude of the straightness deviation. Then the average loss of (329) becomes

$$\langle A_{[m]}(\lambda) \rangle = \frac{L C_{t[m]}^2 X_0}{\lambda^2} = L C_{t[m]}^2 X_0. \quad (333)$$

The loss fluctuation of course remains as given by (331).

The total TE_{01} average loss $\langle A \rangle$ and mean square loss fluctuations $\langle (\delta A)^2 \rangle$ are given by (320a) and (322) respectively, summing over the contributions of all the propagating TE_{1m}^+ spurious modes. We have

$$\langle A \rangle = L X_0 \sum_{[m]} C_{t[m]}^2. \quad (334)$$

$$\langle (\delta A)^2 \rangle = \frac{1}{2} L^2 X_0^2 \sum_{[m]} C_{t[m]}^4. \quad (335)$$

Substituting numerical values from Appendix A and summing over the 9 propagating TE_{1m}^+ modes, for a frequency of 55 kmc and a 1-inch guide radius (334) and (335) become

$$\langle A \rangle = 111.24 L X_0 \quad (336)$$

$$\langle (\delta A)^2 \rangle = 3721.64 L^2 X_0^2. \quad (337)$$

From (333) and (334) the average TE_{01} loss $\langle A \rangle$ is inversely proportional to λ^2 , or directly proportional to f^2 . This is approximately in agreement with the experimental results of A. P. King and G. D. Mandeville for one type of copper guide³⁹ and provides the reason for the particular choice of power spectrum for the straightness deviation that has been made here.

Finally, the rms straightness deviation in a given range is obtained by integrating the straightness deviation power spectrum over the appropriate range. For the x -component alone, we have for the mean square straightness deviation for mechanical frequencies lying between ζ_a and ζ_b , or equivalently mechanical wavelengths $\frac{1}{\zeta}$ lying between $\frac{1}{\zeta_a}$ and $\frac{1}{\zeta_b}$

$$\langle x^2 \rangle = \left[\int_{-\zeta_b}^{-\zeta_a} + \int_{\zeta_a}^{\zeta_b} \right] X(\zeta) d\zeta, \quad (338)$$

where we recall that the straightness deviation power spectrum $X(\xi)$ has been defined for both positive and negative mechanical frequencies. Substituting the particular power spectrum given in (332),

$$\langle x^2 \rangle = \frac{2X_0}{3(2\pi)^4} \left(\frac{1}{\xi_a^3} - \frac{1}{\xi_b^3} \right). \quad (339)$$

From (325) and the fact that x and y are independent, we have for the total mean square straightness deviation in this range

$$\langle x^2 \rangle + \langle y^2 \rangle = \frac{X_0}{12\pi^4} \left(\frac{1}{\xi_a^3} - \frac{1}{\xi_b^3} \right). \quad (340)$$

4.4.3 Numerical Example

We assume an added average TE_{01} loss in 2-inch I.D. guide due to mode conversion of 1 db per mile at 55 kmc. From Section 3.4 and (336), we determine X_0 .

$$X_0 = \frac{1}{111.24 \times 8.6859 \times 5280} = 1.960 \times 10^{-7} \text{ ft}^{-1}. \quad (341)$$

We will specify the rms straightness deviation (including both the x and y components of the displacement of the guide axis) for components having wavelengths less than five feet, as discussed above. From (340),

$$\sqrt{\langle x^2 \rangle + \langle y^2 \rangle} = 1.737 \text{ mils}; \quad \frac{1}{\xi} < 5 \text{ feet}. \quad (342)$$

For purposes of illustration we shall emphasize the 50–60 kmc band. In this band the beat wavelength range for TE_{12}^+ and TE_{11}^+ , the most important spurious modes, is 2–3 feet. Since these components of straightness deviation are the principal contributors to the TE_{01} loss in the 50–60 kmc band, it is of interest to give the rms straightness deviation lying in the 2–3 foot region. Again from (340),

$$\sqrt{\langle x^2 \rangle + \langle y^2 \rangle} = 0.677 \text{ mils}; \quad 2 \text{ feet} < \frac{1}{\xi} < 3 \text{ feet}. \quad (343)$$

Table VI presents the transmission behavior of a 20-mile, 2-inch I.D. guide for two mode filter spacings, 200 feet and about 1000 feet, calculated from the present results and those of Section 3.4. In addition to the total average loss and rms loss fluctuation, the contributions of the individual spurious modes are given. The average loss is of course the same for both cases; the rms total loss fluctuation is reduced for the shorter mode filter spacing.

TABLE VI — TE₀₁ LOSS STATISTICS FOR STRAIGHTNESS DEVIATION WITH A FLAT CURVATURE POWER SPECTRUM

$L_{total} = 20$ miles, total guide length. $8.6859 \frac{\langle \Delta A_{total} \rangle}{L_{total}} = 1$ db/mile, additional average loss at 55 kmc. $f = 55$ kmc, midband frequency. $a = 1$ inch, guide radius.							
Case 1	$L = 200$ feet, mode filter spacing. $M = 528$, number of mode filters. $8.6859 \sqrt{\langle (\delta A_{total})^2 \rangle} = 0.4773$ db, rms total loss fluctuation for 20-mile line.						
Case 2	$L = 996.23$ feet, mode filter spacing. $M = 106$, number of mode filters. $8.6859 \sqrt{\langle (\delta A_{total})^2 \rangle} = 1.0652$ db, rms total loss fluctuation for 20-mile line.						
				Case 1		Case 2	
Spurious Mode	Beat Wave-length Range for 50-60 Kmc Band	RMS Straightness Deviation in Beat Wave-length Range	Added Average Loss	RMS Total Loss Fluctuation for 20 Mile Line	$\Delta f_{3,db}$: 3-db Bandwidth of Power Spectrum of Loss Fluctuation	RMS Total Loss Fluctuation for 20 Mile Line	$\Delta f_{3,db}$: 3-db Bandwidth of Power Spectrum of Loss Fluctuation
	feet	mils	db/mile	db	mc	db	mc
TE ₁₁ ⁺	2.453-2.949	0.513	0.2649	0.1630	1486	0.3638	298
TE ₁₂ ⁺	1.997-2.408	0.381	0.7290	0.4486	1212	1.0012	243
TE ₁₃ ⁺	0.464-0.562	0.043	0.0055	0.0034	282	0.0076	57
TE ₁₄ ⁺	0.215-0.263	0.014	0.0005	0.0003	131	0.0007	26
TE ₁₅ ⁺	0.123-0.152	0.006	0.0001	0.0001	76	0.0001	15

4.4.4 Discussion

The above results show that for the assumed mechanical power spectrum, the principal contributors to both the additional average loss and to the loss fluctuations arise from the TE₁₂⁺ and the TE₁₁⁺ spurious modes, as has been observed experimentally. As discussed above, the average loss measured over a very wide frequency band by A. P. King and G. D. Mandeville³⁹ provides the primary experimental data on the power spectrum of the straightness deviations.

For some guides both the experimental transmission data and consideration of the manufacturing process indicate that the power spectrum of the second derivative of the straightness deviation should be more or less flat up to some high mechanical frequency (short wavelength). As discussed above, this leads to a power spectrum for the straightness deviation itself with infinite power, so that the displacement of the guide

axis from a perfect straight line will not be a stationary random process. However, in practice additional forces are imposed in laying the guide, so that the straightness deviation must become stationary. In order for this to be so, the power spectrum for $x''(z)$ and $y''(z)$ and for the corresponding coupling coefficients must now fall to zero as ζ approaches 0, at least as fast as ζ^4 . However, it seems reasonable to assume that the modification in the coupling coefficient power spectrum will take place only for very small values of ζ (long wavelengths). In the important spectral region corresponding to the TE_{11}^+ and TE_{12}^+ beat wavelengths, for $\frac{1}{\zeta}$ less than a few feet, the power spectrum should be modified very little; therefore, very little change will take place in the TE_{01} transmission statistics.

Since only components of the straightness deviation having wavelengths between about 1.4 and 4.4 feet will significantly affect the TE_{01} loss (in a band from about 35–90 kmc), it is clear that random straightness deviations arising in the laying of the guide, or manufacturing imperfections such as long bows, will have very little effect on the TE_{01} transmission, because such straightness deviations will have their principal components at much longer wavelengths (e.g., perhaps greater than ten feet). One model that is readily analyzed is the "random bow line," made of pipes of identical length with uniform bows, screwed together at random. The x and y components of the coupling coefficient are simply random square waves, whose power spectrum is well known. The allowable tolerance is several orders of magnitude more lenient than the tolerance on short wavelength straightness deviations.

Finally, it is possible that quite different types of power spectra than those discussed here could arise for different manufacturing processes. For example, a process that resulted in a periodic straightness deviation in the beat wavelength range would result in a rather broadly peaked band-pass power spectrum for the coupling coefficient. Such things are of course to be avoided. In practice, different manufacturing processes have produced quite different straightness deviation power spectra.

4.5 *TE₀₁ Loss Statistics for Random Diameter Variations, Ellipticity, and Higher-Order Deformations*

4.5.1 *Introduction*

In the present section we apply the results of Section 4.3 to random diameter variations, ellipticity, and higher-order deformations of the cross-section of the guide, using Morgan's coupling coefficients $\Xi_{[nm]}$.

Random straightness deviations may of course also be treated in this way, but are omitted since they have been discussed in Section 4.4.

We take the same model for the guide as in Section 4.4, i.e., a 20-mile total guide length of 2-inch I.D. copper with equally spaced mode filters, spaced either 200 feet or about 1000 feet apart. The mode filters are assumed to have zero loss for the TE_{01} signal mode, zero loss for the TE_{0m} spurious modes, and infinite loss for all other TE_{nm} spurious modes. The differential loss is assumed small in the distance between mode filters for all spurious modes; in addition, for the TE_{0m} spurious modes the present analysis forces us to assume that the differential loss is negligible for the total guide length, 20 miles in the present example.

The present analysis should provide a reasonable approximation for ellipticity and for higher-order deformations. The differential loss in these cases will not affect significantly the average TE_{01} loss, but will reduce the TE_{01} loss fluctuations somewhat below the values computed for zero differential loss in those cases where the differential loss is not completely negligible in the distance between mode filters.

However, for diameter variations the TE_{0m} differential loss is certainly not negligible in 20 miles, as required by the analysis. While this approximation will not lead to an appreciable error for the average TE_{01} loss, it will certainly lead to serious error for the mean square TE_{01} loss fluctuation, which is the really significant quantity, and for Δf_{3ab} . The actual TE_{01} loss fluctuations will be much smaller than those computed here.

The power spectrum for straightness deviations is known, at least approximately, over a moderately wide range from TE_{01} transmission measurements, as discussed in Section 4.4. A very elementary consideration of the manufacturing process suggests the same shape for this power spectrum as is actually observed in certain cases. Unfortunately, for other types of deformation neither of these approaches has suggested the proper form for the power spectrum. Spurious modes, other than TE_{12}^+ and TE_{11}^+ , have not been observed at high enough levels to permit an estimate of the power spectrum of the corresponding mechanical imperfection to be made. As yet, no simple picture of the manufacturing processes has yielded a guess as to the shape of the power spectra. About all that can be said is that the power spectra must eventually fall off in some manner for high enough mechanical frequencies ζ (short enough mechanical wavelengths), since the mean square values of the various cross-sectional tolerances are certainly bounded.

Since the power spectra of the various deformations are not known, the numerical examples presented below are less specific than the example

for straightness deviation in Section 4.4. For example, ellipticity generates the TE_{2m}^+ spurious modes. For *each* of these spurious modes the rms ellipticity in the mechanical wavelength range corresponding to the beat wavelength for the 50–60 kmc band is chosen to yield an additional average TE_{01} loss of 1 db per mile, assuming a flat power spectrum in this range. The same is done for trifoil and higher-order deformations.

Diameter variations are treated separately, partly because these results cannot be taken too seriously, as discussed above, and partly because the equations are somewhat different for this case. Here for *each* of the spurious TE_{0m}^+ modes we choose the rms diameter variation in the mechanical wavelength range corresponding to the beat wavelength for the 50–60 kmc band to yield a 1-db rms loss fluctuation for the TE_{01} loss component due to the spurious mode, for the entire 20-mile line.

4.5.2 Random Diameter Variations

From (179) to (181) the radius of the guide is given by

$$r = a + a_0(z). \quad (344)$$

From (207) the coupling coefficient is given by

$$c_{[0m]}(z) = -\Xi_{[0m]}a_0(z). \quad (345)$$

Let $K_0(\zeta)$ be the power spectrum of $a_0(z)$, i.e.,

$$K_0(\zeta) = \int_{-\infty}^{\infty} \langle a_0(z)a_0(z + \tau) \rangle e^{-j2\pi\zeta\tau} d\tau. \quad (346)$$

Then the power spectrum $S_{[0m]}(\zeta)$ of the coupling coefficient $c_{[0m]}(z)$ is

$$S_{[0m]}(\zeta) = \Xi_{[0m]}^2 \cdot K_0(\zeta). \quad (347)$$

Remembering that the spurious mode now has only a single polarization, we have from (302) and (304b)

$$\langle A_{[0m]} \rangle = \sqrt{\langle (\delta A_{[0m]})^2 \rangle} = \frac{L}{2} \Xi_{[0m]}^2 \cdot K_0\left(\frac{D_{[0m]}}{2\pi} \lambda\right), \quad (348)$$

where the constant $D_{[0m]}$ is related to the beat wavelength $B_{[0m]}$ by

$$\frac{D_{[0m]}}{2\pi} \lambda = \frac{1}{B_{[0m]}} = \frac{\beta_{01} - \beta_{0m}}{2\pi}. \quad (349)$$

The coupling coefficient $\Xi_{[0m]}$ is approximately proportional to the free-space wavelength λ .

Finally, the mean square radius variation in the range of mechanical frequencies ζ_a to ζ_b (or the range of mechanical wavelengths $\frac{1}{\zeta}$ from $\frac{1}{\zeta_a}$

to $\frac{1}{\zeta_b}$) is given by

$$\langle a_0^2(z) \rangle = \left[\int_{-\zeta_b}^{-\zeta_a} + \int_{\zeta_a}^{\zeta_b} \right] K_0(\zeta) d\zeta, \tag{350}$$

again integrating over positive and negative frequencies. Assuming that the power spectrum is flat over the range of integration,

$$K_0(\zeta) = K_0, \quad \zeta_a < |\zeta| < \zeta_b, \tag{351}$$

(350) becomes

$$\langle a_0^2(z) \rangle = 2K_0 \cdot (\zeta_b - \zeta_a). \tag{352}$$

These formulas are used to calculate the results in Table VII.

The present analysis could alternately have been carried out in terms of the coupling coefficients $C_{d[m]}$, as discussed in Section 2.3.9.

TABLE VII — RADIUS VARIATION YIELDING 1 DB RMS TE_{01} LOSS FLUCTUATION AND 1 DB AVERAGE TE_{01} LOSS FOR THE LOSS COMPONENT $A_{[0m]}$ DUE TO EACH OF THE TE_{0m}^+ SPURIOUS MODES, AT 55 KMC

$L_{total} = 20$ miles, total guide length.
 $8.6859\sqrt{\langle(\delta A_{[0m]})^2\rangle} = 1$ db, rms loss fluctuation for each component, for 20-mile line.
 $8.6859\langle A_{[0m]} \rangle = 1$ db, additional average loss for each component, for 20-mile line.
 $f = 55$ kmc, midband frequency.
 $a = 1$ inch, guide radius.

Differential loss assumed small over total guide length of 20 miles.

Spurious Mode	Beat Wavelength Range for 50-60 Kmc Band	RMS Radius Variation in Beat Wavelength Range	K_0 : Spectral Density of Radius Variation	Δf_{3db} : 3-db Bandwidth of Power Spectrum of Loss Fluctuation
	feet	mils	mils ² /foot ⁻¹	mc
TE_{02}^+	0.7886-0.9532	0.087	1.7291×10^{-2}	0.907
TE_{03}^+	0.3003-0.3654	0.097	0.7942×10^{-2}	0.347
TE_{04}^+	0.1588-0.1953	0.102	0.4397×10^{-2}	0.185
TE_{05}^+	0.0964-0.1206	0.105	0.2671×10^{-2}	0.113
TE_{06}^+	0.0627-0.0805	0.109	0.1691×10^{-2}	0.075
TE_{07}^+	0.0418-0.0563	0.115	0.1064×10^{-2}	0.051
TE_{08}^+	0.0259-0.0402	0.129	0.0609×10^{-2}	0.035

4.5.3 Random *n*-foils

We now consider random cross-sectional deformations of higher order. For an “*n*-foil” the radius of the guide is given by (179) to (181) as

$$r = a + a_n(z) \cos n\phi + b_n(z) \sin n\phi. \tag{353}$$

The $n = 0$ case corresponds to diameter variations, studied in Section 4.5.2. The $n = 1$ case corresponds to straightness deviations, studied in Section 4.4; the quantities $a_1(z)$ and $b_1(z)$ are equal to $x(z)$ and $y(z)$, the rectangular components of straightness deviation. The $n = 2$ case corresponds to ellipticity, the $n = 3$ case has been designated as "trifoil," etc. All formulas in the present section hold *only* for $n \geq 1$; the $n = 0$ case has been treated in the preceding section.

The magnitude of the n -foil distortion at a given position z along the axis is specified by the maximum departure from a perfect guide, $r = a$. From (353) we have for an n -foil

$$|r - a|_{\max} = \sqrt{a_n^2(z) + b_n^2(z)}. \quad (354)$$

For diameter variations this definition yields the change in guide radius; for straightness deviations it yields the total displacement (in the x - y plane) of the guide axis. For ellipticity and higher-order deformations, (354) yields the maximum deviation from a perfect circle $r = a$. Note that for $n = 2$ this is only one-quarter as large as a commonly accepted definition of ellipticity, the maximum diameter minus the minimum diameter.

The coupling coefficients for the two polarizations of each of the spurious modes are given from (207) by

$$c_{[nm]}^{\parallel}(z) = -\Xi_{[nm]} a_n(z), \quad (355)$$

$$c_{[nm]}^{\perp}(z) = -\Xi_{[nm]} b_n(z). \quad (356)$$

Let $K_n(\xi)$ be the power spectrum of each of the two components $a_n(z)$ and $b_n(z)$, i.e.,

$$K_n(\xi) = \int_{-\infty}^{\infty} \left[\begin{array}{l} \langle a_n(z) a_n(z+\tau) \rangle \\ \langle b_n(z) b_n(z+\tau) \rangle \end{array} \right] e^{-j2\pi\xi\tau} d\tau. \quad (357)$$

Then $S_{[nm]}(\xi)$, the power spectrum of each of the two components of the coupling coefficient, is given by

$$S_{[nm]}(\xi) = \Xi_{[nm]}^2 \cdot K_n(\xi). \quad (358)$$

Using the results of (314) and (316b) for two polarizations, we have

$$\langle A_{[nm]} \rangle = \sqrt{2 \langle (\delta A_{[nm]})^2 \rangle} = L \cdot S_{[nm]} \left(\frac{D_{[nm]}}{2\pi} \lambda \right), \quad (359)$$

where

$$\frac{D_{[nm]}}{2\pi} \lambda = \frac{1}{B_{[nm]}} = \frac{\beta_{01} - \beta_{nm}}{2\pi}. \quad (360)$$

The coupling coefficients $\Xi_{[nm]}$ are approximately proportional to the free-space wavelength λ .

Finally, the mean square n -foil magnitude (defined as the maximum departure from perfect circularity at a given cross section) in the range of mechanical frequencies ζ_a to ζ_b (or the range of mechanical wavelengths $\frac{1}{\zeta}$ from $\frac{1}{\zeta_a}$ to $\frac{1}{\zeta_b}$) is given from (354) by

$$\begin{aligned} \langle |r - a|_{\max}^2 \rangle &= \langle a_n^2(z) \rangle + \langle b_n^2(z) \rangle \\ &= 2 \left[\int_{\zeta_b}^{-\zeta_a} + \int_{\zeta_a}^{\zeta_b} \right] K_n(\zeta) d\zeta. \end{aligned} \quad (361)$$

Assuming that the power spectrum is flat in the range of integration, $K_n(\zeta) = K_n$,

$$\langle |r - a|_{\max}^2 \rangle = 4K_n \cdot (\zeta_b - \zeta_a). \quad (362)$$

These formulas are used to calculate the results in Tables VIII through XII for values of n ranging from 2 to 6. The case of straightness deviations, $n = 1$, is omitted; the same results as those of Section 4.4 would of course be obtained.

4.5.4 Discussion

Neither TE_{01} transmission measurements nor mechanical measurements have thus far yielded information on the shape of the power spectra for diameter variations, ellipticity, and higher-order cross-sectional deformations. This is true principally because these effects are very small in present guides. We are thus unable to predict what the relative contributions of the various spurious modes to the average TE_{01} loss and to the TE_{01} loss fluctuation might be.

In the numerical examples presented in this section, the spectral density of the geometric imperfection has been chosen in such a way that for each type of imperfection the contributions of each of the spurious modes are equal. Plots of the logarithm of the spectral density K_n vs the logarithm of the mechanical frequency ζ show that in each case $K_n(\zeta)$ falls off approximately as $\frac{1}{\zeta}$. Thus, if the spectral density of the imperfection is flat, the higher modes will become progressively more important as the second mode index increases. This is quite different from the observed behavior for straightness deviations, discussed in Section 4.4, where only TE_{12}^+ and TE_{11}^+ have an appreciable effect on

TABLE VIII ($n = 2$) — ELLIPTICITY YIELDING 1 DB/MILE ADDITIONAL AVERAGE TE_{01} LOSS FOR THE LOSS COMPONENT $A_{[2m]}$ DUE TO EACH OF THE TE_{2m}^+ SPURIOUS MODES, AT 55 KMC

$$L_{total} = 20 \text{ miles, total guide length.}$$

$$8.6859 \frac{\langle A_{[2m]total} \rangle}{L_{total}} = 1 \text{ db/mile, additional average loss for each component.}$$

$$f = 55 \text{ kmc, midband frequency.}$$

$$a = 1 \text{ inch, guide radius.}$$

Case 1	$L = 200$ feet, mode filter spacing. $M = 528$, number of mode filters. $8.6859 \sqrt{\langle (\delta A_{[2m]total})^2 \rangle} = 0.6155$ db, rms total loss fluctuation for each component, for 20-mile line.
Case 2	$L = 996.23$ feet, mode filter spacing. $M = 106$, number of mode filters. $8.6859 \sqrt{\langle (\delta A_{[2m]total})^2 \rangle} = 1.3736$ db, rms total loss fluctuation for each component, for 20-mile line.

Spurious Mode	Beat Wavelength Range for 50-60 Kmc Band	RMS Ellipticity in Beat Wavelength Range	K_2 Spectral Density of Ellipticity	Case 1	Case 2
				Δf_{3db} : 3-db Bandwidth of Power Spectrum of Loss Fluctuation	Δf_{3db} : 3-db Bandwidth of Power Spectrum of Loss Fluctuation
	feet	mils	mils ² /foot ⁻¹	mc	mc
TE_{21}^+	5.1623-6.2110	0.3737	1.0676	3128	628
TE_{22}^+	0.9006-1.0880	0.5142	0.3457	547	110
TE_{23}^+	0.3154-0.3836	0.5992	0.1592	192	39
TE_{24}^+	0.1631-0.2005	0.6349	0.0882	100	20
TE_{25}^+	0.0982-0.1226	0.6603	0.0536	61	12
TE_{26}^+	0.0636-0.0815	0.6861	0.0340	40	8
TE_{27}^+	0.0423-0.0569	0.7216	0.0214	27	5.5
TE_{28}^+	0.0263-0.0405	0.8103	0.0123	19	3.8

the TE_{01} loss; there, however, the straightness deviation power spectrum falls off very rapidly, as $\frac{1}{\zeta^4}$.

In any practical case the power spectrum $K_n(\zeta)$ of the geometric deformation must eventually fall off as ζ increases (for $n \neq 1$). If the derivative of the imperfection exists (which seems a reasonable requirement), $K_n(\zeta)$ must eventually fall off faster than $\frac{1}{\zeta^2}$ as ζ becomes large; of course the real question is how large ζ must be for this behavior to dominate. The higher spurious modes have very short beat wavelengths — in the range of an inch or less; if the power spectrum $K_n(\zeta)$ has begun to fall off appreciably at such wavelengths, the contribution of the higher-order spurious modes will be small.

TABLE IX ($n = 3$) — TRI-FOIL YIELDING 1 DB/MILE ADDITIONAL AVERAGE TE_{01} LOSS FOR THE LOSS COMPONENT $A_{[3m]}$ DUE TO EACH OF THE TE_{3m}^+ SPURIOUS MODES, AT 55 KMC

$$L_{total} = 20 \text{ miles, total guide length.}$$

$$8.6859 \frac{\langle A_{[3m]total} \rangle}{L_{total}} = 1 \text{ db/mile, additional average loss for each component.}$$

$$f = 55 \text{ kmc, midband frequency.}$$

$$a = 1 \text{ inch, guide radius.}$$

Case 1	$L = 200 \text{ feet, mode filter spacing.}$ $M = 528, \text{ number of mode filters.}$ $8.6859 \sqrt{\langle (\delta A_{[3m]total})^2 \rangle} = 0.6155 \text{ db, rms total loss fluctuation for each component, for 20-mile line.}$
--------	--

Case 2	$L = 996.23 \text{ feet, mode filter spacing.}$ $M = 106, \text{ number of mode filters.}$ $8.6859 \sqrt{\langle (\delta A_{[3m]total})^2 \rangle} = 1.3736 \text{ db, rms total loss fluctuation for each component, for 20-mile line.}$
--------	---

Spurious Mode	Beat Wavelength Range for 50-60 Kmc Band	RMS Tri-Foil in Beat Wavelength Range	K_3 Spectral Density of Tri-Foil	Case 1	Case 2
				Δf_{3db} : 3-db Bandwidth of Power Spectrum of Loss Fluctuation	Δf_{3db} : 3-db Bandwidth of Power Spectrum of Loss Fluctuation
	feet	mils	mils ² /foot ⁻¹	mc	mc
TE_{31}^+	9.2836-11.1800	0.1876	0.4817	5628	1130
TE_{32}^+	0.5464-0.6617	0.5364	0.2257	332	67
TE_{33}^+	0.2315-0.2827	0.6045	0.1168	141	28
TE_{34}^+	0.1285-0.1590	0.6389	0.0684	79	16
TE_{35}^+	0.0801-0.1011	0.6661	0.0428	50	10
TE_{36}^+	0.0526-0.0687	0.6964	0.0272	34	6.7
TE_{37}^+	0.0346-0.0485	0.7449	0.0167	23	4.6

For ellipticity and higher-order deformations, the additional average loss in db per mile and the total loss fluctuation for a 20-mile guide are roughly comparable, for reasonable mode filter spacings. Thus if the additional average loss is small, as it is in present guides, the loss fluctuation will also be small.

For diameter variations, the total additional average loss for a 20-mile guide is roughly comparable with the total loss fluctuation. Here the total loss fluctuation can remain serious even though the additional average loss in db per mile remains small, as it does in present guides.

The results for diameter variations given here are pessimistic, since the differential loss to the TE_{0m}^+ spurious modes is neglected over the total guide length of 20 miles in the present analysis. The loss fluctuations in practice will be much smaller than those given here. The results

TABLE X ($n = 4$) — 4-FOIL YIELDING 1 DB/MILE ADDITIONAL AVERAGE TE_{01} LOSS FOR THE LOSS COMPONENT $A_{[4m]}$ DUE TO EACH OF THE TE_{4m}^+ SPURIOUS MODES, AT 55 KMC

$$8.6859 \frac{\langle A_{[4m]total} \rangle}{L_{total}} = 1 \text{ db/mile, additional average loss for each component.}$$

$L_{total} = 20 \text{ miles, total guide length.}$
 $f = 55 \text{ kmc, midband frequency.}$
 $a = 1 \text{ inch, guide radius.}$

Case 1	$L = 200 \text{ feet, mode filter spacing.}$ $M = 528, \text{ number of mode filters.}$ $8.6859 \sqrt{\langle (\delta A_{[4m]total})^2 \rangle} = 0.6155 \text{ db, rms total loss fluctuation for each component, for 20-mile line.}$
Case 2	$L = 996.23 \text{ feet, mode filter spacing.}$ $M = 106, \text{ number of mode filters.}$ $8.6859 \sqrt{\langle (\delta A_{[4m]total})^2 \rangle} = 1.3736 \text{ db, rms total loss fluctuation for each component, for 20-mile line.}$

Spurious Mode	Beat Wavelength Range for 50-60 kmc Band	RMS 4-Foil in Beat Wavelength Range	K_4 Spectral Density of 4-Foil	Case 1	Case 2
				$\Delta f_{3\text{db}}$: 3-db Bandwidth of Power Spectrum of Loss Fluctuation	$\Delta f_{3\text{db}}$: 3-db Bandwidth of Power Spectrum of Loss Fluctuation
	feet	mils	mils ² /foot ⁻¹	mc	mc
TE ₄₁ ⁺	2.0190-2.4344	0.2991	0.2647	1225	246
TE ₄₂ ⁺	0.3757-0.4562	0.5433	0.1571	229	46
TE ₄₃ ⁺	0.1782-0.2186	0.6059	0.0885	109	22
TE ₄₄ ⁺	0.1039-0.1295	0.6414	0.0540	64	13
TE ₄₅ ⁺	0.0663-0.0847	0.6719	0.0344	42	8.4
TE ₄₆ ⁺	0.0438-0.0586	0.7091	0.0218	28	5.7
TE ₄₇ ⁺	0.0275-0.0416	0.7890	0.0127	19	3.9

for other cross-sectional deformations are of course valid, since the differential loss will be small in the short distance between mode filters.

4.6 Conclusions

The TE_{01} loss of a long waveguide has been treated as a random process, and its statistics determined in terms of the statistics of the geometric imperfections. A statistical analysis is necessary because it would be impractical to make mechanical measurements of the complete geometry of any great length of guide, even if this were possible.

The numerical results show that rms tolerances of the order of 1 mil are required, for any of the various types of imperfections, to yield an additional average TE_{01} loss of the order of 1 db per mile. The rms di-

TABLE XI ($n = 5$) — 5-FOIL YIELDING 1 DB/MILE ADDITIONAL AVERAGE TE_{01} LOSS FOR THE LOSS COMPONENT $A_{[5m]}$ DUE TO EACH OF THE TE_{5m}^+ SPURIOUS MODES, AT 55 KMC

$L_{total} = 20$ miles, total guide length. $8.6859 \frac{\langle A_{[5m]total} \rangle}{L_{total}} = 1$ db/mile, additional average loss for each component. $f = 55$ kmc, midband frequency. $a = 1$ inch, guide radius.					
Case 1	$L = 200$ feet, mode filter spacing. $M = 528$, number of mode filters. $8.6859 \sqrt{\langle (\delta A_{[5m]total})^2 \rangle} = 0.6155$ db, rms total loss fluctuation for each component, for 20-mile line.				
Case 2	$L = 996.23$ feet, mode filter spacing. $M = 106$, number of mode filters. $8.6859 \sqrt{\langle (\delta A_{[5m]total})^2 \rangle} = 1.3736$ db, rms total loss fluctuation for each component, for 20-mile line.				
Spurious Mode	Beat Wavelength Range for 50-60 kmc Band	RMS 5-Foil in Beat Wavelength Range	K_5 Spectral Density of 5-Foil	Case 1	Case 2
				Δf_{3db} : 3-db Bandwidth of Power Spectrum of Loss Fluctuation	Δf_{3db} : 3-db Bandwidth of Power Spectrum of Loss Fluctuation
	feet	mils	mils ² /foot ⁻¹	mc	mc
TE_{51}^+	1.0317-1.2458	0.3297	0.1632	626	126
TE_{52}^+	0.2771-0.3375	0.5439	0.1144	169	34
TE_{53}^+	0.1417-0.1748	0.6057	0.0687	87	17
TE_{54}^+	0.0855-0.1076	0.6436	0.0432	53	11
TE_{55}^+	0.0554-0.0719	0.6787	0.0277	35	7.1
TE_{56}^+	0.0363-0.0503	0.7269	0.0172	24	4.8

iameter variation of present copper waveguide is of the order of 0.1 mil. Consequently the diameter, ellipticity, tri-foil, and higher-order deformations must have comparable or smaller tolerances, so that they should have a negligible effect on present TE_{01} transmission measurements. Experimental observations support this conclusion. The only spurious modes ever observed in measurements on relatively short (i.e., a few hundred feet in length) waveguides are TE_{12}^+ and TE_{11}^+ , arising from straightness deviations. TE_{0m}^+ , TE_{2m}^+ , TE_{3m}^+ , and higher TE_{nm}^+ modes have never been observed with significant magnitudes.⁸

Straightness deviation is the one tolerance about which in the past we have had no information at all, via mechanical measurements; it is the only significant tolerance in present measurements. Random straightness deviations are believed to account for substantially all of the addi-

TABLE XII ($n = 6$) — 6-FOIL YIELDING 1 DB/MILE ADDITIONAL AVERAGE TE_{01} LOSS FOR THE LOSS COMPONENT $A_{[6m]}$ DUE TO EACH OF THE TE_{6m}^+ SPURIOUS MODES, AT 55 KMC

$L_{total} = 20$ miles, total guide length. $8.6859 \frac{\langle A_{[6m]total} \rangle}{L_{total}} = 1$ db/mile, additional average loss for each component. $f = 55$ kmc, midband frequency. $a = 1$ inch, guide radius.					
Case 1	$L = 200$ feet, mode filter spacing. $M = 528$, number of mode filters. $8.6859 \sqrt{\langle (\delta A_{[6m]total})^2 \rangle} = 0.6155$ db, rms total loss fluctuation for each component, for 20-mile line.				
Case 2	$L = 996.23$ feet, mode filter spacing. $M = 106$, number of mode filters. $8.6859 \sqrt{\langle (\delta A_{[6m]total})^2 \rangle} = 1.3736$ db, rms total loss fluctuation for each component, for 20-mile line.				
Spurious Mode	Beat Wavelength Range for 50-60 kmc Band	RMS 6-Foil in Beat Wavelength Range	K_6 Spectral Density of 6-Foil	Case 1	Case 2
				Δf_{3db} : 3-db Bandwidth of Power Spectrum of Loss Fluctuation	Δf_{3db} : 3-db Bandwidth of Power Spectrum of Loss Fluctuation
	feet	mils	mils ² /foot ⁻¹	mc	mc
TE_{61}^+	0.6532-0.7902	0.3394	0.1085	397	80
TE_{62}^+	0.2138-0.2614	0.5418	0.0861	131	26
TE_{63}^+	0.1154-0.1433	0.6047	0.0543	71	14
TE_{64}^+	0.0714-0.0907	0.6461	0.0349	45	9
TE_{65}^+	0.0465-0.0617	0.6873	0.0223	30	6
TE_{66}^+	0.0295-0.0434	0.7582	0.0133	20	4.1

tional loss observed in present copper waveguides, and to account for a substantial part of the additional loss in helix as well.

Diameter variations in both copper and helix guide give rise to the TE_{0m} spurious modes, which cannot be satisfactorily attenuated by existing structures. It may not be sufficient to have a good enough diameter tolerance to yield a small additional average TE_{01} loss, since the TE_{01} loss fluctuations may still remain objectionable. The diameter tolerance must thus be substantially better than the tolerance for ellipticity, tri-foil, etc., in copper guide. However, the present results for diameter variations are too pessimistic, since the differential loss over the entire length of guide was neglected.

Tolerances on diameter, ellipticity, tri-foil, etc. for drawn copper guide are tolerances at a single cross section, and are controlled primarily by the accuracy of the die through which the guide is drawn. A

good straightness tolerance requires accurate alignment between different cross sections separated by substantial distances (equal to the beat wavelength of TE_{12}^+ , TE_{11}^+). This depends on many factors other than the dimensional accuracy of the dies; for example, random variations in hardness or wall thickness may cause the axis of the guide to curve as it is being drawn. Consequently, a good straightness tolerance is more difficult to attain than any of the other cross-sectional tolerances, in drawing copper guide. Numerous other manufacturing processes for copper guide are currently under study.⁴²

The variation of the TE_{01} loss statistics with mode filter spacing is illustrated in the examples given above. The average loss is unaffected by the mode filter spacing, but the rms loss fluctuation is inversely proportional to the square root of the number of mode filters. As discussed in Section 3.4, the total TE_{01} loss will be approximately a Gaussian random process.

While the present analysis applies only to copper waveguide with the differential loss neglected, further study⁴¹ shows that adding loss to the spurious mode has an effect similar to that of increasing the number of mode filters; for moderate values, as $\Delta\alpha$ increases the average loss changes very little, while the loss fluctuation will be progressively reduced. Thus, accurate tolerances will be important in helix (or in copper guide with a lossy dielectric lining) as well as copper guide, although in helix the principal effect of poor tolerances will be an increased average loss, the loss fluctuations remaining small (diameter variations excluded).

Finally, the shape of the power spectra of the different mechanical imperfections is all-important in determining the resulting TE_{01} loss due to mode conversion. Only components of the mechanical imperfection in the beat wavelength range of the important spurious modes have any effect on the TE_{01} transmission. The short-wavelength straightness deviations "built in" to the guide in the manufacturing process will be principally responsible for additional loss due to mode conversion to modes such as TE_{12}^+ and TE_{11}^+ in copper guide; long bows or random straightness deviations due to imperfect laying of the guide will have only a very small effect.

REFERENCES

1. Miller, S. E., Waveguide as a Communication Medium, B.S.T.J., **33**, Nov. 1954, pp. 1209-1266.
2. Morgan, S. P., Theory of Curved Circular Waveguide Containing an Inhomogeneous Dielectric, B.S.T.J., **36**, Sept. 1957, pp. 1209-1251.
3. Unger, H. G., Helix Waveguide Theory and Application, B.S.T.J., **37**, Nov. 1958, pp. 1599-1647.

4. Schelkunoff, S. A., Conversion of Maxwell's Equations into Generalized Telegraphist's Equations, *B.S.T.J.*, **34**, Sept. 1955, pp. 995-1043.
5. Morgan, S. P., Mode Conversion Losses in Transmission of Circular Electric Waves Through Slightly Noncylindrical Guides, *J. Appl. Phys.*, **21**, April 1950, pp. 329-338.
6. Jouguet, M., Effects of the Curvature on the Propagation of Electromagnetic Waves in Guides of Circular Cross-Section, Cables and Trans., **1**, No. 2, July 1947, pp. 133-153.
7. Miller, S. E., Notes on Methods of Transmitting the Circular Electric Wave around Bends, *Proc. I.R.E.*, **40**, Sept. 1952, pp. 1104-1113.
8. Rowe, H. E., and Warters, W. D., Transmission Deviations in Waveguide Due to Mode Conversion: Theory and Experiment, *Proc. I.E.E.*, **106**, Part B, Supplement No. 13, Sept. 1959, pp. 30-36; Warters, W. D., and Rowe, H. E., The Effects of Mode Conversion in Long Circular Waveguide, *I.R.E. Wescon Convention Record*, 1958, Part 1, pp. 13-20.
9. Unger, H. G., Normal Mode Bends for Circular Electric Waves, *B.S.T.J.*, **36**, Sept. 1957, pp. 1292-1307.
10. Unger, H. G., Circular Waveguide Taper of Improved Design, *B.S.T.J.*, **37**, July 1958, pp. 899-912.
11. Morgan, S. P., and Young, J. A., Helix Waveguide, *B.S.T.J.*, **35**, Nov. 1956, pp. 1347-1384.
12. Unger, H. G., Round Waveguide with Lossy Lining, *Proc. of the Symposium on Millimeter Waves*, Polytechnic Inst. of Brooklyn, 1959, pp. 535-541.
13. Unger, H. G., Round Waveguide with Double Lining, *B.S.T.J.*, **39**, Jan. 1960, pp. 161-168.
14. Unger, H. G., Circular Electric Wave Transmission in a Dielectric-Coated Waveguide, *B.S.T.J.*, **36**, Sept. 1957, pp. 1253-1278.
15. Morgan, S. P., unpublished memoranda, 1951.
16. Iiguchi, S., Mode Conversion in the Transmission of TE_{01} Waves Through a Slight Tilt and a Slight Offset of Waveguide, *J. Inst. Elec. Comm. Engrs., Japan*, **40**, 1957, pp. 870-876, 1095-1102.
17. Miller, S. E., Coupled Wave Theory and Waveguide Applications, *B.S.T.J.*, **33**, May 1954, pp. 661-720.
18. Rowe, H. E., Approximate Solutions for the Coupled Line Equations, *B.S.T.J.* this issue, pp. 1011.
19. Unger, H. G., Noncylindrical Helix Waveguide, *B.S.T.J.*, **40**, Jan. 1961, pp. 233-254.
20. Unger, H. G., Mode Conversion in Metallic and Helix Waveguide, *B.S.T.J.*, **40**, March 1961, pp. 613-626.
21. Unger, H. G., Winding Tolerances in Helix Waveguide, *B.S.T.J.*, **40**, March 1961, pp. 627-643.
22. Unger, H. G., Normal Modes and Mode Conversion in Helix Waveguide, *B.S.T.J.*, **40**, Jan. 1961, pp. 255-280.
23. Montgomery, C. G., Dicke, R. H., and Purcell, E. M., *Principles of Microwave Circuits*, McGraw-Hill Book Co., Inc., New York, N. Y.; 1948.
24. Bolinder, E. F., Fourier Transforms and Tapered Transmission Lines, *Proc. I.R.E.*, **44**, April 1956, p. 557.
25. Klopfenstein, R. W., A Transmission Line Taper of Improved Design, *Proc. I.R.E.*, **44**, Jan. 1956, pp. 31-35.
26. Collin, R. F., The Optimum Tapered Transmission Line Matching Section, *Proc. I.R.E.*, **44**, April 1956, pp. 539-548.
27. Kaufman, H., Bibliography of Nonuniform Transmission Lines, *I.R.E. Trans. on Antennas and Propagation*, Vol. **AP-3**, Oct. 1955, pp. 218-220.
28. Louisell, W. H., Analysis of the Single Tapered Mode Coupler, *B.S.T.J.*, **34**, July 1955, pp. 853-870.
29. Mertz, P., and Pfeiffer, K. W., Irregularities in Broadband Wire Transmission Circuits, *B.S.T.J.*, **16**, Oct. 1937, pp. 541-559.
30. Kaden, H., Advances Made in the Statistics of Impedance Irregularities of Television Cables, *Archiv. der Electr. Über.*, **8**, 1954, pp. 523-529.
31. Ince, E. L., *Ordinary Differential Equations*, Dover, New York, N. Y.; 1956.
32. Bellman, R., *Stability Theory of Differential Equations*, McGraw-Hill Book Co., Inc., New York, N. Y.; 1953.

33. Shannon, C. E., and Weaver, W., *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, Ill.; 1949; see p. 53.
34. Shannon, C. E., Communication in the Presence of Noise, Proc. I.R.E., **37**, Jan. 1949, pp. 10-21.
35. Unger, H. G., Circular Electric Wave Transmission Through Serpentine Bends, B.S.T.J., **36**, Sept. 1957, pp. 1279-1291.
36. Franklin, P., *Methods of Advanced Calculus*, McGraw-Hill Book Co., Inc., New York, N. Y.; 1944; see Chapter III.
37. Davenport, W. B., and Root, W. L., *Random Signals and Noise*, McGraw-Hill Book Co., Inc., New York, N. Y.; 1958.
38. Bellman, R., *Introduction to Matrix Analysis*, McGraw-Hill Book Co., Inc., New York, N. Y.; 1960.
39. King, A. P., and Mandeville, G. D., The Observed 33-90 Kmc. Attenuation of Two Inch Improved Waveguide, B.S.T.J., **40**, Sept. 1961, pp. 1323-1330.
40. Rice, S. O., Mathematical Analysis of Random Noise; *Noise and Stochastic Processes*, Nelson Wax, Editor, Dover, New York, N. Y., 1954.
41. Rowe, H. E., unpublished work.
42. Beck, A. C., and Rose, C. F. P., Waveguide for Circular Electric Mode Transmission, Proc. I. E. E., **106**, Part B, Supplement No. 13, Sept. 1959, pp. 159-162.

APPENDIX A

Coupling Coefficients for Tilts, Offsets, and Diameter Changes^{15, 16}

General formulas for the coupling coefficients from TE₀₁ to the first-order spurious modes at offsets, tilts, and diameter changes, as determined by S. P. Morgan, are given in Table XIII. Numerical values are tabulated at 55 kmc for 2-inch diameter guide in Tables XIV, XV and XVI; the computations here and in Appendix D were performed by Mrs. C. L. Beattie.

Notation

- $C_{o[m]}^{\pm}$ — coupling coefficient between TE₀₁ and forward (+) or backward (-) TE_{1m} for an offset in copper guide.
- $C_{t[m]}^{\pm}$ — coupling coefficient between TE₀₁ and forward (+) or backward (-) TE_{1m} for a tilt in copper guide.
- $C_{t(11)}^{+}$ — coupling coefficient between TE₀₁ and forward TM₁₁ for a tilt in copper guide.
- $C_{d[m]}^{\pm}$ — coupling coefficient between TE₀₁ and forward (+) or backward (-) TE_{0m} for a diameter change in copper or helix guide.
- a — guide radius.
- λ — free-space wavelength.
- k_{nm} — Bessel root given by $J_n'(k_{nm}) = 0$.
- $\nu_{nm} = \frac{k_{nm}\lambda}{2\pi a}$, cutoff factor for the TE_{nm} mode.
- $\beta_{nm} = \frac{2\pi}{\lambda} \sqrt{1 - \nu_{nm}^2}$, propagation constant for the TE_{nm} mode.

TABLE XIII

Coupling Coefficient	Spurious Mode	Equation
$C_{o[m]}^{\pm} = \frac{1}{\sqrt{2}} \frac{k_{01}k_{1m}^2}{a(k_{1m}^2 - k_{01}^2)} \frac{\beta_{01} \pm \beta_{1m}}{\sqrt{k_{1m}^2 - 1} \sqrt{\beta_{01}\beta_{1m}}}$	TE _{1m}	(A-1)
$C_{l[m]}^{\pm} = \frac{a}{\sqrt{2}} \frac{k_{01}k_{1m}^2}{(k_{01}^2 - k_{1m}^2)^2 \sqrt{k_{1m}^2 - 1}} \frac{(\beta_{01} \pm \beta_{1m})^2}{\sqrt{\beta_{01}\beta_{1m}}}$	TE _{1m}	(A-2)
$C_{l(11)}^{+} = \frac{\sqrt{2}\pi a}{k_{01}\lambda}$	TM ₁₁	(A-3)
$C_{d[m]}^{\pm} = \frac{1}{a} \frac{k_{01}k_{0m}}{k_{0m}^2 - k_{01}^2} \frac{\beta_{0m} \pm \beta_{01}}{\sqrt{\beta_{01}\beta_{0m}}}, \quad m \neq 1.$	TE _{0m}	(A-4)

TABLE XIV — COUPLING COEFFICIENTS FOR OFFSET
(f = 55 kmc; a = 1 inch):

Spurious Mode	$C_{o[m]}^{+}$	$C_{o[m]}^{-}$
	inch ⁻¹	inch ⁻¹
TE ₁₁	-1.052295	0.0035023
TE ₁₂	2.140300	0.0087996
TE ₁₃	0.800615	0.0143221
TE ₁₄	0.520766	0.0204174
TE ₁₅	0.392376	0.0274958
TE ₁₆	0.317601	0.0362493
TE ₁₇	0.269338	0.0480943
TE ₁₈	0.238323	0.0668142
TE ₁₉	0.230068	0.1105813

TABLE XV — COUPLING COEFFICIENTS FOR TILT
(f = 55 kmc; a = 1 inch)

Spurious Mode	$C_{l[m]}^{+}$	$C_{l[m]}^{-}$
	radian ⁻¹	radian ⁻¹
TE ₁₁	5.428116	0.000060129
TE ₁₂	9.004674	0.000152211
TE ₁₃	0.784756	0.000251132
TE ₁₄	0.237781	0.000365504
TE ₁₅	0.103215	0.000506837
TE ₁₆	0.0534059	0.000695704
TE ₁₇	0.0306226	0.000976413
TE ₁₈	0.0187490	0.00147362
TE ₁₉	0.0121822	0.00282165

	$C_{l(11)}^{+}$	$C_{l(11)}^{-}$
	radian ⁻¹	radian ⁻¹
TM ₁₁	5.403	0.

TABLE XVI — COUPLING COEFFICIENTS FOR DIAMETER CHANGE
($f = 55 \text{ kmc}$; $a = 1 \text{ inch}$)

Spurious Mode	$C_d[m]^+$	$C_d[m]^-$
	inch ⁻¹	inch ⁻¹
TE ₀₂	1.556796	-0.0162900
TE ₀₃	0.878135	-0.0244495
TE ₀₄	0.627935	-0.0336831
TE ₀₅	0.493929	-0.0447267
TE ₀₆	0.410428	-0.0589493
TE ₀₇	0.355528	-0.0795021
TE ₀₈	0.324255	-0.1168700
TE ₀₉	0.425098	-0.329569

APPENDIX B

Geometry of Discrete Tilts

Let unit vectors directed along the guide axes in the two guide sections adjacent to a discrete tilt be \mathbf{t}_1 and \mathbf{t}_2 . Then

$$\mathbf{t}_1 = ix_1 + jy_1 + kz_1; \quad |\mathbf{t}_1| = 1. \quad (\text{B-1})$$

$$\mathbf{t}_2 = ix_2 + jy_2 + kz_2; \quad |\mathbf{t}_2| = 1. \quad (\text{B-2})$$

\mathbf{i} , \mathbf{j} , and \mathbf{k} are unit vectors along the x , y , and z axes respectively. Since \mathbf{t}_1 and \mathbf{t}_2 are unit vectors,

$$x_1^2 + y_1^2 + z_1^2 = 1, \quad (\text{B-3})$$

$$x_2^2 + y_2^2 + z_2^2 = 1. \quad (\text{B-4})$$

Let the tilt have angle α , orientation θ as defined in Section 2.1.2; further let the corresponding angles of the projections of the guide axes on the x - z and y - z planes be α_x and α_y , as in (38). Then by taking appropriate dot products we have:

$$\cos \alpha = x_1x_2 + y_1y_2 + z_1z_2, \quad (\text{B-5})$$

$$\cos \alpha_x = \frac{x_1x_2 + z_1z_2}{\sqrt{x_1^2 + z_1^2} \sqrt{x_2^2 + z_2^2}}, \quad (\text{B-6})$$

$$\cos \alpha_y = \frac{y_1y_2 + z_1z_2}{\sqrt{y_1^2 + z_1^2} \sqrt{y_2^2 + z_2^2}}, \quad (\text{B-7})$$

where (B-3) and (B-4) of course hold true. If the angular deviation of the guide axis from the z -axis is small, we have

$$x_1 \ll 1, \quad y_1 \ll 1, \quad x_2 \ll 1, \quad y_2 \ll 1. \quad (\text{B-8})$$

Under these conditions (B-5) to (B-7) yield to first order

$$\alpha \approx \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \quad (\text{B-9})$$

$$\alpha_x \approx x_2 - x_1, \quad (\text{B-10})$$

$$\alpha_y \approx y_2 - y_1, \quad (\text{B-11})$$

where we have made use of (B-3) and (B-4).

Now define the unit vector \mathbf{p} as follows:

$$\mathbf{p} = \frac{\mathbf{t}_2 - \mathbf{t}_1}{|\mathbf{t}_2 - \mathbf{t}_1|} = \frac{\mathbf{t}_2 - \mathbf{t}_1}{2 \sin(\alpha/2)} \quad (\text{B-12})$$

\mathbf{p} lies in the plane of the tilt, i.e., in the $\varphi = \theta$ plane, and bisects the angle made by the guide axes on the two sides of the tilt. The vectors $\mathbf{t} = (\mathbf{t}_1 + \mathbf{t}_2)/2$ and \mathbf{p} are the analogs for the discrete case of the tangent and principal normal vectors of differential geometry, introduced in Section 2.4, in the treatment of the continuous case. Denote the transverse component of \mathbf{p} by $\mathbf{p}_{x,y}$; then by (B-12), (B-1) and (B-2)

$$\mathbf{p}_{x,y} = \frac{1}{2 \sin(\alpha/2)} [\mathbf{i}(x_2 - x_1) + \mathbf{j}(y_2 - y_1)]. \quad (\text{B-13})$$

Now if the angular deviation of the guide axis from the z -axis is small (on both sides of the tilt), as assumed in (B-8), and if a unit vector perpendicular to the guide axis and lying in the $\varphi = 0$ plane is almost parallel to the x -axis, then the angle of $\mathbf{p}_{x,y}$ with respect to the x -axis will be approximately equal to the orientation θ of the tilt; under these conditions we have from (B-13)

$$\cos \theta \approx \frac{x_2 - x_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}, \quad (\text{B-14})$$

$$\sin \theta \approx \frac{y_2 - y_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}. \quad (\text{B-15})$$

From (B-9) to (B-11), (B-14) and (B-15) we then have

$$\alpha \cos \theta \approx \alpha_x, \quad (\text{B-16})$$

$$\alpha \sin \theta \approx \alpha_y, \quad (\text{B-17})$$

as stated in (38).

APPENDIX C

Energy Relations for Guides with Real Coupling Coefficients

Consider the coupled line equations given in (85)

$$I_0'(z) = -\Gamma_0 I_0(z) + jc(z) I_1(z) \quad (\text{C-1})$$

$$I_1'(z) = jc(z) I_0(z) - \Gamma_1 I_1(z)$$

$$\Gamma_0 = \alpha_0 + j\beta_0; \quad \Gamma_1 = \alpha_1 + j\beta_1. \quad (\text{C-2})$$

We assume in this appendix that $c(z)$ is pure real;

$$\text{Im } c(z) = 0. \quad (\text{C-3})$$

Consider first the case of ideal metallic guide, for which $\alpha_0 = \alpha_1 = 0$. The total power $P(z)$ flowing in the guide at the point z is simply

$$P(z) = |I_0(z)|^2 + |I_1(z)|^2 = I_0(z)I_0^*(z) + I_1(z)I_1^*(z). \quad (\text{C-4})$$

Now in general, for guide whose walls are not perfect conductors, the α 's will not be identically zero, and neither (C-3) nor (C-4) hold true. Helix waveguide furnishes an interesting example. Here the coupling coefficients are complex, so that (C-3) is not valid; further the powers in the various modes are not orthogonal, so that (C-4) is untrue. In this appendix we consider cases where $\alpha \neq 0$ but where (C-3) holds true, so that $c(z)$ is pure real. We define a quantity $P(z)$ by (C-4); however only for ideal metallic guide, where the α 's are equal to zero, are we assured that $P(z)$ really represents the total power. If the α 's are not zero we have no reason to think that $P(z)$ should be the total power; however the results given below render this plausible when $c(z)$ is real.

From (C-4) we write

$$P'(z) = \frac{dP(z)}{dz} = I_0(z)I_0'^*(z) + I_0'(z)I_0^*(z) + I_1(z)I_1'^*(z) + I_1'(z)I_1^*(z). \quad (\text{C-5})$$

Substituting (C-1) into (C-5), making use of (C-2) and (C-3), we find

$$P'(z) = -2\alpha_0 |I_0(z)|^2 - 2\alpha_1 |I_1(z)|^2. \quad (\text{C-6})$$

If $P(z)$ is the total power flowing along the guide, (C-6) has a simple physical interpretation. It says that each mode contributes to the de-

crease in power along the guide in proportion to the product of its attenuation constant and the power it carries. Equation (C-6) may be extended to any number of modes via straightforward matrix techniques.

Finally, consider the case where

$$\alpha_0 = \alpha_1 \equiv \alpha, \quad \Delta\alpha \equiv \alpha_0 - \alpha_1 = 0. \quad (\text{C-7})$$

From the transformation of (88)–(89), (C-4) becomes

$$P(z) = e^{-2\alpha z} [|G_0(z)|^2 + |G_1(z)|^2]. \quad (\text{C-8})$$

Similarly from (C-6), (C-7), and (C-4), we find

$$P'(z) = -2\alpha P(z), \quad (\text{C-9})$$

which has the solution

$$P(z) = e^{-2\alpha z} P(0) = e^{-2\alpha z} [|I_0(0)|^2 + |I_1(0)|^2]. \quad (\text{C-10})$$

Assuming as usual that the guide is excited by a unit TE_{01} wave so that the initial conditions of (87) apply, i.e.,

$$I_0(0) = 1, \quad I_1(0) = 0, \quad (\text{C-11})$$

(C-10) becomes

$$P(z) = e^{-2\alpha z}. \quad (\text{C-12})$$

From (C-8) and (C-12) we have finally

$$|G_0(z)|^2 + |G_1(z)|^2 = 1, \quad (\text{C-13})$$

subject of course to the following conditions:

$$G_0(0) = 1, \quad G_1(0) = 0. \quad (\text{C-14})$$

$$\alpha_0 = \alpha_1, \quad \Delta\alpha = \alpha_0 - \alpha_1 = 0. \quad (\text{C-15})$$

A similar treatment may of course be given for the coupled line equations of (93).

APPENDIX D (See Section 2.3.8)

Coupling Coefficients $\Xi_{[nm]}$ for General Continuous Deformations, and Beat Wavelengths $B_{[nm]}^+$, for Metallic Guide (Guide Diameter = 2 Inches)

n	m	Frequency = 50 Kmc		Frequency = 55 Kmc		Frequency = 60 Kmc	
		B _[nm] ⁺ feet	Ξ _[nm] inch ⁻²	B _[nm] ⁺ feet	Ξ _[nm] inch ⁻²	B _[nm] ⁺ feet	Ξ _[nm] inch ⁻²
0							
	2	0.78861	1.033665	0.87104	0.935818	0.95318	0.855165
	3	0.30028	1.531489	0.33298	1.380849	0.36543	1.258058
	4	0.15879	2.072235	0.17717	1.855771	0.19531	1.682663
	5	0.09641	2.689180	0.10863	2.380811	0.12055	2.142344
	6	0.06272	3.452634	0.07182	2.992239	0.08053	2.658456
	7	0.04176	4.574033	0.04936	3.771644	0.05630	3.271486
	8	0.02590	7.815578	0.03405	4.986672	0.04021	4.077331
	9	CUT OFF		0.02066	10.765911	0.02842	5.422487
1							
	1	-2.45264	0.224648	-2.70093	0.203996	-2.94898	0.186837
	2	1.99720	0.561114	2.20291	0.508716	2.40812	0.465363
	3	0.46389	0.903729	0.51323	0.816787	0.56232	0.745451
	4	0.21503	1.268585	0.23908	1.140518	0.26289	1.036959
	5	0.12293	1.673662	0.13773	1.491619	0.15228	1.348056
	6	0.07762	2.151892	0.08805	1.888713	0.09815	1.690415
	7	0.05134	2.784072	0.05953	2.368870	0.06723	2.084309
	8	0.03377	3.901930	0.04119	3.029237	0.04763	2.572121
	9	CUT OFF		0.02776	4.333118	0.03402	3.272356
	10	CUT OFF		CUT OFF		0.02321	4.869801
2							
	1	-5.16225	0.414888	-5.68694	0.376608	-6.21100	0.344833
	2	0.90059	0.730808	0.99444	0.661825	1.08797	0.604918
	3	0.31537	1.081401	0.34959	0.975358	0.38358	0.888839
	4	0.16314	1.462636	0.18196	1.310395	0.20053	1.188501
	5	0.09816	1.897242	0.11054	1.680662	0.12263	1.512888
	6	0.06357	2.433890	0.07274	2.111386	0.08153	1.876880
	7	0.04225	3.217515	0.04987	2.659138	0.05685	2.308706
	8	0.02631	5.387660	0.03437	3.507377	0.04054	2.874896
	9	CUT OFF		0.02108	7.134864	0.02865	3.812804
3							
	1	9.28358	0.617958	10.23270	0.560645	11.18001	0.513138
	2	0.54642	0.905720	0.60417	0.819106	0.66165	0.747918
	3	0.23147	1.265707	0.25719	1.138773	0.28266	1.035917
	4	0.12849	1.667518	0.14385	1.487579	0.15895	1.345276
	5	0.08006	2.141126	0.09072	1.881991	0.10105	1.685882
	6	0.05262	2.763190	0.06091	2.357577	0.06872	2.077242
	7	0.03459	3.837927	0.04200	3.006912	0.04848	2.560340
	8	CUT OFF		0.02835	4.252809	0.03457	3.248210
	9	CUT OFF		CUT OFF		0.02366	4.759583
4							
	1	2.01904	0.834246	2.22697	0.756350	2.43441	0.691899
	2	0.37568	1.087248	0.41603	0.981657	0.45615	0.895260
	3	0.17821	1.458404	0.19854	1.308247	0.21863	1.187591
	4	0.10386	1.886313	0.11679	1.673870	0.12945	1.508455
	5	0.06627	2.412943	0.07568	2.099151	0.08471	1.868999
	6	0.04377	3.169891	0.05147	2.636799	0.05856	2.295676
	7	0.02754	5.053602	0.03538	3.454117	0.04157	2.851181
	8	CUT OFF		0.02221	6.258648	0.02936	3.751777
	9	CUT OFF		CUT OFF		0.01672	24.592342

n	m	Frequency = 50 Kmc		Frequency = 55 Kmc		Frequency = 60 Kmc	
		B _[nm] ⁺ feet	Ξ _[nm] inch ²	B _[nm] ⁺ feet	Ξ _[nm] inch ²	B _[nm] ⁺ feet	Ξ _[nm] inch ²
5							
	1	1.03171	1.063496	1.13893	0.963362	1.24582	0.880700
	2	0.27706	1.276544	0.30740	1.150302	0.33750	1.047565
	3	0.14174	1.661453	0.15842	1.485036	0.17484	1.344734
	4	0.08552	2.123151	0.09669	1.871474	0.10755	1.679382
	5	0.05538	2.725027	0.06388	2.337156	0.07193	2.064780
	6	0.03629	3.724150	0.04371	2.965282	0.05029	2.538314
	7	CUT OFF		0.02954	4.111601	0.03571	3.202782
	8	CUT OFF		CUT OFF		0.02456	4.570756
6							
	1	0.65318	1.305633	0.72180	1.181454	0.79015	1.079233
	2	0.21378	1.474795	0.23771	1.325829	0.26139	1.205391
	3	0.11540	1.877243	0.12947	1.670557	0.14326	1.508273
	4	0.07137	2.383934	0.08123	2.083255	0.09074	1.859682
	5	0.04651	3.099496	0.05437	2.603525	0.06168	2.276663
	6	0.02954	4.663254	0.03715	3.374330	0.04340	2.814772
	7	CUT OFF		0.02386	5.492384	0.03060	3.660164
	8	CUT OFF		CUT OFF		0.01927	7.278403
7							
	1	0.45817	1.560845	0.50693	1.410610	0.55543	1.287362
	2	0.17024	1.683378	0.18977	1.509104	0.20906	1.369320
	3	0.09560	2.108882	0.10773	1.866529	0.11957	1.679271
	4	0.06011	2.677695	0.06899	2.313116	0.07746	2.051409
	5	0.03903	3.582532	0.04652	2.910415	0.05327	2.509531
	6	0.02148	10.369537	0.03140	3.937248	0.03755	3.141452
	7	CUT OFF		CUT OFF		0.02593	4.344516
8							
	1	0.34163	1.829591	0.37853	1.651012	0.41518	1.505110
	2	0.13874	1.903988	0.15512	1.701141	0.17125	1.540005
	3	0.08022	2.360640	0.09090	2.075128	0.10125	1.858998
	4	0.05092	3.019453	0.05908	2.566671	0.06674	2.257265
	5	0.03243	4.298900	0.03988	3.279521	0.04625	2.770706
	6	CUT OFF		0.02603	4.903585	0.03244	3.549922
	7	CUT OFF		CUT OFF		0.02122	5.845324
9							
	1	0.26536	2.112618	0.29451	1.903044	0.32342	1.732653
	2	0.11507	2.138800	0.12910	1.903182	0.14287	1.718219
	3	0.06797	2.638691	0.07752	2.299234	0.08671	2.049035
	4	0.04320	3.437147	0.05087	2.852480	0.05792	2.480954
	5	0.02597	5.866004	0.03410	3.755552	0.04026	3.071952
	6	CUT OFF		0.01929	11.251662	0.02784	4.113642
10							
	1	0.21225	2.410992	0.23602	2.167296	0.25955	1.970313
	2	0.09673	2.390696	0.10898	2.116784	0.12093	1.904895
	3	0.05796	2.952529	0.06666	2.542807	0.07494	2.251409
	4	0.03650	3.992360	0.04393	3.184628	0.05052	2.727785
	5	CUT OFF		0.02886	4.456392	0.03506	3.434148
	6	CUT OFF		CUT OFF		0.02344	5.078573

n	m	Frequency = 50 Kmc		Frequency = 55 Kmc		Frequency = 60 Kmc	
		B _[nm] ⁺ feet	Ξ _[nm] inch ⁻²	B _[nm] ⁺ feet	Ξ _[nm] inch ⁻²	B _[nm] ⁺ feet	Ξ _[nm] inch ⁻²
11							
	1	0.17355	2.726150	0.19341	2.444582	0.21303	2.218565
	2	0.08217	2.663620	0.09302	2.343950	0.10356	2.101187
	3	0.04961	3.317918	0.05768	2.811536	0.06524	2.468802
	4	0.03040	4.869555	0.03794	3.588797	0.04422	3.005794
	5	CUT OFF		0.02364	5.918920	0.03042	3.899372
	6	CUT OFF		CUT OFF		0.01796	10.206978
12							
	1	0.14434	3.059998	0.16128	2.735980	0.17796	2.478048
	2	0.07035	2.963187	0.08012	2.587323	0.08953	2.308543
	3	0.04248	3.763972	0.05011	3.114056	0.05711	2.704883
	4	0.02390	7.400976	0.03262	4.119666	0.03877	3.328012
	5	CUT OFF		CUT OFF		0.02614	4.574673
13							
	1	0.12168	3.415043	0.13636	3.042884	0.15078	2.749591
	2	0.06057	3.297766	0.06949	2.850485	0.07800	2.528816
	3	0.03621	4.353961	0.04363	3.464397	0.05020	2.964872
	4	CUT OFF		0.02769	4.929759	0.03396	3.717676
	5	CUT OFF		CUT OFF		0.02185	5.901255
14							
	1	0.10368	3.794611	0.11660	3.367083	0.12924	3.034244
	2	0.05233	3.680631	0.06060	3.138458	0.06838	2.764429
	3	0.03043	5.271752	0.03797	3.887661	0.04425	3.256582
	4	CUT OFF		0.02257	6.850328	0.02962	4.222458
15							
	1	0.08911	4.203178	0.10063	3.710887	0.11183	3.333332
	2	0.04526	4.134654	0.05305	3.458593	0.06025	3.018632
	3	0.02433	7.682039	0.03290	4.435549	0.03905	3.592421
	4	CUT OFF		CUT OFF		0.02555	4.966220
16							
	1	0.07711	4.646904	0.08750	4.077300	0.09755	3.648526
	2	0.03905	4.704352	0.04654	3.822244	0.05330	3.295920
	3	CUT OFF		0.02820	5.243552	0.03443	3.993906
	4	CUT OFF		CUT OFF		0.02139	6.491317
17							
	1	0.06707	5.134532	0.07655	4.470288	0.08566	3.981936
	2	0.03343	5.495075	0.04085	4.248284	0.04727	3.602748
	3	CUT OFF		0.02341	6.919841	0.03025	4.503227
18							
	1	0.05856	5.679004	0.06731	4.895211	0.07564	4.336278
	2	0.02801	6.872935	0.03579	4.771480	0.04199	3.948853
	3	CUT OFF		CUT OFF		0.02634	5.221466
19							
	1	0.05122	6.300569	0.05941	5.359516	0.06710	4.715079
	2	0.02050	22.459224	0.03116	5.466895	0.03730	4.349893
	3	CUT OFF		CUT OFF		0.02244	6.509164

n	m	Frequency = 50 Kmc		Frequency = 55 Kmc		Frequency = 60 Kmc	
		$B_{[nm]}^+$ feet	$\Xi_{[nm]}$ inch ⁻²	$B_{[nm]}^+$ feet	$\Xi_{[nm]}$ inch ⁻²	$B_{[nm]}^+$ feet	$\Xi_{[nm]}$ inch ⁻²
20	1	0.04481	7.033398	0.05257	5.873928	0.05974	5.123026
	2	CUT OFF		0.02675	6.549116	0.03308	4.833374
21	1	0.03910	7.941831	0.04659	6.454660	0.05335	5.566509
	2	CUT OFF		0.02197	9.289222	0.02920	5.454350
22	1	0.03388	9.169691	0.04130	7.127805	0.04774	6.054504
	2	CUT OFF		CUT OFF		0.02552	6.349417
23	1	0.02886	11.158150	0.03654	7.939276	0.04277	6.600202
	2	CUT OFF		CUT OFF		0.02175	8.047971
24	1	0.02315	17.444425	0.03217	8.981077	0.03832	7.224001
25	1	CUT OFF		0.02804	10.482049	0.03428	7.959870
26	1	CUT OFF		0.02382	13.335010	0.03057	8.870500
27	1	CUT OFF		CUT OFF		0.02707	10.091403
28	1	CUT OFF		CUT OFF		0.02364	12.010406
29	1	CUT OFF		CUT OFF		0.01979	16.923252

APPENDIX E

Geometry of Continuous Bends

Let $x(z)$ and $y(z)$ be the transverse displacements of the guide axis from the z -axis, in a rectangular co-ordinate system. Then we write

$$\mathbf{r}(z) = i x(z) + j y(z) + k z, \quad (\text{E-1})$$

where \mathbf{i} , \mathbf{j} , and \mathbf{k} are unit vectors along the x , y , and z axes respectively. Assuming that (215) holds true, from (216a) we have

$$s \approx z. \quad (\text{E-2})$$

Thus, arc length along the bent guide axis is approximately equal to distance measured along the z -axis. From (211a) and (E-2),

$$\mathbf{t} = \frac{d\mathbf{r}}{ds} \approx \frac{d\mathbf{r}}{dz} = \mathbf{i}x'(z) + \mathbf{j}y'(z) + \mathbf{k}, \quad (\text{E-3})$$

where primes indicate differentiation with respect to z . From (211b)

$$\frac{1}{\rho} \mathbf{p} = \frac{d\mathbf{t}}{ds} \approx \frac{d\mathbf{t}}{dz} \approx \mathbf{i}x''(z) + \mathbf{j}y''(z). \quad (\text{E-4})$$

Therefore the curvature is approximately

$$\frac{1}{\rho} \approx \sqrt{x''^2(z) + y''^2(z)} \quad (\text{E-5})$$

so that

$$\mathbf{p} \approx \frac{\mathbf{i}x''(z) + \mathbf{j}y''(z)}{\sqrt{x''^2(z) + y''^2(z)}}. \quad (\text{E-6})$$

Then from (211c)

$$\mathbf{b} = \mathbf{t} \times \mathbf{p} \approx \mathbf{k} \times \mathbf{p} \approx \frac{\mathbf{j}x''(z) - \mathbf{i}y''(z)}{1/\rho}. \quad (\text{E-7})$$

Since \mathbf{p} and \mathbf{b} are approximately transverse, i.e., their z -components are small, from (211) and (212)

$$\theta \approx \angle \mathbf{b} - \frac{\pi}{2} \approx \angle \mathbf{p}, \quad (\text{E-8})$$

since by (211c) $\mathbf{b} \perp \mathbf{p}$. From (E-4)

$$\tan \theta \approx \frac{y''(z)}{x''(z)}. \quad (\text{E-9})$$

Thus,

$$\cos \theta \approx \frac{x''(z)}{\sqrt{x''^2(z) + y''^2(z)}} \approx \rho \cdot x''(z), \quad (\text{E-10})$$

$$\sin \theta \approx \frac{y''(z)}{\sqrt{x''^2(z) + y''^2(z)}} \approx \rho \cdot y''(z). \quad (\text{E-11})$$

From (E-10) and (E-11) we finally obtain the approximation of (216);

$$\frac{\cos \theta}{\rho} \approx x''(z), \quad (\text{E-12})$$

$$\frac{\sin \theta}{\rho} \approx y''(z). \quad (\text{E-13})$$

The above analysis has been a little crude; it is helpful to get a precise estimate on the error in the approximate result of (E-4) for the vector $\frac{1}{\rho} \cdot \mathbf{p}$, which in the above approximation is assumed to be purely transverse. We write

$$\mathbf{r}(z) = i x(z) + j y(z) + \mathbf{k} z \quad (\text{E-14})$$

as before. Next,

$$ds = \sqrt{dx^2 + dy^2 + dz^2}, \quad (\text{E-15})$$

$$\frac{ds}{dz} = \sqrt{1 + x'^2(z) + y'^2(z)}, \quad (\text{E-16})$$

where as before we reserve the prime to denote differentiation with respect to z . Then

$$\begin{aligned} \mathbf{t} &= \frac{d\mathbf{r}}{ds} = \frac{d\mathbf{r}}{dz} \frac{dz}{ds} \\ &= \frac{1}{\sqrt{1 + x'^2(z) + y'^2(z)}} [i x'(z) + j y'(z) + \mathbf{k}]. \end{aligned} \quad (\text{E-17})$$

Further

$$\begin{aligned} \frac{1}{\rho} \cdot \mathbf{p} &= \frac{d\mathbf{t}}{ds} = \frac{d\mathbf{t}}{dz} \frac{dz}{ds} = \frac{1}{1 + x'^2(z) + y'^2(z)} \cdot \left\{ i x''(z) + j y''(z) \right. \\ &\quad \left. - \frac{x'(z)x''(z) + y'(z)y''(z)}{1 + x'^2(z) + y'^2(z)} [i x'(z) + j y'(z) + \mathbf{k}] \right\}. \end{aligned} \quad (\text{E-18})$$

Now we may write (E-18) as follows:

$$\frac{1}{\rho} \cdot \mathbf{p} = i x''(z) + j y''(z) - \mathbf{A} - \mathbf{B}. \quad (\text{E-19})$$

$$\mathbf{A} = \frac{x'^2(z) + y'^2(z)}{1 + x'^2(z) + y'^2(z)} [i x''(z) + j y''(z)]. \quad (\text{E-20})$$

$$\mathbf{B} = \frac{x'(z)x''(z) + y'(z)y''(z)}{[1 + x'^2(z) + y'^2(z)]^2} [i x'(z) + j y'(z) + \mathbf{k}]. \quad (\text{E-21})$$

The first two terms of (E-19) are identical to the approximation of (E-4); the vectors \mathbf{A} and \mathbf{B} represent correction terms that we shall show to be small compared to the first two terms.

From (E-20)

$$\frac{|\mathbf{A}|}{\sqrt{x''^2(z) + y''^2(z)}} = \frac{x'^2(z) + y'^2(z)}{1 + x'^2(z) + y'^2(z)} \leq x'^2(z) + y'^2(z). \quad (\text{E-22})$$

From (E-21)

$$\begin{aligned} |\mathbf{B}| &= \frac{|x'(z)x''(z) + y'(z)y''(z)|}{[1 + x'^2(z) + y'^2(z)]^2} [1 + x'^2(z) + y'^2(z)]^{\frac{1}{2}} \\ &\leq |x'(z)x''(z) + y'(z)y''(z)| \\ &\leq \sqrt{x'^2(z) + y'^2(z)} \sqrt{x''^2(z) + y''^2(z)}, \end{aligned} \quad (\text{E-23})$$

where the last step follows from the Schwarz inequality. Then

$$\frac{|\mathbf{B}|}{\sqrt{x''^2(z) + y''^2(z)}} \leq \sqrt{x'^2(z) + y'^2(z)}. \quad (\text{E-24})$$

If (215) is satisfied, (E-22) and (E-24) show that the correction terms \mathbf{A} and \mathbf{B} of (E-19) are small compared to the first two terms, so that the approximation of (E-4) will be valid.

APPENDIX F

Rigorous Treatment of TE_{01} Loss Statistics for the Discrete Case

In treating the TE_{01} loss as a Fourier series with random coefficients, the frequency dependence of the $\Delta\alpha$'s and the C 's in (218) was neglected, since the principal frequency dependence occurs through the $\Delta\beta$'s. While this provides a simple and accurate analysis, a rigorous treatment of the TE_{01} loss statistics as a function of frequency is of interest.

We consider only the case of independent offsets or tilts, treated in Sections 3.3.1 and 3.3.2 respectively. From (218), (235), and (236)

$$A = \tilde{A} + \delta A \quad (\text{F-1})$$

$$\tilde{A} = \frac{1}{2}A_0 \quad (\text{F-2})$$

$$\delta A = \sum_{k=1}^{N-1} A_k \cos k\Delta\beta l_0 \quad (\text{F-3})$$

$$A_k = e^{k\Delta\alpha l_0} \sum_{i=1}^{N-k} (x_i^{\parallel} x_{i+k}^{\parallel} + x_i^{\perp} x_{i+k}^{\perp}) \quad (\text{F-4})$$

$$\langle A_0 \rangle = \hat{x}^2 N, \quad \langle A_0^2 \rangle = \hat{x}^4 N(N+1) \quad (\text{F-5a})$$

$$\langle A_k \rangle = 0, \quad \langle A_k^2 \rangle = \frac{\hat{x}^4}{2} (N-k) e^{k^2\Delta\alpha l_0}, \quad 1 \leq k \leq N-1. \quad (\text{F-5b})$$

$$\langle A_k A_l \rangle = 0, \quad k \neq l. \quad (\text{F-5c})$$

The new quantity \tilde{A} is defined by (F-2); δA remains the same as before.

We have for the expected values of the various losses:

$$\langle \tilde{A} \rangle = \frac{1}{2} \langle A_0 \rangle, \quad \langle \delta A \rangle = 0. \quad (\text{F-6})$$

Therefore

$$\langle A \rangle = \langle \tilde{A} \rangle = \frac{1}{2} \langle A_0 \rangle = \frac{\hat{x}^2}{2} N. \quad (\text{F-7})$$

Equation (F-7) gives the expected value of the TE_{01} loss as a function of frequency, since the rms conversion coefficient \hat{x} will in general vary with frequency (this variation will be small for offsets and diameter changes, approximately inversely with the free-space wavelength λ for tilts). We may now average over wavelength (indicated by a bar) and obtain instead of the result given in (237):

$$\langle \bar{A} \rangle = \overline{\langle A \rangle} = \overline{\frac{\hat{x}^2}{2}} N. \quad (\text{F-8})$$

Thus, the average of \hat{x}^2 over the band should be used instead of the value of \hat{x}^2 at the middle of the band. Over reasonable bandwidths the error will be small.

Next from (F-5c), \tilde{A} and δA are easily shown to be uncorrelated.

$$\langle \tilde{A}(\delta A) \rangle = 0. \quad (\text{F-9})$$

We next find the mean square value of δA . From (F-3), (F-5b), and (F-5c):

$$\begin{aligned} \langle (\delta A)^2 \rangle &= \sum_{k=1}^{N-1} \sum_{l=1}^{N-1} \langle A_k A_l \rangle \cos k\Delta\beta l_0 \cos l\Delta\beta l_0 \\ &= \sum_{k=1}^{N-1} \langle A_k^2 \rangle \cos^2 k\Delta\beta l_0 \\ \langle (\delta A)^2 \rangle &= \frac{\hat{x}^4}{2} \sum_{k=1}^{N-1} (N-k) e^{k2\Delta\alpha l_0} \cos^2 k\Delta\beta l_0. \end{aligned} \quad (\text{F-10})$$

The summation of (F-10) is easily written in closed form, but its general behavior is much more easily seen by examining the usual two special cases, small and large differential loss.

1. *Small differential loss over total length $L_N = Nl_0$.*

$$\langle (\delta A)^2 \rangle = \frac{\hat{x}^4}{8} \left[N(N-2) + \left(\frac{\sin N\Delta\beta l_0}{\sin \Delta\beta l_0} \right)^2 \right]; \quad -N2\Delta\alpha l_0 \ll 1. \quad (\text{F-11})$$

2. Large differential loss over total length L_N , small differential loss over pipe length l_0 .

$$\langle (\delta A)^2 \rangle = \frac{\hat{x}^4}{4} \frac{N}{-\Delta\alpha l_0} \frac{1 + \frac{1}{2} \left(\frac{\sin \Delta\beta l_0}{-\Delta\alpha l_0} \right)^2}{1 + \left(\frac{\sin \Delta\beta l_0}{-\Delta\alpha l_0} \right)^2}; \quad -N2\Delta\alpha l_0 \gg 1, \quad (\text{F-12})$$

$$-2\Delta\alpha l_0 \ll 1.$$

Sketches of the general behavior of $\langle (\delta A)^2 \rangle$ vs $\Delta\beta l_0$ (proportional to the free-space wavelength λ) are given in Fig. 11 for a single period, these functions being periodic of period π . In both cases $\langle (\delta A)^2 \rangle$ is almost constant except in narrow bands centered at $\Delta\beta l_0 = m\pi$, where it becomes twice as large; these peaks occur because we have assumed equally spaced mode converters. The half-width of these peaks initially decreases as $1.39/N$ as long as the differential loss remains small, approaching a limiting value of $(-\Delta\alpha l_0)$ when the differential loss becomes large. Since these peaks are narrow, they may be neglected in averaging over λ (or $\Delta\beta l_0$), yielding for small and large differential loss respectively:

$$\langle \overline{(\delta A)^2} \rangle = \overline{\langle (\delta A)^2 \rangle} = \frac{\hat{x}^4}{8} N(N-2); \quad -N2\Delta\alpha l_0 \ll 1 \quad (\text{F-13})$$

$$\langle \overline{(\delta A)^2} \rangle = \overline{\langle (\delta A)^2 \rangle} = \left(\frac{\hat{x}^4}{-\Delta\alpha} \right) \frac{N}{8l_0}; \quad -N2\Delta\alpha l_0 \gg 1, \quad (\text{F-14})$$

$$-2\Delta\alpha l_0 \ll 1.$$

Comparing with (241) and (242), we see that the average of \hat{x}^4 or of $\frac{\hat{x}^4}{-\Delta\alpha}$ over the band should be used rather than their values at the middle of the band. Again, over reasonable bandwidths the error will be small.

The minor difference between (F-13) and (241) for \hat{x}^4 independent of frequency — i.e., the factor $(N-2)$ instead of $(N-1)$ — arises because of our approximate integration of the function of Fig. 11(a), in which the narrow peaks and small ripples were ignored and the function set equal to $N(N-2)$. It is clear that this yields a result that is too small. An exact integration of (F-11) yields a result identical to (241).

The above calculations have considered only a single mode, the mode subscript being omitted as usual. For the total δA , including all first order spurious modes, we have

$$\delta A = \sum_{[m]} \delta A_{[m]}, \quad (\text{F-15})$$

$$\langle (\delta A)^2 \rangle = \sum_{[m]} \sum_{[n]} \langle \delta A_{[m]} \delta A_{[n]} \rangle. \quad (\text{F-16})$$

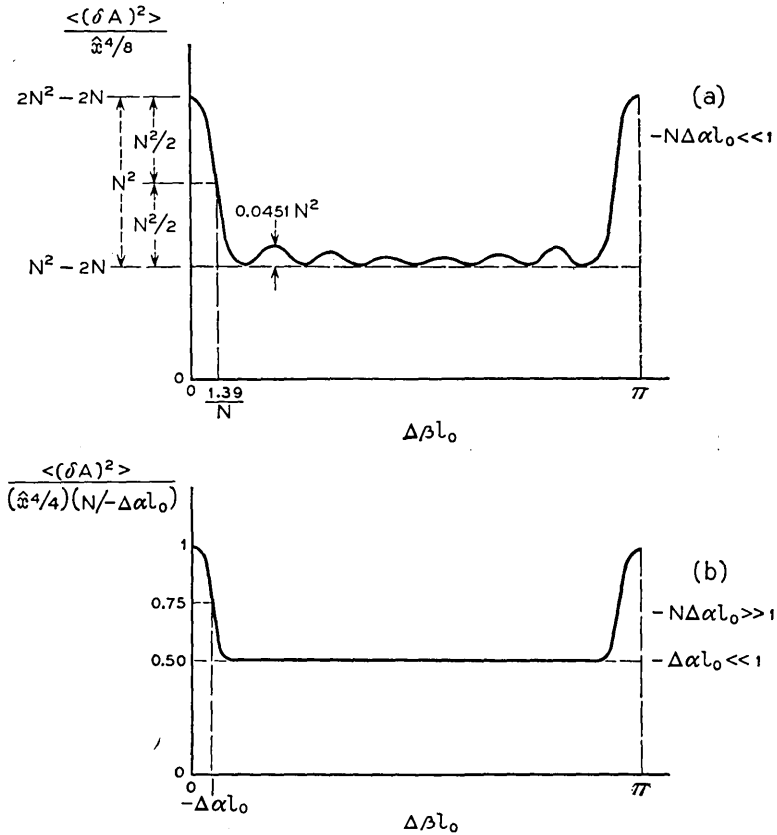


Fig. 11 — General behavior of $\langle(\delta A)^2\rangle$ vs $\Delta\beta l_0$.

In violation of our usual convention that [] indicate only TE spurious modes, in the present appendix this notation includes in addition the TM_{11}^+ spurious mode in the case of tilts.

The $\langle(\delta A_{[m]})^2\rangle$ have been discussed above; we need in addition the cross terms $\langle\delta A_{[m]}\delta A_{[n]}\rangle$. These cross terms did not appear in the approximate analysis, so we expect to find them negligible here. The different $\delta A_{[m]}$ are of course not independent. From (218), (221) and (173) we see that $\delta A_{[m]}(\lambda_1) \propto \delta A_{[n]}(\lambda_2)$ where $\lambda_1 \propto \lambda_2$; i.e., the TE_{01} loss component due to the m^{th} spurious mode at one frequency is proportional to the TE_{01} loss component due to the n^{th} spurious mode at a widely separated frequency. However, at the same frequency $\delta A_{[m]}$ and $\delta A_{[n]}$ are almost uncorrelated, so that the cross terms in (F-16) may be neglected.

We have

$$\langle \delta A_{[m]} \delta A_{[n]} \rangle = \sum_{k=1}^{N-1} \sum_{l=1}^{N-1} \langle A_{k[m]} A_{l[n]} \rangle \cos k \Delta \beta_{[m]} l_0 \cos l \Delta \beta_{[n]} l_0 \quad (\text{F-17})$$

where

$$\begin{aligned} \langle A_{k[m]} A_{k[n]} \rangle &= \frac{\hat{x}_{[m]}^2 \hat{x}_{[n]}^2}{2} (N - k) e^{k(\Delta \alpha_{[m]} + \Delta \alpha_{[n]}) l_0}, \\ \langle A_{k[m]} A_{l[n]} \rangle &= 0, \quad k \neq l. \end{aligned} \quad (\text{F-18})$$

Setting

$$\begin{aligned} \cos k \Delta \beta_{[m]} l_0 \cos k \Delta \beta_{[n]} l_0 &= \frac{1}{2} \cos k(\Delta \beta_{[m]} - \Delta \beta_{[n]}) l_0 \\ &\quad + \frac{1}{2} \cos k(\Delta \beta_{[m]} + \Delta \beta_{[n]}) l_0 \end{aligned}$$

(F-17) now becomes

$$\begin{aligned} \langle \delta A_{[m]} \delta A_{[n]} \rangle &= \frac{\hat{x}_{[m]}^2 \hat{x}_{[n]}^2}{4} \sum_{k=1}^{N-1} (N - k) e^{k(\Delta \alpha_{[m]} + \Delta \alpha_{[n]}) l_0} \\ &\quad \cdot [\cos k(\Delta \beta_{[m]} - \Delta \beta_{[n]}) l_0 + \cos k(\Delta \beta_{[m]} + \Delta \beta_{[n]}) l_0] \end{aligned} \quad (\text{F-19})$$

This summation is of the same general type as (F-10). In the special cases of small and large differential loss:

$$\begin{aligned} \langle \delta A_{[m]} \delta A_{[n]} \rangle &= \frac{\hat{x}_{[m]}^2 \hat{x}_{[n]}^2}{8} \left[-2N + \left(\frac{\sin \frac{1}{2} N (\Delta \beta_{[m]} - \Delta \beta_{[n]}) l_0}{\sin \frac{1}{2} (\Delta \beta_{[m]} - \Delta \beta_{[n]}) l_0} \right)^2 \right. \\ &\quad \left. + \left(\frac{\sin \frac{1}{2} N (\Delta \beta_{[m]} + \Delta \beta_{[n]}) l_0}{\sin \frac{1}{2} (\Delta \beta_{[m]} + \Delta \beta_{[n]}) l_0} \right)^2 \right]; \quad -N(\Delta \alpha_{[m]} + \Delta \alpha_{[n]}) l_0 \ll 1. \end{aligned} \quad (\text{F-20})$$

$$\begin{aligned} \langle \delta A_{[m]} \delta A_{[n]} \rangle &= \frac{\hat{x}_{[m]}^2 \hat{x}_{[n]}^2}{4} \frac{N}{-(\Delta \alpha_{[m]} + \Delta \alpha_{[n]}) l_0} \\ &\quad \cdot \left[\frac{1 - \frac{2}{-(\Delta \alpha_{[m]} + \Delta \alpha_{[n]}) l_0} [\sin \frac{1}{2} (\Delta \beta_{[m]} - \Delta \beta_{[n]}) l_0]^2}{1 + \left(\frac{2}{-(\Delta \alpha_{[m]} + \Delta \alpha_{[n]}) l_0} \right)^2 [\sin \frac{1}{2} (\Delta \beta_{[m]} - \Delta \beta_{[n]}) l_0]^2} \right. \\ &\quad \left. + \frac{1 - \frac{2}{-(\Delta \alpha_{[m]} + \Delta \alpha_{[n]}) l_0} [\sin \frac{1}{2} (\Delta \beta_{[m]} + \Delta \beta_{[n]}) l_0]^2}{1 + \left(\frac{2}{-(\Delta \alpha_{[m]} + \Delta \alpha_{[n]}) l_0} \right)^2 [\sin \frac{1}{2} (\Delta \beta_{[m]} + \Delta \beta_{[n]}) l_0]^2} \right]; \quad (\text{F-21}) \\ &\quad - N(\Delta \alpha_{[m]} + \Delta \alpha_{[n]}) l_0 \gg 1, \quad -(\Delta \alpha_{[m]} + \Delta \alpha_{[n]}) l_0 \ll 1. \end{aligned}$$

Equations (F-20) and (F-21) again exhibit narrow peaks of the same general type as illustrated in Fig. 11; away from these peaks the functions will be quite small. Thus if $(\Delta\beta_{[m]} \mp \Delta\beta_{[n]})l_0 \approx m\pi$, from (F-20) and (F-21) for either small or large differential loss:

$$\langle \delta A_{[m]} \delta A_{[n]} \rangle = -\frac{\hat{x}_{[m]}^2 \hat{x}_{[n]}^2}{4} N; \quad -(\Delta\alpha_{[m]} + \Delta\alpha_{[n]})l_0 \ll 1, \quad (\text{F-22})$$

$$(\Delta\beta_{[m]} \mp \Delta\beta_{[n]})l_0 \approx m\pi.$$

From (F-22) the correlation coefficient of $\delta A_{[m]}$ and $\delta A_{[n]}$ will be small for moderately large values of N . Therefore, for those modes that make a significant contribution to the total $\langle (\delta A)^2 \rangle$, the cross terms in (F-16) will be negligible.

APPENDIX G

Correlation Coefficient of TE₀₁ Loss Components due to Different Spurious Modes for the Continuous Case

Consider the ac components $\delta A_{[m]}$ and $\delta A_{[n]}$ of the total TE₀₁ loss, due to two different spurious modes, each with two polarizations, generated by the same type of geometric imperfection. The geometric imperfection and thus the coupling coefficients are assumed to have white power spectra.

As a specific example, consider random deviations of the guide axis from perfect straightness, which generate principally the forward TE₁₂ and TE₁₁ spurious modes. The coupling coefficients to the two polarizations of the TE_{1m} mode are given in (308) in terms of the second derivatives of the rectangular co-ordinates of the guide axis, $x''(z)$ and $y''(z)$. We assume that $x''(z)$ and $y''(z)$ have white power spectra.

We expand the geometric imperfection to which the coupling coefficients are proportional in a Fourier series. For random straightness deviations:

$$x''(z) = \sum_{n=-\infty}^{\infty} \gamma_n^{\parallel} e^{j2\pi n z/L};$$

$$y''(z) = \sum_{n=-\infty}^{\infty} \gamma_n^{\perp} e^{j2\pi n z/L}; \quad (\text{G-1})$$

$$\gamma_n^{\parallel} = \alpha_n^{\parallel} + j\beta_n^{\parallel}, \quad \gamma_n^{\perp} = \alpha_n^{\perp} + j\beta_n^{\perp}.$$

The complex Fourier coefficients c_n in Section 4.1 are simply proportional to the corresponding γ_n . For straightness deviations, for the TE_{1m} mode

we have:

$$\begin{aligned} c_{n[m]}^{\parallel} &= C_{t[m]} \gamma_n^{\parallel}, & \text{TE}_{1m}^{\parallel}; \\ c_{n[m]}^{\perp} &= C_{t[m]} \gamma_n^{\perp}, & \text{TE}_{1m}^{\perp}. \end{aligned} \tag{G-2}$$

Consequently, the γ_n 's have the same statistical properties as the c_n 's, given in Section 4.2. Since $x''(z)$ and $y''(z)$ have white power spectra,

$$\langle |\gamma_n^{\parallel}|^2 \rangle = \langle |\gamma_n^{\perp}|^2 \rangle = \hat{\gamma}^2. \tag{G-3}$$

Further, the different γ 's are strictly independent.

We next define for convenience the following quantities:

$$\begin{bmatrix} g_{\alpha}^{\parallel}(t) \\ g_{\beta}^{\parallel}(t) \\ g_{\alpha}^{\perp}(t) \\ g_{\beta}^{\perp}(t) \end{bmatrix} = \sum_{n=-\infty}^{\infty} \begin{bmatrix} \alpha_n^{\parallel} \\ \beta_n^{\parallel} \\ \alpha_n^{\perp} \\ \beta_n^{\perp} \end{bmatrix} (-1)^n \frac{\sin \pi(t-n)}{\pi(t-n)}, \quad t \equiv \frac{\Delta\beta L}{2\pi} = \frac{L}{B}. \tag{G-4}$$

These quantities are proportional to the real or imaginary parts of I in (285b). These four quantities are approximately independent stationary band-limited Gaussian random processes in the practical case where the length L is large compared to the beat wavelength B , so that $|t| \gg 1$. The autocorrelation function of each of these quantities, found directly from (G-4), is therefore

$$r(\tau) \equiv \langle g_{\alpha,\beta}^{\parallel,\perp}(t) g_{\alpha,\beta}^{\parallel,\perp}(t+\tau) \rangle = \frac{1}{2} \hat{\gamma}^2 \frac{\sin \pi\tau}{\pi\tau}. \tag{G-5}$$

Next we define for convenience the quantity $g^2(t)$ as

$$g^2(t) = g_{\alpha}^{\parallel 2}(t) + g_{\beta}^{\parallel 2}(t) + g_{\alpha}^{\perp 2}(t) + g_{\beta}^{\perp 2}(t). \tag{G-6}$$

The autocorrelation function of $g^2(t)$ is easily found in terms of the autocorrelation function of the individual g 's, given in (G-5). If x is a stationary Gaussian random process with autocorrelation $R_x(\tau)$ and $y = x^2$, then the autocorrelation of y , $R_y(\tau)$, is given by^{37, 40}

$$R_y(\tau) = R_x^2(0) + 2R_x^2(\tau), \tag{G-7}$$

where the first term corresponds to the dc component, the second to the ac component of y . Since the individual quantities whose squares appear on the right-hand side of (G-6) are independent random variables, we have from (G-7) and (G-5) for the autocorrelation function $R(\tau)$ of $g^2(t)$

$$R(\tau) \equiv \langle g^2(t)g^2(t + \tau) \rangle = 4\gamma^4 + 2\gamma^4 \left(\frac{\sin \pi\tau}{\pi\tau} \right)^2, \quad (\text{G-8})$$

where again the first term corresponds to the dc component, the second to the ac component of $g^2(t)$.

From the results of Section 4.1 we may now write the TE_{01} loss due to the two polarizations of the m^{th} spurious mode (TE_{1m} for straightness deviations) in terms of the function $g^2(t)$ defined in (G-6):

$$A_{[m]} = \frac{C_{t_{[m]}}^2 \cdot L^2}{2} g^2(t_{[m]}), \quad t_{[m]} = \frac{|\Delta\beta_{[m]}| L}{2\pi}. \quad (\text{G-9})$$

Equation (G-9) is appropriate for our present purposes because it places in evidence the relation between the different TE_{01} loss components $A_{[m]}$. From (G-9), (G-8), and the results of Section 4.1, the normalized correlation coefficient ρ_{mn} of $\delta A_{[m]}$ and $\delta A_{[n]}$, the ac components of TE_{01} loss due to the m^{th} and n^{th} spurious modes, may be written in the following form:

$$\rho_{mn} \equiv \frac{\langle \delta A_{[m]} \delta A_{[n]} \rangle}{\sqrt{\langle (\delta A_{[m]})^2 \rangle} \sqrt{\langle (\delta A_{[n]})^2 \rangle}} = \left(\frac{\sin \pi\tau_{mn}}{\pi\tau_{mn}} \right)^2, \quad (\text{G-10})$$

$$\tau_{mn} = \frac{(|\Delta\beta_{[m]}| - |\Delta\beta_{[n]}|)L}{2\pi}.$$

Thus,

$$\rho_{mn} < \left(\frac{1}{\pi\tau_{mn}} \right)^2. \quad (\text{G-11})$$

Since the different spurious modes have substantially different beat wavelengths (Appendix D), $\rho_{mn} \ll 1$ for moderate values of length L . As a numerical example consider TE_{12}^+ and TE_{11}^+ , for a total length between mode filters $L = 200$ feet. From (C-11), $\rho < 0.00036$. In practical cases the TE_{01} loss contributions of the different spurious modes will be almost uncorrelated, so that $\langle (\delta A)^2 \rangle$ will be given by (322).

Contributors to This Issue

HAROLD E. CURTIS, B.S. and M.S., 1929, Massachusetts Institute of Technology; A.T.&T. Co., 1929-34; Bell Telephone Laboratories, 1934—. He has specialized in work on transmission problems related to multi-channel carrier telephony, including microwave radio relay, coaxial cable and waveguide systems.

JAMES L. FLANAGAN, B.S., 1948, Mississippi State University; S.M., 1950 and Sc.D., 1955, Massachusetts Institute of Technology; faculty, Mississippi State University, 1950-52; Rockefeller Foundation Fellow, 1952-53; Air Force Cambridge Research Center, 1954-57; Bell Telephone Laboratories, 1957—. He has specialized in work on speech communication over narrow bandwidths, including studies of acoustical, physiological and psychological phenomena related to speech and speech perception. Since March 1961, he has headed the Speech and Auditory Research Department. Fellow Acoustical Society of America; member I.R.E., Kappa Mu Epsilon, Sigma Xi, Tau Beta Pi.

H. L. FRISCH, B.A., 1947, Williams College; Ph.D., 1952, Polytechnic Institute of Brooklyn; faculty, Syracuse University, 1952-54; faculty, University of Southern California, 1954-56; Bell Telephone Laboratories, 1956—. At the Laboratories, he has engaged in research in statistical mechanics. Member American Chemical Society, American Association for the Advancement of Science, American Physical Society, Sigma Xi.

S. B. GORDON, B.A., 1957, Vassar College; Bell Telephone Laboratories, 1957-61. Mrs. Gordon first did programming for IBM 704 and 7090 computers, in connection with the Laboratories' evaluation of the Air Force Air Defense System (SAGE). More recently she was concerned with the basic IBM 7090 computer program, and wrote the percolation program for this computer.

GEORGE B. GUCKER, B.E.E., 1945, Cooper Union; New York Telephone Company, 1925-26; Bell Telephone Laboratories, 1926—. At the Laboratories, Mr. Gucker has been primarily concerned with the de-

velopment of electron tubes. He presently supervises a group engaged in this work. Member Tau Beta Pi.

J. M. HAMMERSLEY, B.A., 1947, and M.A., 1949, Cambridge University; M.A., 1949, Oxford University; D.Sc., 1959, Oxford University; Sc.D., 1959, Cambridge University; Fellow of Trinity College, Oxford, 1961—. Dr. Hammersley has been a consulting physicist at Bell Telephone Laboratories on two occasions. He is Senior Research Officer, Institute of Statistics at Oxford University, and a consultant to the United Kingdom Atomic Energy Authority at Harwell, England. Fellow, Cambridge Philosophical Society, Institute of Mathematical Statistics, Royal Statistical Society; member Mathematical Association, International Statistical Institute.

RICHARD LINDNER, B.M.E., 1953, Cooper Union; M.S., 1954, Massachusetts Institute of Technology; Bell Telephone Laboratories, 1954—. As a member of the switching apparatus group he first was engaged in contact studies. Since 1956, he has been engaged in basic surface studies of semiconductors in the transistor development department. Member I.R.E., Tau Beta Pi, Pi Tau Sigma.

MARVIN C. PAULL, B.E.E., 1952, Clarkson College of Technology; Bell Telephone Laboratories, 1953—. At the Laboratories his work has included studies in the use of magnetic cores in computer design, research and development in the logical design of computer circuitry and computer programs. Member I.R.E., Eta Kappa Nu, Tau Beta Pi.

TIBOR RADO, Ph.D., 1922, University of Szeged (Hungary); Professor, Ohio State University, 1930—. Dr. Rado, Research Professor of Mathematics at Ohio State University, is the author of 5 books and more than 100 research papers on pure and applied mathematics. He was a visiting lecturer at the Laboratories in 1961. Member American Mathematical Society, Mathematical Association of America, American Association for the Advancement of Science, Sigma Xi, Pi Mu Epsilon.

HARRISON E. ROWE, B.S., 1948, M.S., 1950, and Sc.D., 1952, M.I.T.; Bell Telephone Laboratories, 1952 —. He was initially associated with a group engaged in systems research. He later worked on mode conversion problems arising in multimode waveguides. Presently he is concerned with problems relating to optical masers. Member I.R.E., Sigma Xi, Tau Beta Pi, Eta Kappa Nu.

IRWIN W. SANDBERG, B.E.E., 1955, M.E.E., 1956, and D.E.E., 1958, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1958—. He has been concerned with analysis of military systems, particularly radar systems, and with synthesis and analysis of active and time-varying networks. Member I.R.E., Eta Kappa Nu, Sigma Xi, Tau Beta Pi.

A. J. SCHEPIS, B.M.E., 1952, City College of New York; M.M.E., 1960, New York University; Western Electric Co., 1956–57; Bell Telephone Laboratories, 1957—. At the Laboratories he was first responsible for guidance antenna design, waveguide design and liaison with aircraft companies for Laboratories' equipment installation. He later engaged in mathematical analysis of heat, vibration and stress. Recently he has been concerned with determining orbital elements from doppler frequency shifts obtained from a single satellite pass. Member American Rocket Society, American Society of Mechanical Engineers.

V. A. VYSSOTSKY, B.A., 1950, and M.S., 1956, University of Chicago; Bell Telephone Laboratories, 1956–61. At the Laboratories he engaged in advanced computer programming and research on speech analysis, using a digital computer. He was also concerned with automatic coding.

WILLIAM D. WARTERS, A.B., Harvard College, 1949; M.S., 1950, and Ph.D., 1953, California Institute of Technology; Bell Telephone Laboratories, 1953—. Since joining the Laboratories, Mr. Warters has engaged in studies of circular waveguides for long distance, broadband transmission. Currently, as head of the repeater research department, he is working on high-speed PCM repeaters for use in a waveguide transmission system. Member I.R.E., American Physical Society, Sigma Xi.

