The Bell System Technical Journal

Vol. XXVIII *January, 1949* *No. 1*

Propagation of TE_{01} Waves in Curved Wave Guides

By W. J. ALBERSHEIM

TE_{01} waves transmitted through curve wave guides lose power by conversion to other modes, especially to TM_{11}.

This power transfer to coupled modes is explained by the theory of coupled transmission lines. It is shown that the power interchange between coupled lines and their propagation constants can be derived from a single coupling discriminant.

Earlier calculations of TE_{01} conversion loss in circular wave guide bends are confirmed and extended to S-shaped bends.

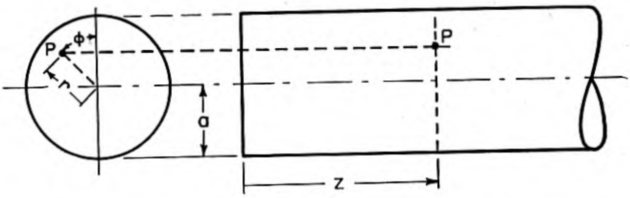
Tolerance limits for random deflections from an average straight course are given.

THE TE_{01} mode of propagation in circular wave guides has great potential value for the transmission of wide-band signals because its attenuation decreases with frequency. In order to take full advantage of this property one must use sufficiently large wave guides to operate well above the cutoff of the lowest transmitted frequency. The difficulty of this transmission method lies in the fact that TE_{01} is not the dominant mode and that energy may be lost by transfer to the many other modes capable of transmission in the wave guide. In an ideal wave guide, which is perfectly straight, perfectly circular and perfectly conducting, the propagation is undisturbed; but slight imperfections and especially a slight curvature of the wave guide axis may produce serious disturbances.

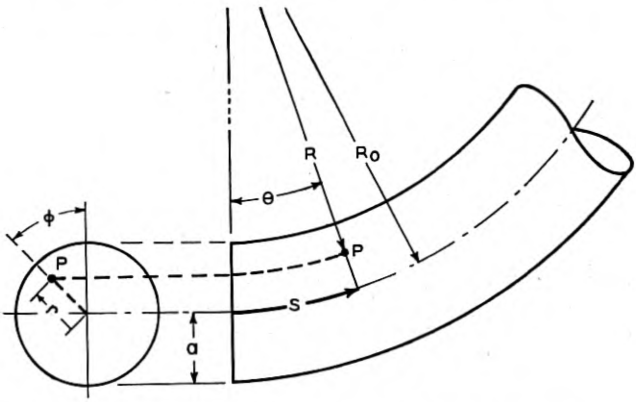
The character of these disturbances has been investigated in several publications by Prof. M. Jouguet[1] and in unpublished work by Mr. S. O. Rice of the Bell Telephone Laboratories. Both Jouguet and Rice use the method of perturbations, which is a form of calculus invented by astronomers to compute the deviations from the exact elliptical orbits of the planets which are caused by the disturbing influences of their fellow planets. Although the above-mentioned authors obtained valuable results, the interpretation of their solutions is difficult due to this rather abstract mathematical formulation. To most engineers the understanding of a physical problem is greatly helped if it is possible to use a method of analysis which is elementary in character and easily interpreted in familiar physical terms. The familiar concept on which the present treatment will be based is that of coupled circuits.

[1] See References 2 and 3, listed on page 7.

It has been stated by the earlier authors that the curvature of the wave guide produces a coupling between modes. Before going into a detailed analysis one may estimate by inspection the nature of this coupling and the kind of modes that are most strongly coupled to each other. Figure 1a shows the cross section and the longitudinal section of a straight cylindrical wave guide. The location of every point inside the wave guide is determined by three coordinates: the radial distance r from the cylinder axis; the azimuth angle φ from an arbitrary 0 line and the axial distance z from the



a - CYLINDRICAL CO-ORDINATES IN STRAIGHT WAVEGUIDE



b - TOROIDAL CO-ORDINATES IN CURVED WAVEGUIDE

Fig. 1

origin. If the wave guide is bent as shown on Fig. 1b, but a wave front at right angles to the cylinder axis is to be maintained, the waves must be shortened at the inside of the bend and lengthened at the outside of the bend. Regarding compression as a positive and expansion as a negative deformation, one sees that the distortion of the wave shape is proportional to the curvature of the wave guide multiplied by the cosine of the azimuth angle. It is natural to assume that the coupling between modes is proportional to this distortion.

Now it is known that all modes of propagation in a circular wave guide can be derived from functions $J_n(\chi r)$ cos $n\varphi$. In these functions, n is called the aximuthal index because it indicates the type of symmetry around the circumference of the wave guide. When these characteristic functions are multiplied by the distortion factor cosine φ, the resulting expressions are proportional to the sum of cosine $(n + 1)\,\varphi$ and cosine $(n - 1)\,\varphi$. This means that the bending of the wave guide couples mainly those modes which differ by ± 1 in azimuth index. Since the TE_{01} mode has the azimuthal index 0, it is coupled to all modes of the type TE_{1m} and TM_{1m}.

In the above qualitative discussion we have claimed that coupling exists without defining the physical coupling parameters and their effects. We must now supply this definition and show that the TE_{01} mode is particularly susceptible to coupling losses.

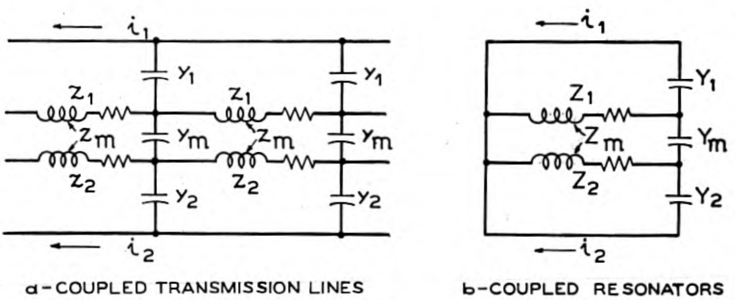


Fig. 2

Our investigation is guided by S. A. Schelkunoff's statement[2] that a wave guide mode has the same equation of propagation as a high-pass transmission line. Schelkunoff further points out[3] that the high-pass character of circular wave guide modes can be interpreted as the effect of interfering plane waves whose directions of propagation deviate from the wave guide axis by a constant slanting angle.

We therefore approach the problem of coupled wave guide modes by studying the behavior of two coupled transmission lines such as shown on Fig. 2a. Each transmission line is schematically shown as an array of small ladder sections. The series impedances per unit length of the lines are z_1 and z_2; their shunt admittances per unit length, y_1 and y_2. The two lines are loosely coupled by small mutual series impedances per unit length (z_m) and by small mutual shunt admittances per unit length (y_m).

A network of coupled ladder sections is more tractable than a wave guide structure, but still somewhat complicated. Let us therefore carry the analogy one step further. Figure 2b shows two resonant circuits, each

[2] Ref. 4, pp. 378 and 381 of the book.
[3] Ref. 4, p. 410 of the book.

consisting of single capacity C, and an impedance Z which includes an inductance L and a damping resistance R. The resonators are coupled by a small mutual inductance Z_m and by a small mutual capacity Y_m.

The behavior of coupled resonators is very well known to radio engineers. They occur as tuned transformers in amplifier circuits, as band-pass filters and as "tank circuits" in radio transmitters. Even before the advent of radio, their acoustical equivalents were studied in the form of resonant tuning forks. The mathematical aspects of this problem were already clearly set forth in a paper by Wien written in 1897[4]. He showed that the interaction between the free vibrations of two tuned circuits depends on the coupling coefficient and on the ratio of their complex resonance frequencies. The closer the two frequencies are to each other, the less coupling is needed to transfer energy between the two circuits. The reason is that the individual free vibrations of two nearly synchronous circuits remain in step long enough to accumulate the small energy transfer impulses of many vibrations.

Now consider the two transmission lines of Fig. 2a and assume that a constant frequency signal is impressed upon the input of one or both of them. The signals are carried along the two lines as traveling waves. Again it is true that loosely coupled signals affect each other strongly if they remain in step. With traveling waves "remaining in step" means that they must travel with approximately equal phase velocities. We conclude that the phase velocities or phase constants of coupled transmission lines play a similar role as the resonant frequencies of coupled tuned circuits. This intuitive reasoning is confirmed by analysis (see Section 1 of the analytical part of this paper).

We thus find that we must expect trouble for TE_{01} wave guide transmission if a mode with an azimuth index 1 has a propagation constant close to that of the TE_{01}. It so happens that there exists one mode, the TM_{11}, which in an ideal wave guide has exactly the same propagation constant as the TE_{01}. This then should be the principal source of trouble—and from previous work it is known that such is the case.

Our discussion of coupled transmission lines has shown that the interaction effects are functions of their relative uncoupled propagation constants and of the coupling coefficient. The propagation constants of the TE_{01} and TM_{11} wave guide modes are known but their coupling coefficient remains to be found.

Since the energy of the transmission modes is located in the dielectric inside the wave guide, we consider first the coupling between the plane "slant wave" groups from which the modes are built up.

[4] Reference 5.

As shown in the analytical part, the coupling coefficient of these slant waves may be defined as the energy interchanged between the modes per unit length of line divided by the geometric mean of the energies per unit length stored up in each of the modes.

From the coupling coefficient of the slant waves the coupling coefficient of the wave guide modes is derived.

On the basis of the above physical interpretation the analysis is carried out and the properties of TE_{01} propagation through curved wave guides of various shapes are derived in the analytical part of this paper which is subdivided into the following nine sections:

Section 1 develops an approximate theory of loosely coupled, weakly damped circuits. The theory is first derived for coupled resonators which are familiar to communication engineers, and then applied in similar form to coupled transmission lines. It is shown that the important interaction properties of coupled lines are functions of a single coupling discriminant. The relative energy content of the two lines in each of the two possible coupled modes is plotted as a function of the coupling discriminant.

Section 2 contains the field equations of a straight circular wave guide and their modification by a toroidal bend.

Section 3 gives the solutions of the field equations for the uncoupled TE_{01} and TM_{11} modes in wave guides with infinite, and with small but finite conductivity.

Section 4 applies the coupling theory to the TE_{01} and TM_{11} modes in circular wave guide bends. The coupling coefficient, coupling discriminant and energy division between the two modes are derived as functions of the wave guide diameter bending radius and conductivity and of the signal frequency.

Section 5 derives the critical bending radius and the attenuation of TE_{01} waves in long wave guides of constant curvature. Two numerical examples are given.

Section 6 shows that in a curved section of wave guide which follows a long straight section or other source of pure TE_{01} the energy fluctuates back and forth between a condition of pure TE_{01} and of predominant TM_{11}. The length and magnitude of the fluctuations are derived.

Section 7 computes the increase in average attenuation caused by serpentine bends of regular shapes. Numerical examples are tabulated.

Section 8 shows that the results of Section 7 can be applied to helical bends and to small two-dimensional random deviations from a straight course.

Section 9 shows that for any given statistical distribution of random angular deviations the average attenuation is minimized by an optimum wave guide radius for each signal wave length and by an optimum signal wave length for each wave guide radius.

Numerical examples are given for sinusoidal bends.

Summary of Results

1. The energy loss of TE_{01} waves in curved wave guides by conversion into the TM_{11} mode is interpreted as a case of coupling between resonant transmission lines.
2. In a pair of coupled lines the energy cannot be confined entirely to a single line but travels through both in one or both of two possible combination modes.
3. All important properties of coupled circuits, including wave guide modes, are functions of a single discriminant.
4. When the discriminant is much smaller than one, most of the energy can be carried in one line or component mode.
5. When the discriminant is much larger than one, the energy flow is nearly equally divided between the two lines or component modes.
6. In wave guides of typical dimensions the coupling discriminant becomes one for a "critical" bending radius greater than a mile. For all sharper bends, that is for most practical installations, the discriminant is greater than one.
7. In a long wave guide section with more than critical curvature the average attenuation constant is the arithmetic mean between those of the TE_{01} and the TM_{11} modes.
8. If a wave guide region carrying pure TE_{01} is followed by a curved region, the energy in the curved region fluctuates back and forth between pure TE_{01} and predominant TM_{11}. The location of TE_{01} minima and maxima is a function of the signal frequency, the wave guide diameter and the total bending angle.
9. For highly supercritical curvatures the bending angles at which minima and maxima occur are nearly independent of the curvature and approach the limiting values previously computed by Jouguet and Rice. The minima approach zero. When the bending radius approaches or exceeds the critical value, the maxima and minima become shallower and their spacing is increased by a function of the coupling discriminant.
10. For regular serpentine bends or random angular deviations from an average straight course which are much smaller than the first extinction angle, the percentage increase in average attenuation is proportional to the square of the maximum deviation and to the fourth power of wave guide diameter and signal frequency.
11. Wave guide installations of practical dimensions for frequencies now

attainable are tolerant to random angular deviations of the order of 1 degree.

12. For any expected distribution of random angular deviations there exists an optimum wave guide radius for each signal wave length and an optimum signal wave length for each wave guide radius, which minimize the average attenuation.

REFERENCES

1. Jahnke & Emde, Tables of Functions, *Dover Publications*, New York, 1943.
2. M. Jouguet, Propagation dans les tujaux courbés, *Comptes Rendus—Académie des Sciences, Paris*, Feb. 18, 1946, March 4, 1946 and Jan. 6, 1947.
3. M. Jouguet, Effets de la courbure dans un guide à section circulaire, *Cables & Transmission*, 1 No. 2, July 1947, pp. 133–153.
4. S. A. Schelkunoff, Electromagnetic Waves, D. Van Nostrand Company, Inc., New York, 1943.
5. M. Wien, Ueber die Rueckwirkung eines resonierenden Systems, *Ann. d. Physik*, 1897, Vol. 61, pp. 151–189.

ANALYSIS

1. Interaction of Coupled Circuits

1.1 *Free Oscillations of Coupled Resonators (Fig. 1B)*

The circuits are coupled according to the following four equations:

$$e_1 = -Z_1 i_1 + Z_m i_2 \qquad 1.1\text{-}1$$

$$i_1 = Y_1 e_1 + Y_m e_2 \qquad 1.1\text{-}2$$

$$e_2 = -Z_2 i_2 + Z_m i_1 \qquad 1.1\text{-}3$$

$$i_2 = Y_2 e_2 + Y_m e_1 \qquad 1.1\text{-}4$$

where index $_1$ refers to circuit 1, index $_2$ to circuit 2 and index $_m$ to the mutual coupling impedance and admittance. The coupled oscillations have the solution:

$$e_1 = E_{1a}\epsilon^{p_a t} + E_{1b}\epsilon^{p_b t} \qquad 1.1\text{-}5$$

$$e_2 = E_{2a}\epsilon^{p_a t} + E_{2b}\epsilon^{p_b t} \qquad 1.1\text{-}6$$

In the limiting case of zero coupling ($Y_m = 0, Z_m = 0$) the obvious solution shows independent oscillations in the two separate circuits:

$$e_{10} = K_1 i_{10} = E_{10}\epsilon^{p_1 t} \qquad 1.1\text{-}7$$

$$e_{20} = K_2 i_{20} = E_{20}\epsilon^{p_2 t} \qquad 1.1\text{-}8$$

The wave impedance K_1 of the primary circuit is found by dividing equation 1.1-1 by 1.1-2

$$K_1 = \frac{e_1}{i_1} = \sqrt{\frac{Z_1}{Y_1}}$$

Similarly,

$$K_2 = \frac{e_2}{i_2} = \sqrt{\frac{-Z_2}{Y_2}}$$

By multiplying equation 1.1-1 by 1.1-2 one finds

$$-Z_1 Y_1 = 1 \qquad 1.1\text{-}9$$

from which one can compute the exponent p_1. In the specific circuits shown in Fig. 1b

$$Z_1 = L_1 p_1 + R_1 \quad \text{and} \qquad 1.1\text{-}10$$

$$Y_1 = C_1 p_1 \qquad 1.1\text{-}11$$

From 1.1-9, 10 and 11

$$p_1 = -\delta_1 + j\omega_1 = -\frac{R_1}{2L_1} + j\sqrt{\frac{1}{L_1C_1} - \frac{R_1^2}{4L_1^2}} \qquad 1.1\text{-}12$$

and by analogy

$$p_2 = -\delta_2 + j\omega_2 = -\frac{R_2}{2L_2} + j\sqrt{\frac{1}{L_2C_2} - \frac{R_2^2}{4L_2^2}} \qquad 1.1\text{-}13$$

In equations 1.1-7 and 1.1-8, E_{10} and E_{20} are amplitude constants determined by boundary conditions. In equations 1.1-12 and 1.1-13, δ_1 and δ_2 are the decay or damping constants, ω_1 and ω_2 the radian frequencies.

With finite but loose coupling and small damping the circuits can oscillate with either or both of the two frequencies.

$$p_a = \frac{p_1 + p_2}{2} + \frac{p_1 - p_2}{2}\sqrt{1 + \kappa^2} = p_1 + 0.5\, p_2(1 - \sqrt{1 + \kappa^2}) \qquad 1.1\text{-}14$$

$$p_b = \frac{p_1 - p_2}{2} - \frac{p_1 - p_2}{2}\sqrt{1 + \kappa^2} = p_2 + 0.5\, p_1(1 - \sqrt{1 + \kappa^2}) \qquad 1.1\text{-}15$$

In the last two equations, the symbol κ, defined by $\kappa = \sqrt{\frac{p_1 p_2}{p_1 - p_2}} \cdot k$, may be called the coupling discriminant. The first term of the product on the right side of this expression is the reciprocal of the fractional difference between the uncoupled frequencies; the second term k is the "coupling coefficient."

When there is only one coupling impedance, the coupling coefficient is usually defined as the mutual circuit impedance divided by the geometric mean of the separate circuit impedances. A broader definition which applies to all combinations of mutual impedances and admittances is

$$k = \frac{P_{12}}{\sqrt{P_1 P_2}} = \frac{P_{21}}{\sqrt{P_1 P_2}} \qquad 1.1\text{-}16$$

In this equation P_1 is the energy stored in circuit 1, P_2 the energy stored in circuit 2 and P_{12} is the energy transferred from one circuit to the other. One finds

$$P_1 = \frac{e_1^2}{2K} + \frac{i_1^2 K}{2} = \frac{e_1^2}{K_1} = i_1^2 K_1 \qquad 1.1\text{-}17$$

$$P_2 = \frac{e_2^2}{K_2} = i_2^2 K_2 \qquad 1.1\text{-}18$$

$$P_{12} = \frac{e_{12}\, e_2}{K_2} = i_{12}\, i_2\, K_2 = \frac{e_{21}\, e_1}{K_1} + i_{21}\, i_1\, K_1 \qquad 1.1\text{-}19$$

Equations 1.1-5 and 1.1-6 contain four amplitude constants. Two of these, for instance E_{1a} and E_{2b}, can be adjusted to satisfy boundary conditions. The other two are fixed by the equation

$$\frac{E_{2a}^2 K_1}{E_{1a}^2 K_2} = \frac{p_a - p_1}{p_a - p_2} = \frac{p_b - p_2}{p_b - p_1} = \frac{E_{1b}^2 K_2}{E_{2b}^2 K_1}$$

The square root of this expression,

$$\frac{E_{2a}}{E_{1a}} \sqrt{\frac{K_1}{K_2}} = \frac{E_{1b}}{E_{2b}} \sqrt{\frac{K_2}{K_1}} = A$$

may be called the *normalized amplitude ratio.* It is a vector quantity denoting the amplitude ratio and phase relation of each oscillation frequency in the two circuits, assuming that they have been normalized to equal resistances by an ideal transformer. The absolute value

$$\left| \frac{E_{2a}^2 K_1}{E_{1a}^2 K_2} \right| = W_a = \frac{1}{W_b}$$

is the ratio of the energies stored in the two circuits oscillating at frequencies p_a and p_b respectively.

From 1.1-14, 15, 16 and 18

$$W_a = \left| \frac{\sqrt{1 + \kappa^2} - 1}{\sqrt{1 + \kappa^2} + 1} \right| = W$$

$$A_a = \sqrt{1 + \kappa^{-2}} - \kappa^{-1} = A$$

When the indexes are left off, $W < 1$ and $| A | < 1$ by definition. One sees that energy, amplitude and phase relations between the coupled circuits at each oscillating frequency are governed by the coupling discriminant. This also applies to the damping coefficients and frequencies of the coupled oscillations. It can be shown by combining and transforming equations 1.1-14, 15, 19, that the coupled damping coefficients are

$$\delta_a = \frac{\delta_1 + \delta_2 W}{1 + W} = \delta_1 \frac{W_{1a}}{W_{\text{total}}} + \delta_2 \frac{W_{2a}}{W_{\text{total}}}$$

$$\delta_b = \frac{\delta_1 W + \delta_2}{1 + W} = \frac{\delta_1 W_{1b}}{W_{\text{total}}} + \frac{\delta_2 W_{2b}}{W_{\text{total}}}$$

The damping constants of two coupled resonances are found by combining the uncoupled damping constants in the same proportion as the energies oscillating in the two resonators.

The coupled frequencies are

$$\omega_a = \frac{\omega_1 - W_{\omega 2}}{1 - W} \text{ and}$$

$$\omega_b = \frac{\omega_2 - W_{\omega 1}}{1 - W}$$

1.2 *Forced traveling waves in coupled transmission lines* (*Fig. 1A*).

The two lines are coupled according to the four equations

$$\Gamma e_1 = z_1 i_1 + z_m i_2$$

$$\Gamma i_1 = y_1 e_1 + y_m e_2$$

$$\Gamma e_2 = z_2 i_2 + z_m i_1$$

$$\Gamma i_2 = y_2 e_2 + y_m e_1$$

which may be compared to the corresponding equations of section 1.1. There is a dimensional difference because in transmission lines the series impedances z are measured in ohm/meter and the shunt reactances y in mho/meter. Γ is the propagation constant of the wave traveling in the $+s$ direction. If a sinusoidal signal with the radian frequency ω is impressed upon the input of the lines, the coupled waves have the solution

$$e_1 = E_{1a} \epsilon^{j\omega t - \Gamma_a s} + E_{1b} \epsilon^{j\omega t - \Gamma_b s} \quad 1.2\text{-}1$$

$$e_2 = E_{2a} \epsilon^{j\omega t - \Gamma_a s} + E_{2b} \epsilon^{j\omega t - \Gamma_b s} \quad 1.2\text{-}2$$

For zero coupling one finds, in analogy to Section 1.1

$$e_{10} = K_1 i_1 = E_{10} \epsilon^{j\omega t - \Gamma_1 s}$$

$$e_{20} = K_2 i_2 = E_{20} \epsilon^{j\omega t - \Gamma_2 s}$$

it

$$K_1 = \sqrt{\frac{z_1}{y_1}}$$

$$K_2 = \sqrt{\frac{z_2}{y_2}}$$

and

$$\Gamma_1 = \sqrt{y_1 z_1}$$
$$\Gamma_2 = \sqrt{y_2 z_2}$$

E_{10} and E_{20} are independent integration constants. For finite but loose coupling and small attenuation constants one finds in analogy to 1.1-14 and 1.1-15

$$\Gamma_a = \frac{\Gamma_1 + \Gamma_2}{2} + \frac{\Gamma_1 - \Gamma_2}{2}\sqrt{1 + \kappa^2} = \Gamma_1 + 0.5\,\Gamma_2(1 - \sqrt{1 + \kappa^2})$$

$$\Gamma_b = \frac{\Gamma_1 + \Gamma_2}{2} - \frac{\Gamma_1 - \Gamma_2}{2}\sqrt{1 + \kappa^2} = \Gamma_2 + 0.5\,\Gamma_1(1 - \sqrt{1 + \kappa^2})$$

where

$$\kappa = \frac{k}{\Gamma_1 - \Gamma_2}\sqrt{\Gamma_1 \Gamma_2} \qquad 1.2\text{-}3$$

is the coupling discriminant. Just as in Section 1.1 the coupling coefficient k is defined by the equation

$$k = \frac{P_{12}}{\sqrt{P_1 P_2}} = \frac{P_{21}}{\sqrt{P_1 P_2}} \qquad 1.2\text{-}4$$

P_1 is the energy per unit length stored in line 1; P_2, the energy per unit length stored in line 2, and $P_{12} = P_{21}$, the energy per unit length interchanged between the lines. The waves can travel in the coupled lines with either or both of two transmission constants. Two of the amplitude vectors in equations 1.2-1 and 1.2-2, for instance E_{1a} and E_{2b}, are free to satisfy boundary conditions; the other two are determined by the equation

$$\frac{E_{2a}^2 K_1}{E_{1a}^2 K_2} = \frac{\Gamma_a - \Gamma_1}{\Gamma_a - \Gamma_2} = \frac{\Gamma_b - \Gamma_2}{\Gamma_b - \Gamma_1} = \frac{E_{1b}^2 K_2}{E_{2b}^2 K_1}$$

$$\frac{E_{2a}}{E_{1a}}\sqrt{\frac{K_1}{K_2}} = \frac{E_{1b}}{E_{2b}}\sqrt{\frac{K_2}{K_1}} = A_a = -\frac{1}{A_b} \qquad 1.2\text{-}5$$

A is the *normalized amplitude and phase ratio* for two lines transformed to equal wave impedances.

$$\left|\frac{E_{2a}^2 K_1}{E_{1a}^2 K_2}\right| = W_a = \frac{1}{W_b}$$

$$W_a = \left|\frac{\sqrt{1 + \kappa^2} - 1}{\sqrt{1 + \kappa^2} + 1}\right| = W$$

is the ratio of energy flow in the two lines. At the propagation constant Γ_a,

$$A_a = \sqrt{1 + \kappa^{-2}} - \kappa^{-1} = A \qquad 1.2\text{-}6$$

When the indexes are left off, $W < 1$ and $| A | < 1$ by definition.
In a manner analogous to that of Section 1.1 it can be shown that the coupled attenuation constants are

$$\alpha_a = \frac{\alpha_1 + \alpha_2 W}{1 + W} = \alpha_1 \frac{W_{1a}}{W_{\text{total}}} + \alpha_2 \frac{W_{2a}}{W_{\text{total}}} \quad 1.2\text{-}7$$

$$\alpha_b = \frac{\alpha_2 + \alpha_1 W}{1 + W} = \alpha_1 \frac{W_{1b}}{W_{\text{total}}} + \alpha_2 \frac{W_{2b}}{W_{\text{total}}} \quad 1.2\text{-}8$$

The attenuation constants of the coupled waves are found by combining the uncoupled attenuation constants in the same proportion as the energies traveling in the two lines.
The coupled phase constants are

$$\beta_a = \frac{\beta_1 - W\beta_2}{1 - W} \text{ and} \quad 1.2\text{-}9a$$

$$\beta_b = \frac{\beta_2 - W\beta_1}{1 - W} \quad 1.2\text{-}9b$$

From equations 1.2-5 to 1.2-8 one sees that the coupled propagation constants are conveniently described in terms of the power ratio W. W itself is a known function of the complex coupling determinant κ which is shown on the attached Fig. 4 for the following three special cases:

Case 1.

The two lines have *equal phase* constants and *different attenuation* constants: $\beta_2 = \beta_1 \quad \alpha_2 \gtrless \alpha_1$

κ is an imaginary number.

W changes its character abruptly at the critical coupling.

$| \kappa \text{ critical} | = 1$

For $| \kappa | < 1$

$$W < 1; \quad \alpha_b \gtrless \alpha_a; \quad \beta_b = \beta_a$$

For $| \kappa | \geqq 1$

$$W = 1; \quad \alpha_b = \alpha_a; \quad \beta_b \gtrless \beta_a$$

Case 2.

The lines have *different phase* constants and *equal attenuation* constants.

κ is a real number

W changes asymptotically from

$$W_0 = 0 \text{ to}$$

$$W_1 = 0.172 \text{ and to}$$

$$W_\infty = 1$$

Case 3.

The *phase* and *attenuation* constants *differ by equal amounts.* As shown below, in section 4, this case applies to the coupling between the TE_{01} and TM_{11} modes in curved circular wave guides with finite conductivity.

κ^2 is an imaginary number.

W changes asymptotically from

$$W_0 = 0 \text{ to}$$

$$W_1 = 0.217 \text{ and}$$

$$W_\infty = 1$$

For $\kappa \gg 1$ all three cases approach the limit

$$\beta_\infty \doteq \frac{\beta_1 + \beta_2}{2}\left(1 \mp \frac{k}{2}\right) \quad 1.2\text{-}10$$

$$\alpha_\infty \doteq \frac{\alpha_1 + \alpha_2}{2}\left(1 \mp \frac{k}{2}\right) \doteq \frac{\alpha_1 + \alpha_2}{2}$$

2. Derivation of Field Equations

Consider a straight circular cylinder with an inside radius such as shown on Fig. 1 A. Let the radial coordinate equal r, the azimuthal coordinate equal φ and the longitudinal coordinate equal z. Let the dielectric losses inside the cylinder be negligible.

The field equations inside the cylinder are[5]

$$\frac{\partial E_z}{r\partial\varphi} - \frac{\partial E_\varphi}{\partial z} = -j\omega\mu\, H_r$$

$$\frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} = -j\omega\mu r\, H_\varphi$$

$$\frac{\partial(rE_\varphi)}{\partial r} - \frac{\partial E_r}{\partial\varphi} = -j\omega\mu r\, H_z$$

and

$$\frac{\partial H_z}{r\partial\varphi} - \frac{\partial H_\varphi}{\partial z} = j\omega\epsilon\, E_r$$

$$\frac{\partial H_r}{\partial z} - \frac{\partial H_z}{\partial r} = j\omega\epsilon\, E_\varphi$$

$$\frac{\partial(rH_\varphi)}{\partial r} - \frac{\partial Hr}{\partial\varphi} = j\omega\epsilon r\, E_z$$

[5] See Ref. 4, pg. 94 of the book.

The natural transmission modes which satisfy these equations have the form

$$E = f_n(r)\,\epsilon^{jn(\varphi+\varphi_0)} \cdot \epsilon^{j\omega t - \Gamma z} \qquad \text{2-0}$$

Each of these modes conforms to the same equations as a wave traveling in a transmission line with an impedance and phase velocity dependent upon the mode. In a straight cylinder with perfectly conducting walls, there exists no coupling between the different modes so that any and all can exist without interacting. If the conductivity of the walls in a straight circular cylinder is finite, it produces a resistive coupling between modes of equal azimuthal index (n in equation 2-0). In copper tubing and at the frequencies now obtainable ($\omega < 10^{12}$) this coupling effect is negligible.

A stronger coupling may be caused by deviations of the wave guide from the shape of a straight circular cylinder. The deformation considered in the present analysis consists in a circular bend of the axis, as shown schematically on Fig. 1b.

In such a circular bend the longitudinal coordinate is replaced by the product of the bending radius R by the bending angle θ:

$$z = R\theta$$

This transforms the first two component equations of curl E into

$$\frac{\partial(RE_\theta)}{Rr\partial\varphi} - \frac{\partial E_\varphi}{R\partial\theta} = -j\omega\mu\, H_r$$

$$\frac{\partial E_r}{R\partial\theta} - \frac{\partial(RE_\theta)}{R\partial r} = -j\omega\mu\, H_\varphi$$

The variable R can be eliminated by the relation

$$R = R_0 - r\cos\varphi$$

where R_0 is the bending radius of the cylinder axis. The coordinate θ can be replaced by a longitudinal coordinate s, measured along the cylinder axis. Hence,

$$s = \theta R_0$$

The progressive modes which we investigate have the approximate form

$$E = f_n(r)\,\epsilon^{jn(\varphi+\varphi_0)}\,\epsilon^{j\omega t - \Gamma s}$$

Hence

$$\frac{\partial}{\partial\theta} = R_0\frac{\partial}{\partial s} = -\,R_0\Gamma$$

$$\frac{\partial}{\partial\varphi} = jn \quad \text{for all field components.}$$

$\dfrac{\partial}{\partial r}$ may be expressed by a prime:

$$\frac{\partial F}{\partial r} = F'$$

Thus the equations with curl E become

$$\frac{jnE_s}{r} + \frac{E_s \sin\varphi}{R_0 - r\cos\varphi} + \frac{R_0 \Gamma E_\varphi}{R_0 - r\cos\varphi} = -j\omega\mu\, H_r$$

$$\frac{-R_0 \Gamma E_r}{R_0 - r\cos\varphi} - E_s' + \frac{E_s \cos\varphi}{R_0 - r\cos\varphi} = -j\omega\mu\, H_\varphi$$

$$E_\varphi + rE_\varphi' = jnE_r = -j\omega\mu r\, H_s$$

For gradual bends

$$R_0 \gg a > r$$

One may therefore approximate

$$\frac{R_0}{R_0 - r\cos\varphi} \doteq 1 + \frac{r}{R_0}\cos\varphi$$

It is convenient to introduce the symbol

$$c = \frac{a}{R_0}$$

which is proportional to the coupling coefficient. All powers of c greater than the first will be neglected. One can now write the approximate field equations in the curved cylinder:

$$\frac{jnE_s}{r} + \Gamma E_\varphi + \frac{c}{a} E_s \sin\varphi + c\Gamma \frac{r}{a} E_\varphi \cos\varphi = -j\omega\mu\, H_r \qquad 2\text{-}1$$

$$-\Gamma E_r - E_s' - c\Gamma \frac{r}{a} E_r \cos\varphi + \frac{c}{a} E_s \cos\varphi = -j\omega\mu\, H_\varphi \qquad 2\text{-}2$$

$$E_\varphi + rE_\varphi' - jnE_r = -j\omega\mu r H_s \qquad 2\text{-}3$$

$$\frac{jn\, H_s}{r} + \Gamma H_\varphi + \frac{c}{a} H_s \sin\varphi + c\Gamma \frac{r}{a} H_\varphi \cos\varphi = j\omega\epsilon\, E_r \qquad 2\text{-}4$$

$$-\Gamma H_r - H_s' - c\Gamma \frac{r}{a} H_r \cos\varphi + \frac{c}{a} H_s \cos\varphi = j\omega\epsilon\, E_\varphi \qquad 2\text{-}5$$

$$H_\varphi + rH_\varphi' - jnH_r = j\omega\epsilon r\, E_s \qquad 2\text{-}6$$

The coupling terms all contain the factor $\cos\varphi$ or $\sin\varphi$. This means that every transmission mode is coupled only to modes with an azimuth index differing from its own by ± 1.

3. Characteristic Equations of TE_{01} and TM_{11} Modes

The mode which one desires to propagate through the wave guide is the TE_{01} mode. In a straight wave guide with perfectly conducting walls it is characterized by the following equations:

$$n = 0 \quad \text{3-1}$$

$$E_{r1} = E_{z1} = H_{\varphi 1} = 0 \quad \text{3-2}$$

$$E_{\varphi 1} = E_1 \epsilon^{j\omega t - \Gamma_1 z} J_1(y) = e_1 J_1(y) \quad \text{3-3}$$

$$H_{r1} = -\frac{e_1 \Gamma_1}{j\beta_0 \eta} J_1(y) \quad \text{3-4}$$

$$H_{z1} = -\frac{e_1 \chi}{j\beta_0 \eta} J_0(y) \quad \text{3-5}$$

In these equations

$$\eta = \sqrt{\frac{\mu}{\epsilon}} = 377 \text{ ohm (intrinsic free space resistance)} \quad \text{3-6}$$

$$\beta_0 = \omega\sqrt{\epsilon\mu} = \frac{2\pi}{\lambda_0} \quad \text{3-7}$$

$$y = \chi r \quad \text{3-8}$$

$$\Gamma_1 = \sqrt{\chi^2 - \beta_0^2} \quad \text{3-9}$$

In a perfectly conducting wave guide

$$\chi_0 = \frac{3.832}{a} \quad \text{3-10}$$

$$\nu = \frac{\chi_0}{\beta_0} = \frac{.61\lambda_0}{a} \text{ (cutoff factor)} \quad \text{3-11}$$

$$\Gamma_1 = j\beta_0\sqrt{1 - \nu^2} = j\beta_1 \quad \text{3-12}$$

If the wave guide has the conductivity of g mho/m, its intrinsic high frequency impedance is

$$Z_i = (1 + j)R_i = (1 + j)34.4 \sqrt{\frac{\mu}{g\lambda_0}} \,* \quad \text{3-13}$$

This changes χ to

$${}_1\chi \doteq \chi_0 - \frac{(1 - j)R_i\nu}{\eta a} \quad \text{and} \quad \text{3-14}$$

$$\Gamma_1 \doteq j\beta_1 + \frac{\nu^2}{\sqrt{1 - \nu^2}} \frac{(1 + j)R_i}{a\eta} \dagger \quad \text{3-15}$$

* Ref. 4 pg. 83.
† Compare Ref. 4, pg. 390.

Due to the curvature of the guide, this desired mode is coupled to all modes which have the azimuthal index number 1.

However, for low curvatures, this coupling is very loose and only causes appreciable effects if it can act over a great length of wave guide without phase interference.

This means that the disturbing mode must have nearly the same phase velocity as the desired mode. It so happens that in a perfectly conducting circular cylinder there exists one mode, the TM_{11}, which has exactly the same velocity as the TE_{01}. Such a coincidence is called "degeneracy."

In the analysis of very gradual bends, only this TM_{11} mode need be considered. It is characterized in a straight guide by the following equations:

$$n = 1 \qquad 3\text{-}16$$

$$E_{\varphi 2} = E_2 \epsilon^{j\omega\tau - \Gamma_2 z} \cdot \frac{J_1(y)}{y} \cos(\varphi + \varphi_0) = e_2 \frac{J_1(y)}{y} \cos\varphi \qquad 3\text{-}17$$

The TM_{11} mode can be polarized in all directions. But since only the component directed toward

$$\varphi_0 = 0$$

is excited by the wave guide curvature, φ_0 has been omitted in the last term of eq. (3-17).

$$E_{r2} = e_2 \frac{dJ_1(y)}{dy} \sin\varphi = e_2 \dot{J}_1(y) \sin\varphi, \quad \text{where } \dot{J} = \frac{dJ}{dy}$$

$$E_{z2} = \frac{-\chi_2 e_2}{\Gamma_2} J_1(y) \sin\varphi$$

$$H_{\varphi 2} = \frac{e_2 j\beta_0}{\eta\Gamma_2} \dot{J}_1(y) \sin\varphi$$

$$H_{r2} = -\frac{e_2 j\beta_0}{\eta\Gamma_2} \frac{J_1(y)}{y} \cos\varphi$$

$$H_{z2} = 0$$

In a perfectly conducting wave guide the χ defined by eq. 3-8 is

$$\chi = \chi_2 = \chi_0 \text{ and}$$

$$\Gamma_2 = \Gamma_1 = j\beta_1$$

In a wave guide with an intrinsic impedance per 3-13,

$$\chi_2 \doteq \chi_0 - \frac{(1-j)R_i}{\eta a \nu} \quad \text{and} \qquad 3\text{-}18$$

$$\Gamma_2 \doteq j\beta_1 + \frac{(1+j)R_i \dagger}{\sqrt{1-\nu^2}\, a\eta} \qquad 3\text{-}19$$

From 3-15 and 3-19 one finds $\alpha_1 = \alpha_2\nu^2$ 3–20

4. Interaction Between TE_{01} and TM_{11} Modes

Since the separate modes of propagation behave like traveling waves in transmission lines[2], their interaction can be derived from the coupling equations derived in Section 1 of this analysis.

The uncoupled propagation constants Γ_1 and Γ_2 are known (equations 3-15 and 3-19). In order to find the coupling discriminant one must derive the coupling coefficient from the field equations 2-1 to 2-6. The coupling coefficient is defined as

$$k = \frac{P_{12}}{\sqrt{P_1 P_2}} = \frac{P_{21}}{\sqrt{P_1 P_2}} \qquad \text{(eq. 1.2-4)}$$

In computing the coupling coefficient one may neglect the small attenuation constant. The energy stored by the TE_{01} wave per unit length is

$$P_1 = \int_0^a \frac{E_1^2}{\eta} \cdot 2\pi r\, dr = \int_0^a H_1^2 \eta \cdot 2\pi r\, dr \qquad 4\text{-}1$$

This expression is not affected by the cutoff factor ν because one may consider the field inside the guide as composed of slanting plane waves with the electric field strength E_1. The energy stored by the TM_{11} wave per unit length is

$$P_2 = \int_0^{2\pi} \int_0^a \frac{E_2^2}{\eta} r\, d\varphi\, dr = \int_0^{2\pi} \int_0^a H_2^2 \eta r\, d\varphi\, dr$$

The interchanged energy:

$$P_{21} = \int_0^{2\pi} \int_0^a \frac{E_{21} E_1}{\eta}\, d\varphi\, dr + \int_0^{2\pi} \int_0^a H_{21} H_1 \eta\, d\varphi\, dr$$

Combining equation (4-1) with 3-3, 3-8 and 3-10

$$P_1 = \frac{2\pi a^2 e_1^2}{3.832^2 \eta} \int_0^{3.832} y\delta_1^2 J_1(y)\, dy$$

From reference 1, page 146 of the book

$$\int_0^{3.832} y\delta_1^2 J_1(y)\, dy = \frac{3.832^2}{2} J_0(3.832)$$

Hence

$$P_1 = \frac{0.51 e_1^2 a^2}{\eta} \qquad 4\text{-}2$$

In a similar manner one finds

$$P_2 = \frac{0.51 e_2^2 a^2}{2(1 - \nu^2)\eta} \qquad 4\text{-}3$$

[2] Loc. cit.

and

$$P_{12} = P_{21} = \frac{0.51 ce_1 e_2 a^2}{3.832\eta} \qquad 4\text{-}4$$

Substituting the values of 4-2, 4-3 and 4-4 into 1.2-4 one finds for the coupling coefficient between the two groups of plane waves traveling at a slant to the wave guide axis

$$k_s = \frac{c\sqrt{2}\sqrt{1-\nu^2}}{3.83} = \frac{0.369a}{R_0}\sqrt{1-\nu^2}$$

The coupling coefficient k between the TE_{01} and TM_{11} modes which are the resultants of their slant wave groups is greater than k_s according to the following reasoning:

From 1.2-10

$$\beta \doteq \beta_0(1 \pm 0.5\, k_s)$$

This makes the cutoff factors of the coupled modes

$$\nu = \frac{x}{\beta} = \frac{\nu_0}{1 \pm 0.5k_s}$$

and the coupled propagation constants

$$\Gamma = \beta\sqrt{1-\nu^2} = \beta_0\sqrt{(1 \pm 0.5k_s)^2 - \nu_0^2}$$

For $k_s \ll 1$

$$\Gamma \doteq \beta_0\sqrt{1-\nu_0^2}\left(1 \pm \frac{0.5k_s}{1-\nu_0^2}\right) = \Gamma_0(1 - 0.5k)$$

Hence, in view of eq. 1.2-10, the effective coupling coefficient of the wave guide modes is

$$k = \frac{k_s}{1-\nu^2} = \frac{0.369a}{R_0\sqrt{1-\nu^2}} \qquad 4\text{-}5$$

From equations 3-15 and 3-19

$$\Gamma_1 - \Gamma_2 = -\frac{1+j}{a\eta} R_i\sqrt{1-\nu^2} \qquad 4\text{-}6$$

$$\frac{\sqrt{\Gamma_1\Gamma_2}}{\Gamma_1 - \Gamma_2} \doteq \frac{j\beta_0}{\Gamma_1 - \Gamma_2} \doteq \frac{-2\pi\eta a}{(1-j)R_i\lambda_0\sqrt{1-\nu^2}} \qquad 4\text{-}7$$

From 1.2-3, 4-5 and 4-7 one obtains the coupling discriminant

$$\kappa = \frac{-\,0.369\cdot 2\pi\cdot 377a^2}{R_0R_i\lambda_0(1-j)(1-\nu^2)}$$

Since

$$R_i = 0.00452\,\lambda_0^{-0.5}\cdot\rho_r\,* \qquad 4\text{-}8$$

* Loc. cit.

with the relative high frequency resistivity

$$\rho_r = \sqrt{\frac{\rho}{\rho_{\text{copper}}}}$$

$$\kappa = -\frac{9.65 \cdot 10^4(1 + j)a^2\lambda^{-0.5}}{R_0\,\rho_r(1 - \nu^2)}$$

Its absolute value

$$|\,\kappa\,| = \frac{1.366 \cdot 10^5 a^2\lambda^{-0.5}}{R_0\,\rho_r(1 - \nu^2)}$$

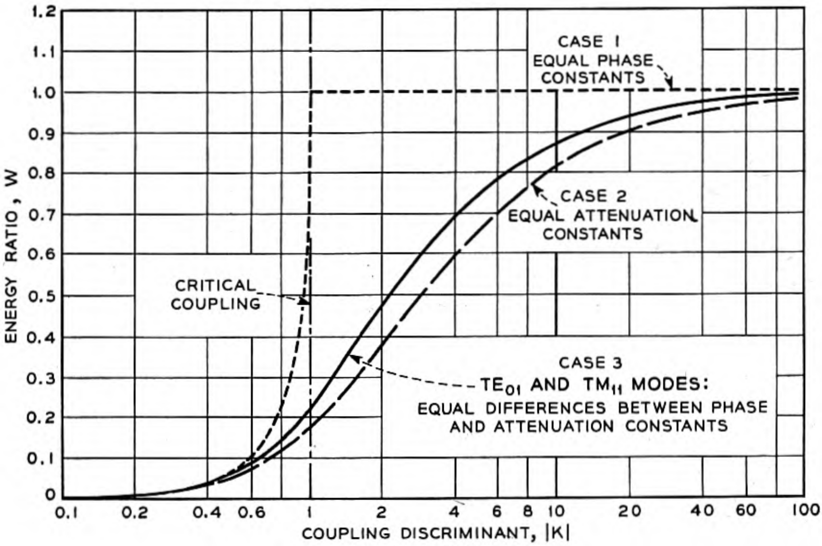


Fig. 3

As shown in equation 4-6, the differences between the propagation constants of the TM_{11} and TE_{01} waves are proportional to the intrinsic skin impedance R_i which contains the factor $(1 + j)$. This means that the phase and attenuation constants of the two waves differ by equal amounts in accordance with "*Case 3*" of Section 1. For this case the power ratio per 1.2-20 may be written

$$W = \left|\frac{\sqrt{i + j\,|\,\kappa\,|^2} - 1}{\sqrt{1 + j\,|\,\kappa\,|^2} + 1}\right|$$

The numerical value of this function is plotted on Fig. 3. It can be computed conveniently by means of the following auxiliary parameters:

$$\sqrt{1 + j\,|\,\kappa\,|^2} = p + jq = \cosh x + j \sinh x \qquad \text{4-9}$$

$$p = \sqrt{0.5 + 0.5\sqrt{1 + |\kappa|^4}} = \cosh x$$

$$q = \sqrt{-0.5 + 0.5\sqrt{1 + |\kappa|^4}} = \sinh x$$

$$W = \frac{p - 1}{p} \tanh \frac{x}{2}$$

In analogy to *Case 1* of Section 3, the condition of $|\kappa| = 1$ may be called *critical coupling.* It occurs at the *critical radius of curvature*

$$R_{0\,\mathrm{cr}} = \frac{1.366 \cdot 10^5 \; a^2 \lambda^{-0.5}}{\rho_r(1 - v^2)} \text{ meters} \tag{4-10}$$

For subcritical coupling ($R_0 \gg R_{11}$) W approaches

$$W_{\mathrm{subcr.}} \doteq \left| \frac{\kappa^2}{4} \right| \longrightarrow 0$$

For supercritical coupling ($R_0 \ll R_{\mathrm{cr}}$),

$$W_{\mathrm{supercr.}} \doteq 1 - \frac{\sqrt{2}}{|\kappa|} \longrightarrow 1$$

From the above results, it is possible to predict the behavior of waves originating either as TE_{01} or as TM_{11} modes, in any given wave guide configuration. This is done for some typical cases in the following sections.

5. Propagation in Long Waveguides with Constant Curvature

It has been shown in Section 3 that for each curvature there exist two modes of propagation.

In one,

$$W_a = \frac{P_{\mathrm{TM11}}}{P_{\mathrm{TE01}}} < 1 \tag{5-1}$$

and the attenuation

$$\alpha_a < \frac{\alpha_{\mathrm{TE}} + \alpha_{\mathrm{TM}}}{2} \tag{5-2}$$

In the other,

$$W_b = \frac{1}{W_a} > 1 \tag{5-3}$$

and the attenuation

$$\alpha_b > \frac{\alpha_{\mathrm{TE}} + {\mathrm{TM}}}{2} \tag{5-4}$$

In a long wave guide the "*b*" mode will die down due to its greater attenuation, no matter how much of it was initially present, so that one need only consider the "*a*" mode.

This mode has a phase velocity slightly smaller than that of the uncoupled TE_{01} wave and an attenuation nearer to that of the TE_{01} than the TM_{11} wave.

The magnitude of the critical radius is illustrated by the two examples of Table I.

TABLE I
CHARACTERISTIC VALUES

Parameter	Symbol	Equation	Example 1	Example 2
Wave guide radius	a		.05 m	.05 m
Free space wave length	λ_0		.03 m	.01 m
Cutoff ratio	ν	3-11	.366	.122
Attenuation constant	$\alpha_{TE(cu)}$	3-15, 4-17	2.04×10^{-4} neper/m	3.58×10^{-5} neper/m
Attenuation constant	$\alpha_{TM(cu)}$	3-19	1.53×10^{-3} neper/m	2.41×10^{-3} neper/m
Critical Radius	R_{crit}	4-10	2.12 km.	3.44 km.

TABLE II
RELATIVE ATTENUATION VERSUS RADIUS OF CURVATURE

General formulae				Example 1		Example 2	
κ	R_0/R_{cr}	W	α/α_0	R_0km	α/α_0	R_0km	α/α_0
0	∞	0	1	∞	1.00	∞	1.00
0.1	10	0.0025	$1 + 0.0025\,(\nu^{-2} - 1)$	19.7	1.02	34.15	1.17
0.2	5	0.01	$1 + 0.01$ "	9.85	1.06	17.08	1.66
0.5	2	0.06	$1 + 0.057$ "	3.84	1.38	6.83	4.45
1	1	0.22	$1 + 0.18$ "	1.97	2.16	3.42	12.94
2	0.5	0.48	$1 + 0.32$ "	0.98	3.11	1.71	22.2
5	0.2	0.75	$1 + 0.43$ "	0.38	3.83	0.68	26.6
10	0.1	0.87	$1 + 0.46$ "	0.20	4.05	0.34	27.9
∞	0	1.00	$1 + 0.50$ "	0	4.30	0	34.1

The increase of attenuation in long wave guides with uniform curvature is shown on Table II, with numerical values for the same examples as in Table I.

6. PROPAGATION IN A UNIFORMLY CURVED SECTION OF WAVE GUIDE FOLLOWING A LONG STRAIGHT SECTION. (FIG. 4)

No matter what mixture of modes may prevail at the beginning of the wave guide, all modes except the TE_{01} die down in the long straight section due to their higher attenuation, so that the wave form at the beginning of the curved section is pure TE_{01}.

Since it has been shown in Sections 1 and 4 that each of the two possible modes of propagation in a curved wave guide consists of both TE_{01} and TM_{11} waves, it follows that both modes must be superimposed in such a manner that at the transition point the TM components cancel each other by interference.

Let the relative amplitudes of the two TE components equal a and b; then the corresponding amplitudes of the TM modes are aA_a and bA_b

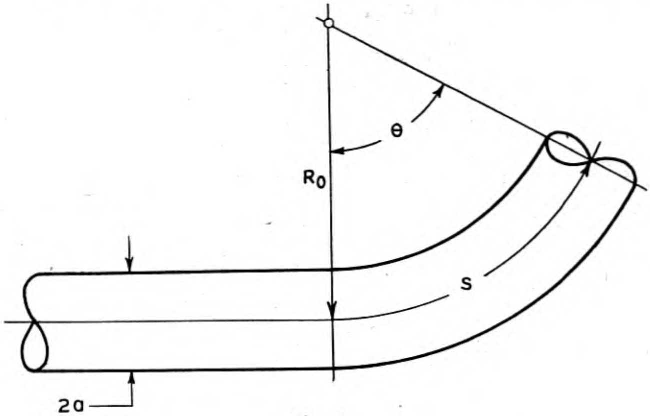


Fig. 4

respectively, where A_a and A_b are the normalized voltage ratios per 1.2-5. At the beginning of the curved section

$$a + b = 1 \qquad \text{(TE amplitude)}$$

$$aA_a + bA_b = aA_a - b/A_a = 0 \qquad \text{(TM amplitude)}$$

Hence

$$a = \frac{1}{1 + A_a^2}$$

$$b = \frac{A_a^2}{1 + A_a^2}$$

$$\left|\frac{b}{a}\right| = \left|A_a^2\right| = W$$

The two waves have different phase velocities and therefore interfere with each other. According to 1.2-9, the difference between the phase constants is

$$\beta_b - \beta_a = (\beta_2 - \beta_1)\,\frac{1 + W}{1 - W}$$

By means of equations 4-9, this can be transformed into

$$\beta_b - \beta_a = (\beta_2 - \beta_1)(p + q) \qquad \text{6-1}$$

The length of one complete interference cycle is

$$s_{2\pi} = \frac{2\pi}{(p + q)(\beta_2 - \beta_1)}$$

If both components had equal attenuation, the beats superimposed on the decaying envelope of the TE_{01} wave would correspond to an amplitude ratio

$$\frac{e_0 \max}{e_0 \min} = \frac{1 + W}{1 - W} = p + q$$

However, during the progress of the mixed wave through the curved section, the intensity of the fluctuations is reduced by the greater attenuation constant of the faster and weaker "b" component. In one complete interference cycle, the differential attenuation reduces the weaker component to

$$\frac{A_{2\pi}}{A_0} = \epsilon^{-(\alpha_2 - \alpha_1)s_{2\pi}} = \epsilon^{-(2\pi/p+q)}$$

Approximation for Weak Coupling

For $|\kappa| \ll 1$

$\beta_b - \beta_a \doteq (\beta_2 - \beta_1)(1 + 0.5 |\kappa|^2)$

From 3-15 and 3-19

$$\beta_b - \beta_a \doteq \frac{R_i}{a\eta}\sqrt{1 - \nu^2}$$

$$\doteq \frac{1.2 \times 10^{-5}}{a}\sqrt{\frac{(1 - \nu^2)g_{cu}}{\lambda_0 g}}\,(1 + 0.5 |\kappa|^2) \text{ radian}/m$$

For intermediate coupling, 6-1 may be transformed into

$$\beta_b - \beta_a = k\beta_1 f(\kappa)$$

$$\text{with } f_{(\kappa)} = \frac{p + q}{\sqrt{2}|\kappa|} = \frac{\sqrt{1 + \sqrt{1 + |\kappa|^4}} + \sqrt{-1 + \sqrt{1 + |\kappa|^4}}}{2|\kappa|}$$

Approximation for Strong Coupling

For $|\kappa| \gg 1$

$$f_{(\kappa)} \doteq 1 + 0.125 |\kappa|^{-4} \quad \beta_b - \beta_a \doteq k\beta_1(1 + 0.125 |\kappa|^{-4})$$

Substituting the value of k from 4-13 and transforming,

$$\beta_b - \beta_a = \frac{2.32a}{\lambda_g R_0} f_{(\kappa)} = \frac{2.32a}{\lambda_0 R_0}(1 + 0.125 |\kappa|^{-4}) \qquad 6\text{-}2$$

The phase difference between the two components is

$$\Psi_{b-a} = \frac{2.32as}{\lambda_0 R_0} f_{(\kappa)} = \frac{2.32a\theta}{\lambda_0} f_{(\kappa)} = M\theta \qquad 6\text{-}3$$

where θ is the bending angle of the wave guide. The power carried by the TE_{01} wave is

$$P_{TE} = \cos^2 \frac{M\theta}{2}. \qquad 6\text{-}4$$

Minima of TE_{01} occur when this phase difference is an odd multiple of π. Hence, the bending angles producing minima of TE_{01} amplitudes are:

$$\theta_{\min} \doteq (2n + 1) \cdot \frac{1.36\lambda_0}{a(1 + 0.125 \,|\, \kappa \,|^{-4})} \doteq \frac{(2n + 1)2.22\nu}{f_{(\kappa)}} \qquad 6\text{-}5$$

The initial fluctuation ratio approaches

$$\frac{e_{1\,\max}}{e_{1\,\min}} = p + q \doteq \sqrt{2}\,|\,\kappa\,|$$

which is a large value tending to infinity.

The relative attenuation of the slightly weaker component during one beat cycle is

$$\frac{A_{2\pi}}{A_0} = \epsilon^{-2\pi/p+q} = \epsilon^{-\sqrt{2}\pi/|\kappa|} \doteq 1 - \frac{4.44}{|\,\kappa\,|}.$$

which is a small reduction tending to zero. Hence, the fluctuations persist through a large number of beats. The power is transformed back and forth between the TE_{01} and the TM_{11} modes.

In Section 5, it was shown that in a long, uniformly curved wave guide the attenuation is intermediate between that of the TE_{01} and TM_{11} modes. But from equations 1.2-7 and 8 it follows that the two modes contribute to the attenuation in proportion to their relative power flow. Since at the beginning of the bend the power of the TM_{11} component is zero, it is to be expected that the *initial* rate of attenuation equals that of the TE_{01} wave alone. This is proved by differentiating with regard to s. One finds for all values of κ that

$$\frac{d}{ds}\left|\, a\epsilon^{-\Gamma_a s} + b\epsilon^{-\Gamma_b s} \,\right|_{s=0} = -\alpha_1$$

Discussion of Results

Equation 6-2 corresponds directly to an equation derived by S. O. Rice and, after allowing for the different choice of variables, to M. Jouguet's equation (75)[6]. It differs from the results of these earlier calculations by the factor $f(\kappa) \doteq 1 + 0.125\,|\,\kappa\,|^{-4}$ which is a reminder that the simplified form of the equations given by the earlier authors is an extrapolation to infinite conductivity or infinite curvature of the wave guide.

[6] Reference 3, pg. 150 of *Cables and Transmission,* July 1947.

From equations 6-3 and 6-4 it is seen that the TE_{01} wave is recovered by bends which are an even multiple of θ_{min}. But such bends are efficient transmitters of TE_{01} waves only over a narrow frequency range since θ_{min} varies with frequency.

If the circular bend is followed by a long straight section, the TE_{01} and TM_{11} components existing at the end of the bend are carried over into the straight section, but the TM_{11} component dies down due to its greater attenuation and constitutes a total loss.

Numerical examples for first extinction angle.

Using the same dimensions as in Table I of Section 5, one finds from eq. 6-5 for:

$$\text{Example 1: } \theta_{min} \doteq 0.816 \text{ Radians} \doteq 46.8^\circ$$

$$\text{Example 2: } \theta_{min} \doteq 0.272 \text{ Radians} \doteq 15.6^\circ$$

7. Serpentine Bends

Sections 5 and 6 dealt with bends continued with uniform curvature over large angles. The present section considers the small random deviations from a straight course which are unavoidable in field installations.

Actual deviations are expected to be random both with regard to maximum deflection angle and to curvature; they are likely to approximate a sinusoidal shape. For purposes of computation, the following analysis assumes as a first case *circular* S-*bends* which consist of alternate regions of equal lengths and equal but opposite curvatures. An exaggerated schematic of such S-bends is shown on Fig. 5A.

Each circular bend tends to produce a single mode with an attenuation per equation 1.2-7. However, the discontinuous reversals of curvature at the inflexion points produce mixed modes, and the initial part of each region reduces the amplitude of the TM components produced in the previous region.

Each region may be treated as a discrete 4-terminal section of a transmission network. Regardless of the wave composition at the input terminal, differential attenuation will establish in a long serpentine wave guide a steady state condition. In this steady state each region produces equal attenuation. This attenuation per region and the resulting average attenuation constant will now be derived.

The TE_{01} and TM_{11} waves each consist of "a" and "b" components with separate amplitude ratios and propagation constants, as derived in Sections 1 and 4. In the first region (between points 1 and 2 of Fig. 5A) R_0 is taken as positive, and

$$A = \sqrt{1 + \kappa^{-2}} - \kappa^{-1} \qquad 1.2\text{-}6$$

In the second region, between points 2 and 3, the polarity of R_0, and consequently of κ and the ratio of TM to TE amplitudes, are reversed.

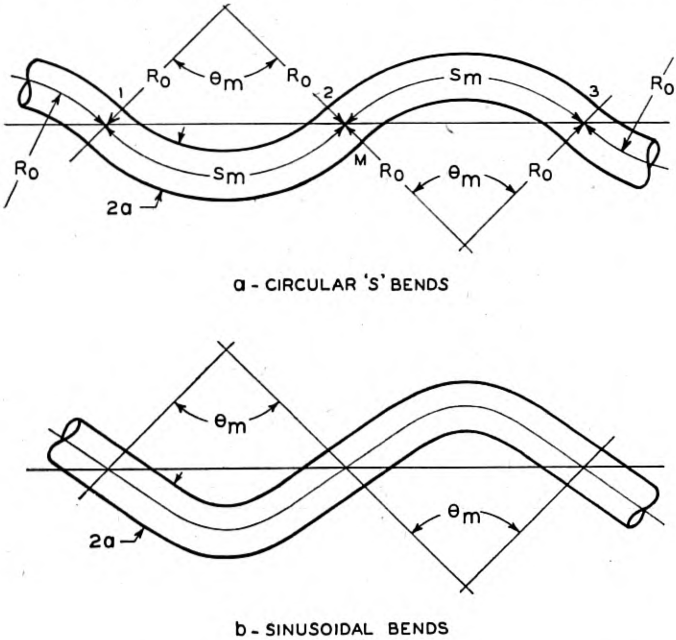


Fig. 5

Except for this change of polarity the amplitude ratios at points 1 and 2 are equal. Introducing the symbols

$$g = \epsilon^{-\Gamma sm}$$
$$g_m = \tfrac{1}{2}(g_a + g_b)$$
$$g_d = \tfrac{1}{2}(g_b - g_a)$$

one can tabulate

point	e_{TE}	e_{TM}	
1	$a + b$	$aA - \dfrac{b}{A}$	7-1
2	$g_m\left(\dfrac{a}{g_d} + bg_d\right)$	$g_m\left(\dfrac{aA}{g_d} - \dfrac{b}{A}g_d\right)$	7-2

Calling $\dfrac{b}{a} = y$, one finds

$$\frac{1 + yg_d^2}{1 + y} = \frac{-A^2 + yg_d^2}{A^2 - y} = \frac{g_d}{g_m} \cdot g_{\text{average}} = \frac{g_{\text{average}}}{g_a}$$

$$\Gamma_{\text{average}} = \Gamma_a + \frac{1}{s_m} \log \frac{1 + y}{1 + yg_d^2} \text{ with}$$

$$y = \frac{A}{2} \kappa^{-1}(1 + g_d^{-2}) + A \sqrt{g_d^{-2} + \frac{\kappa^{-2}}{4} (1 + g_d^{-2})^2}$$

This formal solution is hard to evaluate. It can be greatly simplified for the subcritical and supercritical cases.

1. *Subcritical Curvature* $|\kappa| \ll 1$

$$y \doteq \frac{\kappa^2}{2 + 2_d^2}$$

$$\Gamma_{\text{average}} = \Gamma_a \left(1 + \frac{\kappa^2}{2} \frac{1 - g_d^2}{1 + g_d^2}\right) \doteq \Gamma_a \doteq \Gamma_1$$

For very low curvatures, the average attenuation approaches that of the "a" mode, and this in turn approaches that of the TE_{01} wave.

2. *Supercritical Curvature.* $|\kappa| \gg 1$

The differential attenuation constant is small compared to the differential phase constant.

$|y| \doteq |A| \doteq 1$

Substituting these values into 7-1 and 2, one finds

$y \doteq \epsilon^{-0.5j\theta_m}$

Expressed as a function of θ:

$y_\theta = \cos \psi + j \sin \psi$ with

$\psi = M(\theta - 0.5\ \theta_m)$ 7-3

M has the value per eq. 6-18.

The power ratio of the combined TM_{11} and TE_{01} waves is

$W_\theta \doteq \tan^2 \psi/2$

In view of equation 1.2-7 the instantaneous rate of energy loss is

$$\alpha_\theta = \alpha_1 \cos^2 \frac{\psi}{2} + \alpha_2 \sin^2 \frac{\psi}{2} = \alpha_1 + (\alpha_2 - \alpha_1) \sin^2 \frac{\psi}{2} \qquad \text{7-4}$$

$$\alpha_{\text{average}} = \frac{1}{s_m} \int_0^{s_m} \alpha_s\, d_s = \frac{1}{\theta_m} \int_0^{\theta_m} \alpha_\theta\, d\theta$$

$$\alpha_{\text{average}} = \alpha_1 + (\alpha_2 - \alpha_1) \left(\frac{1}{2} - \frac{\sin M\theta_m}{2M\theta_m}\right)$$

In view of 3-20

$$\alpha_{\text{average}} = \alpha_1 \left[\frac{\nu^{-2} + 1}{2} - \frac{\nu^{-2} - 1}{2} \frac{\sin M\theta_m}{M\theta_m}\right]$$

For small deflection angles, $M\theta_m \ll 2$

$$\alpha_{\text{average}} \doteq \alpha_1 \left[1 + \frac{\nu^{-2} - 1}{6} M^2\theta_m^2 \right]$$

If a $p\%$ increase in attenuation is the tolerance limit

$$\theta_m \leqq \frac{1}{10M} \sqrt{\frac{6p}{\nu^{-2} - 1}}.$$ Substituting the value of M from 6-3,

$$\theta_m = \frac{0.105\sqrt{p}\,\nu\lambda_0}{a\sqrt{1 - \nu^2}}.$$

The maximum deflection equals

$$\Delta_\theta = 0.5\,\theta_m.$$ Hence, in view of 3-11

$$\Delta_\theta \leqq \frac{0.032\sqrt{p}\lambda_0^2}{a^2\sqrt{1 - \nu^2}} \text{ radians} = \frac{1.84\lambda_0^2}{a^2} \sqrt{\frac{p}{1 - \nu^2}} \text{ degrees}$$

3. *Sinusoidal Bends with Predominantly Supercritical Curvature.*

Sinusoidal bends cannot be supercritically curved over their entire length because at the inflection points the curvature drops to zero. For sufficiently short bends, however, no great error is caused by treating the entire length as supercritical. In that case, equations 7-3 and 7-4 remain valid. θ takes the new value

$$\theta = \frac{\theta_m}{2} + \frac{\theta_m}{2} \sin \frac{\pi(s - s_m)}{2s_m}$$

Hence

$$\psi = \frac{M\theta_m}{2} \sin \frac{\pi(s - s_m)}{2s_m}$$

$$\alpha_{\text{average}} = \alpha_1 + \frac{\alpha_2 - \alpha_1}{s_m} \int_0^{s_m} \sin^2 \left[\frac{M\theta_m}{2} \sin \frac{\pi(s - s_m)}{2s_m} \right] ds$$

For small deflection angles, $M\theta_m \ll 2$

$$\alpha_{\text{average}} = \alpha_1 + \frac{\alpha_2 - \alpha_1}{s_m} \frac{M^2\theta_m^2}{4} \int_0^{s_m} \sin^2 \frac{\pi(s - s_m)}{2s_m} ds$$

$$= \alpha_1 + (\alpha_2 - \alpha_1) \frac{M^2\theta_m^2}{4} = \alpha_1 \left[1 + \frac{\nu^{-2} - 1}{4} M^2\theta_m^2 \right]$$

$$\Delta_\theta = \frac{0.026\lambda_0^2\sqrt{p}}{a^2\sqrt{1 - \nu^2}} \text{ radians} = \frac{1.49\lambda_0^2}{a^2} \sqrt{\frac{p}{1 - \nu^2}} \text{ degrees}$$

The tolerance limit for sinusoidal deflections is 20% smaller than for circular S bends.

The effect of supercritical but shallow circular and sinusoidal S bends is illustrated by the following *numerical examples*.

TABLE III
INCREASE OF ATTENUATION IN S BENDS

Attenuation Increase p%	Maximum deflection $\Delta\Theta$ (in degrees)			
	Example 1 (ν = .366)		Example 2 (ν = .122)	
	Circular	Sinusoidal	Circular	Sinusoidal
10	2.25	1.82	0.23	0.19
20	3.18	2.58	0.33	0.27
30	3.89	3.15	0.41	0.33
40	4.50	3.64	0.47	0.38
50	5.03	4.07	0.52	0.42

8. Helical Bends and Random Two-Dimensional Deviations

A helical bend may be treated as a bend which has a constant absolute magnitude, but a changing direction of curvature. As indicated in eq. 3-17, the TM_{11} wave can be polarized in all directions. At any differential element of wave guide length, the TM_{11} component polarized in the local bending plane is coupled to the TE_{01} wave; the TM_{11} component polarized at right angles is not coupled and persists unchanged. By requiring that the absolute magnitude of the TM_{11}/TE_{11} amplitude ratio remain constant, a steady state solution can be found.

Shallow helical bends of small curvatures may be treated as the superposition of two sinusoidal bends offset by 90° in the longitudinal direction and in the bending plane. The increases in attenuation due to these two sinusoidal bends are computed from eq. 7-5 and added.

It is believed that random deviations from a straight course approach sinusoidal shape more closely than circular shape, hence equation 7-5 may be used to establish a tolerance limit for such random deviations. For quantitative results the statistical distribution of the squared deviation maxima must be taken into consideration.

9. Optima of Wave Guide Radius, Signal Wave Length and Attenuation as a Function of Angular Deviation

In a straight wave guide the attenuation decreases with wave guide radius and signal frequency. However, the deterioration due to wave guide curvature increases with wave guide radius and frequency. Hence, for a given tolerance limit to angular deviation from the straight course there exists an optimum radius for each wave length and an optimum wave length for each radius. This will be shown for the case of uniform sinusoidal bends, under the simplifying assumption that the cutoff radio $\nu \ll 1$.

Solving 7-5 for p one obtains

$$p \doteq 0.45\Delta^2 a^4 \lambda^{-4} \qquad 9\text{-}1$$

where p si the percentage increase in attenuation, and Δ the deviation angle in degrees.

Hence the average attenuation

$$\alpha_\Delta = \alpha(1 + 0.01p) \qquad 9\text{-}2$$

From 3–15 and 3–11

$$\alpha \doteq \frac{\nu^2 R_i}{a\eta} \doteq 10^{-3} R_i \lambda^2 a^{-3} \qquad 9\text{-}3$$

Introducing the R_i value from 4-8

$$\alpha \doteq 4.5\ 10^{-5} \rho_1 \lambda^{1.5} a^{-3} \qquad 9\text{-}4$$

where ρ is the high-frequency resistance of the wave guide relative to copper. From 9-1, 2 and 4

$$\alpha_\Delta = 4.5\ 10^{-6} \rho \lambda^{1.5} a^{-3}(1 + q\lambda^{-4} a^4) \qquad 9\text{-}5$$

with

$$q = 4.5\ 10^{-3} \Delta^2$$

The attenuation reaches a minimum when

$$f(\lambda, a) = \lambda^{1.5} a^{-3} + q\lambda^{-2.5} a = \text{minimum}$$

Case 1. *λ is given*

$$\delta f/\delta a = -3\lambda^{1.5} a^{-4} + q\lambda^{-2.5} = 0$$

$$a_{opt} = 1.32\lambda q^{-0.25} = 5.2\lambda\Delta^{-0.5}$$

From 9-5

$$\alpha_{\Delta opt} = 4\alpha = 1.29\ 10^{-7} \rho \lambda^{-1.5} \Delta^{1.5}$$

Case 2. *a is given*

$$\delta f/\delta\lambda = 1.5\lambda^{0.5} a^{-3} - 2.5 q\lambda^{-3.5} a = 0$$

$$\lambda_{opt} = 1.14 a q^{0.25} = 0.294 a \Delta^{0.5}$$

From 9-5

$$\alpha_{\Delta opt} = 1.6\alpha = 1.15\ 10^{-6} \rho a^{-1.5} \Delta^{0.75}$$

Numerical Example

Let $\Delta = 0.42°$
$a = 0.05\ m$
$\lambda = 0.01\ m$

From Table III

$$\alpha_\Delta = 1.50\,\alpha = 5.4\ 10^{-5}\rho \text{ neper}/m$$

Case 1: λ fixed at 0.01 m

$$a_{opt} = 0.08\ m$$

$$\alpha_{opt} = 2.76\ 10^{-5}\rho \text{ neper/meter}$$

Case 2: a fixed at 0.05 m

$$\lambda_{opt} = 0.0097\ m$$

$$\alpha_{opt} = 5.36\ 10^{-5}\rho \text{ neper}/m$$

Assuming sinusoidal bends with a 0.42° maximum deviation, the attenuation of centimeter waves can be reduced to one half by increasing the wave guide radius from 5 to 8 cm. For a 5 cm wave guide radius, 1 centimeter wavelength is close to the optimum.

A New Type of High-Frequency Amplifier

By J. R. PIERCE and W. B. HEBENSTREIT

This paper describes a new amplifier in which use is made of an electron flow consisting of two streams of electrons having different average velocities. When the currents or charge densities of the two streams are sufficient, the streams interact to give an increasing wave. Conditions for an increasing wave and the gain of the increasing wave are evaluated for a particular geometry of flow.

1. Introduction

IN CENTIMETER range amplifiers involving electromagnetic resonators or transmission circuits as, in klystrons and conventional traveling-wave tubes, it is desirable to have the electron flow very close to the metal circuit elements, where the radio-frequency field of the circuit is strong, in order to obtain satisfactory amplification. It is, however, difficult to confine the electron flow close to metal circuit elements without an interception of electrons, which entails both loss of efficiency and heating of the circuit elements. This latter may be extremely objectionable at very short wavelengths for which circuit elements are small and fragile.

In this paper the writers describe a new type of amplifier. In this amplifier the gain is not obtained through the interaction of electrons with the field of electromagnetic resonators, helices or other circuits. Instead, an electron flow consisting of two streams of electrons having different average velocities is used. When the currents or charge densities of the two streams are sufficient, the streams interact so as to give an increasing wave. Electromagnetic circuits may be used to impress a signal on the electron flow, or to produce an electromagnetic output by means of the amplified signal present in the electron flow. The amplification, however, takes place in the electron flow itself, and is the result of what may be termed an electromechanical interaction.[1,2]

While small magnetic fields are necessarily present because of the motions of the electrons, these do not play an important part in the amplification.

[1] Some electro-mechanical waves with a similar amplifying effect are described in "Possible Fluctuations in Electron Streams Due to Ions," J. R. Pierce, *Jour. App. Phys.*, Vol. 19, pp. 231–236, March 1948.

[2] While this paper was in preparation a classified report by Andrew V. Haeff entitled "The Electron Wave Tube—A Novel Method of Generation and Amplification of Microwave Energy" was received from the Naval Research Laboratory. Dr. Haeff's report (now declassified) contains a similar analysis of interaction of electron streams and in addition gives experimental data on the performance of amplifying tubes built in accordance with the new principle. We understand that similar work has been done at the RCA Laboratories.

The important factors in the interaction are the electric field, which stores energy and acts on the electrons, and the electrons themselves. The charge of the electrons produces the electric field; the mass of the electrons, and their kinetic energy, serve much as do inductance and stored magnetic energy in electromagnetic propagation.

By this sort of interaction, a traveling wave which increases as it travels, i.e., a traveling wave of negative attenuation, may be produced. To start such a wave, the electron flow may be made to pass through a resonator or a short length of helix excited by the input signal. Once initiated, the wave grows exponentially in amplitude until the electron flow is terminated or until non-linearities limit the amplitude. An amplified output can be obtained by allowing the electron flow to act on a resonator, helix or other output circuit at a point far enough removed from the input circuit to give the desired gain.

There are several advantages of such an amplifier. Because the electrons interact with one another, the electron flow need not pass extremely close to complicated circuit elements. This is particularly advantageous at very short wavelengths. Further, if we make the distance of electron flow between the input and output circuits long enough, amplification can be obtained even though the input and output circuits have very low impedance or poor coupling to the electron flow. Even though the region of amplification is long, there is no need to maintain a close synchronism between an electron velocity and a circuit wave velocity, as there is in the usual traveling-wave tube.

A companion paper by Dr. A. V. Hollenberg of these laboratories describes an experimental "double stream" amplifier tube consisting of two cathodes which produce concentric electron streams of somewhat different average velocity, and short helices serving as input and output circuits. No further physical description of double stream amplifiers will be given in this paper. Rather, a theoretical treatment of such devices will be presented.

2. Simple Theory

For simplicity we will assume that the flow consists of coincident streams of electrons of d-c. velocities u_1 and u_2 in the x direction. It will be assumed that there is no electron motion normal to the x direction. The treatment will be a small-signal or perturbation theory, in which products of a-c. quantities are neglected. M.K.S. units will be used. All quantities will be assumed to vary with time and distance $\exp j(\omega t - \beta x)$. The wavelength in the stream, λ_s, is then related to β by

$$\beta = 2\pi/\lambda_s \tag{1}$$

The following additional nomenclature will be used:

ϵ_0	dielectric constant of vacuum $\epsilon_0 = 8.85 \times 10^{-12}$ farad/meter
η	charge-to-mass ratio of the electron $\eta = 1.76 \times 10^{11}$ coulomb/kilogram
J_1, J_2	d-c. current densities
u_1, u_2	d-c. velocities
ρ_{01}, ρ_{02}	d-c. charge densities $\rho_{01} = -J_1/u_1$, $\rho_{02} = -J_2/u_2$
ρ_1, ρ_2	a-c. charge densities
v_1, v_2	a-c. velocities
V_1, V_2	d-c. voltages with respect to the cathode
V	a-c. potential

$\beta_1 = \omega/u_1$, $\beta_2 = \omega/u_2$

Although the small-signal equations relating charge density to voltage V have been derived many times, it seems well to present them for the sake of completeness. For one stream of electrons the first-order force equation is

$$\frac{dv_1}{dt} = \frac{\partial v_1}{\partial t} + \frac{\partial v_1}{\partial x} u_1 = \eta \frac{\partial V}{\partial x}$$

$$(\omega - \beta u_1)v_1 = -\eta\beta V$$

$$v_1 = \frac{-\eta\beta V}{u_1(\beta_1 - \beta)} \tag{2}$$

From the conservation of charge we obtain to the first order

$$\frac{\partial \rho_1}{\partial t} = -\frac{\partial}{\partial x}(\rho_{01}v_1 + \rho_1 u_1)$$

$$\omega\rho_1 = \rho_{01}\beta v_1 + u_1\beta\rho_1$$

$$\rho_1 = \frac{\rho_{01}\beta v_1}{u_1(\beta_1 - \beta)}$$

$$\rho_1 = -\frac{J_1\beta v_1}{u_1^2(\beta_1 - \beta)} \tag{3}$$

From (2) and (3) we obtain

$$\rho_1 = \frac{\eta J_1 \beta^2 V}{u_1^3(\beta_1 - \beta)^2} \tag{4}$$

We would find similarly

$$\rho_2 = \frac{\eta J_2 \beta^2 V}{u_2^3(\beta_2 - \beta)^2} \tag{5}$$

It will be convenient to call the fractional velocity separation b, so that

$$b = \frac{2(u_1 - u_2)}{u_1 + u_2} \tag{6}$$

It will also be convenient to define a sort of mean velocity u_0

$$u_0 = \frac{2u_1 u_2}{u_1 + u_2} \tag{7}$$

We may also let V_0 be the potential drop specifying a velocity u_0, so that

$$u_0 = \sqrt{2\eta V_0} \tag{8}$$

It is further convenient to define a phase constant based on u_0

$$\beta_0 = \frac{\omega}{u_0} \tag{9}$$

We see from (6), (7) and (9) that

$$\beta_1 = \beta_0(1 - b/2) \tag{10}$$

$$\beta_2 = \beta_0(1 + b/2) \tag{11}$$

We shall treat only a special case, that in which

$$\frac{J_1}{u_1^3} = \frac{J_2}{u_2^3} = \frac{J_0}{u_0^3} \tag{12}$$

Here J_0 is a sort of mean current which, together with u_0, specifies the ratios J_1/u_1^3 and J_2/u_2^3, which appear in (4) and (5).

In terms of these new quantities, the expression for the total a-c. charge density ρ is, from (4) and (5) and (8)

$$\rho = \rho_1 + \rho_2 = \frac{J_0\beta^2}{2u_0 V_0} \cdot \left[\frac{1}{\left[\beta_0\left(1 - \frac{b}{2}\right) - \beta\right]^2} + \frac{1}{\left[\beta_0\left(1 + \frac{b}{2}\right) - \beta\right]^2}\right] V \tag{13}$$

Equation (13) is a *ballistical* equation telling what charge density ρ is produced when the flow is bunched by a voltage V. To solve our problem, that is, to solve for the phase constant β, we must associate (13) with a *circuit* equation which tells us what voltage V the charge density produces. We assume that the electron flow takes place in a tube too narrow to propagate a wave of the frequency considered. Further, we assume that the wave velocity is much smaller than the velocity of light. Under these circumstances the circuit problem is essentially an electrostatic problem. The a-c. voltage will be of the same sign as, and in phase with, the a-c. charge density ρ. In other words, the "circuit effect" is purely capacitive.

Let us assume at first that the electron stream is very narrow compared with the tube through which it flows, so that V may be assumed to be constant over its cross section. We can easily obtain the relation between

V and ρ in two extreme cases. If the wavelength in the stream, λ_s, is very short (β large), so that transverse a-c. fields are negligible, then from Poisson's equation we have

$$\rho = -\epsilon_0 \frac{\partial^2 V}{\partial x^2}$$

$$\rho = \epsilon_0 \beta^2 V \tag{14}$$

If, on the other hand, the wavelength is long compared with the tube radius (β small) so that the fields are chiefly transverse, the lines of force running from the beam outward to the surrounding tube, we may write

$$\rho = CV \tag{15}$$

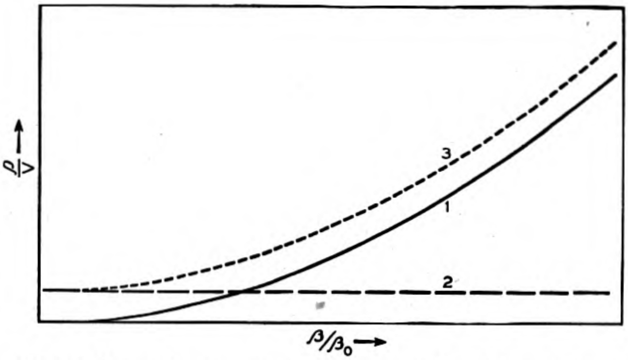


Fig. 1—A "circuit" curve for a narrow electron stream in a tube. The ratio of the a-c. charge density ρ to the a-c. voltage V produced by the charge density is plotted vs. a parameter β/β_0, which is inversely proportional to the wavelength λ_s in the flow. Curve 1 holds for very large values of β/β_0; curve 2 holds for very small values of β/β_0, and curve 3 over-all shows approximately how ρ/V varies for intermediate values of β/β_0.

Here C is a constant expressing the capacitance per unit length between the region occupied by the electron flow and the tube wall.

We see from (14) and (15) that if at some particular frequency we plot ρ/V vs. β/β_0 for real values of β, ρ/V will be constant for small values of β and will rise as β^2 for large values of β, approximately as shown in Fig. 1. For another frequency, β_0 would be different and, as ρ/V is a function of β, the horizontal scale of the curve would be different.

Now, we have assumed that the charge is produced by the action of the voltage, according to the ballistical equation (10). This relation is plotted in Fig. 2, for a relatively large value of J_0/u_0V_0 (curve 1) and for a smaller value of J_0/u_0V_0 (curve 2). There are poles at $\beta/\beta_0 = 1 \pm \frac{b}{2}$, and a minimum between the poles. The height of the minimum increases as J_0/u_0V_0 is increased.

A circuit curve similar to that of Fig. 1 is also plotted on Fig. 2. We see that for the small-current case (curve 2) there are four intersections, giving *four real* values of β and hence *four unattenuated* waves. However, for the larger current (curve 1) there are only two intersections and hence two unattenuated waves. The two additional values of β satisfying both the circuit equation and the ballistical equation are complex conjugates, and represent waves traveling at the same speed, but with equal positive and negative attenuations.

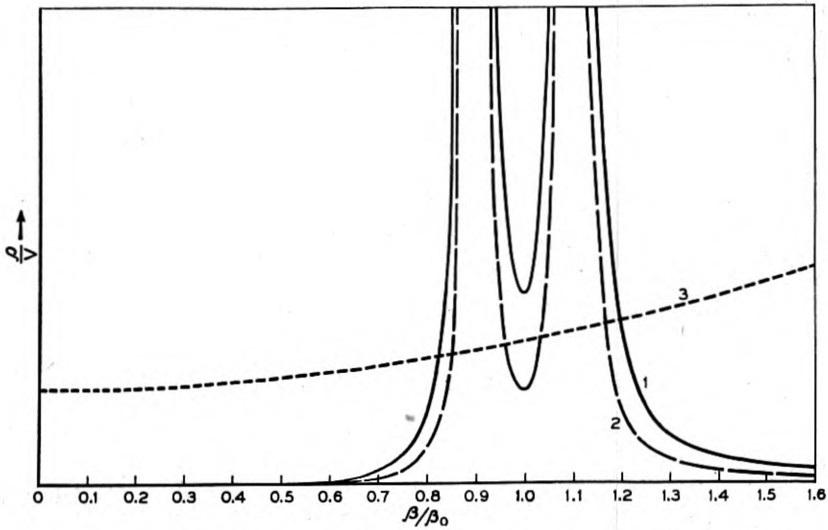


Fig. 2—Curve 3 is a circuit curve similar to that of Fig. 1. Curves 1 and 2 are based on a ballistical equation telling how much charge density ρ is produced when the voltage V acts to bunch a flow consisting of electrons of two velocities. The abscissa, β/β_0, is proportional to phase constant. Intersections of the circuit curve with a ballistical curve represent waves. Curve 2 is for a relatively small current. In this case intersections occur for four real values of β, so the four waves are unattenuated. For a larger current (curve 1) there are two intersections (two unattenuated waves). For the other two waves β is complex. There are an increasing and a decreasing wave.

Thus we deduce that, as the current densities in the electron streams are raised, a wave with negative attenuation appears for current densities above a certain critical value.

We can learn a little more about these waves by assuming an approximate expression for the circuit curve of Fig. 1. Let us merely assume that over the range of interest (near $\beta/\beta_0 = 1$) we can use

$$\rho = \alpha^2 \epsilon_0 \beta^2 V \tag{16}$$

Here α^2 is a factor greater than unity, which merely expresses the fact that the charge density corresponding to a given voltage is somewhat greater than if there were field in the x direction only and equation (11) were valid. Combining (16) with (13) we obtain

$$\frac{1}{\left(\beta_0\left(1-\frac{b}{2}\right)-\beta\right)^2}+\frac{1}{\left(\beta_0\left(1+\frac{b}{2}\right)-\beta\right)^2}=\frac{1}{\beta_0^2 U^2} \tag{17}$$

where

$$U = \frac{J_0}{2\alpha^2 \epsilon_0 \beta_0^2 u_0 V_0} \tag{18}$$

In solving (17) it is most convenient to represent β in terms of β_0 and a new variable δ

$$\beta = \beta_0(1 + \delta) \tag{19}$$

Thus, (14) becomes

$$\frac{1}{\left(\delta-\frac{b}{2}\right)^2}+\frac{1}{\left(\delta+\frac{b}{2}\right)^2}=\frac{1}{U^2} \tag{20}$$

Solving for δ, we obtain

$$\delta = \pm\left(\frac{b}{2}\right)\left[\left(\frac{2U}{b}\right)^2 + 1 \pm\left(\frac{2U}{b}\right)\sqrt{\left(\frac{2U^2}{b}\right)+4}\right]^{1/2}. \tag{21}$$

The positive sign inside of the brackets always gives a real value of δ and hence unattenuated waves. The negative sign inside the brackets gives unattenuated waves for small values of U/b. However, when

$$\left(\frac{U}{b}\right)^2 > \frac{1}{8} \tag{22}$$

there are two waves with a phase constant β_0 and with equal and opposite attenuation constants.

Suppose we let U_M be the minimum value of U for which there is gain. From (22),

$$U_{M^2} = b^2/8 \tag{23}$$

From (21) we have for the increasing wave

$$\delta = j\frac{b}{2}\left[\frac{1}{2}\left(\frac{U}{U_M}\right)^2\left(\sqrt{1+8\left(\frac{U}{U_M}\right)^2}-1\right)-1\right]^{1/2} \tag{24}$$

The gain in db/wavelength is

$$\begin{aligned} \text{db/wavelength} &= 20(2\pi)\log_{10} e\,|\,\delta\,| \\ &= 54.6\,|\,\delta\,| \end{aligned} \tag{25}$$

We see that by means of (24) and (25) we can plot db/wavelength per unit b vs. $(U/U_M)^2$. This is plotted in Fig. 3. Because U^2 is proportional to current, the variable $(U/U_M)^2$ is the ratio of the actual current to the current which will just give an increasing wave. If we know this ratio, we can obtain the gain in db/wavelength by multiplying the corresponding ordinate from Fig. 3 by b.

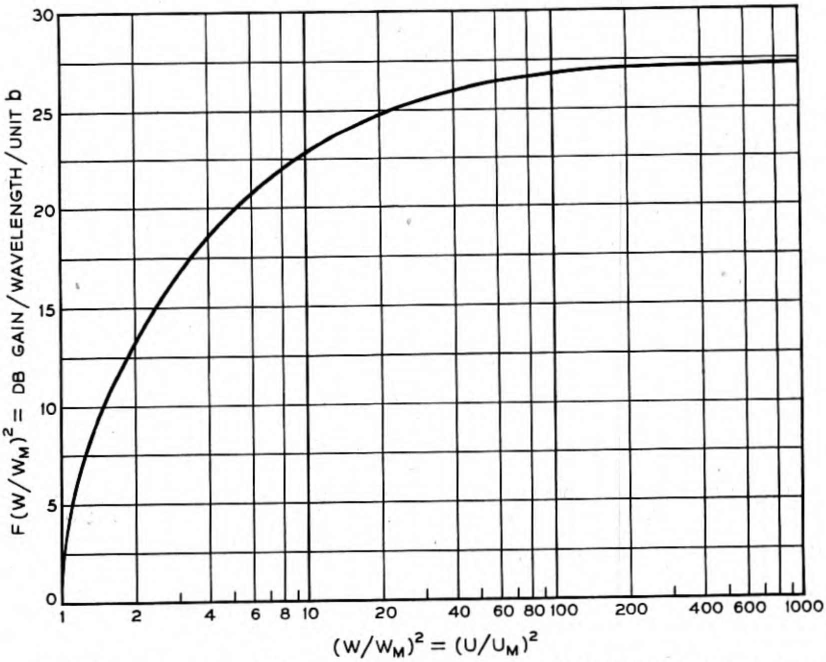


Fig. 3—The parameter $(W/W_M)^2 = ((U/U_M)^2$ is proportional to current. As the current is increased above a critical value for which $(W/W_M)^2 = 1$, there is an increasing wave of increasing gain. In this curve the gain per wavelength per unit b, called $F(W/W_M)^2$, is plotted vs. $(W/W_M)^2$. For large values of $(W/W_M)^2$, $F(W/W_M)^2$ approaches 27.3 and the gain per wavelength approaches 27.3 b.

We see that, as the current is increased, the gain per wavelength at first rises rapidly and then rises more slowly, approaching a value of $27.3b$ db/wavelength for very large values of $(U/U_M)^2$.

We now have some idea of the variation of gain per wavelength with velocity separation b and with current $(U/U_M)^2$. A more complete theory would require the evaluation of the lower limiting current for gain (or of U_M^2) in terms of physical dimensions and an investigation of the boundary conditions to show how strong an increasing wave is set up by a given input signal. The latter problem will not be considered in this paper; the former is dealt with in the third section and in the appendix.

3. Design Curves

It is proposed to present in this section material for actually evaluating the gain of the increasing wave for a particular geometry of electron flow. In this section there is some repetition from earlier sections, so that the material presented can be used without referring unduly to section 2. In order to avoid confusion, much of the mathematical work on which the section is based has been put in the appendix.

The flow considered is one in which electrons of two velocities, u_1 and u_2, corresponding to accelerating voltages V_1 and V_2, are intermingled, the corresponding current densities J_1 and J_2 being constant over the flow. The flow occupies a cylindrical space of radius a. It is assumed that the surrounding cylindrical conducting tube is so remote as to have negligible effect on the a-c. fields.

It will be assumed, according to (12), that the current densities and the voltages V_1 and V_2 are specified in terms of a "mean" current J_0 and a "mean" voltage V_0 corresponding to a velocity u_0, by

$$\frac{J_1}{V_1^{3/2}} = \frac{J_2}{V_2^{3/2}} = \frac{J_0}{V_0^{3/2}} \tag{12a}$$

The gain will depend on the beam radius, the free-space wavelength λ, and on J_0 and V_0, and on the fractional velocity separation

$$b = \frac{2(u_1 - u_2)}{u_1 + u_2} \tag{6}$$

The wavelength in the beam, λ_s, which is associated with the voltage V_0 is given by

$$\lambda_s = \lambda \frac{u_0}{c} = \lambda \frac{\sqrt{2\eta V_0}}{c}$$

$$\lambda_s = 1.98 \times 10^{-3} \lambda \sqrt{V_0} \tag{26}$$

Here c is the velocity of light.

A dimensionless parameter W is defined to be

$$W^2 = \frac{\omega_e^2}{\omega^2} = \frac{J_0}{\epsilon_0 u_0 \omega^2} \tag{27}$$

$$W^2 = 8.52 \times 10^6 \frac{J_0}{f\sqrt{V_0}} \tag{28}$$

Here ω_e is the electron plasma frequency associated with the average space charge density J_0/u_0, and ω is the radian frequency corresponding to the wavelength λ. In (28), the constant is adjusted so that J_0 is expressed in amperes per square centimeter rather than in amperes per square meter, while f is expressed in megacycles.

Below a minimum value of W, which will be called W_M, there is no gain. W_M is a function of the velocity separation b and of the ratio of the beam radius a to the beam wavelength, λ_s. A plot of $(W_M/b)^2$ as a function of (a/λ_s) is shown in Fig. 4.

The variation of gain in the interval, $W_M \gtrless W < \infty$, is shown in Fig. 3 where "Decibels gain/wavelength/unit b" is plotted as a function of $(W/W_M)^2$. This is the same curve which was derived in section 2. The

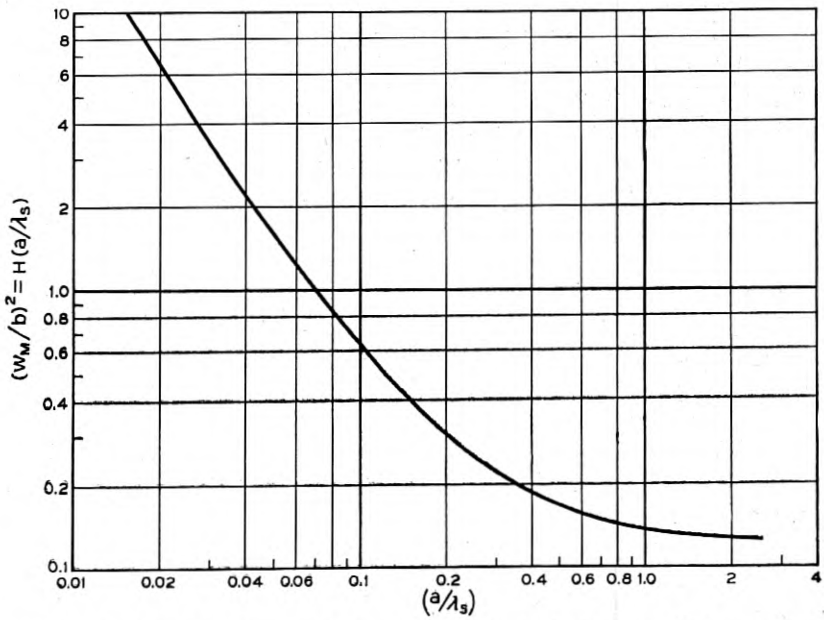


Fig. 4—As the ratio of beam radius a to wavelength in the beam, λ_s, is increased, the critical value of W, W_M, decreases and less current is needed in order to obtain gain. Here $(W_M/b)^2$, which is called $H(a/\lambda_s)$, is plotted vs. (a/λ_s).

ratio $(W/W_M)^2$ is the same as the parameter $(U/U_M)^2$ used there, although U and W are not the same.

The curve in Fig. 3 is useful in that it reduces the interdependence of a large number of parameters to a single curve. However, there are cases as, for example, when one is computing the bandwidth of an amplifier, in which it would be more convenient to have the curve in Fig. 3 broken up into a family of curves. We can do this by the following means:

We can write the gain in db/wavelength in the form

$$\text{db/wavelength} = bF(W/W_M)^2 \qquad (29)$$

Here $F(W/W_M)^2$ is the function plotted in Fig. 3. If ℓ is the total length of the flow, the total gain in db, G, will thus be

$$G = \frac{\ell b}{\lambda_s} F(W/W_M)^2 \tag{30}$$

We will now express $(W/W_M)^2$ in such a form as to indicate its dependence on wavelength in the beam, λ_s. We can write from (27)

$$W = \frac{\omega_e^2}{\omega^2} = \frac{\lambda_s^2}{\lambda_e^2} \tag{31}$$

Here λ_e is a "plasma wavelength," defined by the relation

$$\lambda_e = \frac{u_0}{(\omega_e/2\pi)} \tag{32}$$

We further have

$$W_M^2 = b^2 H(a/\lambda_s) \tag{33}$$

Here $H(a/\lambda_s)$ is the function of (a/λ_s) which is plotted in Fig. 5.

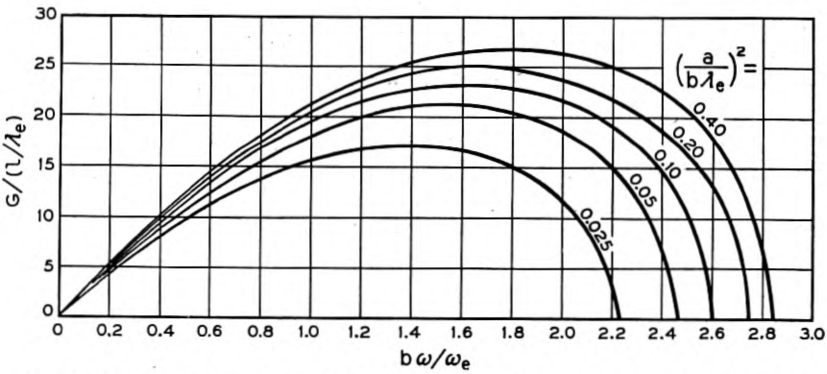


Fig. 5—In these curves the total gain in db, G, divided by the ratio of the length ℓ to the plasma wavelength λ_e, is plotted vs. $b\omega/\omega_e$, which is proportional to frequency, for several values of the parameter $(a/b\lambda_e)^2$. Changing b, the velocity seperation, changes both the parameter and the frequency scale.

Now, from (26), (27), and (29) we can write

$$G = \left(\frac{\ell}{\lambda_e}\right)\left(\frac{b\lambda_e}{a}\right)\left(\frac{a}{\lambda_s}\right) F\left[\left(\frac{\lambda_s}{a}\right)^2 \left(\frac{a}{b\lambda_e}\right)^2 \frac{1}{H(a/\lambda_s)}\right] \tag{34}$$

For a given tube the parameters (ℓ/λ_e) and $(a/b\lambda_e)$ do not vary with frequency, while (a/λ_s) is proportional to frequency. Hence, we can construct universal frequency curves by plotting $G/(\ell/\lambda_e)$ vs. (a/λ_s) for various values of the parameter $(a/b\lambda_e)$. It is more convenient, however, to use as an abscissa $b\lambda_e/\lambda_s = b\omega/\omega_e$, and this has been done in Fig. 5.

In order to use these curves it is necessary to express the parameters $b\omega/\omega_e$, λ_e and $(a/b\lambda_e)^2$ in terms of convenient physical quantities. We obtain

$$\begin{aligned} b\omega/\omega_e &= .545 \times 10^{-10} bV^{1/4}\omega/J_0^{1/2} \\ \lambda_e &= 2.04 \times 10^{-2} V_0^{3/4}/J_0^{1/2} \\ (a/b\lambda_e)^2 &= 767\ I_0/b^2V^{3/2} \end{aligned} \tag{35}$$

Here I_0 is current in amperes and J_0 is in amperes / cm.^{2}

The broadness of the frequency response curves of Fig. 5 is comparable to that of curves for helix-type traveling-wave tubes.

It is interesting to note that the maximum value of $G/(\ell/\lambda_e)$ varies little for a considerable range of the parameter $a/b\lambda_e$, approaching a constant for large values of the parameter. This means that, with a beam of given length, velocity and charge density, one can obtain almost the same optimum gain over a wide range of frequencies simply by adjusting the velocity-separation parameter b.

4. Concluding Remarks

There is a great deal of room for extension of the theory of double-stream amplifiers. This paper has not dealt with the setting up of the increasing wave, nor with other geometries than that of a cylindrical beam in a very remote tube, nor with the effect of physical separation of the electron streams of two velocities nor with streams of many velocities or streams with continuous velocity distributions.

This last is an interesting subject in that it may provide a means for dealing with problems of noise in multivelocity electron streams. Indeed, it was while attempting such a treatment that the writers were distracted by the idea of double-stream amplification.

APPENDIX

Derivation of Results Used in Section 3

Consider a double-stream electron beam whose axis coincides with the z-axis of a system of cylindrical coordinates (r, φ, z) and which is subject to an infinite, longitudinal, d-c. magnetic field. The radius of the beam is a and each of the streams is characterized by d-c. velocities, u_1 and u_2, which are vectors in the positive z direction, and d-c. space charge densities, ρ_{02} and ρ_{02}. All d-c. quantities are assumed to be independent of the coordinates and time, except, of course, for the discontinuities at the surface of the beam. Small a-c. disturbances are superimposed upon these d-c. quantities and they are small enough so that their cross products can be neglected compared with the products of d-c. quantities and a-c. quantities. It is further assumed that only those a-c. quantities are allowed which have no azimuthal variation, that is, $\frac{\partial}{\partial\varphi} = 0$. Fig. 6 shows the electron beam.

Outside the beam the appropriate Maxwell's equations are

$$\frac{1}{r}\frac{\partial}{\partial r}(rH_\varphi) = j\frac{k}{\eta_0}E_2 \tag{A-1}$$

$$\frac{\partial H_\varphi}{\partial z} = -j\frac{k}{\eta_0}E_r \tag{A-2}$$

$$\frac{\partial E_z}{\partial r} - \frac{\partial E_r}{\partial z} = jk\eta_0 H_\varphi \tag{A-3}$$

where

$$k = \frac{\omega}{c} \tag{A-4}$$

$$\eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 377 \text{ ohms} \tag{A-5}$$

Inside the beam, equations (A-2) and (A-3) remain the same, but instead of equation (A-1) we have

$$\frac{1}{r}\frac{\partial}{\partial r}(rH_\varphi) = j\frac{k}{\eta_0}E_z + q_1 + q_2 \tag{A-6}$$

where q_1 and q_2 are the first order a-c. convection current densities of the two streams. These quantities can be calculated from the force equation and the equation for the conservation of charge. Assuming that all a-c. quantities vary as $\exp j(\omega t - \beta z)$, the force equation is (for stream number one, say)

$$j\omega v_1 - j\beta u_1 v_1 = -(e/m)E_z \tag{A-7}$$

and the equation for the conservation of charge is

$$j\beta\rho_{01}v_1 + j\beta u_1\rho_1 = +j\omega\rho_1 \tag{A-8}$$

Equations (A-7) and (A-8) can be solved for v_1 and ρ_1 :

$$v_1 = \frac{-(e/m)E_z}{j\omega\left(1 - \frac{\beta}{\beta_1}\right)} \tag{A-7a}$$

$$\rho_1 = \frac{\beta\rho_{01}}{\omega\left(1 - \frac{\beta}{\beta_1}\right)} v_1 \tag{A-8a}$$

where

$$\beta_1 = \frac{\omega}{u_1}$$

Combining equations (A-7a) and (A-8a) one has

$$\rho_1 = \frac{j\beta\rho_{01}(e/m)E_z}{\omega^2\left(1 - \frac{\beta}{\beta_1}\right)^2} \tag{A-9}$$

The first order a-c. convection current density is given by

$$q_1 = \rho_{01}v_1 + \rho_1 u_1 \tag{A-10}$$

which, by combining with (A-7a) and (A-8b), becomes

$$q_1 = \frac{j(k/\eta_0)(\rho_{01}/m\epsilon_0)E_z}{\omega^2\left(1 - \frac{\beta}{\beta_1}\right)^2} \tag{A-11}$$

Similarly

$$q_2 = \frac{j\,\frac{k}{\eta_0}\,\rho_{01}\,\frac{e}{m\epsilon_0}\,E_z}{\omega^2\left(1 - \frac{\beta}{\beta_2}\right)^2} \tag{A-12}$$

If we now define

$$\beta_0 = \tfrac{1}{2}(\beta_1 + \beta_2) \tag{A-13}$$

$$B_1 = \frac{\beta_1}{\beta_0}; \qquad B_2 = \frac{\beta_2}{\beta_0} \tag{A-14}$$

and let

$$Z = \frac{\beta}{\beta_0} \tag{A-15}$$

$$W_1 = \frac{\omega_{e1}}{\omega}; \qquad W_2 = \frac{\omega_{e2}}{\omega} \tag{A-16}$$

where ω_e, the plasma-electron angular frequency given by

$$\omega_{e1}^2 = -\frac{e\rho_{01}}{m\epsilon_0}, \text{ etc.} \tag{A-17}$$

Equations (11) and (12) become

$$q_1 = \frac{-j(k/\eta_0)W_1^2 B_1^2}{(Z - B_1)^2}\,E_z \tag{A-18}$$

$$q_2 = \frac{-i(k/\eta_0)W_2^2 B_2^2}{(Z - B_2)^2} \tag{A-19}$$

Thus equation (A-6) becomes

$$\frac{1}{r}\frac{\partial}{\partial r}(rH_\varphi) = j\,\frac{k}{\eta_0}\,LE_z \tag{A-6a}$$

where

$$L = 1 - \frac{W_1^2 B_1^2}{(Z - B_1)^2} - \frac{W_2^2 B_2^2}{(Z - B_2)^2} \tag{A-20}$$

If we assume that the tube which surrounds the beam be taken as infinitely remote, the appropriate solutions outside the beam are

$$\hat{H}_{\varphi 0} = A_0 K_1(\gamma r) \tag{A-21}$$

$$\hat{E}_{z0} = j \frac{\eta_0 \gamma}{k} A_0 K_0(\gamma r) \tag{A-22}$$

and inside the beam

$$\hat{H}_{\varphi i} = A_i I_1(\xi r) \tag{A-23}$$

$$\hat{E}_{zl} = -j \frac{\eta_0 \gamma}{\sqrt{L}} A_i I_0(\xi r) \tag{A-24}$$

where

$$\begin{aligned} \gamma^2 &= \beta^2 - k^2 \approx \beta^2 \\ \xi^2 &= \gamma^2 L \end{aligned} \tag{A-25}$$

The I's and K's in equations (A-21)–(A-24) are modified Bessel functions.[3] At the surface of the beam ($r = a$), one has the following two independent boundary conditions

$$\hat{H}_{\varphi i} = \hat{H}_{\varphi 0} \tag{A-26}$$

$$\hat{E}_{zi} = \hat{E}_{z0} \tag{A-25a}$$

which, using equations (A-21)–(A-24), yield

$$\frac{I_0(\xi_a)}{\sqrt{L}\, I_1(\xi_a)} = -\frac{K_\eta(\gamma a)}{K_1(\gamma a)} \tag{A-27}$$

From equations (A-13), (A-14), (A-15) and (A-24) one has

$$\xi a = Z\beta_0 a \sqrt{L} \tag{A-28}$$

$$\gamma a = Z\beta_0 a \tag{A-29}$$

If we now define a beam wavelength, λ_s, by the relations

$$\beta_0 = \frac{2\pi}{\lambda_s} \tag{A-30}$$

and assume for the purpose of simplifying the calculation that in the expression for L in (A-20)

$$W_1^2 B_1^2 = W_2^2 B_2^2 = W^2 \tag{A-31}$$

[3] See A Treatise on the Theory of Bessel Functions, G. N. Watson, Chapter 3.

We easily see that

$$W^2 = (\omega_e/\omega)^2 \tag{A-32}$$

where

$$\omega_e = \frac{e}{m} J_0/\epsilon_0 u_0 \tag{A-33}$$

We obtain from (A-20), (A-28), (A-29) and (A-30)

$$\left[\frac{K_1\left(\frac{2\pi a Z}{\lambda_s}\right) I_0\left(\sqrt{L}\,\frac{2\pi a Z}{\lambda_s}\right)}{K_0\left(\frac{2\pi a Z}{\lambda_s}\right) I_1\left(\sqrt{L}\,\frac{2\pi a Z}{\lambda_s}\right)}\right]^2 = L$$
$$= 1 - \left[\frac{W^2}{(Z - B_1)^2} + \frac{W^2}{(Z - B_2)^2}\right] \tag{A-34}$$

Equation (A-31) is equivalent to Equation (12) of the text or to

$$\frac{J_1}{V_1^{3/2}} = \frac{J_2}{V_2^{3/2}} \tag{A-35}$$

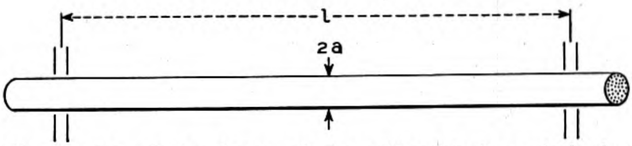


Fig. 6—The diameter of the electron flow considered is $2a$, and the length is ℓ.

Letting $Y = -L$ and making use of the following well known relations between the Bessel functions

$$I_0(jx) = J_0(x)$$
$$I_1(jx) = jJ_1(x) \tag{A-36}$$

Equation (A-34) becomes

$$Y = \left[\frac{K_1\left(\frac{2\pi a}{\lambda_s} Z\right) J_0\left(\sqrt{Y}\,\frac{2\pi a}{\lambda_s} Z\right)}{K_0\left(\frac{2\pi a}{\lambda_s} Z\right) J_1\left(\sqrt{Y}\,\frac{2\pi a}{\lambda_s} Z\right)}\right]^2$$
$$= \frac{W^2}{(Z - B_1)^2} + \frac{W^2}{(Z - B_2)^2} - 1 \tag{A-37}$$

Let the right-hand side of equation (A-37) be denoted by $F_1(Z)$ and the middle of $F_2(Z)$. In order to find the real roots of equation (A-37) one can plot F_1 and F_2 as functions of Z on the same chart. The abscissae of the intersections of the two curves will then be the real roots. In Fig. 7, F_1 is plotted as a function of Z for $B_1 = 0.9$ and $B_2 = 1.1$.

In view of the definitions in equations (A-13) and (A-14), both B_1 and B_2 are uniquely defined by a single parameter, namely, the fractional velocity separation, b. That is

$$b = 2(u_1 - u_2)/(u_1 + u_2) = 2(\beta_2 - \beta_1)/(\beta_2 + \beta_1) = B_2 - B_1 \qquad \text{(A-38)}$$

and

$$B_1 = 1 - (b/2) \qquad B_2 = 1 + (b/2) \qquad \text{(A-39)}$$

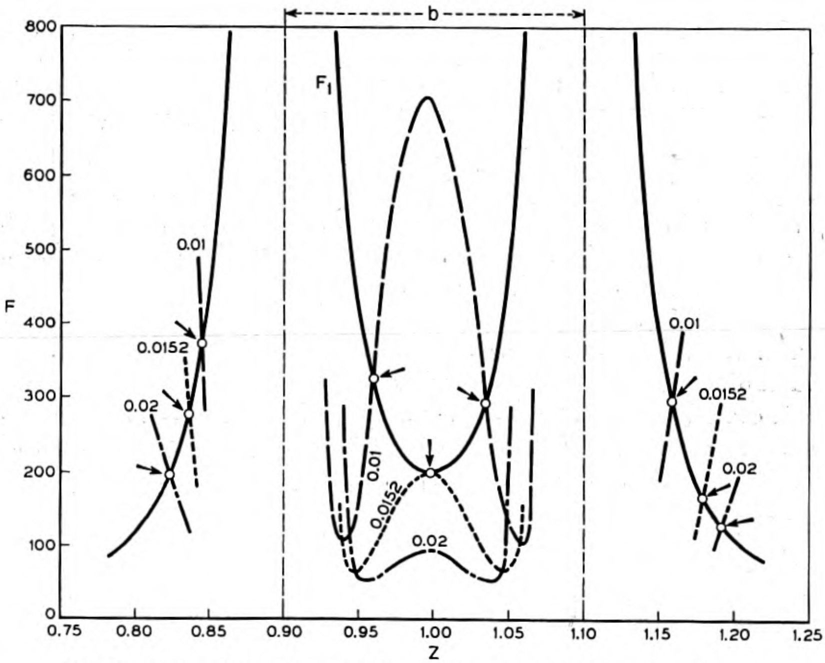


Fig. 7—A curve illustrating conditions giving rise to various types of roots.

A complete plot of F_2, for any value of the parameters W and (a/λ_s), would show that equation (A-37) has an infinite number of real solutions. A real solution of equation (A-37) means an unattenuated wave. Thus there are an infinite number of unattenuated waves possible. The waves which will actually be excited in any given case, however, depend upon the boundary conditions at the input and output of the tube. Ordinarily only those waves will be excited which do not have a reversal in phase of the longitudinal E vector, say, as r varies from 0 to a. Attention, therefore, will be given only to those waves for which E_z does not change sign over a cross-section of the beam. By inspection of equation (A-23), it is evident that this requirement is automatically satisfied if $L > 0$. On the other hand, if L is negative, one has

$$E_{zi} \sim J_0\left(\sqrt{Y}\,\frac{2\pi r}{\lambda_s}\,Z\right) \qquad \text{(A-23a)}$$

Thus attention will be limited to those roots which satisfy

$$\sqrt{Y}\,\frac{2\pi a}{\lambda_s}\,Z < 2.405 \qquad \text{(A-40)}$$

where 2.405 is the first zero of the Bessel function in equation (A-21a).

Returning to Fig. 7, portions of three different F_2 curves are plotted: one for $W^2 = 0.01$, one for $W^2 = 0.0152$ and one for $W^2 = 0.02$. All three curves are for $(a/\lambda_s) = 0.16$. The intersections which represent roots which satisfy the inequality (A-40) are marked with arrows. Evidently there are either four real roots of this type or there are two real roots and a complex conjugate pair, the distinction being determined by the value of W. Thus there is a critical value of W^2 (in this case it is 0.0152) for which two of the real roots are identical. This identical pair is indicated by two arrows near the minimum of the F_1 curve at $Z = 1$.

A pair of conjugate complex roots means that there are an increasing wave and a decreasing wave. Thus for each value of b and (a/λ_s) there is a least value of W^2 below which the tube will have no gain.

It can be shown that the critical tangency of the F_1 and F_2 curves occurs at a value of Z which is less than b^2 away from unity. Very little error will be incurred, then, by assuming that this critical point occurs at $Z = 1$ if b is small.

Letting $Z = 1$ in equation (A-37), and using equations (A-39) one has

$$8(W_M/b)^2 - 1 = \left(\frac{K_1(2\pi a/\lambda_s)J_0\left(\sqrt{8(W_M/b)^2 - 1}\;2\pi a/\lambda_s\right)}{K_0(2\pi a/\lambda_s)J_1\left(\sqrt{8(W_M/b)^2 - 1}\;2\pi a/\lambda_s\right)}\right) \qquad \text{(A-41)}$$

where W_M is the critical value of W. Equation (A-41) determines $(W_M/b)^2$ as af unction of (a/λ_s). This relationship is plotted in Fig. 4.

We will find that there will be an increasing wave in the range $W_M \leq W < \infty$. The calculation of the gain in this interval would be very laborious since Bessel functions of complex argument would be involved. However, a good approximation can be made when b is small. The real part of Z will always be near unity and the imaginary part will be found to be less than $b/2$. Therefore one can let $Z = 1$ in equation (A-37) where it multiplies the factor $(2\pi a/\lambda_a)$ in the argument of the Bessel functions and let $Z - 1 = U$ in the right-hand side of Equation (A-37). With these assumptions Y can be determined as a function of (a/λ_s) and U can be determined as a function of Y. We have from Equation (A-37)

$$\frac{1}{(U + b/2)^2} + \frac{1}{(U - b/2)^2} = \frac{1 + Y}{W^2} \tag{A-37a}$$

When $U = 0$, $W^2 = W_M^2 = W_M^2$, so that

$$1 + Y = 8(W_M/b)^2 \tag{A-42}$$

and equation (A-37a) becomes

$$\frac{1}{(U + b/2)^2} + \frac{1}{(U - b/2)^2} = (8/b^2)(W_M/W)^2 \tag{A-37b}$$

the solution of which, for the increasing wave, is

$$U = j(b/2)[(1/2)(W/W_M)^2(\sqrt{1 + 8(W_M/W)^2} - 1) - 1]^{\frac{1}{2}} \tag{A-43}$$

and the gain will be given by

$$\text{Gain}/b = 27.3[(1/2)(W/W_M)^2(\sqrt{1 + 8(W_M/W)^2} - 1) - 1]^{\frac{1}{2}} \tag{A-44}$$

$$\text{db/wavelength/unit } b$$

"Decibels gain/wavelength/unit b" is plotted against $(W/W_M)^2$ in Fig. 3.

As $(W/W_M)^2$ becomes very large, the gain per wavelength approaches 27.3 b db.

Experimental Observation of Amplification by Interaction Between Two Electron Streams

By A. V. HOLLENBERG

The construction and performance of an amplifier employing the interaction between two streams of electrons having different average velocities are described. Gain of 33 db at a center frequency of 255 Mc has been observed with bandwidth of 110 Mc between 3 db points.

1. Introduction

A NEW type of amplifier in which the gain is obtained by an interaction between streams of electrons of two or more average velocities is proposed in a companion paper by Pierce and Hebenstreit.[1] This amplifier contains input and output portions in which signals are impressed on and extracted from the electron flow by electromagnetic circuits and a central portion in which gain occurs purely by interaction between streams of electrons without any circuits being involved. A small signal theory for coincident electron streams of two d-c. velocities is presented in Pierce and Hebenstreit's paper.

In this paper a description of the construction and operation of an amplifier of this kind will be presented. Departures of the actual conditions in the amplifier from the assumptions of the theory limit the expectations of quantitative agreement. It is believed, however, that the evidence for gain arising from the interaction between two streams of electrons is clear, and that the broad frequency response predicted by the theory has been confirmed.

2. Description of Amplifier

The frequency range near 200 Mc was chosen for the first experimental test of the proposed method of amplification for reasons of convenience. The theory indicates that current density requirements increase with frequency, but that these requirements become severe only at the higher microwave frequencies. Availability of circuit parts and test equipment, rather than anticipated difficulties at higher frequencies, led to the choice that was made.

The essential features of one of the double-stream amplifier tubes which has been constructed and operated are represented in Fig. 1. The output helix was identical with the input helix in construction and connection

[1] A New Type of High Frequency Amplifier, J. R. Pierce and W. B. Hebenstreit, this issue of the *B. S. T. J.*

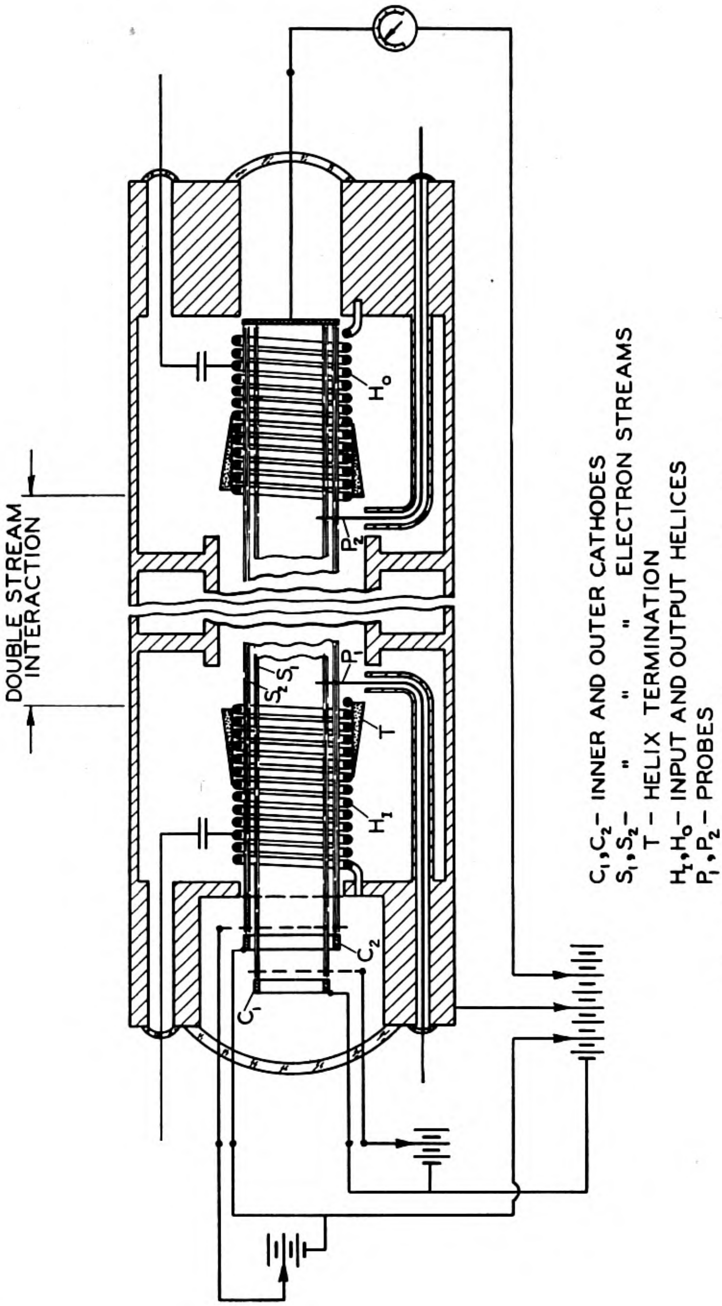


Fig. 1—Representation of double-stream amplifier with helix output.

to the coaxial line. The two identical probes p_1, and p_2, extending from coaxial lines into the two electron streams at the beginning and end of the central portion of the tube between the two helices were inserted for comparison of the signal amplitudes at the beginning and end of the region in which no circuit is present.

A similar tube containing an output gap in place of the output helix section is represented in Fig. 2.

In both cases concentric tubular electron streams originate at the ring-shaped emitting surfaces of the two cathodes at potentials V_1 and V_2, pass through their respective control grids and through a common accelerating grid. An axial magnetic field of approximately 700 gauss is applied in order to maintain the definition of the beams. The outer and inner tubular beams

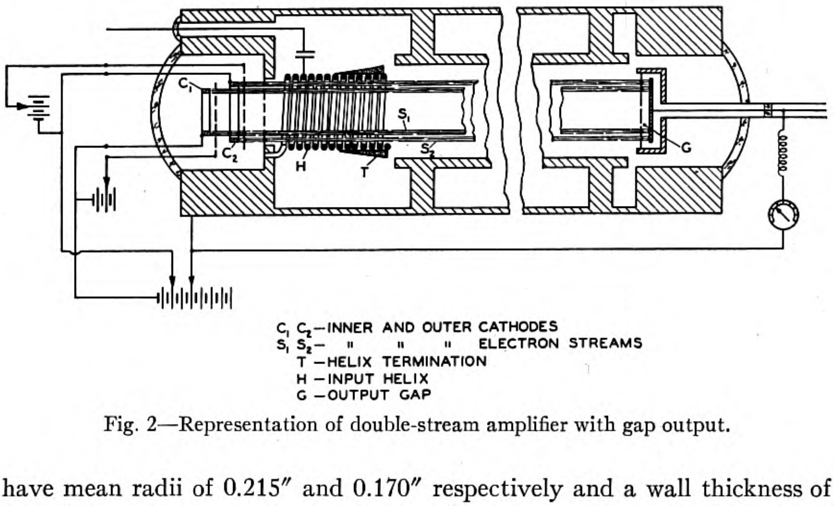


Fig. 2—Representation of double-stream amplifier with gap output.

have mean radii of 0.215″ and 0.170″ respectively and a wall thickness of 0.030″ in each case.

The short sections of helix which are used for input and output are wound of 0.013″ diameter molybdenum wire, 44 t.p.i., and mean diameter of 0.500″. The axial velocity of signal propagation along this helix is equal to that of 54-volt electrons. The helix sections are each 2″ long. Ceramic supporting rods on each helix section are sprayed with aquadag, over $1\frac{1}{2}$″ of their length on the end nearest the center of the tube, for terminating purposes. The thickness of the spray coating increases toward the center of the tube. The distance between helices is 8.7″.

The gain produced by the electronic interaction depends upon a difference in velocity between the two electron streams. The signal is impressed upon one of the streams by the helix when its velocity is that at which traveling wave amplifier interaction between the stream and the helix occurs. It is required, therefore, for this helix that one of the streams travel at a velocity corresponding to a potential in the neighborhood of 54 volts. Useful interaction occurs from 50 to 60 volts. The inner stream is adjusted for helix interaction in this amplifier, and the outer stream travels at a lower velocity to bring about the interaction between the two streams. Operation about a mean voltage of about 50 volts was planned in designing the amplifier, and in estimating its expected performance. The amplifier is 16 wavelengths long in terms of the wavelength associated with a mean voltage of 50 volts and a frequency of 200 Mc. Eleven of these wavelengths are in the center portion between the helices.

The conditions in the amplifier tube differ from those assumed in the derivation of the theory of the double-stream interaction in the following significant ways:

1. The beams are separated in space and not completely intermingled. Calculations on the effect of this separation have been made. Numerical examples of the calculated magnitude of the effect on gain will be given below.
2. Hollow tubular beams are used, instead of "solid" beams of uniform current density over their cross-sectional area. The theory indicates that, for the beam dimensions and currents used here, the parameters which depend upon beam radius and total current in the beam are nearly the same whether the current is concentrated in an infinitely thin cylindrical shell or uniformly distributed over the cross-section of a cylinder of the same radius.
3. The metal wall surrounding the beams is not infinitely remote. Its diameter was chosen as a compromise between the requirements of preventing serious d-c. space charge depression of potential in the beam and of being far enough removed from the beam to prevent a large effect on the interaction due to its presence. Its proximity tends to increase the minimum current required for producing gain, and therefore to reduce the ratio of actual to critical current on which the gain depends.
4. The beams are not perfectly confined to hollow cylinders of the dimensions given. There is evidence that some spreading outside of these dimensions occurs. The currents reaching the collector can be measured and these are used as "beam currents" in the discussion to follow and in comparisons between theory and experiment. Somewhat larger currents than these were initially launched, and the lost fraction may have contributed to the interaction before striking the walls.

Although the assumptions of the theory are not fulfilled in the actual amplifier, estimates of its performance were first made without correction for the discrepancies. With voltages of 40 and 60 volts on the outer and device was very large, for the velocity of the outer beam was far from that at which interaction with the helices occurs.

A signal from the probe at the end of the central portion of the tube 23 db greater than that from the probe at the beginning of this section was observed in a comparison of the second type. This can probably be taken as a measure of the increase in signal in this portion of the tube due to the double-stream interaction alone, although the probe arrangement may also be subject to some remaining complicating effects. Overall gain for the device in this measurement was 32 db. Further interaction of the same kind occurs in the portions of the tube outside of the space between the probes.

Measurements of the gain of an amplifier with helix output as a function of velocity separation between the streams have been made. For fixed mean voltage and current, theory predicts an increase in gain from zero db at zero separation to a maximum and then a decrease to zero as the velocity separation is further increased. A maximum gain was observed experimentally as velocity separation was varied, and in the neighborhood of the predicted optimum value of velocity separation for the current used.

In the amplifier tube with gap output it was possible to evaluate the a-c. component of current in the electron stream produced by the amplified signal since the impedance across the gap was known. The power output from this tube at saturation was 0.1 mw, a little less than the maximum shown in Fig. 3. For 75 ohm output impedance this power corresponds to 1.15 milliampere r.m.s., or about one third of the total d-c. current to the collector in both streams. The output power, although relatively low, is thus of the right order of magnitude for the currents used.

Acknowledgment

The writer wishes to acknowledge his indebtedness to J. R. Pierce for valuable suggestions and discussion, and for supplying unpublished calculations concerning the relation between hollow and solid beams, the effect of the proximity of the conducting wall, and the effect of the separation of the beams in space.

Thanks are also due to A. R. Strnad for assistance in mechanical design and to R. E. Azud for construction of the amplifier tubes.

The Synthesis of Two-Terminal Switching Circuits

By CLAUDE. E. SHANNON

PART I: GENERAL THEORY

1. Introduction

THE theory of switching circuits may be divided into two major divisions, analysis and synthesis. The problem of analysis, determining the manner of operation of a given switching circuit, is comparatively simple. The inverse problem of finding a circuit satisfying certain given operating conditions, and in particular the *best* circuit is, in general, more difficult and more important from the practical standpoint. A basic part of the general synthesis problem is the design of a two-terminal network with given operating characteristics, and we shall consider some aspects of this problem.

Switching circuits can be studied by means of Boolean Algebra.[1,2] This is a branch of mathematics that was first investigated by George Boole in connection with the study of logic, and has since been applied in various other fields, such as an axiomatic formulation of Biology,[3] the study of neural networks in the nervous system,[4] the analysis of insurance policies,[5] probability and set theory, etc.

Perhaps the simplest interpretation of Boolean Algebra and the one closest to the application to switching circuits is in terms of propositions. A letter X, say, in the algebra corresponds to a logical proposition. The sum of two letters $X + Y$ represents the proposition "X or Y" and the product XY represents the proposition "X and Y". The symbol X' is used to represent the negation of proposition X, i.e. the proposition "not X". The constants 1 and 0 represent truth and falsity respectively. Thus $X + Y = 1$ means X or Y is true, while $X + YZ' = 0$ means X or (Y and the contradiction of Z) is false.

The interpretation of Boolean Algebra in terms of switching circuits[6,8,9,10] is very similar. The symbol X in the algebra is interpreted to mean a make (front) contact on a relay or switch. The negation of X, written X', represents a break (back) contact on the relay or switch. The constants 0 and 1 represent closed and open circuits respectively and the combining operations of addition and multiplication correspond to series and parallel connections of the switching elements involved. These conventions are shown in Fig. 1. With this identification it is possible to write an algebraic

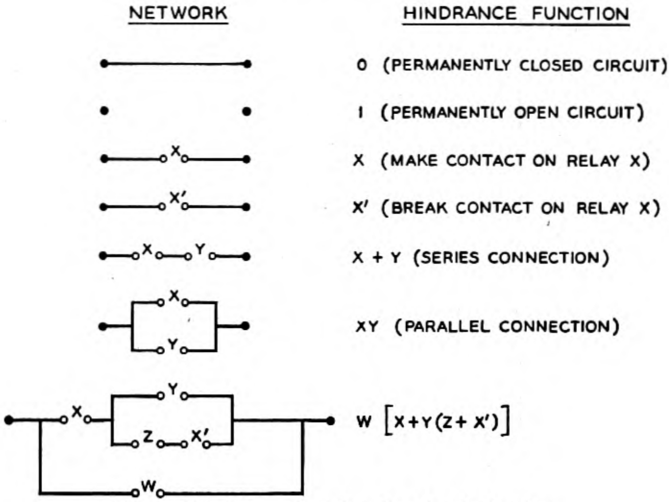


Fig. 1—Hindrance functions for simple circuits.

expression corresponding to a two-terminal network. This expression will involve the various relays whose contacts appear in the network and will be called the hindrance or hindrance function of the network. The last network in Fig. 1 is a simple example.

Boolean expressions can be manipulated in a manner very similar to ordinary algebraic expressions. Terms can be rearranged, multiplied out, factored and combined according to all the standard rules of numerical algebra. We have, for example, in Boolean Algebra the following identities:

$$0 + X = X$$

$$0 \cdot X = 0$$

$$1 \cdot X = X$$

$$X + Y = Y + X$$

$$XY = YX$$

$$X + (Y + Z) = (X + Y) + Z$$

$$X(YZ) = (XY)Z$$

$$X(Y + Z) = XY + XZ$$

The interpretation of some of these in terms of switching circuits is shown in Fig. 2.

There are a number of further rules in Boolean Algebra which allow

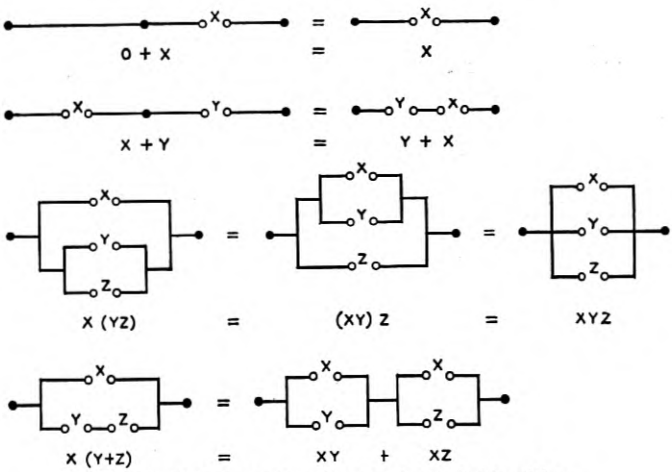


Fig. 2—Interpretation of some algebraic identities.

simplifications of expressions that are not possible in ordinary algebra. The more important of these are:

$$X = X + X = X + X + X = \text{etc.}$$

$$X = X \cdot X = X \cdot X \cdot X = \text{etc.}$$

$$X + 1 = 1$$

$$X + YZ = (X + Y)(X + Z)$$

$$X + X' = 1$$

$$X \cdot X' = 0$$

$$(X + Y)' = X'Y'$$

$$(XY)' = X' + Y'$$

The circuit interpretation of some of these is shown in Fig. 3. These rules make the manipulation of Boolean expressions considerably simpler than ordinary algebra. There is no need, for example, for numerical coefficients or for exponents, since $nX = X^n = X$.

By means of Boolean Algebra it is possible to find many circuits equivalent in operating characteristics to a given circuit. The hindrance of the given circuit is written down and manipulated according to the rules. Each different resulting expression represents a new circuit equivalent to the given one. In particular, expressions may be manipulated to eliminate elements which are unnecessary, resulting in simple circuits.

Any expression involving a number of variables $X_1, X_2, \cdots, X_n$ is called a *function* of these variables and written in ordinary function notation, $f(X_1, X_2, \cdots, X_n)$. Thus we might have $f(X, Y, Z) = X + Y'Z + XZ'$. In Boolean Algebra there are a number of important general theorems which hold for any function. It is possible to *expand* a function about one or more of its arguments as follows:

$$f(X_1, X_2, \cdots, X_n) = X_1 f(1, X_2, \cdots, X_n) + X' f(0, X_2, \cdots, X_n)$$

This is an expansion about X_1. The term $f(1, X_2, \cdots, X_n)$ is the function

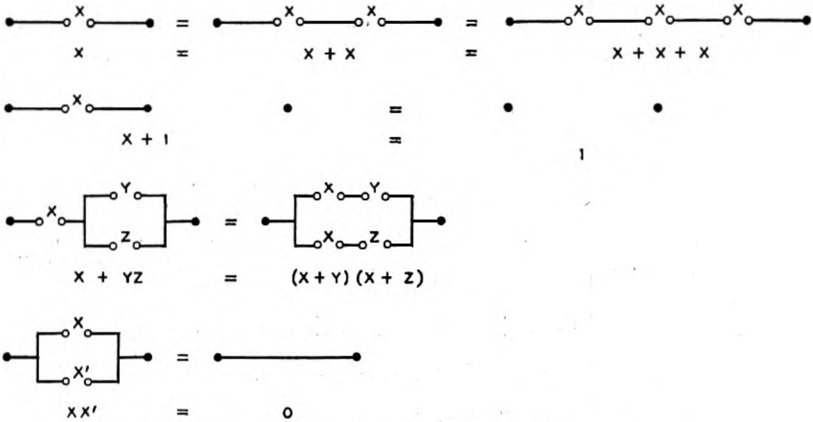


Fig. 3—Interpretation of some special Boolean identities.

$f(X_1, X_2, \cdots, X_n)$ with 1 substituted for X, and 0 for X', and conversely for the term $f(0, X_2, \cdots, X_n)$. An expansion about X_1 and X_2 is:

$$f(X_1, X_2, \cdots, X_n) = X_1 X_2 f(1, 1, X_3, \cdots, X_n) + X_1 X_2' f(1, 0, X_3, \cdots, X_n)$$
$$+ X_1' X_2 f(0, 1, X_3, \cdots, X_n) + X_1' X_2' f(1, 1, X_3, \cdots, X_n)$$

This may be continued to give expansions about any number of variables. When carried out for all n variables, f is written as a sum of 2^n products each with a coefficient which does not depend on any of the variables. Each coefficient is therefore a constant, either 0 or 1.

There is a similar expansion whereby f is expanded as a product:

$$\begin{aligned} f(X_1, X_2, \cdots, X_2) & \\ &= [X_1 + f(0, X_2, \cdots, X_n)]\,[X_1' + f(1, X_2, \cdots, X_n)] \\ &= [X_1 + X_2 + f(0, 0, \cdots, X_n)]\,[X_1 + X_2' + f(0, 1, \cdots, X_n)] \\ &\quad [X_1' + X_2 + f(1, 0, \cdots, X_n)]\,[X_1' + X_2' + f(1, 1, \cdots, X_n)] \\ &= \text{etc.} \end{aligned}$$

The following are some further identities for general functions:

$$X + f(X, Y, Z, \cdots) = X + f(0, Y, Z, \cdots)$$

$$X' + f(X, Y, Z, \cdots) = X' + f(1, Y, Z, \cdots)$$

$$Xf(X, Y, Z, \cdots) = Xf(1, Y, Z, \cdots)$$

$$X'f(X, Y, Z, \cdots) = X'f(0, Y, Z, \cdots)$$

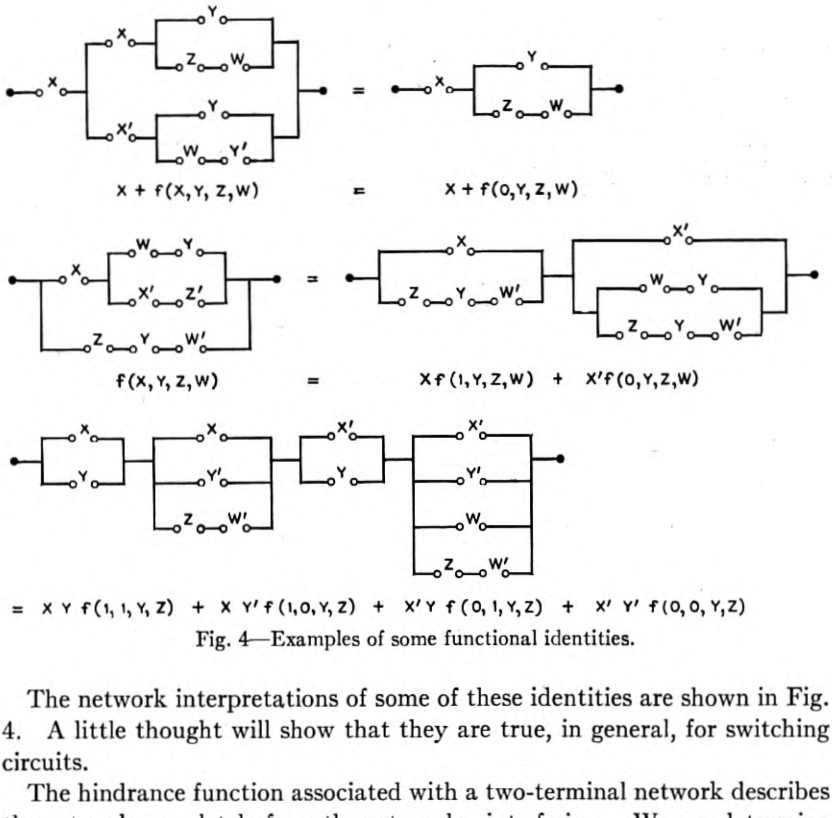


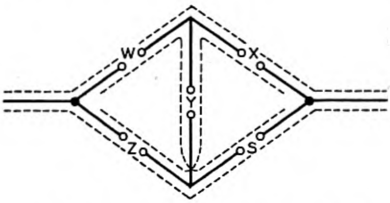
Fig. 4—Examples of some functional identities.

The network interpretations of some of these identities are shown in Fig. 4. A little thought will show that they are true, in general, for switching circuits.

The hindrance function associated with a two-terminal network describes the network completely from the external point of view. We can determine from it whether the circuit will be open or closed for any particular position of the relays. This is done by giving the variables corresponding to operated relays the value 0 (since the make contacts of these are then closed and the break contacts open) and unoperated relays the value 1. For example, with the function $f = W[X + Y(Z + X')]$ suppose relays X and Y operated and Z and W not operated. Then $f = 1[0 + 0(1 + 1)] = 0$ and in this condition the circuit is closed.

A hindrance function corresponds explicitly to a series-parallel type of circuit, i.e. a circuit containing only series and parallel connections. This is because the expression is made up of sum and product operations. There is however, a hindrance function representing the operating characteristics (conditions for open or closed circuits between the two terminals) for any network, series-parallel or not. The hindrance for non-series-parallel networks can be found by several methods of which one is indicated in Fig. 5 for a simple bridge circuit. The hindrance is written as the product of a set of factors. Each factor is the series hindrance of a possible path between the two terminals. Further details concerning the Boolean method for switching circuits may be found in the references cited above.

This paper is concerned with the problem of synthesizing a two-terminal circuit which represents a given hindrance function $f(X_1, \cdots, X_n)$. Since any given function f can be realized in an unlimited number of different



$$f = (W+X)(Z+S)(W+Y+S)(Z+Y+X)$$

Fig. 5—Hindrance of a bridge circuit.

ways, the particular design chosen must depend upon other considerations. The most common of these determining criteria is that of economy of elements, which may be of several types, for example:

(1) We may wish to realize our function with the least total number of switching elements, regardless of which variables they represent.

(2) We may wish to find the circuit using the least total number of relay springs. This requirement sometimes leads to a solution different from (1), since contiguous make and break elements may be combined into transfer elements so that circuits which tend to group make and break contacts on the same relay into pairs will be advantageous for (2) but not necessarily for (1).

(3) We may wish to distribute the spring loading on all the relays or on some subset of the relays as evenly as possible. Thus, we might try to find the circuit in which the most heavily loaded relay was as lightly loaded as possible. More generally, we might desire a circuit in which the loading on the relays is of some specified sort, or as near as possible to this given distribution. For example, if the relay X_1

must operate very quickly, while X_2 and X_3 have no essential time limitations but are ordinary U-type relays, and X_4 is a multicontact relay on which many contacts are available, we would probably try to design a circuit for $f(X_1, X_2, X_3, X_4)$ in such a way as, first of all, to minimize the loading on X_1, next to equalize the loading on X_2 and X_3 keeping it at the same time as low as possible, and finally not to load X_4 any more than necessary. Problems of this sort may be called *problems in spring-load distribution.*

Although all equivalent circuits representing a given function f which contain only series and parallel connections can be found with the aid of Boolean Algebra, the most economical circuit in any of the above senses will often not be of this type. The problem of synthesizing non-series-parallel circuits is exceedingly difficult. It is even more difficult to show that a circuit found in some way is the *most* economical one to realize a given function. The difficulty springs from the large number of essentially different networks available and more particularly from the lack of a simple mathematical idiom for representing these circuits.

We will describe a new design method whereby any function $f(X_1, X_2, \cdots, X_n)$ may be realized, and frequently with a considerable saving of elements over other methods, particularly when the number of variables n is large. The circuits obtained by this method will not, in general, be of the series-parallel type, and, in fact, they will usually not even be planar. This method is of interest theoretically as well as for practical design purposes, for it allows us to set new upper limits for certain numerical functions associated with relay circuits. Let us make the following definitions:

$\lambda(n)$ is defined as the least number such that any function of n variables can be realized with not more than $\lambda(n)$ elements.* Thus, any function of n variables can be realized with $\lambda(n)$ elements and at least one function with no less.

$\mu(n)$ is defined as the least number such that given any function f of n variables, there is a two-terminal network having the hindrance f and using not more than $\mu(n)$ elements on the most heavily loaded relay.

The first part of this paper deals with the general design method and the behaviour of $\lambda(n)$. The second part is concerned with the possibility of various types of spring load distribution, and in the third part we will study certain classes of functions that are especially easy to synthesize, and give some miscellaneous theorems on switching networks and functions.

2. Fundamental Design Theorem

The method of design referred to above is based on a simple theorem dealing with the interconnection of two switching networks. We shall first

* An *element* means a make or break contact on one relay. A *transfer element* means a make-and-break with a common spring, and contains two *elements*.

state and prove this theorem. Suppose that M and N (Fig. 6) are two $(n + 1)$ terminal networks, M having the hindrance functions U_k $(k = 1, 2, \cdots n)$ between terminals a and k, and N having the functions V_k between b and k. Further, let M be such that $U_{jk} = 1 (j, k = 1, 2, \cdots, n)$. We will say, in this case, that M is a *disjunctive* network. Under these conditions we shall prove the following:

Theorem 1: If the corresponding terminals 1, 2, $\cdots$, n *of* M *and* N *are connected together, then*

$$U_{ab} = \prod_{k=1}^{n} (U_k + V_k) \tag{1}$$

where U_{ab} *is the hindrance from terminal* a *to terminal* b.

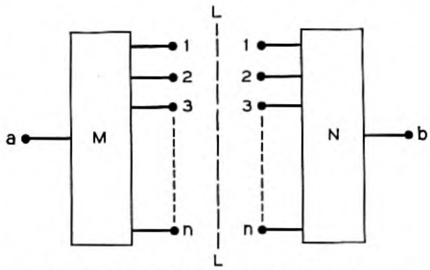


Fig. 6—Network for general design theorem.

Proof: It is known that the hindrance U_{ab} may be found by taking the product of the hindrances of all possible paths from a to b along the elements of the network.[6] We may divide these paths into those which cross the line L once, those which cross it three times, those which cross it five times, etc. Let the product of the hindrances in the first class be W_1, in the second class W_3, etc. Thus

$$U_{ab} = W_1 \cdot W_3 \cdot W_5 \cdots \tag{2}$$

Now clearly

$$W_1 = \prod_{1}^{n} (U_k + V_k)$$

and also

$$W_3 = W_5 = \cdots = 1$$

since each term in any of these must contain a summand of the type U_{jk} which we have assumed to be 1. Substituting in (2) we have the desired result.

The method of using this theorem to synthesize networks may be roughly described as follows: The function to be realized is written in the form of a product of the type (1) in such a way that the functions U_k are the same for a large class of functions, the V_k determining the particular one under consideration. A basic disjunctive network M is constructed having the functions U_k between terminals a and k. A network N for obtaining the functions V_k is then found by inspection or according to certain general rules. We will now consider just how this can be done in various cases.

3. Design of Networks for General Functions—Behavior of $\lambda(n)$.

a. Functions of One, Two and Three Variables:

Functions of one or two variables may be dismissed easily since the number of such functions is so small. Thus, with one variable X, the possible functions are only:

$$0, 1, X, X'$$

and obviously $\lambda(1) = 1$, $\mu(1) = 1$.

With two variables X and Y there are 16 possible functions:

$$\begin{array}{lllllll} 0\ X\ Y & XY & XY' & X'Y & X'Y' & XY' + X'Y \\ 1\ X'\ Y' & X + Y & X + Y' & X' + Y & X' + Y' & XY + X'Y' \end{array}$$

so that $\lambda(2) = 4$, $\mu(2) = 2$.

We will next show that any function of three variables $f(X, Y, Z)$ can be realized with not more than eight elements and with not more than four from any one relay. Any function of three variables can be expanded in a product as follows:

$$f(X, Y, Z) = [X + Y + f(0, 0, Z)][X + Y' + f(0, 1, Z)]$$
$$[X' + Y + f(1, 0, Z)]\,[X' + Y' + f(1, 1, Z)].$$

In the terminology of Theorem 1 we let

$$\begin{array}{ll} U_1 = X + Y & V_1 = f(0, 0, Z) \\ U_2 = X + Y' & V_2 = f(0, 1, Z) \\ U_3 = X' + Y & V_3 = f(1, 0, Z) \\ U_4 = X' + Y' & V_4 = f(1, 1, Z) \end{array}$$

so that

$$U_{ab} = f(X, Y, Z) = \prod_{k=1}^{4} (U_k + V_k)$$

The above U_k functions are realized with the network M of Fig. 7 and it is easily seen that $U_{jk} = 1$ $(j, k = 1, 2, 3, 4)$. The problem now is to construct a second network N having the V_k functions V_1, V_2, V_3, V_4. Each of these is a function of the one variable Z and must, therefore, be one of the four possible functions of one variable:

$$0, 1, Z, Z'.$$

Consider the network N of Fig. 8. If any of the V's are equal to 0, connect the corresponding terminals of M to the terminal of N marked 0; if any are equal to Z, connect these terminals of M to the terminal of N marked Z, etc. Those which are 1 are, of course, not connected to anything. It is clear from Theorem 1 that the network thus obtained will realize the function $f(X, Y, Z)$. In many cases some of the elements will be superfluous, e.g., if one of the V_i is equal to 1, the element of M connected to terminal i can

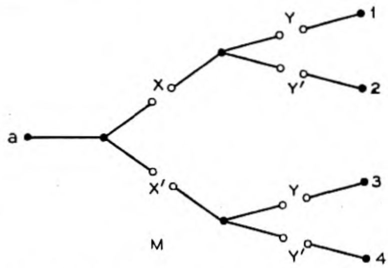


Fig. 7—Disjunctive tree with two bays.

be eliminated. At worst M contains six elements and N contains two. The variable X appears twice, Y four times and Z twice. Of course, it is completely arbitrary which variables we call X, Y, and Z. We have thus proved somewhat more than we stated above, namely,

Theorem 2: Any function of three variables may be realized using not more than 2, 2, *and* 4 *elements from the three variables in any desired order. Thus* $\lambda(3) \leq 8$, $\mu(3) \leq 4$. *Further, since make and break elements appear in adjacent pairs we can obtain the distribution* 1, 1, 2, *in terms of transfer elements.*

The theorem gives only upper limits for $\lambda(3)$ and $\mu(3)$. The question immediately arises as to whether by some other design method these limits could be lowered, i.e., can the $\leq$ signs be replaced by $<$ signs. It can be shown by a study of special cases that $\lambda(3) = 8$, the function

$$X \oplus Y \oplus Z = X(YZ + Y'Z') + X'(YZ' + Y'Z)$$

requiring eight elements in its most economical realization. $\mu(3)$, however, is actually 3.

It seems probable that, in general, the function

$$X_1 \oplus X_2 \oplus \cdots \oplus X_n$$

requires $4(n - 1)$ elements, but no proof has been found. Proving that a certain function cannot be realized with a small number of elements is somewhat like proving a number transcendental; we will show later that almost all* functions require a large number of elements, but it is difficult to show that a particular one does.

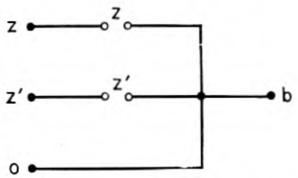


Fig. 8—Network giving all functions of one variable.

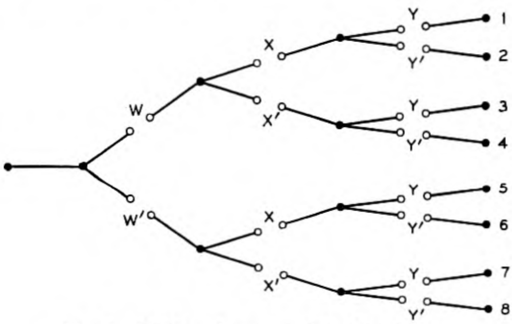


Fig. 9—Disjunctive tree with three bays.

b. Functions of Four Variables:

In synthesizing functions of four variables by the same method, two courses are open. First, we may expand the function as follows:

$$\begin{aligned} f(W, X, Y, Z) = {} & [W + X + Y + V_1(Z)] \cdot [W + X + Y' + V_2(Z)]. \\ & [W + X' + Y + V_3(Z)] \cdot [W + X' + Y' + V_4(Z)]. \\ & [W' + X + Y + V_5(Z)] \cdot [W' + X + Y' + V_5(Z)]. \\ & [W' + X' + Y + V_7(Z)] \cdot [W' + X' + Y' + V_8(Z)]. \end{aligned}$$

By this expansion we would let $U_1 = W + X + Y$, $U_2 = W + X + Y'$, $\cdots$, $U_8 = W' + X' + Y'$ and construct the M network in Fig. 9. N would

* We use the expression "almost all" in the arithmetic sense: e.g., a property is true of almost all functions of n variables if the fraction of all functions of n variables for which it is not true $\to 0$ as $n \to \infty$.

again be as in Fig. 8, and by the same type of reasoning it can be seen that $\lambda(4) \leq 16$.

Using a slightly more complicated method, however, it is possible to reduce this limit. Let the function be expanded in the following way:

$$f(W, X, Y, Z) = [W + X + V_1(Y, Z)] \cdot [W + X' + V_2(Y, Z)]$$
$$[W' + X + V_3(Y, Z)] \cdot [W' + X' + V_4(Y, Z)].$$

We may use a network of the type of Fig. 7 for M. The V functions are now functions of two variables Y and Z and may be any of the 16 functions:

$$A\begin{cases}0\\1\end{cases} \quad B\begin{cases}Y\\Y'\\Z\\Z'\end{cases} \quad C\begin{cases}YZ\\Y'Z\\YZ'\\Y'Z'\end{cases} \quad D\begin{cases}Y+Z\\Y+Z'\\Y'+Z\\Y'+Z'\end{cases} \quad E\begin{cases}Y'Z+YZ'\\YZ+Y'Z'\end{cases}$$

We have divided the functions into five groups, A, B, C, D and E for later reference. We are going to show that any function of four variables can

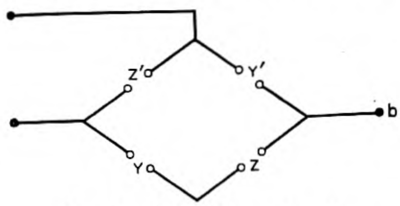


Fig. 10—Simplifying network.

be realized with not more than 14 elements. This means that we must construct a network N using not more than eight elements (since there are six in the M network) for any selection of four functions from those listed above. To prove this, a number of special cases must be considered and dealt with separately:

(1) If all four functions are from the groups, A, B, C, and D, N will certainly not contain more than eight elements, since eight letters at most can appear in the four functions.

(2) We assume now that just one of the functions is from group E; without loss of generality we may take it to be $YZ' + Y'Z$, for it is the other, replacing Y by Y' transforms it into this. If one or more of the remaining functions are from groups A or B the situation is satisfactory, for this function need require no elements. Obviously 0 and 1 require no elements and Y, Y', Z or Z' may be "tapped off" from the circuit for $YZ' + Y'Z$ by writing it as $(Y + Z)(Y' + Z')$. For example, Y' may be obtained with the circuit of Fig. 10. This leaves four elements, certainly a sufficient number for any two functions from A, B, C, or D.

(3) Now, still assuming we have one function, $YZ' + Y'Z$, from E, suppose at least two of the remaining are from D. Using a similar "tapping off" process we can save an element on each of these. For instance, if the functions are $Y + Z$ and $Y' + Z'$ the circuit would be as shown in Fig. 11.

(4) Under the same assumption, then, our worst case is when two of the functions are from C and one from D, or all three from C. This latter case is satisfactory since, then, at least one of the three must be a term of $YZ' + Y'Z$ and can be "tapped off." The former case is bad only when the two functions from C are YZ and $Y'Z'$. It may be seen that the only

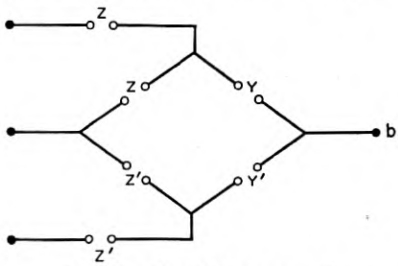


Fig. 11—Simplifying network.

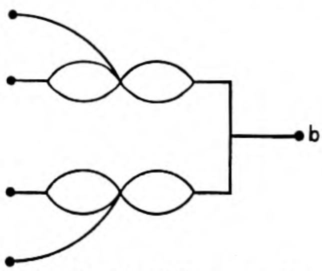


Fig. 12—Simplifying network.

essentially different choices for the function from D are $Y + Z$ and $Y' + Z$. That the four types of functions f resulting may be realized with 14 elements can be shown by writing out typical functions and reducing by Boolean Algebra.

(5) We now consider the cases where two of the functions are from E. Using the circuit of Fig. 12, we can tap off functions or parts of functions from A, B or D, and it will be seen that the only difficult cases are the following: (a) Two functions from C. In this case either the function f is symmetric in Y and Z or else both of the two functions may be obtained from the circuits for the E functions of Fig. 12. The symmetric case is handled in a later section. (b) One is from C, the other from D. There is only one unsymmetric case. We assume the four functions are $Y \oplus Z$, $Y \oplus Z'$, YZ and $Y + Z'$. This gives rise to four types of functions f, which can all be reduced by algebraic methods. This completes the proof.

Theorem 3: Any function of four variables can be realized with not more than 14 *elements.*

c. Functions of More Than Four Variables:

Any function of five variables may be written

$$f(X_1, \cdots, X_5) = [X_5 + f_1(X_1, \cdots, X_4)] \cdot [X_5' + f_2(X_1, \cdots, X_4)]$$

and since, as we have just shown, the two functions of four variables can be realized with 14 elements each, $f(X_1, \cdots X_5)$ can be realized with 30

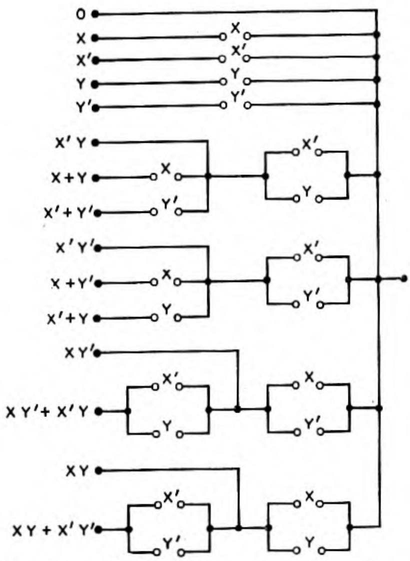


Fig. 13—Network giving all functions of two variables.

Now consider a function $f(X_1, X_2, \cdots, X_n)$ of n variables. For $5 < n \leq 13$ we get the best limit by expanding about all but two variables.

$$f(X_1, X_2, \cdots, X_n) = [X_1 + X_2 + \cdots + X_{n-2} + V_1(X_{n-1}, X_n)] \cdot \cdots \cdot [X_1' + X_2' + \cdots + X_{n-2}' + V_S(X_{n-1}, V_n)] \qquad (4)$$

The V's are all functions of the variables X_{n-1}, X_n and may be obtained from the general N network of Fig. 13, in which *every* function of two variables appears. This network contains 20 elements which are grouped into five transfer elements for one variable and five for the other.* The M network for (4), shown in Fig. 14, requires in general $2^{n-1} - 2$ elements. Thus we have:

* Several other networks with the same property as Fig. 13 have been found, but they all require 20 elements.

Theorem 4. $\lambda(n) \leq 2^{n-1} + 18$

d. Upper Limits for $\lambda(n)$ with Large n.

Of course, it is not often necessary to synthesize a function of more than say 10 variables, but it is of considerable theoretical interest to determine as closely as possible the behavior of $\lambda(n)$ for large n.

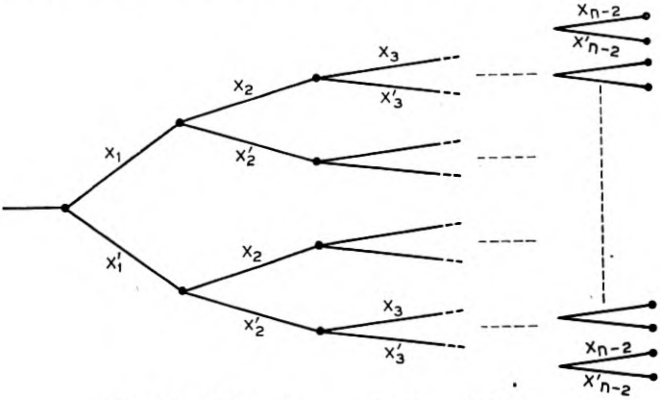


Fig. 14—Disjunctive tree with $(n - 2)$ bays.

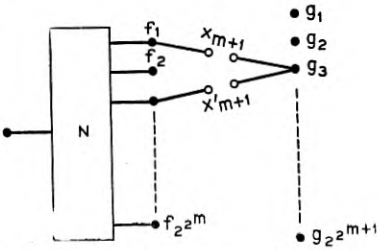


Fig. 15—Network giving all functions of $(m + 1)$ variables constructed from one giving all functions of m variables.

We will first prove a theorem placing limits on the number of elements required in a network analogous to Fig. 13 but generalized for m variables.

Theorem 5. An N network realizing all 2^{2^m} *functions of m variables can be constructed using not more than* $2 \cdot 2^{2^m}$ *elements, i.e., not more than two elements per function. Any network with this property uses at least* $(\frac{3}{2} - \epsilon)$ *elements per function for any* $\epsilon > 0$ *with n sufficiently large.*

The first part will be proved by induction. We have seen it to be true for $m = 1, 2$. Suppose it is true for some m with the network N of Fig. 15. Any function of $m + 1$ variables can be written

$$g = [X_{m+1} + f_a][X'_{m+1} + f_b]$$

where f_a and f_b involve only m variables. By connecting from g to the corresponding f_a and f_b terminals of the smaller network, as shown typically for g_3, we see from Theorem 1 that all the g functions can be obtained. Among these will be the 2^{2^m} f functions and these can be obtained simply by connecting across to the f functions in question without any additional elements. Thus the entire network uses less than

$$(2^{2^{m+1}} - 2^{2^m})2 + 2 \cdot 2^{2^m}$$

elements, since the N network by assumption uses less than $2 \cdot 2^{2^m}$ and the first term in this expression is the number of added elements.

The second statement of Theorem 7 can be proved as follows. Suppose we have a network, Fig. 16, with the required property. The terminals can be divided into three classes, those that have one or less elements di-

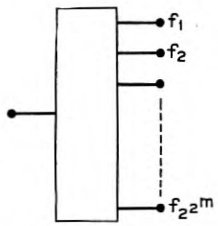


Fig. 16—Network giving all functions of m variables.

rectly connected, those with two, and those with three or more. The first set consists of the functions 0 and 1 and functions of the type

$$(X + f) = X + f_{X=0}$$

where X is some variable or primed variable. The number of such functions is not greater than $2m \cdot 2^{2^{m-1}}$ for there are $2m$ ways of selecting an "X" and then $2^{2^{m-1}}$ different functions $f_{X=0}$ of the remaining $m - 1$ variables. Hence the terminals in this class as a fraction of the total $\rightarrow 0$ as $m \rightarrow \infty$. Functions of the second class have the form

$$g = (X + f_1)(Y + f_2)$$

In case $X \neq Y'$ this may be written

$$XY + XY'g_{X=1,Y=0} + X'Yg_{X=0,Y=1} + X'Y'g_{X=0,Y=0}$$

and there are not more than $(2m)(2m - 2)[2^{2^{m-2}}]^3$ such functions, again a vanishingly small fraction. In case $X = Y'$ we have the situation shown in Fig. 17 and the XX' connection can never carry ground to another terminal since it is always open as a series combination. The inner ends of these elements can therefore be removed and connected to terminals corresponding to functions of less than m variables according to the equation

$$g = (X + f_1)(X' + f_2) = (X + f_{1X=0})(X' + f_{2X=1})$$

if they are not already so connected. This means that all terminals of the second class are then connected to a vanishingly small fraction of the total terminals. We can then attribute two elements each to these terminals and at least one and one-half each to the terminals of the third group. As these two groups exhaust the terminals except for a fraction which $\rightarrow 0$ as $n \rightarrow \infty$, the theorem follows.

If, in synthesizing a function of n variables, we break off the tree at the $(n - m)$th bay, the tree will contain $2^{n-m+1} - 2$ elements, and we can find an N network with not more than $2^{2^m} \cdot 2$ elements exhibiting every function of the remaining m variables. Hence

$$\lambda(n) \leq 2^{n-m+1} - 2 + 2\ 2^{2^m} < 2^{n-m+1} + 2\ 2^{2^m}$$

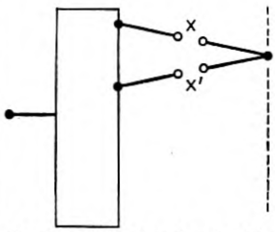


Fig. 17—Possible situation in Fig. 16.

for every integer m. We wish to find the integer $M = M(n)$ minimizing this upper bound.

Considering m as a continuous variable and n fixed, the function

$$f(m) = 2^{n-m+1} + 2^{2^m} \cdot 2$$

clearly has just one minimum. This minimum must therefore lie between m_1 and $m_1 + 1$, where

$$f(m_1) = f(m_1 + 1)$$

i.e.,
$$2^{n-m_1+1} + 2^{2^{m_1}} \cdot 2 = 2^{n-m_1} + 2^{2^{m_1+1}} \cdot 2$$

or
$$2^n = 2^{m_1+1}(2^{2^{m_1+1}} - 2^{2^{m_1}})$$

Now m_1 cannot be an integer since the right-hand side is a power of two and the second term is less than half the first. It follows that to find the *integer* M making $f(M)$ a minimum we must take for M the least integer satisfying

$$2^n \leq 2^{M+1} 2^{2^{M+1}}$$

Thus M satisfies:

$$M + 1 + 2^{M+1} \geq n > M + 2^M \tag{5}$$

This gives:

$n \leq 11$	$M = 2$
$11 < n \leq 20$	$M = 3$
$20 < n \leq 37$	$M = 4$
$37 < n \leq 70$	$M = 5$
$70 < n \leq 135$	$M = 6$
etc.	

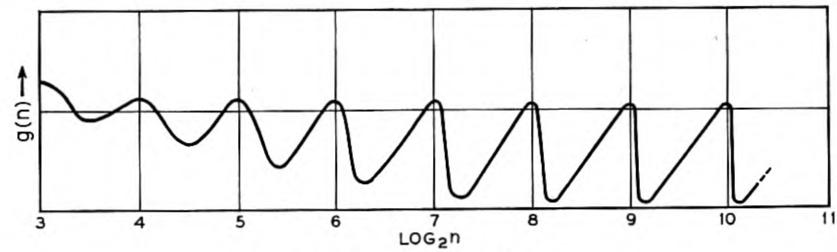


Fig. 18—Behaviour of $g(n)$.

Our upper bound for $\lambda(n)$ behaves something like $\frac{2^{n+1}}{n}$ with a superimposed saw-tooth oscillation as n varies between powers of two, due to the fact that m must be an integer. If we define $g(n)$ by

$$2^{n-M+1} + 2^{2^M}2 = g(n)\,\frac{2^{n+1}}{n},$$

M being determined to minimize the function (i.e., M satisfying (5)), then $g(n)$ varies somewhat as shown in Fig. 18 when plotted against $\log_2 n$. The maxima occur just beyond powers of two, and closer and closer to them as $n \to \infty$. Also, the saw-tooth shape becomes more and more exact. The sudden drops occur just after we change from one value of M to the next. These facts lead to the following:

Theorem 6. (*a*) *For all* n

$$\lambda(n) < \frac{2^{n+3}}{n}.$$

(*b*) *For almost all* n

$$\lambda(n) < \frac{2^{n+2}}{n}.$$

(c) *There is an infinite sequence of n_i for which*

$$\lambda(n_i) < \frac{2^{n+1}}{n}(1+\epsilon) \qquad \epsilon > 0.$$

These results can be proved rigorously without much difficulty.

e. A Lower Limit for $\lambda(n)$ with Large n.

Up to now most of our work has been toward the determination of upper limits for $\lambda(n)$. We have seen that for all n

$$\lambda(n) < B\frac{2^n}{n}.$$

We now ask whether this function $B\frac{2^n}{n}$ is anywhere near the true value of $\lambda(n)$, or may $\lambda(n)$ be perhaps dominated by a smaller order of infinity, e.g., n^p. It was thought for a time, in fact, that $\lambda(n)$ might be limited by n^2 for all n, arguing from the first few values: 1, 4, 8, 14. We will show that this is far from the truth, for actually $\frac{2^n}{n}$ is the correct order of magnitude of $\lambda(n)$:

$$A\frac{2^n}{n} < \lambda(n) < B\frac{2^n}{n}$$

for all n. A closely associated question to which a partial answer will be given is the following: Suppose we define the "complexity" of a given function f of n variables as the ratio of the number of elements in the most economical realization of f to $\lambda(n)$. Then any function has a complexity lying between 0 and 1. Are most functions simple or complex?

Theorem 7: For all sufficiently large n, all functions of n variables excepting a fraction δ require at least $(1-\epsilon)\frac{2^n}{n}$ elements, where ϵ and δ are arbitrarily small positive numbers. Hence for large n

$$\lambda(n) > (1-\epsilon)\frac{2^n}{n}$$

and almost all functions have a complexity $> \frac{1}{4}(1-\epsilon)$. For a certain sequence n_i almost all functions have a complexity $> \frac{1}{2}(1-\epsilon)$.

The proof of this theorem is rather interesting, for it is a pure existence proof. We do not show that any particular function or set of functions requires $(1-\epsilon)\frac{2^n}{n}$ elements, but rather that it is impossible for all functions to require less. This will be done by showing that there are not enough networks with less than $(1 - \epsilon)\frac{2^n}{n}$ branches to go around, i.e., to represent all the 2^{2^n} functions of n variables, taking account, of course, of the different assignments of the variables to the branches of each network. This is only possible due to the extremely rapid increase of the function 2^{2^n}. We require the following:

Lemma: *The number of two-terminal networks with K or less branches is less than* $(6K)^K$.

Any two-terminal network with K or less branches can be constructed as follows: First line up the K branches as below with the two terminals a and b.

a. 1—1′
2—2′
3—3′
4—4′
·
·
·
b. K—K'

We first connect the terminals $a, b, 1, 2, \cdots, K$ together in the desired way. The number of *different* ways we can do this is certainly limited by the number of partitions of $K + 2$ which, in turn, is less than

$$2^{K+1}$$

for this is the number of ways we can put one or more division marks between the symbols $a, 1, \cdots, K, b$. Now, assuming $a, 1, 2, \cdots, K, b$, interconnected in the desired manner, we can connect $1'$ either to one of these terminals or to an additional junction point, i.e., $1'$ has a choice of at most

$$K + 3$$

terminals, $2'$ has a choice of at most $K + 4$, etc. Hence the number of networks is certainly less than

$$2^{K+1}(K + 3)(K + 4)(K + 5) \cdots (2K + 3)$$

$$< (6K)^K \qquad K \geq 3$$

and the theorem is readily verified for $K = 1, 2$.

We now return to the proof of Theorem 7. The number of functions of n variables that can be realized with $\frac{(1 - \epsilon)2^n}{n}$ elements is certainly less than the number of networks we can construct with this many branches multiplied by the number of assignments of the variables to the branches, i.e., it is less than

$$H = (2n)^{(1-\epsilon)(2^n/n)} \left[6(1 - \epsilon) \frac{2^n}{n} \right]^{(1-\epsilon)(2^n/n)}$$

Hence

$$\log_2 H = (1 - \epsilon) \frac{2^n}{n} \log 2n + (1 - \epsilon) \frac{2^n}{n} \log (1 - \epsilon) \frac{2^n}{n} \cdot 6$$

$$= (1 - \epsilon)\, 2^n + \text{terms dominated by this term for large } n.$$

By choosing n so large that $\frac{\epsilon}{2} 2^n$ dominates the other terms of $\log H$ we arrive at the inequality

$$\log_2 H < (1 - \epsilon_1)\, 2^n$$

$$H < 2^{(1-\epsilon_1)2^n}$$

But there are $S = 2^{2^n}$ functions of n variables and

$$\frac{H}{S} = \frac{2^{(1-\epsilon_1)2^n}}{2^{2^n}} \to 0 \quad \text{as} \quad n \to \infty .$$

Hence almost all functions require more than $(1 - \epsilon_1)2^n$ elements.

Now, since for all $n > N$ there is at least one function requiring more than (say) $\frac{1}{2}\frac{2^n}{n}$ elements and since $\lambda(n) > 0$ for $n > 0$, we can say that for *all* n,

$$\lambda(n) > A \frac{2^n}{n}$$

for some constant $A > 0$, for we need only choose A to be the minimum number in the finite set:

$$\frac{1}{2}, \quad \frac{\lambda(1)}{\frac{2^1}{1}}, \quad \frac{\lambda(2)}{\frac{2^2}{2}}, \quad \frac{\lambda(3)}{\frac{2^3}{3}}, \ldots, \frac{\lambda(N)}{\frac{2^N}{N}}$$

Thus $\lambda(n)$ is of the order of magnitude of $\frac{2^n}{n}$. The other parts of Theorem 8 follow easily from what we have already shown.

The writer is of the opinion that almost all functions have a complexity nearly 1, i.e., $> 1 - \epsilon$. This could be shown at least for an infinite sequence n_i if the Lemma could be improved to show that the number of networks is less than $(6K)^{K/2}$ for large K. Although several methods have been used in counting the networks with K branches they all give the result $(6K)^K$.

It may be of interest to show that for large K the number of networks is greater than

$$(6K)^{K/4}$$

This may be done by an inversion of the above argument. Let $f(K)$ be the number of networks with K branches. Now, since there are 2^{2^n} functions of n variables and each can be realized with $(1 + \epsilon) \frac{2^{n+2}}{n}$ elements (n sufficiently large),

$$f\left((1 + \epsilon) \frac{2^{n+2}}{n}\right) (2n)^{(1+\epsilon)(2^{n+2}/n)} > 2^{2^n}$$

for n large. But assuming $f(K) < (6K)^{K/4}$ reverses the inequality, as is readily verified. Also, for an infinite sequence of K,

$$f(K) > (6K)^{K/2}$$

Since there is no obvious reason why $f(K)$ should be connected with powers of 2 it seems likely that this is true for all large K.

We may summarize what we have proved concerning the behavior of $\lambda(n)$ for large n as follows. $\lambda(n)$ varies somewhat as $\frac{2^{n+1}}{n}$; if we let

$$\lambda(n) = A_n \frac{2^{n+1}}{n}$$

then, for large n, A_n lies between $\frac{1}{2} - \epsilon$ and $(2 + \epsilon)$, while for an infinite sequence of n, $\frac{1}{2} - \epsilon < A_n < 1 + \epsilon$.

We have proved, incidentally, that the new design method cannot, in a sense, be improved very much. With series-parallel circuits the best known limit* for $\lambda(n)$ is

$$\lambda(n) < 3.2^{n-1} + 2$$

and almost all functions require $(1 - \epsilon) \frac{2^n}{\log_2 n}$ elements.[7] We have lowered the order of infinity, dividing by at least $\frac{n}{\log_2 n}$ and possibly by n. The best that can be done now is to divide by a constant factor ≤ 4, and for some n, ≤ 2. The possibility of a design method which does this seems, however, quite unlikely. Of course, these remarks apply only to a perfectly general design method, i.e., one applicable to *any* function. Many special *classes* of functions can be realized by special methods with a great saving.

* Mr. J. Riordan has pointed out an error in my reasoning in (6) leading to the statement that this limit is actually reached by the function $X_1 \oplus X_2 \oplus \ldots \oplus X_n$, and has shown that this function and its negative can be realized with about n^2 elements. The error occurs in Part IV after equation 19 and lies in the assumption that the factorization given is the best.

PART II: CONTACT LOAD DISTRIBUTION

4. Fundamental Principles

We now consider the question of distributing the spring load on the relays as evenly as possible or, more generally, according to some preassigned scheme. It might be thought that an attempt to do this would usually result in an increase in the total number of elements over the most economical circuit. This is by no means true; we will show that in many cases (in fact, for almost all functions) a great many load distributions may be obtained (including a nearly uniform distribution) while keeping the total number of elements at the same minimum value. Incidentally this result has a bearing on the behavior of $\mu(n)$, for we may combine this result with

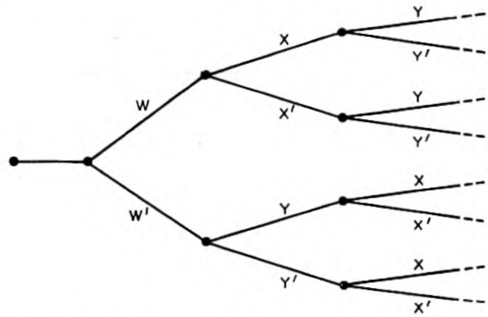


Fig. 19—Disjunctive tree with the contact distribution 1, 3, 3.

preceding theorems to show that $\mu(n)$ is of the order of magnitude of $\frac{2^{n+1}}{n^2}$ as $n \to \infty$ and also to get a good evaluation of $\mu(n)$ for small n.

The problem is rather interesting mathematically, for it involves additive number theory, a subject with few if any previous applications. Let us first consider a few simple cases. Suppose we are realizing a function with the tree of Fig. 9. The three variables appear as follows:

W, X, Y appear

2, 4, 8 times, respectively·

or, in terms of transfer elements*

1, 2, 4.

Now, W, X, and Y may be interchanged in any way without altering the operation of the tree. Also we can interchange X and Y in the lower branch of the tree only without altering its operation. This would give the distribution (Fig. 19)

1, 3, 3

* In this section we shall always speak in terms of transfer elements.

A tree with four bays can be constructed with any of the following distributions

W	X	Y	Z	
1,	2,	4,	8	= 1, 2, 4, + 1, 2, 4
1,	2,	5,	7	= 1, 2, 4 + 1, 3, 3
1,	2,	6,	6	= 1, 2, 4 + 1, 4, 2
1,	3,	3,	8	= 1, 2, 4 + 2, 1, 4
1,	3,	4,	7	= 1, 3, 3 + 2, 1, 4
1,	3,	5,	6	= 1, 4, 2 + 2, 1, 4
1,	4,	4,	6	= 1, 3, 3 + 3, 1, 3
1,	4,	5,	5	= 1, 4, 2 + 3, 1, 3

and the variables may be interchanged in any manner. The "sums" on the right show how these distributions are obtained. The first set of numbers represents the upper half of the tree and the second set the lower half. They are all reduced to the sum of sets 1, 2, 4 or 1, 3, 3 in some order, and these sets are obtainable for trees with 3 bays as we already noted. In general it is clear that if we can obtain the distributions

$$a_1, a_2, a_3, \cdots, a_n$$

$$b_1, b_2, b_3, \cdots, b_n$$

for a tree with n bays then we can obtain the distribution

$$1, a_1 + b_1, a_2 + b_2, \cdots, a_n + b_n$$

for a tree with $n + 1$ bays.

Now note that all the distributions shown have the following property: any one may be obtained from the first, 1, 2, 4, 8, by moving one or more units from a larger number to a smaller number, or by a succession of such operations, without moving any units to the number 1. Thus 1, 3, 3, 8 is obtained by moving a unit from 4 to 2. The set 1, 4, 5, 5 is obtained by first moving two units from the 8 to the 2, then one unit to the 4. Furthermore, every set that may be obtained from the set 1, 2, 4, 8 by this process appears as a possible distribution. This operation is somewhat analogous to heat flow—heat can only flow from a hotter body to a cooler one just as units can only be transferred from higher numbers to lower ones in the above.

These considerations suggest that a disjunctive tree with n bays can be constructed with any load distribution obtained by such a flow from the initial distribution

$$1, 2, 4, 8, \cdots, 2^{n-1}$$

We will now show that this is actually the case.

First let us make the following definition: The symbol $(a_1, a_2, \cdots, a_n)$ represents any set of numbers $b_1, b_2, \cdots, b_n$ that may be obtained from the set $a_1, a_2, \cdots, a_n$ by the following operations:

1. Interchange of letters.

2. A flow from a larger number to a smaller one, no flow, however, being allowed to the number 1. Thus we would write

$$1, 2, 4, 8 = (1, 2, 4, 8)$$

$$4, 4, 1, 6 = (1, 2, 4, 8)$$

$$1, 3, 10, 3, 10 = (1, 2, 4, 8, 12)$$

but $2, 2 \neq (1, 3)$. It is possible to put the conditions that

$$b_1, b_2, \cdots, b_n = (a_1, a_2, \cdots, a_n) \tag{6}$$

into a more mathematical form. Let the a_i and the b_i be arranged as non-decreasing sequences. Then a necessary and sufficient condition for the relation (6) is that

(1) $$\sum_{i=1}^{s} b_i \geq \sum_{1}^{s} a_i \qquad s = 1, 2, \cdots, n,$$

(2) $$\sum_{1}^{n} b_i = \sum_{1}^{n} a_i, \quad \text{and}$$

(3) There are the same number of 1's among the a_i as among the b_i. The necessity of (2) and (3) is obvious. (1) follows from the fact that if a_i is non-decreasing, flow can only occur toward the left in the sequence

$$a_1, a_2, a_3, \cdots, a_n$$

and the sum $\sum_{1}^{s} a_i$ can only increase. Also it is easy to see the sufficiency of the condition, for if $b_1, b_2, \cdots, b_n$ satisfies (1), (2), and (3) we can get the b_i by first bringing a_1 up to b_1 by a flow from the a_i as close as possible to a_1 (keeping the "entropy" low by a flow between elements of nearly the same value), then bringing a_2 up to b_2 (if necessary) etc. The details are fairly obvious.

Additive number theory, or the problem of decomposing a number into the sum of numbers satisfying certain conditions, (in our case this definition is generalized to "sets of numbers") enters through the following Lemma:

Lemma: If $a_1, a_2, \cdots, a_n = (2, 4, 8, \cdots, 2^n)$
then we can decompose the a_i into the sum of two sets

$$a_i = b_i + c_i$$

such that

$$b_1, b_2, \cdots, b_n = (1, 2, 4, \cdots, 2^{n-1})$$

and

$$c_1, c_2, \cdots, c_n = (1, 2, 4, \cdots, 2^{n-1})$$

We may assume the a_i arranged in a non-decreasing sequence, $a_1 \leq a_2 \leq a_3 \leq \cdots \leq a_n$. In case $a_1 = 2$ the proof is easy. We have

$$\begin{array}{ll} 1, 2, 4, \cdots, 2^{n-1} & \text{B} \\ 1, 2, 4, \cdots, 2^{n-1} & \text{C} \\ \hline 2, 4, 8, \cdots, 2^n & \text{A} \end{array}$$

and a flow has occurred in the set

$$4, 8, 16, \cdots, 2^n$$

to give $a_2, a_3, \cdots, a_n$. Now any permissible flow in C corresponds to a permissible flow in either A or B since if

$$c_j = a_j + b_j > c_i = a_i + b_i$$

then either $a_j > a_i$ or $b_j > b_i$

Thus at each flow in the sum we can make a corresponding flow in one or the other of the summands to keep the addition true.

Now suppose $a_1 > 2$. Since the a_i are non-decreasing

$$(n - 1)\, a_2 \leq (2^{n+1} - 2) - a_1 \leq 2^{n+1} - 2 - 3$$

Hence

$$a_2 - 1 \leq \frac{2^{n+1} - 5}{n - 1} - 1 \leq 2^{n-1}$$

the last inequality being obvious for $n \geq 5$ and readily verified for $n < 5$. This shows that $(a_1 - 1)$ and $(a_2 - 1)$ lie between some powers of two in the set

$$1, 2, 4, \cdots, 2^{n-1}$$

Suppose

$$2^{q-1} < (a_1 - 1) \leq 2^q$$

$$2^{p-1} < (a_2 - 1) \leq a^p \qquad q \leq p \leq (n - 1).$$

Allow a flow between 2^q and 2^{q-1} until one of them reaches $(a_1 - 1)$, the other (say) R; similarly for $(a_2 - 1)$ the other reaching S. As the start toward our decomposition, then, we have the sets (after interchanges)

$$\begin{array}{ll|lllll} & & L & & & & \\ (a_1 - 1) & 1 & 2, 4 \cdots 2^{q-2} & R & 2^{q+1} \cdots 2^{p-1}\, 2^p & & 2^{p+1} \cdots 2^{n-1} \\ 1 & a_2 - 1 & 2, 4 \cdots 2^{q-2} & 2^{q-1} & 2^q \;\cdots & 2^{p-2}\, S & 2^{p+1} \cdots 2^{n-1} \\ \hline a_1 & a_2 & 4, 8 \cdots 2^{q-1} & \cdots & & & 2^{p+2} \cdots 2^n \\ & & L & & & & \end{array}$$

We must now adjust the values to the right of $L - L$ to the values $a_3, a_4, \cdots, a_n$. Let us denote the sequence

$$4, 8, \cdots, 2^{q-1}, (2^{q-1} + R), 3\cdot 2^q, 3\cdot 2^{q+1}, \cdots (2^p + S), 2^{p+2}, \cdots, 2^n$$

by $\mu_1, \mu_2, \cdots, \mu_{n-2}$. Now since all the rows in the above addition are non-decreasing to the right of $L - L$, and no 1's appear, we will have proved the lemma if we can show that

$$\sum_{i=1}^{i} \mu_i \leq \sum_{i=3}^{i+3} a_i \qquad i = 1, 2, \cdots, (n-2)$$

since we have shown this to be a sufficient condition that

$$a_3, a_4, \cdots, a_n = (\mu_1, \mu_n, \cdots, \mu_{n-2})$$

and the decomposition proof we used for the first part will work. For $i \leq q - 2$, i.e., before the term $(2^{q-1} + R)$

$$\sum_{i=1}^{i} \mu_i = 4(2^i - 1)$$

and

$$\sum_{3}^{i+3} a_i \geq ia_2 \geq i2^{p-1} \geq i2^{q-1}$$

since

$$q \leq p$$

Hence

$$\sum_{1}^{i} \mu_i \leq \sum_{3}^{i+3} a_i \qquad i \leq q - 2$$

Next, for $(q-1) \leq i \leq (p-3)$, i.e., before the term $(2^p + S)$

$$\sum_{1}^{i} \mu_i = 4(2^{q-1} - 1) + R + 3\cdot 2^q(2^{i-q+1} - 1)$$

$$< 3\cdot 2^{i+1} - 4 \leq 3\cdot 2^{i+1} - 5$$

since

$$R < 2^q$$

also again

$$\sum_{3}^{i+3} a_i \geq i2^{p-1}$$

so that in this interval we also have the desired inequality. Finally for the last interval,

$$\sum_1^i \mu_i = 2^{i-1} - a_1 - a_2 \leq 2^{i+3} - a_1 - a_2 - 2$$

and

$$\sum_3^{i+3} a_i = \sum_1^{i+3} a_i - a_1 - a_2 \geq 2^{i+3} - a_1 - a_2 - 2$$

since

$$a_1, a_2, \cdots, a_n = (2, 4, 8, \cdots, 2^n)$$

This proves the lemma.

5. The Disjunctive Tree

It is now easy to prove the following:

Theorem 8: A disjunctive tree of n bays can be constructed with any distribution

$$a_1, a_2, \cdots, a_n = (1, 2, 4, \cdots, 2^{n-1}).$$

We may prove this by induction. We have seen it to be true for $n = 2, 3, 4$. Assuming it for n, it must be true for $n + 1$ since the Lemma shows that any

$$a_1, a_2, \cdots, a_n = (2, 4, 8, \cdots, 2^n)$$

can be decomposed into a sum which, by assumption, can be realized for the two branches of the tree.

It is clear that among the possible distributions

$$(1, 2, 4, \cdots, 2^{n-1})$$

for the tree, an "almost uniform" one can be found for all the variables but one. That is, we can distribute the load on $(n - 1)$ of them uniformly except at worst for one element. We get, in fact, for

$n = 1$	1
$n = 2$	1, 2
$n = 3$	1, 3, 3
$n = 4$	1, 4, 5, 5,
$n = 5$	1, 7, 7, 8, 8,
$n = 6$	1, 12, 12, 12, 13, 13
$n = 7$	1, 21, 21, 21, 21, 21, 21
etc.	

as nearly uniform distributions.

6. Other Distribution Problems

Now let us consider the problem of load distribution in series-parallel circuits. We shall prove the following:

Theorem 9: Any function $f(X_1, X_2, \cdots, X_n)$ may be realized with a series-parallel circuit with the following distribution:

$$(1, 2, 4, \cdots, 2^{n-2}), 2^{n-2}$$

in terms of transfer elements.

This we prove by induction. It is true for $n = 3$, since any function of three variables can be realized as follows:

$$f(X, Y, Z) = [X + f_1(Y, Z)][X' + f_2(Y, Z)]$$

and $f_1(Y, Z)$ and $f_2(Y, Z)$ can each be realized with one transfer on Y and one on Z. Thus $f(X, Y, Z)$ can be realized with the distribution 1, 2, 2. Now assuming the theorem true for $(n - 1)$ we have

$$f(X_1, X_2, \cdots, X_n) = [X_n + f_1(X_1, X_2, \cdots, X_{n-1})]$$
$$[X'_n + f_2(X_1, X_2, \cdots, X_{n-1})]$$

and

$$\begin{array}{l} 2, 4, 8, \cdots, 2^{n-3} \\ 2, 4, 8, \cdots, 2^{n-3} \\ \hline 4, 8, 16, \cdots, 2^{n-2} \end{array}$$

A simple application of the Lemma thus gives the desired result. Many distributions beside those given by Theorem 9 are possible but no simple criterion has yet been found for describing them. We cannot say any distribution

$$(1, 2, 4, 8, \cdots, 2^{n-2}, 2^{n-2})$$

(at least from our analysis) since for example

$$3, 6, 6, 7 = (2, 4, 8, 8)$$

cannot be decomposed into two sets

$$a_1, a_2, a_3, a_4 = (1, 2, 4, 4)$$

and

$$b_1, b_2, b_3, b_4 = (1, 2, 4, 4)$$

It appears, however, that the almost uniform case is admissible.

As a final example in load distribution we will consider the case of a network in which a number of trees in the same variables are to be realized. A large number of such cases will be found later. The following is fairly obvious from what we have already proved.

Theorem 10: It is possible to construct m different trees in the same n variables with the following distribution:

$$a_1, a_2, \cdots, a_n = (m, 2m, 4m, \cdots, 2^{n-1}m)$$

It is interesting to note that under these conditions the bothersome 1 disappears for $m > 1$. We can equalize the load on all n of the variables, not just $n - 1$ of them, to within, at worst, one transfer element.

7. The Function $\mu(n)$

We are now in a position to study the behavior of the function $\mu(n)$. This will be done in conjunction with a treatment of the load distributions possible for the general function of n variables. We have already shown that any function of three variables can be realized with the distribution

$$1, 1, 2$$

in terms of transfer elements, and, consequently $\mu(3) \leq 4$.

Any function of four variables can be realized with the distribution

$$1, 1, (2, 4)$$

Hence $\mu(4) \leq 6$. For five variables we can get the distribution

$$1, 1, (2, 4, 8)$$

or alternatively

$$1, 5, 5, (2, 4)$$

so that $\mu(5) \leq 10$. With six variables we can get

$$1, 5, 5, (2, 4, 8) \text{ and } \mu(6) \leq 10$$

for seven,

$$1, 5, 5, (2, 4, 8, 16) \text{ and } \mu(7) \leq 16$$

etc. Also, since we can distribute uniformly on all the variables in a tree except one, it is possible to give a theorem analogous to Theorem 7 for the function $\mu(n)$:

Theorem 11: For all n

$$\mu(n) \leq \frac{2^{n+3}}{n^2}$$

For almost all n

$$\mu(n) \leq \frac{2^{n+2}}{n^2}$$

For an infinite number of n_i,

$$\mu(n) \leq (1 + \epsilon) \frac{2^{n+1}}{n^2}$$

The proof is direct and will be omitted.

PART III: SPECIAL FUNCTIONS

8. Functional Relations

We have seen that almost all functions require the order of

$$\frac{2^{n+1}}{n^2}$$

elements per relay for their realization. Yet a little experience with the circuits encountered in practice shows that this figure is much too large. In a sender, for example, where many functions are realized, some of them involving a large number of variables, the relays carry an average of perhaps 7 or 8 contacts. In fact, almost all relays encountered in practice have less than 20 elements. What is the reason for this paradox? The answer, of course, is that the functions encountered in practice are far from being a random selection. Again we have an analogue with transcendental numbers —although almost all numbers are transcendental, the chance of first encountering a transcendental number on opening a mathematics book at random is certainly much less than 1. The functions actually encountered are simpler than the general run of Boolean functions for at least two major reasons:

(1) A circuit designer has considerable freedom in the choice of functions to be realized in a given design problem, and can often choose fairly simple ones. For example, in designing translation circuits for telephone work it is common to use additive codes and also codes in which the same number of relays are operated for each possible digit. The fundamental logical simplicity of these codes reflects in a simplicity of the circuits necessary to handle them.

(2) Most of the things required of relay circuits are of a logically simple nature. The most important aspect of this simplicity is that most circuits can be broken down into a large number of small circuits. In place of realizing a function of a large number of variables, we realize many functions, each of a small number of variables, and then perhaps some function of these functions. To get an idea of the effectiveness of this consider the following example: Suppose we are to realize a function

$$f(X_1, X_2, \cdots, X_{2n})$$

of $2n$ variables. The best limit we can put on the total number of elements necessary is about $\frac{2^{2n+1}}{2n}$. However, if we know that f is a function of two functions f_1 and f_2, each involving only n of the variables, i.e. if

$$f = g(f_1, f_2)$$

$$f_1 = f_1(X_1, X_2, \cdots, X_n)$$

$$f_2 = f_2(X_{n+1}, X_{n+2}, \cdots, X_{2n})$$

then we can realize f with about

$$4 \cdot \frac{2^{n+1}}{n}$$

elements, a much lower order of infinity than $\frac{2^{2n+1}}{2n}$. If g is one of the simpler functions of two variables; for example if $g(f_1, f_2) = f_1 + f_2'$, or in any case at the cost of two additional relays, we can do still better and realize f with about $2\frac{2^{n+1}}{n}$ elements. In general, the more we can decompose a synthesis problem into a combination of simple problems, the simpler the final circuits. The significant point here is that, due to the fact that f satisfies a certain functional relation

$$f = g(f_1, f_2),$$

we can find a simple circuit for it compared to the average function of the same number of variables.

This type of functional relation may be called functional separability. It is often easily detected in the circuit requirements and can always be used to reduce the limits on the number of elements required. We will now show that most functions are not functionally separable.

Theorem 12: The fraction of all functions of n variables that can be written in the form

$$f = g(h(X_1 \cdots X_s), X_{s+1}, \cdots, X_n)$$

where $1 < s < n - 1$ approaches zero as n approaches ∞.

We can select the s variables to appear in h in $\binom{n}{s}$ ways; the function h then has 2^{2^s} possibilities and g has $2^{2^{n-s+1}}$ possibilities, since it has $n - s + 1$ arguments. The total number of functionally separable functions is therefore dominated by

$$\sum_{s=2}^{n-2} \binom{n}{s} 2^{2^s} 2^{2^{n-s+1}}$$

$$\leq (n - 3) \frac{n^2}{2} 2^{2^2} 2^{2^{n-1}}$$

and the ratio of this to $2^{2^n} \rightarrow 0$ as $n \rightarrow \infty$.

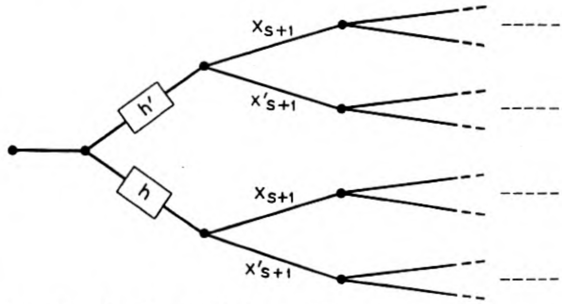


Fig. 20—Use of separability to reduce number of elements.

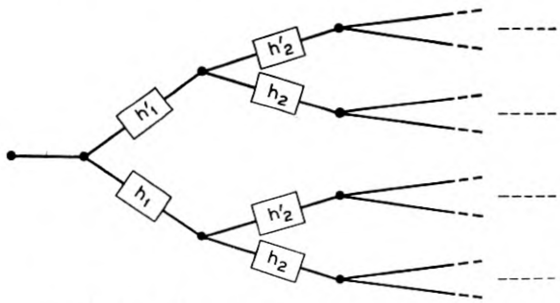


Fig. 21—Use of separability of two sets of variables.

In case such a functional separability occurs, the general design method described above can be used to advantage in many cases. This is typified by the circuit of Fig. 20. If the separability is more extensive, e.g.

$$f = g(h_1(X_1 \cdots X_s), h_2(X_{s+1} \cdots X_t), X_{t+1}, \cdots, X_n)$$

the circuit of Fig. 21 can be used, using for "h_2" either h_1 or h_2, whichever requires the least number of elements for realization together with its negative.

We will now consider a second type of functional relation which often occurs in practice and aids in economical realization. This type of relation may be called group invariance and a special case of it, functions symmetric in all variables, has been considered in (6). A function $f(X_1, \cdots, X_n)$ will be said to be symmetric in X_1, X_2 if it satisfies the relation

$$f(X_1, X_2, \cdots, X_n) = f(X_2, X_1, \cdots, X_n).$$

It is symmetric in X_1 and X'_2 if it satisfies the equation

$$f(X_1, X_2, \cdots, X_n) = f(X'_2, X'_1, X_3, \cdots, X_n)$$

These also are special cases of the type of functional relationships we will consider. Let us denote by

$N_{oo \cdots o} = I$ the operation of leaving the variables in a function as they are,

$N_{1oo \cdots o}$ the operation of negating the first variable (i.e. the one occupying the first position),

$N_{o1o \cdots o}$ that of negating the second variable,

$N_{11o \cdots o}$ that of negating the first two, etc.

So that $N_{101}f(X, Y, Z) = f(X'YZ')$ etc.

The symbols Ni form an abelian group, with the important property that each element is its own inverse; $NiNi = I$ The product of two elements may be easily found — if $N_i N_j = N_k$, k is the number found by adding i and j as though they were numbers in the base two but *without carrying.*

Note that there are 2^n elements to this "negating" group. Now let

$S_{1,2,3,\ldots,n} = I =$ the operation of leaving the variables of a function in the same order

$S_{2,1,3,\ldots,n} =$ be that of interchanging the first two variables

$S_{3,2,1,4,\ldots,n} =$ that of inverting the order of the first three, etc.

Thus

$$S_{312}f(X, Y, Z) = f(Z, X, Y)$$

$$S_{312}f(Z, X, Y) = S_{312}^2 f(X, Y, Z) = f(Y, Z, X)$$

etc. The S_i also form a group, the famous "substitution" or "symmetric" group. It is of order $n!$. It does not, however, have the simple properties of the negating group—it is not abelean ($n > 2$) nor does it have the self inverse property.* The negating group is not cyclic if $n > 2$, the symmetric group is not if $n > 3$.

The outer product of these two groups forms a group G whose general element is of the form $N_i S_j$ and since i may assume 2^n values and j, $n!$ values, the order of G is $2^n n!$

It is easily seen that $S_j N_i = N_k S_j$, where k may be obtained by per-

* This is redundant; the self inverse property implies commutativity for if $XX = I$ then $XY = (XY)^{-1} = Y^{-1}X^{-1} = YX$.

forming on i, considered as an ordered sequence of zero's and one's, the permutation S_j. Thus

$$S_{2314}N_{1100} = N_{1010}S_{2314}.$$

By this rule any product such as $N_i S_j N_k N_l S_m N_n S_p$ can be reduced to the form

$$N_i N_j \cdots N_n S_p S_q \cdots S_r$$

and this can then be reduced to the standard form $N_i S_j$.

A function f will be said to have a non-trivial group invariance if there are elements $N_i S_j$ of G other than I such that identically in all variables

$$N_i S_j f = f.$$

It is evident that the set of all such elements, $N_i S_j$, for a given function, forms a subgroup G_1 of G, since the product of two such elements is an element, the inverse of such an element is an element, and all functions are invariant under I.

A group operator leaving a function f invariant implies certain equalities among the terms appearing in the expanded form of f. To show this, consider a fixed $N_i S_j$, which changes in some way the variables (say) $X_1, X_2, \cdots, X_r$. Let the function $f(X_1, \cdots, X_n)$ be expanded about $X_1, \cdots, X_r$:

$$\begin{aligned} f = {} & [X_1 + X_2 + \cdots + X_r + f_1(X_{r+1}, \cdots, X_n)] \\ & [X_1' + X_2 + \cdots + X_r + f_2(X_{r+1}, \cdots, X_n)] \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & [X_1' + X_2' + \cdots + X_r' + f_{2^r}(X_{r+1}, \cdots, X_n)] \end{aligned}$$

If f satisfies $N_i S_j f = f$ we will show that there are at least $\frac{1}{4}2^r$ equalities between the functions $f_1, f_2, \cdots, f_{2^r}$. Thus the number of functions satisfying this relation is

$$\leq (2^{2^{n-r}})^{\frac{3}{4}2^r} = 2^{\frac{3}{4}2^n}$$

since each independent f_i can be any of just $2^{2^{n-r}}$ functions, and there are at most $\frac{3}{4}2^r$ independent ones. Suppose $N_i S_j$ changes

$$X_1, X_2, \cdots, X_r \qquad \text{A}$$

into

$$X_{a_1}^*, X_{a_2}^*, \cdots, X_{a_r}^* \qquad \text{B}$$

where the *'s may be either primes or non primes, but no $X_{a_i}^* = X_i$. Give

X_1 the value 0. This fixes some element in B namely, X_{a_i} where $a_i = 1$. There are two cases:
(1) If this element is the first term, $a_1 = 1$, then we have

$$0\, X_2\,, \cdots, X_r$$

$$1\, X_{a_1}\,, \cdots, X_{a_r}$$

Letting $X_2, \cdots, X_r$ range through their 2^{r-1} possible sets of values gives 2^{r-1} equalities between different functions of the set f_i since these are really

$$f(X_1, X_2, \cdots, X_r, X_{r+1}, \cdots, X_n)$$

with $X_1, X_2, \cdots, X_r$ fixed at a definite set of values.
(2) If the element in question is another term, say X_{a_2}, we then give X_2 in line A the opposite value, $X_2 = (X_{a_2}^*)' = (X_2^*)'$. Now proceeding as before with the remaining $r - 2$ variables we establish 2^{r-2} equalities between the f_i.

Now there are not more* than $2^n n!$ relations

$$N_i S_j f = f$$

of the group invariant type that a function could satisfy, so that the number of functions satisfying *any* non-trivial relation

$$\leq 2^n\, n!\, 2^{\frac{3}{4}2^n}.$$

Since

$$2^n\, n!\, 2^{\frac{3}{4}2^n} / 2^{2^n} \to 0 \qquad \text{as } n \to \infty$$

we have:

Theorem 13: Almost all functions have no non-trivial group invariance.

It appears from Theorems 12 and 13 and from other results that almost all functions are of an extremely chaotic nature, exhibiting no symmetries or functional relations of any kind. This result might be anticipated from the fact that such relations generally lead to a considerable reduction in the number of elements required, and we have seen that almost all functions are fairly high in "complexity".

If we are synthesizing a function by the disjunctive tree method and the function has a group invariance involving the variables

$$X_1, X_2, \cdots, X_r$$

at least 2^{r-2} of the terminals in the corresponding tree can be connected to

* Our factor is really less than this because, first, we must exclude $N_i S_j = I$; and second, except for self inverse elements, one relation of this type implies others, viz. the powers $(N_i S_j)^p f = f$.

other ones, since at least this many equalities exist between the functions to be joined to these terminals. This will, in general, produce a considerable reduction in the contact requirements on the remaining variables. Also an economy can usually be achieved in the M network. In order to apply this

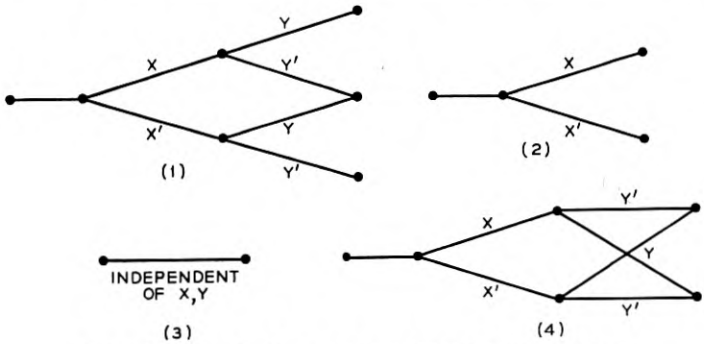


Fig. 22—Networks for group invariance in two variables.

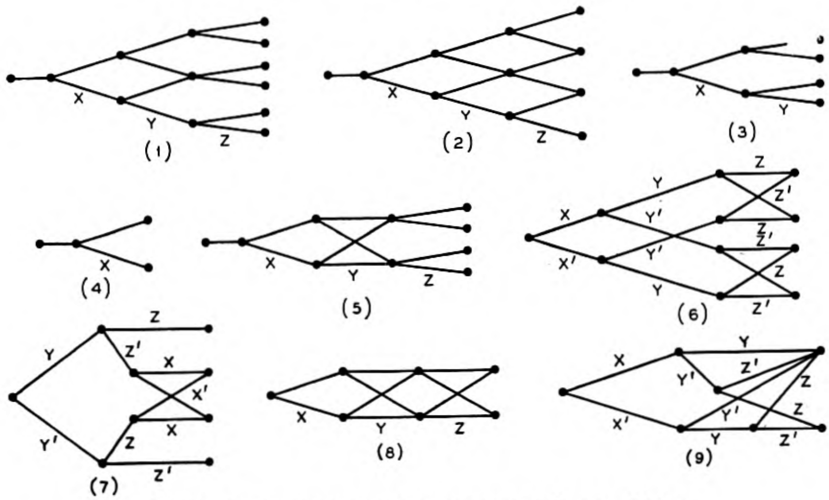


Fig. 23—Networks for group invariance in three variables.

method of design, however, it is essential that we have a method of determining which, if any, of the $N_i S_j$ leave a function unchanged. The following theorem, although not all that might be hoped for, shows that we don't need to evaluate $N_i S_j f$ for all $N_i S_j$ but only the $N_i f$ and $S_j f$.

Theorem 14: A necessary and sufficient condition that $N_i S_j f = f$ *is that* $N_i f = S_j f$.

This follows immediately from the self inverse property of the N_i. Of course, group invariance can often be recognized directly from circuit requirements in a design problem.

Tables I and II have been constructed for cases where a relation exists involving two or three variables. To illustrate their use, suppose we have a function such that

$$N_{111} S_{312} f = f$$

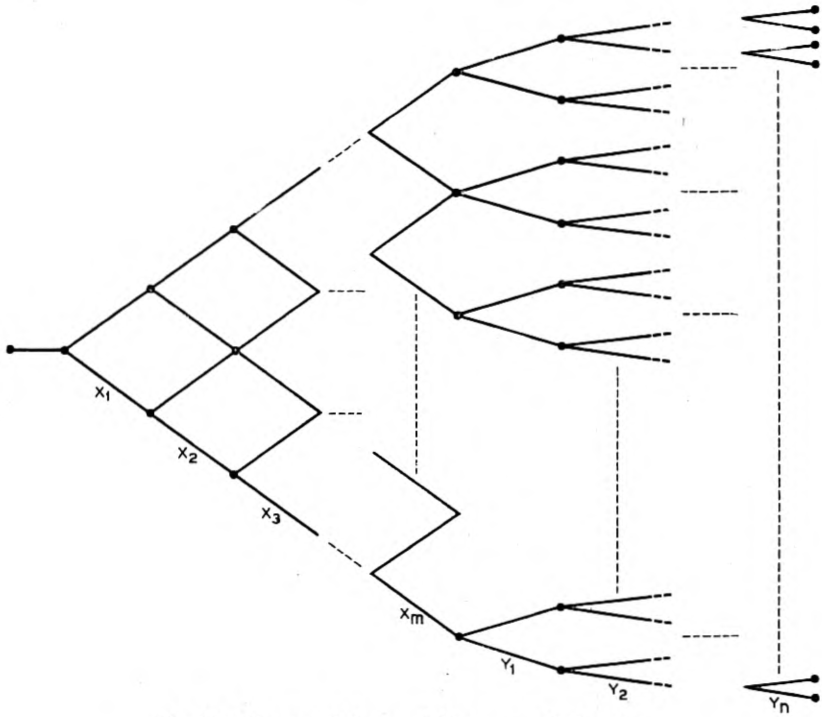


Fig. 24—M network for partially symmetric functions.

The corresponding entry $Z'Y'X$ in the group table refers us to circuit 9 of Fig. 23. The asterisk shows that the circuit may be used directly; if there is no asterisk an interchange of variables is required. We expand f about X, Y, Z and only two different functions will appear in the factors. These two functions are realized with two trees extending from the terminals of the network 9. Any such function f can be realized with (using just one variable in the N network)

$$9 + 2(2^{n-4} - 2) + 2$$

$$= 2^{n-3} + 7 \qquad \text{elements,}$$

a much better limit than the corresponding

$$2^{n-1} + 18$$

for the general function.

TABLE I

GROUP INVARIANCE INVOLVING TWO VARIABLES (SUPERSCRIPTS REFER TO FIG. 22)

	S_{12}	S_{21}
N_{oo}	$(x\ y)$	$(y\ x)^{1*}$
N_{o1}	$(x\ y')^{2*}$	$(y\ x')^{3*}$
N_{10}	$(x'y)^{2}$	$(y'x)^{3*}$
N_{11}	$(x'y')^{4}$	$(y'x')^{1}$

TABLE II

GROUP INVARIANCE INVOLVING THREE VARIABLES (SUPERSCRIPTS REFER TO FIG. 23)

	S_{123}	S_{132}	S_{213}	S_{231}	S_{312}	S_{321}
N_{ooo}	XYZ	XZY^{1}	YXZ^{1*}	YZX^{2*}	ZXY^{2*}	ZYX [illegible]
N_{oo1}	XYZ'^{3*}	XZY'^{4*}	YXZ'^{7}	YZX'^{9}	ZXY'^{9}	ZYX'^{4}
N_{o1o}	$XY'Z^{3}$	$XZ'Y^{4*}$	$YX'Z^{4}$	$YZ'X^{9}$	$ZX'Y^{9}$	$ZY'X^{7}$
N_{o11}	$XY'Z'^{5}$	$XZ'Y'^{1}$	$YX'Z'^{8}$	$YZ'X'^{2}$	$ZX'Y'^{2}$	$ZY'X'^{8}$
N_{1oo}	$X'YZ^{3}$	$X'ZY^{7*}$	$Y'XZ^{4}$	$Y'ZX^{9}$	$Z'XY^{9}$	$Z'YX^{4}$
N_{101}	$X'YZ'^{5}$	$X'ZY'^{8*}$	$Y'XZ'^{8*}$	$Y'ZX'^{2}$	$Z'XY'^{2}$	$Z'YX'^{1}$
N_{110}	$X'Y'Z^{5*}$	$X'Z'Y^{8*}$	$Y'X'Z^{1}$	$Y'Z'X^{2}$	$Z'X'Y^{2}$	$Z'Y'X^{8*}$
N_{111}	$X'Y'Z'^{6}$	$X'Z'Y'^{7}$	$Y'X'Z'^{7}$	$Y'Z'X'^{9}$	$Z'X'Y'^{9}$	$Z'Y'X'^{7}$

9. PARTIALLY SYMMETRIC FUNCTIONS

We will say that a function is "partially symmetric" or "symmetric in a certain set of variables" if these variables may be interchanged at will without altering the function. Thus

$$XYZW + (XY' + X'Y)W + WZ'$$

is symmetric in X and Y. Partial symmetry is evidently a special case of the general group invariance we have been considering. It is known that any function symmetric in all variables can be realized with not more than n^2 elements, where n is the number of variables.[6] In this section we will improve and generalize this result.

Theorem 15: Any function $f(X_1, X_2, \cdots, X_m, Y_1, Y_2, \cdots Y_n)$ *symmetric in* $X_1, X_2, \cdots, X_m$ *can be written*

$$\begin{aligned} & f(X_1, X_2, \cdots, X_m, Y_1, Y_2, \cdots, Y_n) \\ = & [S_o(X_1, X_2, \cdots, X_m) + f_o(Y_1, Y_2, \cdots, Y_n)]. \\ & [S_1(X_1, X_2, \cdots, X_m) + f_1(Y_1, Y_2, \cdots, Y_n)]. \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & [S_m(X_1, X_2, \cdots, X_m) + f_m(Y_1, Y_2, \cdots, Y_n)] \end{aligned} \qquad (6)$$

where

$$\begin{aligned} & f_K(Y_1, Y_2, \cdots, Y_n) \\ = & f(\underbrace{0, 0, \cdots, 0}_{k\ 0\text{'s}}, \underbrace{1, 1, \cdots, 1}_{(m-k)\ 1\text{'s}}, Y_1, Y_2, \cdots, Y_n) \end{aligned}$$

and $S_k(X_1, X_2, \cdots, X_m)$ *is the symmetric function of* $X_1, X_2, \cdots, X_n$ *with* k *for its only a-number.*

This theorem follows from the fact that since f is symmetric in $X_1, X_2, \cdots, X_m$ the value of f depends only on the *number* of X's that are zero and the values of the Y's. If exactly K of the X's are zero the value of f is therefore f_K, but the right-hand side of (6) reduces to f_K in this case, since then $S_j(X_1, X_2, \cdots, X_m) = 1, j \neq K$, and $S_K = 0$.

The expansion (6) is of a form suitable for our design method. We can realize the disjunctive functions $S_K(X_1, X_2, \cdots, X_n)$ with the symmetric function lattice and continue with the general tree network as in Fig. 24, one tree from each level of the symmetric function network. Stopping the trees at Y_{n-1}, it is clear that the entire network is disjunctive and a second application of Theorem 1 allows us to complete the function f with two elements from Y_n. Thus we have

Theorem 16. Any function of $m + n$ *variables symmetric in* m *of them can be realized with not more than the smaller of*

$$(m + 1)(\lambda(n) + m) \text{ or } (m + 1)(2^n + m - 2) + 2$$

elements. In particular a function of n *variables symmetric in* $n - 2$ *or more of them can be realized with not more than*

$$n^2 - n + 2$$

elements.

If the function is symmetric in $X_1, X_2, \cdots, X_m$, and also in $Y_1, Y_2, \cdots, Y_r$, and not in $Z_1, Z_2, \cdots, Z_n$ it may be realized by the same method, using symmetric function networks in place of trees for the Y variables. It should be expanded first about the X's (assuming $m < r$) then about the Y's and finally the Z's. The Z part will be a set of $(m + 1)(r + 1)$ trees.

References

1. G. Birkhoff and S. MacLane, "A Survey of Modern Algebra," Macmillan, 1941.
2. L. Couturat, "The Algebra of Logic," Open Court, 1914.
3. J. H. Woodger, "The Axiomatic Method in Biology," Cambridge, 1937.
4. W. S. McCulloch and W. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bull. Math. Biophysics*, V. 5, p. 115, 1943.
5. E. C. Berkeley, "Boolean Algebra and Applications to Insurance," *Record (American Institute of Actuaries)*, V. 26, p. 373, 1947.
6. C. E. Shannon, "A Symbolic Analysis of Relay and Switching Circuits," *Trans. A. I. E. E.*, V. 57, p. 713, 1938.
7. J. Riordan and C. E. Shannon, "The Number of Two-Terminal Series Parallel Networks," *Journal of Mathematics and Physics*, V. 21, No. 2, p. 83, 1942.
8. A. Nakashima, Various papers in *Nippon Electrical Communication Engineering*, April, Sept., Nov., Dec., 1938.
9. H. Piesch, Papers in Archiv. from *Electrotechnik* XXXIII, p. 692 and p. 733, 1939.
10. G. A. Montgomerie, "Sketch for an Algebra of Relay and Contactor Circuits," *Jour. I. of E. E.*, V. 95, Part III, No. 36, July 1948, p. 303.
11. G. Pólya, "Sur Les Types des Propositions Composées," *Journal of Symbolic Logic*, V. 5, No. 3, p. 98, 1940.

A Method of Measuring Phase at Microwave Frequencies

By SLOAN D. ROBERTSON

A method of measuring microwave phase differences is described in which it is unnecessary to compensate for amplitude inequalities between the signals whose phases are being compared. The apparatus described is also suited for the measurement of the magnitude of a transfer impedance as well as the phase.

WITH the increasing interest in wide-band amplifiers and circuits for microwave communication systems the measurement of the transfer phases of such components has become a necessary procedure. A commonly used technique for measuring phase at microwave frequencies is to sample the signal at the input and output of the device to be measured and to obtain a null balance between the two signals by varying the phase of one signal by a known amount. If the two samples are not of nearly equal amplitudes, it is necessary to attenuate the larger one with an attenuator of known phase shift. The latter operation presents difficulties.

A method of phase measurement has been developed which overcomes these difficulties by permitting measurements to be made with samples of unequal amplitudes. The method uses the homodyne detection principle and operates in the following manner: The output energy of a signal oscillator is divided into two portions. One portion is applied to a balanced modulator where it is modulated by an audio-frequency signal. The suppressed-carrier, double-sideband signal from the modulator is applied to the device to be measured. As before, means are available for sampling the signal at both the input and output of the device. The other portion of the oscillator power is fed through a calibrated phase shifter and is applied to a crystal detector in the manner of a local oscillator in a double-detection receiver. The signal samples are then alternately applied to the crystal detector where they are demodulated by the action of the homodyne carrier. In each case the phase shifter is adjusted so that the audio signal is a minimum in the detector output. This occurs when the phase of the homodyne carrier is in quadrature with the signal sidebands. The difference in phase between the two adjustments of the phase shifter is equal to the phase difference between the two samples.

Figure 1 shows the apparatus used for measuring phase in this manner. Radio frequency power from a suitable oscillator is applied to the H-plane branch of an hybrid junction[1] where it divides and emerges in equal portions

[1] W. A. Tyrrell, "Hybrid Circuits for Microwaves," *Proc. I. R. E.*, Vol. 35, No. 11, pp. 1294–1306; November 1947.

from the two lateral branches. The portion applied to the calibrated variable phase shifter at the top of the figure becomes the homodyne carrier. The remaining portion is applied to a balanced crystal modulator[2] through a second variable phase shifter which need not be calibrated. The latter was introduced in order that the phase of any modulated power reflected due to an imperfect balance in the modulator could be shifted so that it would be in quadrature with the homodyne carrier and would, therefore, not produce an audible signal in the detector.

The portion of the power which enters the modulator is modulated by a signal derived from an audio-frequency oscillator. The suppressed-carrier, double-sideband signal which leaves the modulator is applied, after a certain amount of attenuation, to the input of the device to be measured. Probes are provided at the input and output of the latter for sampling the signal. Provision is made for connecting either probe to a crystal detector of the type used for detecting an amplitude-modulated signal.

The homodyne carrier emerging from the calibrated phase shifter is attenuated to a level of about one milliwatt and is applied to the crystal detector. The output of the detector is connected to an audio-frequency amplifier terminated by a pair of headphones or an output meter. An attenuator may be placed between the amplifier and the detector as an aid in measuring the magnitude of a transfer impedance.

The procedure for adjusting the apparatus and measuring phase is as follows:

With both sampling probes disconnected from the detector the variable phase shifter between the oscillator and modulator is adjusted until the output of the detector is zero. This balances out the effect of any signal reflected by the modulator. The input probe is then connected to the detector and the calibrated phase shifter is adjusted until the signal disappears in the audio output. When this occurs the homodyne carrier is in quadrature with the signal sidebands, and the resultant signal applied to the detector is equivalent to a phase-modulated wave having a low modulation index, and consequently is not demodulated by a detector of the type used here.

The input probe is then disconnected from the detector and the output probe connected. The phase shifter is again adjusted for a null in the audio output. The difference in phase between the two adjustments of the phase shifter is equal to the phase shift between the input and output of the device. If the probes are not located exactly at the input and output terminals of the unknown it may be necessary to make a correction in the meas-

[2] C. F. Edwards, "Microwave Converters," *Proc. I. R. E.*, Vol. 35, No. 11, pp. 1181–1191; November 1947.

ured phase by allowing for the known phase shift in the line between the probes and the actual terminals of the unknown.

So much for the general method. Certain precautions are necessary in order to avoid errors in measurement. In practice the carrier is not completely suppressed in the output of the balanced modulator. It may be at a

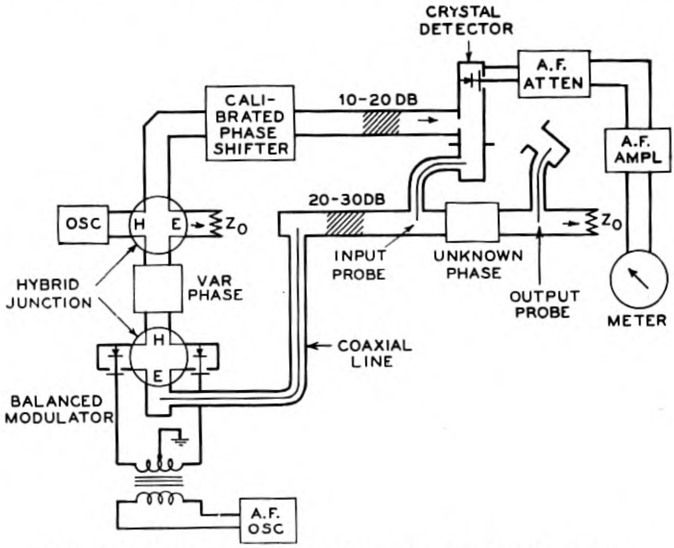


Fig. 1—Schematic circuit for microwave phase measurement.

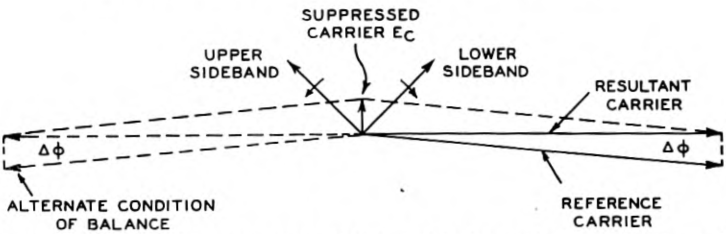


Fig. 2—Vector diagram of balanced condition with the resultant carrier in quadrature with the signal sideband.

level of the order of 10 to 20 decibels below the sidebands. Since the residual carrier will be added to the homodyne carrier in the detector, and since the null adjustment will be reached when the *resultant* carrier is in quadrature with the sidebands, it is desirable that the residual carrier be low in level compared with the homodyne carrier. The error in phase $\Delta\phi$ introduced by the residual carrier is shown in the vector diagram of Fig. 2. A difference in level of about 40 decibels between the homodyne and residual carriers will give an error of not more than half a degree in phase. The homodyne method of detection has all the conversion efficiency of the usual double-detection arrangements and, in addition, has the advantage in this particular application of having a very low noise level due to the relatively narrow band required for the audio signals. The 40-decibel level difference mentioned above is accordingly not a serious handicap.

Other precautions must be observed. The homodyne carrier can be brought in quadrature with the signal for two different phases 180° apart. This is illustrated in Fig. 2. In many applications, where only the variation

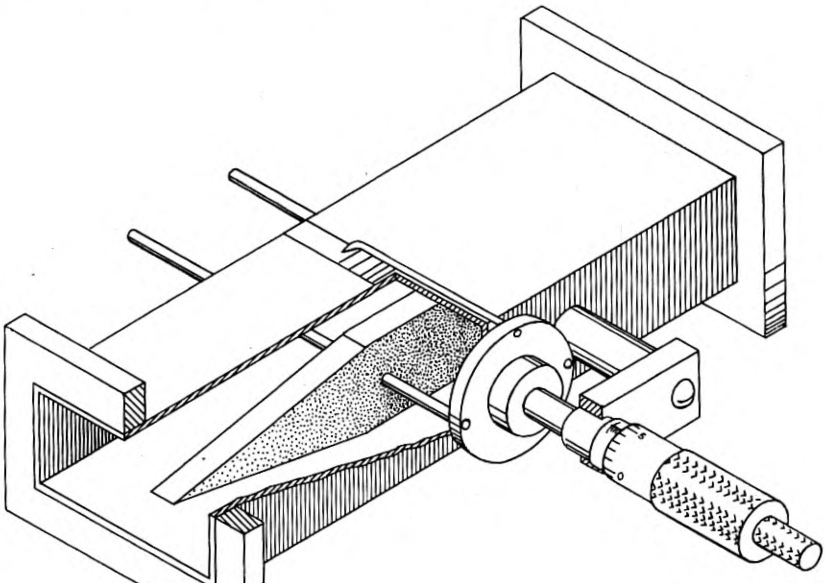


Fig. 3—Variable phase shifter using a polystyrene vane.

in phase difference is of importance, this uncertainty of 180° can be ignored. The correct setting of the homodyne carrier phase can, however, be determined very easily. Assume that the input probe is connected to the receiver and that the phase has been adjusted for a balance. Then disconnect the audio frequency drive from one of the crystals in the balanced modulator. The residual carrier will now no longer be suppressed and the error angle $\Delta\phi$ of Fig. 2 will become larger. Whether the homodyne carrier is lagging or leading the signal carrier can be determined by observing whether more or less phase shift, respectively, must be introduced to restore balance. A similar test performed with the output probe will indicate whether or not it is necessary to add 180° to the measured phase difference. If either probe test indicates a lead, whereas the other probe indicates a lag, then the addition of 180° is indicated.

In microwave circuits it frequently happens that the transfer phase varies quite rapidly with the frequency, particularly if some part of the circuit is at or near resonance. In measuring the phase characteristics of a circuit of this type over a band of frequencies it is necessary, therefore, to take the points of measurement close enough together to avoid phase errors corresponding to multiples of 360°.

When a balance has been established so that the signal is minimized in the detector output, one may observe the presence of the second harmonic of the audio tone. This harmonic is a distortion term generated in the detector. If it is objectionable, it can be eliminated either by a low-pass filter in the audio output, or by using a balanced detector.

In measuring transfer impedances it is desirable to know the ratio of the magnitudes of an output voltage and an input voltage as well as the phase difference. The equipment described here can be used for measuring amplitudes by adjusting the phase shifter for a maximum signal in the audio output. Maximum signal levels can then be compared with the aid of an audio-frequency attenuator and output meter connected as shown in Fig. 1.

The apparatus was assembled with standard 4000-megacycle waveguide components. A satisfactory phase shifter was made of an ordinary vane-type variable attenuator by replacing the resistance strip with a vane of quarter-inch thick polystyrene six inches in length. This phase shifter gave a total shift of about 100°. Constructional details of this phase shifter are shown in Fig. 3. Other phase shifters could have been used with equally satisfactory results. It is desirable, however, that the phase shifter be impedance matched to the line in which it is located in order that reaction back on the oscillator shall be a minimum. In the shifter of Fig. 3 the ends of the polystyrene vane have been tapered two inches at each end to accomplish this result.

The phase shifter can be readily calibrated by using a standing wave detector fitted with a sliding probe as a standard of phase. The standing wave detector is terminated on one end and connected to the modulated signal source on the other. The signal picked up by the sliding probe is applied to the crystal detector. Knowing the guide wavelength in the standing wave detector, known phase shifts can be introduced by sliding the probe along the guide. By adjusting the phase shifter in the homodyne carrier path for balance, calibration points can be established.

The measuring procedure described above has been tested experimentally at 4000 megacycles with very satisfactory results. With ordinary care it was possible to measure phase differences with an accuracy of better than half a degree.

Reflection from Corners in Rectangular Wave Guides—Conformal Transformation*

By S. O. RICE

A conformal transformation method is used to obtain approximate expressions for the reflection coefficients of sharp corners in rectangular wave guides. The transformation carries the bent guide over into a straight guide filled with a non-uniform medium. The reflection coefficient of the transformed system can be expressed in terms of the solution of an integral equation which may be solved approximately by successive substitutions. When the corner angle is small and the corner is not truncated the required integrations may be performed and an explicit expression obtained for the reflection coefficient. Although applied here only to corners, the method has an additional interest in that it is applicable to other types of irregularities in rectangular wave guides.

INTRODUCTION

THE propagation of electromagnetic waves around a rectangular corner has been studied in two recent papers, one by Poritsky and Blewett[1] and the other by Miles[2]. Poritsky and Blewett make use of Schwarz' "alternating procedure" in which a sequence of approximations is obtained by going back and forth between two overlapping regions. Miles derives an equivalent circuit by using solutions of the wave equation in rectangular coordinates. Several papers giving experimental results have been published. Of these, we mention one due to Elson[3] who gives values of reflection coefficients for various types of corners.

Here we shall deal with the more general type of corner shown in Fig. 1 by transforming, conformally, the bent guide (in which the propagation "constant" of the dielectric is constant) into a straight guide in which the propagation "constant" is a function of position—its greatest deviation from the original value being in the vicinity of points corresponding to the corner. This type of corner has been chosen for our example because it possesses a number of features common to problems which may be treated by the transformation method.

The essentials of the procedure used are due to Routh[4] who studied the vibration of a membrane of irregular shape by transforming it into a rectangle. After the transformation the density (analogous to the propagation constant in the guide) was no longer constant but this disadvantage was more than offset by the simplification in shape.

Until this paper was presented at the Symposium I was unaware of any

* Presented at the Second Symposium on Applied Mathematics, Cambridge, Mass., July 29, 1948.

[1] See list of references at end of paper.

other wave guide work based on conformal transformations (as described above) except that of Krasnooshkin[5]. At the meeting I learned that the transformation method had also been discovered (but not yet published) by Levine and by Piloty independently of each other. Levine has studied the same corner, see Fig. 1, as is done here. However, his method of approach is quite different in that he obtains expressions for the elements in the equivalent pi network representing the corner, whereas here the reflection coefficient is considered directly. This is discussed in more detail at the beginning of Section 6. Piloty's work is closely related to the material presented in a companion paper[6] and is discussed in its introduction.

In this paper the partial differential equation resulting from the transformation, together with the boundary conditions, is converted into a rather complicated integral equation. Numerical work indicates that satisfactory values of the reflection coefficient, in which we are primarily interested, may be obtained by solving this integral equation by the method of successive substitutions. However, the question of convergence is not investigated.

Although they are here applied only to corners, the equations of Sections 3, 4 and 5 are quite general. In order to test their generality they were used to check the expression[7] for the reflection coefficient of a gentle circular bend in a rectangular wave guide, E being in the plane of the bend. The work has been omitted because of its length. It was found that the essential parts of the transformation may be obtained by regarding the inner and outer walls of the guide system as the two plates of a condenser, solving the corresponding electrostatic problem (using series of the Fourier type), and utilizing the relation between two-dimensional potentials and the theory of conformal mapping.

When the angle of the corner is small we may obtain the series (7-5) and (7-11) for the reflection coefficients corresponding to simple (i.e. not truncated) E and H corners, respectively (a corner having the electric intensity E in the plane of the bend will be called an E corner or an electric corner. H corners are defined in a similar manner). When the angle of the general E corner shown in Fig. 1 is small we may use the series (7-18).

The series (7-5) and (7-11) giving the reflection from small angle corners are related to the series giving the reflection coefficients for gentle circular bends. In fact, if the radii of curvature of the latter be held constant while the angle of bend is made small, the series for the circular bends reduce to those for the corners.

As for the limitations of the method, note first that it can be used only for wave guide systems in which the dimension normal to the plane of transformation is constant throughout. Moreover, the integral equations of the present paper, except for the work of Appendix III, are derived on the assumption that the dimensions of the guide approach constant values at minus infinity and the same values at plus infinity. When this assumption is not met, a conformal transformation may still be used to carry the system into a straight guide. However, there appears to be some doubt as to the best way of dealing with the resulting partial differential equation. One method, discussed in the companion paper[6], leads to an infinite set of ordinary linear differential equations of the second order. Again, possibly the Green's functions appearing in Sections 3 and 5 may be replaced by suitable approximations.

1. Representation of Field for Corner or Bend in Rectangular Guide

Quite often waves in rectangular wave guides are classed as "transverse electric" or "transverse magnetic". However, for our purposes it is more convenient to class them as "electrically oriented" or "magnetically oriented" waves.[8,9] Thus, the electric and magnetic intensities are obtained by multiplying

$$\begin{aligned} E_x &= \frac{1}{i\omega\epsilon}\frac{\partial^2 A}{\partial x\partial\zeta} - \frac{\partial B}{\partial y} & H_x &= \frac{\partial A}{\partial y} + \frac{1}{i\omega\mu}\frac{\partial^2 B}{\partial x\partial\zeta} \\ E_y &= \frac{1}{i\omega\epsilon}\frac{\partial^2 A}{\partial y\partial\zeta} + \frac{\partial B}{\partial x} & H_y &= -\frac{\partial A}{\partial x} + \frac{1}{i\omega\mu}\frac{\partial^2 B}{\partial y\partial\zeta} \\ E_\zeta &= -i\omega\mu A + \frac{1}{i\omega\epsilon}\frac{\partial^2 A}{\partial\zeta^2} & H_\zeta &= -i\omega\epsilon B + \frac{1}{i\omega\mu}\frac{\partial^2 B}{\partial\zeta^2} \end{aligned} \tag{1-1}$$

by $e^{i\omega t}$ and taking the real part. Here ω, μ, and ϵ are the radian frequency, the permeability of the medium filling the guide ($\mu = 1.257 \times 10^{-6}$ henries per meter for air), and the dielectric constant of the same ($\epsilon = 8.854 \times 10^{-12}$ farads per meter for air), respectively. x, y, and ζ constitute a right-handed set of rectangular coordinates in which the ζ axis is normal to the plane of the bend. Equations (1-1) may be verified by substituting them in Maxwell's equations.

The potentials A and B satisfy the wave equation

$$\begin{gathered} \frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} + \frac{\partial^2 A}{\partial\zeta^2} = \sigma^2 A \\ \sigma = i\omega\sqrt{\mu\epsilon} = i2\pi/\lambda_0 \end{gathered} \tag{1-2}$$

where λ_0 is the wave length in free space corresponding to the radian frequency ω.

When the electric vector lies in the plane of the bend, as shown in Fig. 1, and the incident wave contains only the dominant mode we set

$$A = 0, \qquad B = Q \sin(\pi\zeta/a) \tag{1-3}$$

where a is the wide dimension of the rectangular cross-section, the guide walls normal to the ζ axis are at $\zeta = 0$ and $\zeta = a$, and Q is a function of x and y such that

$$\frac{\partial^2 Q}{\partial x^2} + \frac{\partial^2 Q}{\partial y^2} - \Gamma_{10}^2 Q = 0$$
$$\Gamma_{10} = i2\pi\lambda_0^{-1}(1 - \lambda_0^2 a^{-2}/4).^{1/2} \tag{1-4}$$

The guide walls are assumed to be perfect conductors and hence the tangential component of E must vanish at the walls. This requires the normal derivative of Q to vanish at those walls which are perpendicular to the plane of the bend:

$$\frac{\partial Q}{\partial n} = 0. \tag{1-5}$$

When the magnetic vector lies in the plane of the bend and the incident wave consists of the dominant mode, we set

$$A = P, \qquad B = 0 \tag{1-6}$$

where P is a function of x and y such that

$$\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} - \Gamma_{00}^2 P = 0, \qquad \Gamma_{00} = i2\pi/\lambda_0 \tag{1-7}$$

and

$$P = 0 \tag{1-8}$$

at the walls perpendicular to the plane of the bend. In this case the guide walls parallel to the plane of the bend are at $\zeta = 0$ and $\zeta = b$.

2. Electric Vector in Plane of Bend

Figure 1 shows a section of the bend taken parallel to the electric vector. b is the narrow dimension of the guide. Let the frequency and the wide dimension a of the guide (measured normal to the plane of Fig. 1) be such that only the dominant mode is freely propagated. The position of any point in this section is specified by the complex number $z = x + iy$ where the origin and the orientation of the axes have been chosen somewhat arbitrarily.

The constant k and related propagation constants which appear in the formulas dealing with Q and electric bends are given by

$$k = (2b/\lambda_0)\,[1 - (\lambda_0/2a)^2]^{1/2} = -i\Gamma_{10} b/\pi$$
$$\gamma_m^2 = m^2 - k^2; \qquad m = 0, 1, 2, \cdots; \qquad \gamma_0 = ik \tag{2-1}$$
$$\lambda_0 = \text{free space wavelength}$$

Since, by assumption, only the dominant mode is freely propagated, k and γ_m for $m > 0$ are real and positive.

We imagine an incident wave of unit amplitude coming down from z_5 in the upper left portion of Fig. 1. What are the amplitudes of the reflected wave traveling back toward z_5 and the transmitted wave traveling outward to the right towards z_3? Our task is to find a $Q(x, y)$, satisfying the wave equation (1-4) and the boundary condition (1-5), which represents a disturbance of the assumed type.

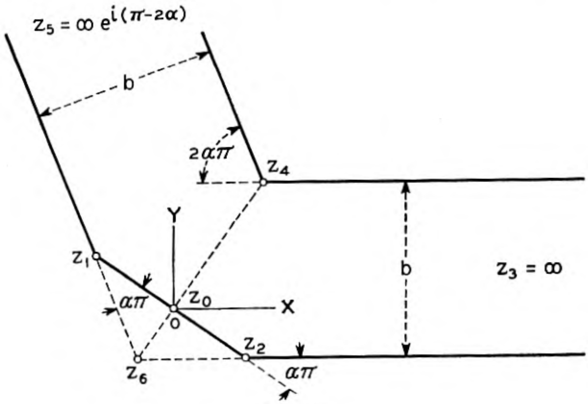


FIG. 1

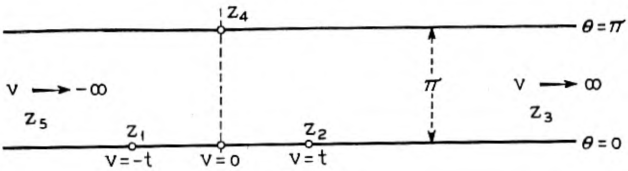


FIG. 2

The first step is to find the conformal transformation

$$z = x + iy = f(v + i\theta) = f(w) \tag{2-2}$$

which carries the bent guide (shown in Fig. 1) in the (x, y) plane over into the straight guide (shown in Fig. 2) in the (v, θ) plane. This may be done by the Schwarz-Christoffel method discussed in Appendix I. This transformation carries the wave equation (1-4) and the boundary condition (1-5) into

$$\frac{\partial^2 Q}{\partial v^2} + \frac{\partial^2 Q}{\partial \theta^2} + [1 + g(v, \theta)]k^2 Q = 0, \tag{2-3}$$

$$\frac{\partial Q}{\partial \theta} = 0 \text{ at } \theta = 0 \text{ and } \theta = \pi \tag{2-4}$$

where the upper and lower guide walls are carried into $\theta = 0$ and $\theta = \pi$, respectively, and $g(v, \theta)$ is given by

$$1 + g(v, \theta) = |f'(v + i\theta)|^2 \pi^2/b^2 \tag{2-5}$$

$$g(v, \theta) = \frac{[ch\, v + \cos\theta]^{2\alpha}}{[ch(v - t) - \cos\theta]^{\alpha}[ch(v + t) - \cos\theta]^{\alpha}} - 1 \tag{2-6}$$

Here ch denotes the hyperbolic cosine, $f'(v + i\theta)$ denotes the first derivative of $f(w)$, and from Appendix I, $2\pi\alpha$ is the total angle of the bend. t is a parameter which depends upon α and the ratio d/d_0 where $d = | z_4 - z_0 |$ and $d_0 = | z_4 - z_6 |$ in Fig. 1. A table giving values of t for a 90° bend ($\alpha = 1/4$) appears in Appendix I.

That the propagation constant is no longer uniform in the transformed guide shows up through the fact that the coefficient of k^2Q in (2-3) is now a function of the coordinates (v, θ). $g(v, \theta)$ measures the deviation of the propagation constant from its value at $v = -\infty$. For example, if we consider a wave front coming down from z_5 we expect it to get past z_4 before it reaches z_0. In Fig. 2 the same wave front is tilted forward corresponding to a high phase-velocity (or small propagation constant) at z_4 where $v = 0$ and $\theta = \pi$. This is in line with the fact that the coefficient of k^2Q in (2-3) vanishes at z_4 by virtue of (2-6). Similar considerations hold at z_1 and z_2.

What is our reflection problem in terms of the transformed guide? In addition to satisfying the two equations (2-3) and (2-4) Q must behave properly at infinity. For large negative values of v, Q must represent an incident wave plus a reflected wave. The incident wave is of unit amplitude and the reflected wave is of the, as yet, unknown value R_E. For large positive values of v Q must represent an outgoing wave. Thus Q must also satisfy the two equations

$$Q = e^{-ikv} + R_E e^{ikv}, \qquad v \to -\infty \tag{2-7}$$

$$Q = T_E e^{-ikv} \qquad , \qquad v \to \infty \tag{2-8}$$

where the subscript E appears on the "reflection coefficient" R_E and the "transmission coefficient" T_E to indicate that here we are dealing with an electric corner.

Our problem is now to take the four equations (2-3, 4, 7, 8) and somehow or other obtain the value of R_E. We are not so much interested in T_E because it does not have the practical importance of the reflection coefficient. There are at least two different ways we may proceed from here. One is to transform the differential equation plus the boundary conditions into an integral equation which may be solved approximately by iteration. Another way is to assume Q to be a Fourier cosine series in θ whose coefficients are functions of v. Substitution of the assumed series in the differential equation (2-3) gives rise to a set of ordinary differential equations having v as the independent variable and the coefficients as the dependent variables. The integral equation method is used in this paper. The second method is discussed in the companion paper.[6]

3. Conversion of Differential Equation into an Integral Equation

The differential equation (2-3) may be converted into an integral equation by using the appropriate Green's function in the conventional manner. The only modifications necessary are essentially those given by Poritsky and Blewett[1] in a similar procedure.

The conversion is based upon Green's theorem in the form

$$\int \left(Q \frac{\partial G}{\partial n} - G \frac{\partial Q}{\partial n} \right) ds = \iint (Q\nabla^2 G - G\nabla^2 Q)\, dv\, d\theta \tag{3-1}$$

where the integration on the right extends over the rectangular region $v_1 < v < v_2$, $0 < \theta < \pi$ (inside the straight guide associated with (v, θ), i.e. the guide of Fig. 2) except for a very small circle surrounding the point (v_0, θ_0). $G \equiv G(v_0, \theta_0; v, \theta)$ is the Green's function corresponding to

$$\frac{\partial^2 V}{\partial v^2} + \frac{\partial^2 V}{\partial \theta^2} + k^2 V = 0 \tag{3-2}$$

in the region $-\infty < v < \infty$, $0 < \theta < \pi$ subject to the boundary condition $\partial V/\partial n = 0$ on the walls ($\partial V/\partial \theta = 0$ at $\theta = 0$ and $\theta = \pi$). G becomes infinite as $-\log r$ when $r \to 0$, r being the distance between the variable point (v, θ) and the fixed point (v_0, θ_0). Poritsky and Blewett* have shown that, in the notation (2-1),

$$G = \sum_{m=0}^{\infty} \epsilon_m \gamma_m^{-1} \cos m\theta_0 \cos m\theta\, \epsilon^{-|v-v_0|\gamma_m} \tag{3-3}$$

$$\epsilon_0 = 1, \ \epsilon_m = 2 \quad \text{for} \quad m = 1, 2, 3 \cdots$$

Equation (3-1) leads to

$$2\pi Q(v_0, \theta_0) + \int_0^{\pi} \left[-Q \frac{\partial G}{\partial v} + G \frac{\partial Q}{\partial v} \right]_{v_1} d\theta + \int_0^{\pi} \left[Q \frac{\partial G}{\partial v} - G \frac{\partial Q}{\partial v} \right]_{v_2} d\theta$$
$$= k^2 \int_{v_1}^{v_2} dv \int_0^{\pi} d\theta\, g(v, \theta) QG \tag{3-4}$$

from which the required integral equation for Q is found to be $Q(v_0, \theta_0) =$

$$e^{-ikv_0} + \frac{k^2}{2\pi} \int_{-\infty}^{\infty} dv \int_0^{\pi} d\theta\, g(v, \theta) Q(v, \theta) \sum_{m=0}^{\infty} \epsilon_m \gamma_m^{-1} \cos m\theta_0 \cos m\theta\, e^{-|v-v_0|\gamma_m} \tag{3-5}$$

* We have replaced their i by $-i$ since here we assume the time to enter through the factor $e^{i\omega t}$ instead of $e^{-i\omega t}$.

where γ_m is given by (2-1). The term e^{-ikv_0} comes from the first integral on the left side of (3-4) as $v_1 \rightarrow -\infty$. Equation (3-5) is a general equation which may be applied to a number of wave guide problems by choosing a suitable function $g(v, \theta)$. For the corner of Fig. 1 $g(v, \theta)$ is given by (2-6).

If $g(v, \theta)$ approaches zero when $|v|$ becomes large, as it does for the corner, expressions for the reflection coefficient R_E and the amplitude T_E of the transmitted wave may be obtained by letting $v_0 \rightarrow \pm\infty$ in (3-5). For very large values of $|v_0|$ the contributions of all the terms in the summation except the first ($m = 0$) vanish. Comparison of the resulting expression for $Q(v_0, \theta_0)$ with the limiting forms (2-7) and (2-8) defining R_E and T_E gives

$$R_E = -\frac{ik}{2\pi}\int_{-\infty}^{\infty} dv \int_0^{\pi} d\theta\, g(v, \theta)\, Q(v, \theta)e^{-ikv} \tag{3-6}$$

$$T_E = 1 - \frac{ik}{2\pi}\int_{-\infty}^{\infty} dv \int_0^{\pi} d\theta\, g(v, \theta)\, Q(v, \theta)e^{ikv} \tag{3-7}$$

Since the integrands involve the as yet unknown $Q(v, \theta)$ these expressions are not immediately applicable. In fact, if we knew $Q(v, \theta)$ it would not be necessary to use these integrals for R_E and T_E—we could simply let $v \rightarrow \pm\infty$ and use (2-7) and (2-8). Nevertheless, (3-6) and (3-7) are useful in obtaining approximations to R_E and T_E when approximations to Q are known.

In Appendix IV it is shown that R_E is the stationary value, with respect to variations of the function Q, of an expression made up of integrals containing Q in their integrands. From the integral equation it follows that when $k \rightarrow 0$, i.e., when the frequency decreases toward the cut-off frequency of the dominant mode, Q becomes approximately $\exp(-ikv)$. Furthermore, R_E approaches zero. This is in contrast to the apparent behavior of R_H which, according to the discussion given in Section 5, may possibly approach -1 under the same circumstances. Thus reflections from the two types of corners, or more generally, irregularities in the E plane and in the H plane, appear to behave quite differently as the cut-off frequency is approached.

R_E and T_E are not independent. Since the energy in the incident wave is equal to the sum of the energies in the reflected and transmitted waves we expect

$$R_E R_E^* + T_E T_E^* = 1, \tag{3-8}$$

where the asterisk denotes the conjugate complex quantity. In addition, there is a relation between R_E and T_E which for a symmetrical irregularity, i.e. for $g(v, \theta)$ an even function of v, states that the phase of R_E differs from that T_E by $\pm\pi/2$. In this special case T_E is determined to within a plus or minus sign when R_E is given. These relations may be proved by substituting various solutions of equation (2-3) for Q and $\hat{Q}$ in the equation

$$\left[Q\frac{\partial\hat{Q}}{\partial v} - \hat{Q}\frac{\partial Q}{\partial v}\right]_{v_1} = \left[Q\frac{\partial\hat{Q}}{\partial v} - \hat{Q}\frac{\partial Q}{\partial v}\right]_{v_2} \tag{3-9}$$

where v_1 and v_2 are large enough (v_1 negative and v_2 positive) to ensure that Q and $\hat{Q}$ have reduced to exponential functions of v. Equation (3-9) follows from Green's theorem. When Q is taken to be the solution for which (2-7) and (2-8) holds and $\hat{Q}$ its conjugate complex Q^*, equation (3-8) is obtained. Keeping the same solution for Q but now letting $\hat{Q}$ denote the solution corresponding to an incident wave of unit amplitude coming in from the right:

$$Q_1 = e^{ikv} + R_1 e^{-ikv}, \qquad v \to \infty$$

$$Q_1 = T_1 e^{ikv} \qquad , \qquad v \to -\infty$$

gives $T = T_1$ where we have dropped the subscript E and have assumed that $g(v, \theta)$ may be unsymmetrical. Taking $\hat{Q}$ to be Q_1^* gives

$$RT_1^* + R_1^*T = 0$$

which is the relation sought. In the symmetrical case $R = R_1$, $R/T + R^*/T^*$ is zero and hence R/T is purely imaginary as was mentioned above. The same relations hold for R_H and T_H. These results are special cases of a more general result which states that the "scattering matrix" is symmetrical and unitary for a lossless junction.[10]

4. *Approximate Solution of Integral Equation*

A first approximation to the solution of the integral equation (3-5) is obtained when we assume that the non-uniformity of the propagation constant has no effect on Q. Thus we put

$$Q^{(1)}(v, \theta) = e^{-ikv} \tag{4-1}$$

in the integral on the right and obtain an expression for the second approximation $Q^{(2)}(v, \theta)$, and so on. Here we shall not go beyond $Q^{(2)}(v, \theta)$.

It is convenient to expand $g(v, \theta)$ in a Fourier cosine series

$$g(v, \theta) = \sum_{n=0}^{\infty} a_n(v) \cos n\theta$$

$$a_n(v) = \frac{\epsilon_n}{\pi}\int_0^{\pi} g(v, \theta) \cos n\theta \, d\theta, \qquad \epsilon_0 = 1; \epsilon_n = 2, n > 0. \tag{4-2}$$

The second approximation, obtained by substituting (4-1) in (3-5), may then be written as

$$Q^{(2)}(v_0, \theta_0) = e^{-ikv_0} + k^2 2^{-1} \sum_{m=0}^{\infty} \gamma_m^{-1} \cos m\theta_0 \cdot \int_{-\infty}^{\infty} a_m(v) e^{-ikv - |v - v_0|\gamma_m} \, dv. \tag{4-3}$$

The nth approximation $R_E^{(n)}$ to the reflection coefficient (when the electric vector lies in the plane of the bend) is defined in terms of $Q^{(n)}$ by

$$\operatorname*{Limit}_{v \to -\infty} \; Q^{(n)}(v, \theta) = e^{-ikv} + R_E^{(n)} e^{ikv} \tag{4-4}$$

$R_E^{(n)}$ is also equal to the integral obtained by replacing Q in (3-6) by $Q^{(n-1)}$. We have

$$R_E^{(1)} = 0, \qquad R_E^{(2)} = -ik2^{-1} \int_{-\infty}^{\infty} a_0(v) e^{-2ikv} \, dv,$$

$$R_E^{(3)} = R_E^{(2)} - ik^3 \sum_{m=0}^{\infty} (4\gamma_m \epsilon_m)^{-1} \cdot \int_{-\infty}^{\infty} dv_0 \, a_m(v_0) \int_{-\infty}^{\infty} dv \, a_m(v) e^{-ik(v+v_0) - |v - v_0|\gamma_m} \tag{4-5}$$

where γ_m is given by (2-1).

The results of this section have the same generality as the integral equation (3-5) in that they are not restricted to corners.

5. *Truncated Corner—Magnetic Vector in Plane of Bend*

When the magnetic vector lies in the plane of the bend the reflection may be calculated by a similar procedure. The wide dimension a of the wave guide now replaces the narrow dimension b in Fig. 1. We shall call the result of making this change the "modified Fig. 1". We again assume the frequency to be such that only the dominant mode is propagated without attenuation. In place of equations (1–3, 4, 5) involving Q we have those of (1–6, 7, 8) involving P.

The conformal transformation which carries the modified Fig. 1 into Fig. 2 leads to

$$\frac{\partial^2 P}{\partial v^2} + \frac{\partial^2 P}{\partial \theta^2} + [1 + g(v, \theta)]\kappa^2 P = 0$$
$$P = 0 \quad \text{at} \quad \theta = 0 \quad \text{and} \quad \theta = \pi \tag{5-1}$$

where

$$\kappa = 2a/\lambda_0 = -i\Gamma_{00}a/\pi, \qquad c = (\kappa^2 - 1)^{1/2} = ak/b$$
$$\delta_m^2 = m^2 - 1 - c^2 = m^2 - \kappa^2, \qquad \delta_1 = ic \tag{5-2}$$

λ_0 = free space wave length, $\qquad m = 1, 2, 3 \cdots$

and

$$|f'_{\text{mod}}(v + i\theta)|^2\pi^2/a^2 = 1 + g(v, \theta). \tag{5-3}$$

Here $f_{\text{mod}}(w)$ pertains to the modified Fig. 1. Since the expression for $f'(w)$ given in Appendix I is proportional to b and since the modified transformation contains a in place of b, it follows that $g(v, \theta)$ for the magnetic corner is exactly the same function, given by (2-6), as for the electric corner.

It is again assumed that the incident wave coming down from the left in the modified Fig. 1 is of unit amplitude and of the dominant mode. At large distances from the corner

$$P = [e^{-icv} + R_H e^{icv}] \sin\theta, \qquad v \to -\infty$$
$$P = T_H e^{-icv} \sin\theta, \qquad v \to +\infty \tag{5-4}$$

which serve to define the coefficients of reflection and transmission. The subscript H on the reflection and transmission coefficients indicate that here we are dealing with a magnetic corner.

The conversion of the differential equation into the integral equation now employs the Green's function

$$G = 2 \sum_{m=1}^{\infty} \delta_m^{-1} \sin m\theta_0 \sin m\theta e^{-|v-v_0|\delta_m} \tag{5-5}$$

which corresponds to

$$\frac{\partial^2 V}{\partial v^2} + \frac{\partial^2 V}{\partial \theta^2} + \kappa^2 V = 0$$
$$V = 0 \quad \text{at} \quad \theta = 0 \quad \text{and } \theta = \pi$$

The integral equation for P is found to be

$$P(v_0, \theta_0) = e^{-icv_0} \sin\theta_0$$
$$+ \frac{\kappa^2}{2\pi} \int_{-\infty}^{+\infty} dv \int_0^{\pi} d\theta \, g(v,\theta) P(v,\theta) \sum_{m=1}^{\infty} 2\delta_m^{-1} \sin m\theta_0 \sin m\theta e^{-|v \quad v_0|\delta_m} \tag{5-6}$$

where the parameters are given by (5-2). This is a general equation. For the corner of the modified Fig. 1 $g(v, \theta)$ is given by (2-6).

By letting $v_0 \to -\infty$ we obtain the exact expression

$$R_H = -\frac{i\kappa^2}{\pi c}\int_{-\infty}^{\infty} dv \int_0^{\pi} d\theta\, g(v, \theta) P(v, \theta) e^{-icv} \sin\theta \tag{5-7}$$

When dealing with the electric corner we saw that $R_E \to 0$ as $k \to 0$. The presence of c in the denominator of (5-7) suggests the possibility that $R_H \to -1$ as $c \to 0$. For R_H must remain finite and this may perhaps come about through $P(v, \theta) \to 0$ in the region, say around $v = 0$, where $g(v, \theta)$ is appreciably different from zero. This and the fact that $P(v, \theta)$ must contain a unit incident wave suggest that for $v < 0$ the dominant portion of $P(v, \theta)$ is $2i \sin cv$ which gives $R_H = -1$. Incidentally, it is apparent that the approximations for $P(v, \theta)$ given below in (5-8) and (5-10) (and therefore also the approximations (5-11) for R_H) fail when c becomes small.

The first approximation to the solution of the integral equation (5-6) is

$$P^{(1)}(v, \theta) = e^{-icv} \sin\theta \tag{5-8}$$

When we introduce the coefficients

$$\begin{aligned} b_n(v) &= \frac{2}{\pi}\int_0^{\pi} g(v, \theta) \sin\theta \sin n\theta\, d\theta \\ \sin\theta g(v, \theta) &= \sum_{n=1}^{\infty} b_n(v) \sin n\theta \end{aligned} \tag{5-9}$$

$$b_1(v) = a_0(v) - a_2(v)/2, \qquad b_n(v) = [a_{n-1}(v) - a_{n+1}(v)]/2, \qquad n > 1$$

we find that the second approximation is

$$P^{(2)}(v_0, \theta_0) = e^{-icv_0} \sin\theta_0 + \kappa^2 2^{-1} \sum_{m=1}^{\infty} \delta_m^{-1} \sin m\theta_0 \cdot \int_{-\infty}^{\infty} b_m(v) e^{-icv - |v - v_0|\delta_m}\, dv. \tag{5-10}$$

The successive approximations to the reflection coefficient are

$$\begin{aligned} R_H^{(1)} &= 0, \qquad R_H^{(2)} = -\frac{i\kappa^2}{2c}\int_{-\infty}^{+\infty} dv\, b_1(v) e^{-2icv} \\ R_H^{(3)} &= R_H^{(2)} - i\kappa^4 \sum_{m=1}^{\infty} (4c\delta_m)^{-1} \int_{-\infty}^{+\infty} dv_0\, b_m(v_0) \cdot \int_{-\infty}^{+\infty} dv\, b_m(v_0) e^{-ic(v+v_0) - |v-v_0|\delta_m}. \end{aligned} \tag{5-11}$$

6. Series for $R^{(2)}$ When Corner Has No Truncation

The integrals which appear in the approximations for the reflection coefficients are difficult to evaluate in general. This section serves to put on record several expressions which have been obtained for $R^{(2)}$ when the corner is not truncated. Corresponding evaluations of $R^{(3)}$ would be welcome since the work of Section 7 for small angle corners indicates that $R^{(3)} - R^{(2)}$ is of the same order as $R^{(2)}$. However, I have been unable to go much beyond the results shown here.

As mentioned in the introduction, H. Levine has studied the effect of a corner in a wave guide by representing it as an equivalent pi network having an inductance for the series element and two equal condensers for the shunt elements. Early in 1947 he derived the following expressions (in our notation) for the elements corresponding to a simpe E corner:*

$$B_a/Y_0 = k\left[\Psi\left(\frac{\beta - 1}{2}\right) - \Psi(-\tfrac{1}{2})\right]$$

$$B_b/Y_0 = (k\pi)^{-1} \cot (\beta\pi/2)$$

where Y_0 is the characteristic admittance of the straight guide, iB_a the admittance of one of the two equal shunt condensers, $-iB_b$ the admittance of the series inductance, $\Psi(x)$ the logarithmic derivative of $\Gamma(x + 1)$, and $\beta\pi$ is the total angle of the simple corner (for no truncation we set $\beta = 2\alpha$).

When the reflection coefficient for the corner is computed from the equivalent network for the case $\beta \to 0$ it is found to lie between the approximate value $R_E^{(2)}$ given by (7-3) and the considerably more accurate value $R_E^{(3)}$ given by (7-5). All three approximations are of the form $A\beta^2 + O(\beta^3)$ where A differs from approximation to approximation but is independent of β, and $O(\beta^3)$ denotes correction terms of order β^3. Since $R_E^{(3)}$ gives the exact value of A, it may be regarded as the standard when the three approximations are compared. If this comparison be taken as a guide, it suggests that the rather cumbersome expressions (6-2) and (6-5) for $R_E^{(2)}$ given below are not as accurate as the simpler expressions resulting from Levine's work. Dr. Levine has also obtained corresponding results for the general E-corner of Fig. 1. It is hoped that his work will be published soon.

When the corner is not truncated it is convenient, as mentioned above, to replace 2α by β so that $\beta\pi$ is the total angle of the bend. For no truncation $t = 0$ and (2-6) becomes

$$g(v, \theta) = \left[\frac{chv + \cos\theta}{chv - \cos\theta}\right]^\beta - 1. \tag{6-1}$$

* I am indebted to Dr. Levine for communicating these expressions to me.

From (4-2) and (4-5), or from (3-6),

$$\begin{aligned} R_E^{(2)} &= -\frac{ik}{2\pi}\int_{-\infty}^{\infty} dv \int_0^{\pi} d\theta g(v,\theta) e^{-2ikv} \\ &= -ik \sum_{n=1}^{\infty} \frac{\Gamma(n-ik)\Gamma(n+ik)}{n!(n-1)!2} \sum_{m=0}^{n} \frac{(-2\beta)_{2m}(\beta)_{n-m}}{(2m)!(n-m)!} \end{aligned} \tag{6-2}$$

where we have expanded $g(v, \theta)$ as given by (6-1) in powers of $\cos\theta/ch\, v$ and integrated termwise. The notation is $(\alpha)_0 = 1$, $(\alpha)_n = \alpha(\alpha+1)\cdots(\alpha+n-1)$.

For a right angle corner $\beta = 1/2$, and a more rapidly convergent series may be obtained by subtracting the sum of the series corresponding to $k = 0$, namely

$$\log 2 = \sum_{n=1}^{\infty} \frac{(1/2)_n}{n!2n} \tag{6-3}$$

Thus for $\beta = 1/2$

$$\begin{aligned} R_E^{(2)} &= -ik\left[\log_e 2 - \sum_{n=1}^{\infty} \frac{(1/2)_n}{n!2n}(1-A_n)\right], \\ A_1 &= \pi k/\sinh \pi k, \qquad A_n = A_1 \prod_{m=1}^{n-1}(1+k^2m^{-2}), \qquad n>1 \end{aligned} \tag{6-4}$$

The rate of convergence of the more general series (6-2) may be increased in a somewhat similar way. It is found that

$$\begin{aligned} R_E^{(2)} &= -\frac{ik}{2}\left[J - 2\beta^2(1-A_1) - \frac{\beta^2}{3}(2+\beta^2)(1-A_2)\right. \\ &\qquad\left. - \frac{2\beta^2}{135}(23+20\beta^2+2\beta^4)(1-A_3) - \cdots\right] \\ J &= K + L \\ K &= \sum_{n=1}^{\infty} \frac{(\beta)_n}{n!n} = \frac{1}{1-\beta} - \Psi(1-\beta) - .5772 \\ L &= \sum_{m=1}^{\infty} \frac{(-2\beta)_{2m}}{(2m)!} \sum_{n=m}^{\infty} \frac{(\beta)_{n-m}}{(n-m)!n} = -\beta \sum_{m=1}^{\infty} \frac{(1/2-\beta)_m}{(1/2)_m m(m-\beta)} \end{aligned} \tag{6-5}$$

where $.5772\cdots$ is Euler's constant, $\Psi(x)$ is the logarithmic derivative of $\Pi(x) = \Gamma(x+1)$, and A_n is given by (6-4).

The results corresponding to $R_H^{(2)}$ are quite similar. When the corner is not truncated

$$R_H^{(2)} = -i\kappa^2 \sum_{n=1}^{\infty} \frac{\Gamma(n - ic)\Gamma(n + ic)}{(n + 1)!(n - 1)!2c} \sum_{m=0}^{n} \frac{(-2\beta)_{2m}(\beta)_{n-m}}{(2m)!(n - m)!}$$

$$= -\frac{i\kappa^2}{2c}\left[\hat{J} - \beta^2(1 - \hat{A}_1) - \frac{\beta^2}{3}(2 + \beta^2)(1 - \hat{A}_2) - \frac{\beta^2}{270}(23 + 20\beta^2 + 2\beta^4)(1 - \hat{A}_3) - \cdots\right] \tag{6-6}$$

$$\hat{J} = 1 - \Psi(1 - \beta) - .5772 - \beta(1 - \beta) \sum_{m=1}^{\infty} \frac{(1/2 - \beta)_m}{(1/2)_m m(m - \beta)(m - \beta + 1)}$$

in which $\hat{A}_n$ is obtained by replacing c by k in the expression (6-4) for A_n.

The evaluation of the integrals for $R_E^{(2)}$ and $R_H^{(2)}$ for general values of t appears to be difficult although it is possible to obtain approximate expressions for the case when t is large.

7. *Reflection from Small Angle Corners*

The expressions for $R^{(2)}$ and $R^{(3)}$ may be evaluated approximately when the angle of the corner is small. It turns out that, for $t = 0$, they are of the same order of magnitude and both of them must be considered. Moreover $R^{(n)}$ for $n > 3$ differs from $R^{(3)}$ by terms of the same order as those neglected in our approximations so that there is no point in going to the higher values of n.

We first obtain the approximation for R_E for a corner with no truncation having the total angle $\pi\beta$. Since β is very small (6-1) may be written as

$$\begin{aligned} g(v, \theta) &= \exp[\beta\varphi] - 1 = \beta\varphi + \beta^2\varphi^2/2! + 0(\beta^3) \\ \varphi &= \log(chv + \cos\theta) - \log(chv - \cos\theta) \end{aligned} \tag{7-1}$$

where $0(\beta^3)$ denotes terms of order β^3. The expression φ becomes very large near the two points (0, 0) and (0, π) (the coordinates being (v, θ)). The following considerations indicate that this does not invalidate our procedure. The remainder, denoted by $0(\beta^3)$, in (7-1) is less than $|\beta\varphi|^3 \exp|\beta\varphi|$. Near (0, 0) φ is approximately equal to $2\log(2/r)$ where $r^2 = v^2 + \theta^2$. Consequently the remainder is less than $(2\beta \log 2/r)^3 (2/r)^{2\beta}$. When the expression (7-1) for $g(v, \theta)$ is set in the integral equation it is seen that all terms, and in particular the remainder term (by virtue of the inequality just stated), of the double integral converge at (0, 0). Hence the contribution of the remainder term is of order β^3 even in the worst case when the

Green's function is replaced by $-\log r$. A similar result holds for the other point in question, namely $(0, \pi)$.

Integrating (7-1) from $\theta = 0$ to $\theta = \pi$ and using equations (A2-1, 3) of Appendix II gives

$$\begin{aligned} a_0(v) &= \beta^2[I_1(v, v) - I_2(v, v)] + 0(\beta^3) \\ &= 4\beta^2 \sum_{n=1,3,5,\cdots}^{\infty} n^{-2} e^{-2nv} + 0(\beta^3) \end{aligned} \tag{7-2}$$

where $v > 0$. We consider only positive values of v since $g(v, \theta)$ and the $a_n(v)$'s are even functions of v. Thus (4-5) yields

$$R_E^{(2)} = -ik2\beta^2 \sum_{n=1,3,5,\cdots} n^{-1}(n^2 + k^2)^{-1} + 0(\beta^3) \tag{7-3}$$

This is an approximation to the exact value given by the double series in (6-2). Comparison of (6-5) and (7-3) when β and k approach zero gives, incidentally,

$$\sum_{m=1}^{\infty} m^{-2} \sum_{n=1}^{m} (n - 1/2)^{-1} = \frac{7}{2} \sum_{m=1}^{\infty} m^{-3}.$$

From (4-2), (7-1) and the expansions (A2-2) of $\log\,(chv \pm \cos\theta)$ it follows that

$$\begin{aligned} a_m(v) &= 4\beta m^{-1}e^{-m|v|} + 0(\beta^2), \qquad m = 1, 3, 5, \cdots \\ a_m(v) &= 0(\beta^2) \qquad\qquad\qquad\quad , \qquad m = 0, 2, 4, 6 \cdots \end{aligned} \tag{7-4}$$

Equations (4-5), (A2-4), the relation $\gamma_m^2 = m^2 - k^2$, and (A2-8) give us the answer we seek:

$$\begin{aligned} R_E^{(3)} &= R_E^{(2)} - ik^3 2\beta^2 \sum_{m=1,3,5,\cdots} \gamma_m^{-1} m^{-2} J(m, m, k, \gamma_m, 0, 0) + 0(\beta^3) \\ &= -ik2\beta^2 \sum_{m=1,3,5,\cdots} \gamma_m^{-1} m^{-2} + 0(\beta^3) \end{aligned} \tag{7-5}$$

It is not necessary to go to $R_E^{(4)}$ because it differs from $R_E^{(3)}$ by only $0(\beta^3)$.

When H lies in the plane of the bend the reflection from a small angle corner with no truncation may be obtained by much the same procedure. For brevity we shall not write down the order of magnitude of the remainder terms. From (5-9), (A2-1), and (A2-3)

$$\begin{aligned} b_1(v) &= a_0(v) - a_2(v)/2 \qquad\qquad (7\text{-}6) \\ &= \beta^2[I_1 - I_2 - (I_3 - I_4)/2] \\ &= \beta^2[2e^{-2v} + 4 \sum_{n=3,5,\cdots} n^{-2}e^{-2nv} - 4 \sum_{n=2,4,\cdots} (n^2 - 1)^{-1} e^{-2nv}] \end{aligned}$$

where we have written I_m for $I_m(v, v)$ and assumed $v > 0$. Then, using (5-11),

$$R_H^{(2)} = -i\kappa^2 c^{-1}\beta^2[\kappa^{-2} + 2 \sum_{n=3,5,\cdots} n^{-1}(n^2 + c^2)^{-1} - 2 \sum_{n=2,4,\cdots} n(n^2 - 1)^{-1}(n^2 + c^2)^{-1}] \tag{7-7}$$

When we put

$$b_n(v) = [a_{n-1}(v) - a_{n+1}(v)]/2, \qquad n > 1 \tag{7-8}$$

$$b_n(v) = 2\beta[(n-1)^{-1}e^{-(n-1)|v|} - (n+1)^{-1}e^{-(n+1)|v|}], \qquad n = 2, 4, 6, \cdots$$

$$b_n(v) = 0(\beta^2), \qquad n = 1, 3, 5, \cdots$$

in (5-11) and use the results of Appendix II we obtain

$$R_H^{(3)} = R_H^{(2)} - i\kappa^4 c^{-1}\beta^2 \sum_{n=2,4,6\cdots} \delta_n^{-1}[(n-1)^{-2} J(n-1, n-1, \delta_n, 0, 0) + (n+1)^{-2} J(n+1, n+1, c, \delta_n, 0, 0) - 2(n^2-1)^{-1} J(n-1, n+1, c, \delta_n, 0, 0)] \tag{7-9}$$

The values of the first two J's, obtained by setting $m = n \pm 1$ in (A2-7), may be simplified by using

$$c^2 + (n \pm 1 + \delta)^2 = 2(n \pm 1)(n + \delta)$$

where we have dropped the subcript n from δ_n. In order to eliminate δ from the denominator we multiply both numerator and denominator by $n - \delta$ and use

$$(n - \delta)(\delta + 2n \pm 2) = (n \pm 1)^2 + c^2 - \delta(n \pm 2)$$

$$n^2 - \delta^2 = 1 + c^2 = \kappa^2$$

Setting in the value, given by (A2-9), of the last J and separating the terms (into those which contain the first power of δ and those which do not) enable us to write the term within the square brackets in (7-9) as

$$\frac{4n^2}{\kappa^4(n^2-1)^2} - \frac{\delta_n}{\kappa^2}\left[\frac{(n-1)-1}{(n-1)^2\{c^2+(n-1)^2\}} + \frac{(n+1)+1}{(n+1)^2\{c^2+(n+1)^2\}} + \frac{2n\{2(n^2+c^2) - \kappa^2(n^2-1)\}}{\kappa^2(n^2-1)^2(n^2+c^2)}\right] \tag{7-10}$$

It is found that when (7-10) is put in (7-9), the contribution of the first two terms within the square bracket of (7-10) exactly cancels the summation which is taken over 3, 5, 7 $\cdots$ in the expression (7-7) for $R_H^{(2)}$. Moreover, if we make use of

$$\sum_{n=2,4,6,\cdots} 4n(n^2 - 1)^{-2} = 1$$

we see that the contribution of the last term within the square brackets of (7-10) cancels the remaining terms in $R_H^{(2)}$. Only the contribution of the first term in (7-10) remains and it gives

$$R_H^{(3)} = -i4\beta^2 c^{-1} \sum_{n=2,4,6,\cdots} n^2(n^2 - 1)^{-2}\delta_n^{-1} + 0(\beta^3) \tag{7-11}$$

The relative simplicity of this result indicates that there may be another method of derivation which avoids the lengthy algebra of our method.

Recently approximate expressions for the reflection coefficient of gentle circular bends have been published[7]. In our present notation these may be written as

$$R_E \approx -ib^2\rho_1^{-2}\left[\frac{\sin u}{24} - 4k\sum_{m=1,3,5,\cdots}\frac{\cos u - e^{-u\gamma_m/k}}{\pi^4 m^4 \gamma_m}\right]$$

$$R_H \approx -ia^2\rho_1^{-2}\left[\frac{\sin u}{8\pi^2 c^2} - 8\sum_{n=2,4,6,\cdots}\frac{\cos u - e^{-u\delta_n/c}}{\pi^4 c\delta_n}\frac{n^2}{(n^2 - 1)^3}\right]$$

where $\beta\pi$ is the angle of the bend, ρ_1 is the radius of curvature of the center line of the guide and u is 2π times the length of the center line in the bend divided by the wavelength in the guide:

$$u = \beta\pi^2 k\rho_1/b = \beta\pi^2 c\rho_1/a$$

The first expression for u is to be used in R_E and the second in R_H. If we now let $\beta \to 0$, keeping ρ_1 fixed, then $u \to 0$. The trigonometric and exponential terms may be approximated by the first few terms in their power series expansions, and part of the series which make their appearance may be replaced by their sums given, for example, by equations (4.1-7) and(4.1-8) of reference[7]. After some cancellation, the above expression for R_E and R_H, which hold for gentle circular bends, reduce to (7-5) and (7-11), respectively, which hold for the sharp corners. In other words, the reflection coefficients for both the sharp and the circular bends approach zero as $\beta \to 0$, and furthermore their ratio approaches unity.

We shall merely outline the derivation of the approximation $R_E^{(3)}$ for a truncated corner. Instead of (7-1) we have from (2-6),

$$g(v, \theta) = \exp[\alpha\varphi] - 1 = \alpha\varphi + \alpha^2\varphi^2/2! + 0(\alpha^3), \tag{7-12}$$

$$\varphi = 2\log[chv + \cos\theta] - \log[ch(v - t) - \cos\theta] - \log[ch(v + t) - \cos\theta]$$

The Fourier coefficients of $g(v, \theta)$ may be obtained by using the results of Appendix II. Assuming $v > 0$, $m > 0$,

$$a_0(v) = 2\alpha(v - t)\psi(v) + 2^{-1}\alpha^2\{4I_1(v, v) + I_1(v - t, v - t)$$
$$+ I_1(v + t, v + t) - 4I_2(v - t, v) - 4I_2(v + t, v) + 2I_1(v - t, v + t)\},$$
$$a_m(v) = 2\alpha m^{-1}[-2(-)^m e^{-m|v|} + e^{-m|v-t|} + e^{-m|v+t|}] \tag{7-13}$$

where $\psi(v) = 1$ when $0 < v < t$ and $\psi(v) = 0$ when $v > t$. Substitution of the values (A2-3) for I_1 and I_2 gives

$$a_0(v) = [2\alpha(v - t) + 2\alpha^2(v - t)^2]\psi(v) \tag{7-14}$$
$$+ \alpha^2 \sum_{n=1}^{\infty} n^{-2}[4e^{-2nv} + e^{-2nv-2nt} - 4(-)^n e^{-2nv-nt}$$
$$+ e^{-2n|v-t|} + 2e^{-n|v-t|-n|v+t|} - 4(-)^n e^{-n|v-t|-nv}]$$

The second approximation to the reflection coefficient is

$$R_E^{(2)} = i\alpha k^{-1}\sin^2 kt - i\alpha^2 k^{-2}2^{-1}(2kt - \sin 2kt)$$
$$- ik\alpha^2 \sum_{n=1}^{\infty} n^{-1}(n^2 + k^2)^{-1}\{2 - (-)^n 2e^{-nt} \tag{7-15}$$
$$+ [1 - 2(-)^n e^{-nt} + e^{-2nt}]\cos 2kt$$
$$+ nk^{-1}[e^{-2nt} - (-)^n 2e^{-nt}]\sin 2kt\}$$

The typical term in the summation (4-5) for $R_E^{(3)}$ is

$$-\frac{ik^3}{4\gamma_m\epsilon_m}\int_{-\infty}^{+\infty} dv_0\, a_m(v_0)\int_{-\infty}^{+\infty} dv\, a_m(v)e^{-ik(v+v_0)-|v-v_0|\gamma_m} \tag{7-16}$$

When $m = 0$, $\epsilon_0 = 1$, $\gamma_0 = ik$, and $a_0(v)$ is $2\alpha(v - t) + 0(\alpha^2)$ for $0 < v < t$ and is $0(\alpha^2)$ for $v > t$. The integral may then be approximated by replacing the upper limit ∞ in (A2-14) by t. The value of (7-16) for $m = 0$ is found to be, to within $0(\alpha^2)$,

$$2^{-1}\alpha^2 t^2(e^{-2ikt} - 1) - (3/4)i\alpha^2 k^{-2}(\sin 2kt - 2kt) \tag{7-17}$$

When $m > 0$, $\epsilon_m = 2$, $\gamma_m^2 = m^2 - k^2$, and the substitution of the value (7-13) for $a_m(v)$ enables us to express (7-16) as the sum of six J's where J is defined by (A2-4). The J's may be evaluated with the help of (A2-7) and (A2-8). Substitution of this value of (7-16) and the value (7-17) for $m = 0$, together with $R_E^{(2)}$ given by (7-15), in the expression (4-5) for $R_E^{(3)}$ gives our final result

$$R_E^{(3)} = i\alpha k^{-1}\sin^2 kt + \alpha^2 t^2 2^{-1}(e^{-2ikt} - 1) \tag{7-18}$$
$$+ i\alpha^2[4^{-1}k^{-2}(2kt - \sin 2kt) - B\sin 2kt + k\sum_{n=1}^{\infty} n^{-2}\gamma_n^{-1}A_n]$$

where

$$B = \sum_{n=1}^{\infty} n^{-2}[e^{-2nt} - 2(-)^n e^{-nt}]$$

$$A_n = \cos 2kt - [2\cos kt - (-)^n e^{-\gamma_n t}]^2$$

Equation (7-18) is an approximation, to within terms of order α^2, for the reflection coefficient of a truncated corner which turns through a small angle $2\pi\alpha$. The electric vector lies in the plane of the bend. When $t = 0$, (7-18) reduces to (7-5) by virtue of $2\alpha = \beta$.

APPENDIX I

Conformal Transformation of Truncated Corner

We shall use a Schwarz-Christoffel transformation* to carry the guide of Fig. 1 into the straight guide of Fig. 2. The first step is to transform the interior of Fig. 1 into the upper half of an auxiliary complex plane which we shall denote by ζ. Let the points z_1, z_2, z_3, z_4, z_5 in Fig. 1 correspond to the points $-h$, h, 1, ∞, -1 in the ζ plane. A suitable transformation is then

$$z = D + E\int_0^{\zeta} (\tau + h)^{-\alpha}(\tau - h)^{-\alpha}(\tau - 1)^{-1}(\tau + 1)^{-1}d\tau \qquad \text{(A1-1)}$$

where D, E and h are to be determined from the geometry of Fig. 1. Because of the symmetry of our transformation about the line joining z_0 and z_4 it follows that $z = z_0$ corresponds to $\zeta = 0$. Hence $D = z_0$. As ζ travels from $1 - \epsilon$ to $1 + \epsilon$, ϵ being very small and positive, along a semicircular indentation above $\zeta = 1$, z as given by (A1-1) increases by

$$E(1 - h^2)^{-\alpha} 2^{-1} \int_{1-\epsilon}^{1+\epsilon} (\tau - 1)^{-1}\, d\tau = \frac{-iE\pi}{2}(1 - h^2)^{-\alpha}$$

while, according to Fig. 1, it increases from $\infty + i0$ to $\infty + ib$. Hence we set the real part of E equal to $-2b\pi^{-1}(1 - h^2)^{\alpha}$. We have tacitly assumed the factors in (A1-1) to have their principal values at $\tau = 1 + \epsilon$ and also that $0 < h < 1$. As z goes from z_1 to z_2, ζ goes from $-h$ to $+h$. In this range $\arg(\tau + h) = 0$ and $\arg(\tau - h) = \pi$.
Consequently, if $|z_2 - z_1| = \ell$, then

$$z_2 - z_1 = \ell e^{-i\alpha\pi} = -Ee^{-i\alpha\pi}\int_{-h}^{h} (h^2 - \tau^2)^{-\alpha}(1 - \tau^2)^{-1}d\tau$$

* See, for example, S. A. Schelkunoff, Electromagnetic Waves, New York (1943) pp. 184–187.

and we see that E is purely real. Hence

$$\ell = 2b\pi^{-1}(1 - h^2)^{\alpha} \int_{-h}^{h} (h^2 - \tau^2)^{-\alpha}(1 - \tau^2)^{-1}\, d\tau$$

is an equation from which h may be determined as a function of ℓ. Setting $\tau^2 = h^2 x$, expanding $(1 - h^2 x)^{-1}$ in powers of h^2 and integrating termwise leads to

$$\begin{aligned} \frac{\ell}{2b} &= \frac{\pi^{-\frac{1}{2}}\Gamma(1 - \alpha)}{\Gamma(\frac{3}{2} - \alpha)} (1 - h^2)^{\alpha} h^{1-2\alpha} F(1, \tfrac{1}{2}; \tfrac{3}{2} - \alpha; h^2) \\ &= \frac{\pi^{-\frac{1}{2}}\Gamma(1 - \alpha)}{\Gamma(\frac{3}{2} - \alpha)} h^{1-2\alpha} F(\tfrac{1}{2} - \alpha, 1 - \alpha; \tfrac{3}{2} - \alpha; h^2) \qquad \text{(A1-2)} \\ &= \frac{1}{\sin \pi\alpha} + \frac{\pi^{-\frac{1}{2}}\Gamma(-\alpha)}{\Gamma(\frac{1}{2} - \alpha)} h^{1-2\alpha}(1 - h^2)^{\alpha} F(1, \tfrac{1}{2}; 1 + \alpha; 1 - h^2) \end{aligned}$$

where we have used relations from the theory of hypergeometric functions. The term $1/\sin \pi\alpha$ is the reduced form of an original term containing a hypergeometric function which has been evaluated by the binomial theorem. The second and third expressions are suited to calculation when $h^2 < 1/2$ and $h^2 > 1/2$ respectively.

Now that the guide of Fig. 1 has been transformed into the upper half of the ζ plane, the next step is to transform this upper half into the straight guide of Fig. 2. We want $\zeta = -1$, i.e. z_5, to go into $v = -\infty$ and $\zeta = 1$, i.e. z_3, to go into $v = +\infty$. Again using the Schwarz-Christoffel formula with $w = v + i\theta$ (the exterior angles at $v = \pm\infty$ are equal to π)

$$w = D_1 + E_1 \int_0^{\zeta} (\tau + 1)^{-1}(\tau - 1)^{-1}\, d\tau \qquad \text{(A1-3)}$$

We take the point z_0 in Fig. 1 to correspond to $v = 0$, $\theta = 0$ in Fig. 2. Since this corresponds to $\zeta = 0$, D_1 must be zero. Also $dw/d\zeta$ is real because w traverses the walls of the guide of Fig. 2 as ζ moves along the real axis in the ζ plane. Hence E_1 is real. As ζ goes from $1 - \epsilon$ to $1 + \epsilon$ around a small circular indentation above $\zeta = 1$, w changes from ∞ to $\infty + i\pi$. Thus

$$i\pi = E_1 2^{-1}(-i\pi) \quad \text{or} \quad E_1 = -2 \qquad \text{(A1-4)}$$

When (A1-3) is integrated, (A1-4) inserted, and the result solved for ζ we obtain

$$\zeta = \tanh w/2 \qquad \text{(A1-5)}$$

The function we require is obtained by differentiating (A1-1) and (A1-3):

$$
\begin{aligned}
f'(v + i\theta) = f'(w) &= \frac{dz}{dw} = \frac{dz}{d\zeta} \Big/ \frac{dw}{d\zeta} \\
&= E(\zeta^2 - h^2)^{-\alpha}(\zeta^2 - 1)^{-1}E_1^{-1}(\zeta^2 - 1) \\
&= (1 - h^2)^{\alpha}(\zeta^2 - h^2)^{-\alpha}b/\pi \\
&= \frac{b}{\pi}\left[\frac{ch^2 w/2}{sh\frac{1}{2}(w - t)\ sh\frac{1}{2}(w + t)}\right]^{\alpha} \\
&= \frac{b}{\pi}\left[\frac{(e^w + 1)^2}{(e^{w-t} - 1)(e^{w+t} - 1)}\right]^{\alpha}
\end{aligned}
\tag{A1-6}
$$

where

$$h = \tanh t/2 \tag{A1-7}$$

For a 90 degree corner $\alpha = 1/4$ and

$$\frac{\ell}{2b} = 2^{1/2}(1 - d/d_0) \tag{A1-8}$$

where, in Fig. 1, $d = | z_4 - z_0 |$ and $d_0 = | z_4 - z_6 |$. In order to obtain the relation between t, defined by (A1-7), and d/d_0 various values of h^2 were picked and the corresponding values of t and d/d_0 (using (A1-2) and (A1-8)) computed. Representative values are given in the following table.

d/d_0	t	d/d_0	t
1.000	0	.5796	1.2302
.9041	.0633	.5385	1.4910
.8565	.1417	.5000	1.7594
.8292	.2007	.4615	2.0634
.7745	.3500	.3727	2.8872
.7196	.5421	.2804	4.0096
.6919	.6549	.1708	5.987
.6273	.9624	.0959	8.294

APPENDIX II

Integrals Associated with Corners of Small Angle

The derivation of the integrals encountered in Sections 7 and 8 will be outlined here. The first ones are

$$I_1(u, v) = \frac{1}{\pi}\int_0^{\pi} \log(ch\ u - \cos\theta)\log(ch\ v - \cos\theta)\,d\theta$$

$$I_2(u, v) = \frac{1}{\pi} \int_0^\pi \log(ch\, u - \cos\theta) \log(ch\, v + \cos\theta)\, d\theta$$

$$I_3(u, v) = \frac{2}{\pi} \int_0^\pi \cos 2\theta \log(ch\, u - \cos\theta) \log(ch\, v - \cos\theta)\, d\theta \tag{A2-1}$$

$$I_4(u, v) = \frac{2}{\pi} \int_0^\pi \cos 2\theta \log(ch\, u - \cos\theta) \log(ch\, v + \cos\theta)\, d\theta$$

Assuming u and v to be positive and using the expansions

$$\log(ch\, u - \cos\theta) = \log(e^u/2) - 2\sum_{n=1}^{\infty} n^{-1}e^{-nu}\cos n\theta$$

$$\log(ch\, u + \cos\theta) = \log(e^u/2) - 2\sum_{n=1}^{\infty} (-)^n n^{-1}e^{-nu}\cos n\theta \tag{A2-2}$$

leads to

$$I_1(u, v) = \log(e^u/2)\log(e^v/2) + 2\sum_{n=1}^{\infty} n^{-2}e^{-nu-nv}$$

$$I_2(u, v) = \log(e^u/2)\log(e^v/2) + 2\sum_{n=1}^{\infty} (-)^n n^{-2}e^{-nu-nv}$$

$$I_3(u, v) = -e^{-2u}\log(e^v/2) - e^{-2v}\log(e^u/2) + 2e^{-u-v} + 2\sum_{n=1}^{\infty} n^{-1}(n+2)^{-1}e^{-nu-nv}(e^{-2u} + e^{-2v}) \tag{A2-3}$$

$$I_4(u, v) = -e^{-2u}\log(e^v/2) - e^{-2v}\log(e^u/2) - 2e^{-v-u} + 2\sum_{n=1}^{\infty} (-)^n n^{-1}(n+2)^{-1}e^{-nu-nv}(e^{-2u} + e^{-2v})$$

When u or v are negative they are to be replaced by their absolute values in the expressions (A2-2, 3).

Now we consider the double integral

$$J(\mu, m, c, \delta; r, s) = \int_{-\infty}^{+\infty} dv_0 \int_{-\infty}^{+\infty} dv \cdot \exp\left[-\mu\,|\,v_0 - r\,| - m\,|\,v - s\,| - ic(v + v_0) - \delta\,|\,v - v_0\,|\right] \tag{A2-4}$$

in which μ, m, c, δ are real and positive and r and s are real. The double integral may be reduced to a single integral by substituting

$$e^{-\delta|v-v_0|} = \frac{\delta}{\pi}\int_{-\infty}^{+\infty} (\delta^2 + x^2)^{-1} e^{ix(v-v_0)}\, dx, \tag{A2-5}$$

interchanging the order of integration, and integrating with respect to v and v_0. Assuming $s - r \geq 0$, the integral is then evaluated by closing the path of integration by an infinite semicircle in the upper half plane and calculating the residues of the integrand at the poles $i\delta$, $c + im$, $-c + i\mu$:

$$J(\mu, m, c, \delta; r, s) = \int_{-\infty}^{\infty} \frac{4\delta\mu m e^{ix(s-r)-ic(r+s)}\,dx}{\pi(\delta^2 + x^2)[\mu^2 + (x+c)^2][m^2 + (x-c)^2]}$$
$$= 4\delta\mu m \left[\frac{e^{-(\delta+ic)s+(\delta-ic)r}}{\delta[\mu^2 + (c+i\delta)^2][m^2 + (c-i\delta)^2]} + \frac{e^{-ms+(m-2ic)r}}{m[\delta^2 + (c+im)^2][\mu^2 + (2c+im)^2]} + \frac{e^{\mu r-(\mu+2ic)s}}{\mu[\delta^2 + (c-i\mu)^2][m^2 + (2c-i\mu)^2]}\right] \quad \text{(A2-6)}$$

Substituting special values for the parameters gives the results required in the text. Thus,

$$\begin{aligned} J(m, m, k, \gamma; t, t) &= e^{-2ikt}\, J(m, m, k, \gamma; 0, 0) \\ J(m, m, k, \gamma; -t, t) &= e^{2ikt}\, J(m, m, k, \gamma; 0, 2t) \\ J(m, m, k, \gamma; -t, 0) &= e^{2ikt}\, J(m, m, k, \gamma; 0, t) \\ J(m, m, c, \delta; 0, 0) &= \frac{2m(\delta + 2m)}{(c^2 + m^2)[c^2 + (m+\delta)^2]} \end{aligned} \quad \text{(A2-7)}$$

which hold irrespective of any relations between the parameters. The derivation of the last result is simplified by setting $\alpha = c + im$, $\bar{\alpha} = c - im$ and factoring the denominators in (A2-6) so as to obtain terms of the form $\bar{\alpha} \pm i\delta$, $\alpha \pm i\delta$.

When $\gamma^2 = m^2 - k^2$ considerable simplification is possible and we obtain

$$\begin{aligned} J(m, m, k, \gamma; 0, 0) &= \frac{\gamma}{k^2}\left[\frac{1}{\gamma} - \frac{m}{m^2 + k^2}\right] \\ J(m, m, k; \gamma, 0, t) &= \frac{\gamma e^{-ikt}}{k^2}\left[\frac{e^{-\gamma t}}{\gamma} - \frac{e^{-mt}(m \cos kt - k \sin kt)}{m^2 + k^2}\right] \end{aligned} \quad \text{(A2-8)}$$

If we put $\mu = n - 1$, $m = n + 1$, and set $\delta^2 = n^2 - 1 - c^2 = n^2 - \kappa^2$ where $\kappa^2 = 1 + c^2$, (A2-6) yields, after some reduction,

$$\begin{aligned} J(n-1, n+1, c, \delta; 0, 0) &= \frac{n^2 - 1}{(cn + i\delta)^2} + \frac{(n-1)\delta}{2(n+1)(1-ic)^2(n-ic)} \\ &\quad + \frac{(n+1)\delta}{2(n-1)(1-ic)^2(n+ic)} \\ &= \frac{c^2n^2 - \delta^2}{\kappa^4(n^2-1)} + \frac{n\delta[2(n^2 + c^2) - \kappa^2(n^2-1)]}{\kappa^4(n^2-1)(n^2+c^2)} \end{aligned} \quad \text{(A2-9)}$$

The form of the final expression has been chosen so as to be suited to the use we shall make of it.

Another double integral which appears in our work is

$$I(k, \gamma) = \int_{-\infty}^{+\infty} dv_0\, a(v_0) \cdot \int_{-\infty}^{\infty} dv\, a(v) \exp\left[-ik(v + v_0) - \gamma \mid v - v_0 \mid\right] \tag{A2-10}$$

where $a(v)$ is an even function of v and is such that all of the integrals encountered converge. We begin our transformation by dividing the interval of integration $(-\infty, \infty)$ for v_0 into $(-\infty, 0)$ and $(0, \infty)$. Making the change of variable $v_0 = -v_0'$, $v = -v'$ in the first interval, dropping the primes and using $a(-v) = a(v)$ leads to

$$I(k, \gamma) = 2 \int_0^{\infty} dv_0\, a(v_0) \int_{-\infty}^{\infty} dv\, a(v) e^{-\gamma|v-v_0|} \cos k(v + v_0) \tag{A2-11}$$

We now split the interval of integration of v in (A2-11) into the intervals $(-\infty, 0)$, $(0, v_0)$, (v_0, ∞). In $(-\infty, 0)$ we change the variable from v to $-v'$, drop the prime, and use $a(-v) = a(v)$. By paying attention to the sign of $v - v_0$ we may remove the absolute value sign. By changing the order of integration in the double integral arising from the third interval (in which $0 \leq v_0 \leq \infty$, $v_0 \leq v \leq \infty$) we may show that it is equal to the double integral arising from the second interval. Thus

$$I(k, \gamma) = 2 \int_0^{\infty} dv_0\, a(v_0) \int_0^{\infty} dv\, a(v) e^{-\gamma v - \gamma v_0} \cos k(v_0 - v) + 4 \int_0^{\infty} dv_0\, a(v_0) \int_0^{v_0} dv\, a(v) e^{-\gamma v_0 + \gamma v} \cos k(v_0 + v) \tag{A2-12}$$

When $a(v)$, γ and k are real we may write (A2-12) as

$$I(k, \gamma) = 2 \left| \int_0^{\infty} dv\, a(v) e^{-\gamma v - ikv} \right|^2 + 4 \text{ Real} \int_0^{\infty} dv_0\, a(v_0) e^{-\gamma v_0 + ikv_0} \int_0^{v_0} dv\, a(v)\, e^{\gamma v + ikv} \tag{A2-13}$$

and when $\gamma = ik$ we have

$$I(k, ik) = 2 \int_0^{\infty} dv_0\, a(v_0) \int_0^{\infty} dv\, a(v) e^{-2ikv} + 2 \int_0^{\infty} dv_0\, a(v_0) \int_0^{v_0} dv\, a(v)[e^{2ikv} + e^{-2ikv_0}]. \tag{A2-14}$$

APPENDIX III

Integral Equation When Guides Entering and Leaving Irregularity Are of Different Sizes

Here we shall indicate how the integral equation method may be extended to cover the case mentioned in the above title. It is supposed that only the dominant mode is propagated freely in both guides.

E in Plane of Irregularity

Let the notation for the guide carrying the incident wave be the same as for the E-corner. b denotes the narrow dimension of the guide and the quantities k and γ_m are given by (2-1). Both guides have the same wide

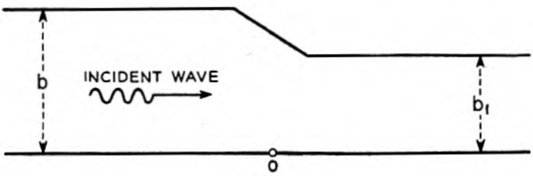


Fig. 3

dimension a. The narrow dimension of the guide shown on the right of Fig. 3 is b_1. We introduce the new quantity

$$k_1 = [(2b_1/\lambda_0)^2 - (b_1/a)^2]^{1/2} \tag{A3-1}$$

to correspond to k. Since, by assumption, only the dominant mode is freely propagated in both guides both k and k_1 are real positive quantities less than unity.

Let $z = f(w)$ carry the system of Fig. 3 into a straight guide of width π in the $w = v + i\theta$ plane (see Fig. 2), and let $g(v, \theta)$ be defined by

$$1 + g(v, \theta) = |f'(w)|^2.$$

The behavior of $g(v, \theta)$ at infinity is shown by the table

v	dz/dw	$g(v, \theta)$
$-\infty$	b/π	0
$+\infty$	b_1/π	$k_1^2k^{-2} - 1$

where $b_1/b = k_1/k$ has been used. It is convenient to introduce the approximation $\hat{g}(v)$ to $g(v, \theta)$. $\hat{g}(v)$ may be chosen at our convenience subject only to the conditions that it be differentiable, $\hat{g}(-\infty) = 0$, and $\hat{g}(\infty) = k_1^2k^{-2} - 1$.

When we define G by equation (3-3) so that, as before, it is the Green's function corresponding to a guide of width b, we may use equation (3-4) to derive the new integral equation

$$Q(v_0, \theta_0) = e^{-ikv_0} + \frac{k^2}{2\pi} \int_{-\infty}^{\infty} dv \int_0^{\pi} d\theta \cdot \{g(v, \theta)Q(v, \theta) - \hat{g}(v)T_E e^{-ik_1 v}\}G(v_0, \theta_0; v, \theta) + T_E F(v_0) \quad \text{(A3-2)}$$

in which

$$\begin{aligned} F(v_0) &= e^{-ik_1 v_0} \hat{g}(v_0)/\hat{g}(\infty) - e^{-ikv_0} N^-(v_0) - e^{ikv_0} N^+(v_0) \\ N^-(v_0) &= 2^{-1} k(k_1 - k)^{-1} \int_{-\infty}^{v_0} \hat{g}'(v) e^{-i(k_1 - k)v} \, dv \\ N^+(v_0) &= 2^{-1} k(k_1 + k)^{-1} \int_{v_0}^{\infty} \hat{g}'(v) e^{-i(k_1 + k)v} \, dv \end{aligned} \quad \text{(A3-3)}$$

Here $\hat{g}'(v)$ denotes $d\hat{g}(v)/dv$. Equation (A3-2) and

$$\operatorname*{Limit}_{v \to \infty} Q(v, \theta) = T_E e^{-ik_1 v} \quad \text{(A3-4)}$$

are to be solved for the unknown function $Q(v, \theta)$ and the unknown quantity T_E. The method of successive approximations may be used in somewhat the same fashion as in the simpler case but we shall not give a general discussion.

The first approximations are found to be

$$T_E^{(1)} = 1/N^-(\infty), \qquad R_E^{(1)} = -N^+(-\infty)/N^-(\infty) \quad \text{(A3-5)}$$

where the N's may be obtained by setting $v_0 = \pm\infty$ in equations (A3-3).

One of the simplest choices for $\hat{g}(v)$ is to let it be zero for negative values of v and to have the value $\hat{g}(\infty) = k_1^2 k^{-2} - 1$ for positive values of v. Then

$$T_E^{(1)} = 2k(k_1 + k)^{-1}, \qquad R_E^{(1)} = (k - k_1)(k + k_1)^{-1} \quad \text{(A3-6)}$$

These are quite similar to the corresponding expressions for a transmission line which have been used extensively in wave guide work.

In working with these formulas, when k is small, it is sometimes convenient to use the result

$$\int_{v_1}^{v_2} dv \int_0^{\pi} d\theta \, g(v, \theta) = \pi^2 b^{-2} \int_{v_1}^{v_2} dv \int_0^{\pi} d\theta \, |f'(w)|^2 - (v_2 - v_1)\pi \quad \text{(A3-7)}$$

where the evaluation of the double integral on the right is made easier by the fact that it represents the area in the original guide (in the (x, y) plane) enclosed by the lines corresponding to $v = v_1$ and $v = v_2$. v_2 and v_1 are chosen to be moderately large positive and negative numbers, respectively. It turns out that, when k_1 and k are very small, this is related to the "excess capacity" localized at the irregularity whose effect must be added to that of the mismatch, indicated by (A3-6).

When the entering and leaving guides are of the same size it is still possible to use the formulas of this appendix. $N^{-}(v_0)$ may be replaced by an expression which now has for its limiting value

$$N^{-}(\infty) = 1 + i(k/2)\int_{-\infty}^{\infty} \hat{g}(v)\, dv \tag{A3-8}$$

H in Plane of Irregularity

Let the figure corresponding to the irregularity be Fig. 3 with b and b_1 replaced by a and a_1, respectively. In addition to the quantities c and κ defined by equations (5-2) we define

$$\kappa_1 = 2a_1/\lambda_0\,, \qquad c_1 = (\kappa_1^2 - 1)^{1/2} \tag{A3-9}$$

where we assume κ and κ_1 to lie between 1 and 2. At $v = -\infty$ $P(v, \theta)$ still consists of the unit incident wave plus the reflected wave given by the first of equations (5-4) and $g(v, \theta)$ is still zero. However, now, at $v = \infty$,

$$\begin{aligned} P(v, \theta) &= T_H e^{-ic_1 v} \sin\theta \\ \hat{g}(\infty) &= \kappa_1^2\kappa^{-2} - 1 = \kappa^{-2}(c_1^2 - c^2) \end{aligned} \tag{A3-10}$$

The integral equation for $P(v, \theta)$ and T_H is

$$\begin{aligned} P(v_0, \theta_0) = e^{-icv_0}\sin\theta_0 + \frac{\kappa^2}{2\pi}\int_{-\infty}^{\infty} dv \int_0^{\pi} \\ \cdot d\theta\{g(v, \theta)P(v, \theta) - \hat{g}(v)T_H e^{-ic_1 v}\sin\theta\}G(v_0, \theta_0; v, \theta) \\ + T_H \sin\theta_0 F_H(v_0) \end{aligned} \tag{A3-11}$$

in which

$$\begin{aligned} F_H(v_0) &= e^{-ic_1 v_0}\hat{g}(v_0)/\hat{g}(\infty) - e^{-icv_0}M^{-}(v_0) - e^{icv_0}M^{+}(v_0) \\ M^{-}(v_0) &= \kappa^2(2c)^{-1}(c_1 - c)^{-1}\int_{-\infty}^{v_0} \hat{g}'(v)e^{-i(c_1-c)v}\, dv \\ M^{+}(v_0) &= \kappa^2(2c)^{-1}(c + c_1)^{-1}\int_{v_0}^{\infty} \hat{g}'(v)e^{-i(c_1+c)v}\, dv \end{aligned} \tag{A3-12}$$

First approximations are

$$T_H^{(1)} = 1/M^{-}(\infty), \qquad R_H^{(1)} = -M^{+}(-\infty)/M^{-}(\infty) \tag{A3-13}$$

which, when we choose $\mathfrak{g}(v)$ to be zero for $v < 0$ and $\kappa_1^2\kappa^{-2} - 1$ for $v > 0$, become

$$T_H^{(1)} = 2c(c_1 + c)^{-1}, \qquad R_H^{(1)} = (c - c_1)(c_1 + c)^{-1} \qquad \text{(A3-14)}$$

which again agrees with results obtained from transmission line considerations. When the entering and leaving guides are the same size we may use

$$M^-(\infty) = 1 + i\kappa^2(2c)^{-1} \int_{-\infty}^{\infty} \mathfrak{g}(v)\, dv \qquad \text{(A3-15)}$$

It seems difficult to give any general rules for the choice of $\mathfrak{g}(v)$. Since for R_H and T_H, the factor $\sin\theta$ reduces the effect of the singularities on the walls of the transformed guide, the choice $\mathfrak{g}(v) = g(v, \pi/2)$ suggests itself. The factor $\sin\theta$ is not present in the formulas for R_E and T_E and regions near the walls are more important. In this case the selection

$$\mathfrak{g}(v) = \pi^{-1} \int_0^{\pi} g(v, \theta)\, d\theta$$

may be useful, especially since it allows us to use the result (A3-7) when k and k_1 become small.

APPENDIX IV

Variational Expressions for Reflection Coefficients

The reflection coefficients are proportional to the stationary values of certain forms associated with the integral equations. In order to obtain these forms we proceed as follows. It is readily seen that the values of x_1 and x_2 which satisfy the symmetrical set of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{12}x_1 + a_{22}x_2 &= b_2 \end{aligned} \qquad \text{(A4-1)}$$

are the ones which make

$$J = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 - 2b_1x_1 - 2b_2x_2 \qquad \text{(A4-2)}$$

stationary when x_1 and x_2 are given small arbitrary increments. This stationary value of J is

$$J_s = -b_1x_1 - b_2x_2$$

If we take the integral equation to be the analogue of the set of linear equations, the reflection coefficient turns out to be proportional to J_s. In order to set down the actual expressions it is convenient to write r for

(v, θ) and dS for the element of area $dv d\theta$ so that the integral equation (3-5) for $Q(v, \theta)$ may be written as

$$Q(r_0) = e^{-ikv_0} + k^2(2\pi)^{-1} \int g(r)Q(r)G(r_0, r)\, dS \qquad \text{(A4-3)}$$

where the integration extends over the interior of the guide and $G(r_0, r)$ denotes the Green's function (3-3).

If the number of equations in the set (A4-1) were increased from two to a large number N, the set of x's would correspond, say, to the values of $Q(r)$ or of $g(r)Q(r)$, and the b's would correspond to the values of $\exp(-ikv_0)$. In any event, we take the analogue of J to be

$$\begin{aligned} J_E = &\int g(r)Q(r)[Q(r) - 2e^{-ikv}]\, dS \\ &- k^2(2\pi)^{-1} \int\int g(r)Q(r)g(r_0)Q(r_0)G(r_0, r)\, dS_0\, dS \end{aligned} \qquad \text{(A4-4)}$$

where the subscript E indicates that we are dealing with an electric corner. It may be verified,* by giving $Q(r)$ a small variation $\delta Q(r)$, that the function $Q(r)$ which makes J_E stationary is the one which satisfies the integral equation (A4-3). Furthermore, when we assume $Q(r)$ to satisfy the integral equation, the expression for J_E reduces to an integral which is proportional to the integral (3-6) for the reflection coefficient R_E. More precisely, R_E is given by

$$R_E = \frac{ik}{2\pi} \text{ [Stationary value of } J_E] \qquad \text{(A4-5)}$$

It follows that if, by some means, we have obtained a fairly good approximation to Q, we may obtain a better approximation to R_E by computing J_E and using the formula

$$R_E \approx ik(2\pi)^{-1}J_E$$

When we use the first approximation $\exp(-ikv)$ for Q to compute J_E it turns out that the above formula gives the third approximation, $R_E^{(3)}$, to the reflection coefficient.

The magnetic corner may be treated in much the same way. The integral equation (5-6) for $P(v, \theta)$ becomes, in the notation of this appendix,

$$P(r_0) = e^{-icv_0} \sin \theta_0 + \kappa^2(2\pi)^{-1} \int g(r)P(r)G(r_0, r)\, dS \qquad \text{(A4-6)}$$

in which the v in $dS = dv d\theta$ is integrated from $-\infty$ to $+\infty$ and θ from 0 to π,

* See Courant and Hilbert, Methoden der Mathematischen Physik, Julius Springer, Berlin (1931), page 176, where a similar problem is treated.

as before, and $G(r_0, r)$ now denotes the Green's function (5-5). We define J_H by

$$J_H = \int g(r)P(r)[P(r) - 2e^{-icv} \sin \theta]dS$$
$$-\kappa^2(2\pi)^{-1} \iint g(r)P(r)g(r_0)P(r_0)G(r_0, r)\, dS_0\, dS. \tag{A4-7}$$

J_H is stationary with respect to small variations in $P(r)$ when $P(r)$ satisfies the integral equation (A4-6). Furthermore, from the integral (5-7) for R_H,

$$R_H = i\kappa^2(\pi c)^{-1} \text{ [Stationary value of } J_H] \tag{A4-8}$$

which may be used in the same way as equation (A4-5) for R_E.

J. Schwinger has used variational methods with considerable success to deal with obstacles in wave guides.* However, his variational equations differ somewhat from those given here. Some light on the relation between Schwinger's equations and the present one may be obtained by returning to the simple algebraic equations (A4-1) and (A4-2). A rough analogue of the expression required to be stationary in Schwinger's theory is

$$(a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2)/(b_1x_1 + b_2x_2)^2 \tag{A4-9}$$

The essential point here is that the stationary value of the expression corresponding to (A4-9) gives the value of an impedance or combination of impedances appearing in some equivalent circuit. Expression (A4-9) may be obtained by expressing J, defined by (A4-2), as a function of x_1 and $y = x_2/x_1$. J is still to be made stationary but now it is a function of x_1 and y. Solving $\partial J/\partial x_1 = 0$ for x_1 and setting this value of x_1 in J gives the following function of y

$$-(b_1 + b_2y)^2\,(a_{11} + 2a_{12}y + a_{22}y^2)^{-1},$$

which is the stationary value of J with respect to variations in x_1 when y is held constant. This function is still required to be stationary with respect to y. The same is true of its reciprocal which becomes (A4-9) when both numerator and denominator are multiplied by x_1^2 and the definition of y used. When (A4-1) is replaced by a larger number of equations similar considerations lead to a generalized form of (A4-9). The expression required to be stationary by Schwinger is obtained when the sums in the generalized form are replaced by integrals.

* An account of the method together with applications is given in "Notes on Lectures by Julian Schwinger: Discontinuities in Waveguides" by David S. Saxon. An account is also given by John W. Miles.[11]

References

1. Poritsky and Blewett, A Method of Solution of Field Problems by Means of Overlapping Regions, *Quart. Jl. Appl. Math.*, *3*, 336–347 (1946).
2. J. W. Miles, The Equivalent Circuit of a Corner Bend in a Rectangular Wave Guide, *Proc. I. R. E.*, *35*, 1313–1317 (1947).
3. N. Elson, Rectangular Wave Guide Systems, Bends, Twists and Junctions, *Wireless Engineer*, *24*, 44–54 (1947).
4. E. J. Routh, Advanced Rigid Dynamics, 6th edition, pp. 461–467, London (1905).
5. P. Krasnooshkin, Acoustic and Electromagnetic Wave Guides of Complicated Shape, *Jl. of Physics* (*Acad. Sci. U.S.S.R.*), 10, 434–445 (1946).
6. S. O. Rice, A Set of Second-Order Differential Equations Associated with Reflections in Rectangular Wave Guides—Application to Guide Connected to Horn, this issue of the B. S. T. J.
7. S. O. Rice, Reflections from Circular Bends in Rectangular Wave Guides—Matrix Theory, *B. S. T. J.*, *27*, 305–349 (1948). In addition to the earlier work of R. E. Marshak referred to there, the interesting, but as yet unpublished, work of H. Levine should be mentioned. He uses variational methods to obtain the values of the impedances appearing in the equivalent network for the bend.
8. S. A. Schelkunoff, On Waves in Bent Pipes, *Quart., Jl. Appl. Math.*, *2*, 171–172 (1944).
9. H. Buchholz, Der Einfluss der Krummung von rechteckigen Hohlleitern auf das Phasenmass ultrakurzer Wellen, *E.N.T.*, *16*, 73–85 (1939).
10. Montgomery, Dicke, and Purcell, Principles of Microwave Circuits, McGraw-Hill (1948), page 149.
11. J. W. Miles, The Equivalent Circuit for a Plane Discontinuity in a Cylindrical Wave Guide, *Proc. I.R.E.*, *34*, 728–742 (1946).

A Set of Second-Order Differential Equations Associated with Reflections in Rectangular Wave Guides—Application to Guide Connected to Horn*

By S. O. RICE

In dealing with corners and similar irregularities in rectangular wave guides it is sometimes helpful to transform the system, conformally, into a straight guide. Propagation in the straight guide may then be studied by an integral equation method, as is done in a companion paper, or by a more general method based upon a certain set of ordinary differential equations. Here the second method is developed and applied to determine the reflection produced at the junction of a straight guide and a sectoral horn—a problem the first method is unable to handle. The *WKB* approximation for a single second-order differential equation is extended to a set of equations and approximate expressions for the reflection coefficient are derived.

IN A companion paper[1] the disturbance produced by a corner in a rectangular wave guide is examined by transforming the system, conformally, into a straight guide. Although the medium in the straight guide is no longer uniform, an integral equation may be set up and approximate solutions obtained.

In that paper the wave guide is assumed to have the same cross-section at $+\infty$ as at $-\infty$. When this is not so, a conformal transformation may still be used to transform the system into a straight guide provided one dimension of the original cross-section is constant. However, now some advantage appears to be gained by replacing the integral equation by a set of differential equations. Since two cases appear, corresponding to E and H corners, there are two sets of equations to be considered.

These two sets of equations are studied in the present paper. After their derivation in Sections 1 and 2 several remarks are made in Section 3 concerning their solution, special emphasis being laid on the problem of determining the reflection coefficient. In the remainder of the paper the general theory is applied to a system formed by joining a rectangular wave guide to a horn (with plane sides) flared in one direction. The reflection coefficients for sectoral horns flared in the planes of the electric and magnetic intensity, respectively, are given approximately by equations (6-1) and (7-1). These approximations assume the angle of flare to be small so that, as it turns out, only the first equations of the respective sets need be considered.

As was mentioned in the companion paper, Robert Piloty has recently made use of conformal transformations in wave guide problems. In his

* Presented at the Second Symposium on Applied Mathematics, Cambridge, Mass., July 29, 1948.

[1]See list of references at end of paper.

method the propagation function $g(v, \theta)$ is derived graphically from the geometry of the wave guide irregularities and the result used in one or the other of two sets of differential equations which are equivalent to those derived below. Piloty's work is scheduled to appear soon in the *Zeitschrift für angewandte Physik* under the title "Ausbreitung el.-magn. Wellen in inhomogenen Rechteckrohren."

1. *Differential Equations when Electric Vector is in (x, y) Plane*

The partial differential equation to be solved is, from equation (2-3) of the companion paper[1],

$$\frac{\partial^2 Q}{\partial v^2} + \frac{\partial^2 Q}{\partial \theta^2} + [1 + g(v, \theta)]k^2 Q = 0 \tag{1-1}$$

where

$$\frac{\partial Q}{\partial \theta} = 0 \text{ at } \theta = 0 \text{ and } \theta = \pi$$

$$1 + g(v, \theta) = 1 + \sum_{n=0}^{\infty} a_n \cos n\theta = |f'(v + i\theta)|^2 \pi^2/b^2 \tag{1-2}$$

$$k = [(2b/\lambda_0)^2 - (b/a)^2]^{\frac{1}{2}}, \qquad \lambda_0 = \text{free space wavelength}$$

In (1-2), $z = x + iy = f(v + i\theta)$ is the transformation which carries the wave guide system in the (x, y) plane into the straight guide of width $\theta = \pi$ in the (v, θ) plane. For the sake of simplicity we shall always assume that far to the left the system becomes a straight wave guide of dimensions a, b $(b < a)$ such that only the dominant mode is propagated without attenuation. This insures that the a_n's (which are functions of v) will approach zero as $v \to -\infty$. The dimension (of our system) normal to the (x, y) plane is a throughout.

Since the normal derivative of Q vanishes on the walls at $\theta = 0$ and $\theta = \pi$ we assume

$$Q = F_0 + F_1 \cos \theta + F_2 \cos 2\theta + \cdots, \tag{1-3}$$

where F_1, F_2, $\cdots$ are functions of v, and substitute it together with the Fourier series (1-2) for $1 + g(v, \theta)$ in (1-1).

The equations obtained by setting the coefficients of the resulting cosine series to zero are

$$F_0'' + (1 + a_0)k^2 F_0 + \frac{k^2}{2} \sum_{n=1}^{\infty} a_n F_n = 0 \tag{1-4}$$

$$F_m'' + [(1 + a_0 + a_{2m}/2)k^2 - m^2]F_m + a_m k^2 F_0 + \frac{k^2}{2} \sum_{n=1}^{\infty}{}' (a_{|n-m|} + a_{n+m})F_n = 0 \tag{1-5}$$

where $m = 1, 2, 3, \cdots$, $F''_m = d^2F_m/dv^2$, and the prime on Σ indicates that the term $n = m$ is to be omitted. In grouping the terms we have assumed that F_0 is the major part of Q.

The principal problem is to solve equations (1-4) and (1-5) when the fundamental mode F_0 is of the form

$$\begin{aligned} F_0 &= e^{-ikv} + R_E e^{ikv}, \qquad v \to -\infty \\ F_0 &= T_E(v), \qquad v \to +\infty \end{aligned} \tag{1-6}$$

in which R_E is a constant and $T_E(v)$ represents a wave traveling towards $v = \infty$. At $v = \pm\infty$ $F_1, F_2, \cdots$ have the form of waves traveling (or being attenuated) away from the region around $v = 0$. As before, we shall be mainly interested in determining the reflection coefficient R.

It is assumed that only the dominant mode is propagated without attenuation in the straight wave guide far to the left and hence $F_1, F_2, \cdots$ all become zero as $v \to -\infty$.

2. *Differential Equations when Magnetic Vector is in (x, y) Plane*

The partial differential equation is now given by equation (5-1) of the companion paper[1]

$$\frac{\partial^2 P}{\partial v^2} + \frac{\partial^2 P}{\partial \theta^2} + [1 + g(v, \theta)]\kappa^2 P = 0 \tag{2-1}$$

where the dimension of the system normal to the (x, y) plane is now b, a is the dimension (in the (x, y) plane) of the straight guide at the far left and

$$P = 0 \text{ at } \theta = 0 \text{ and } \theta = \pi$$

$$1 + g(v, \theta) = 1 + \sum_{n=1}^{\infty} a_n \cos n\theta \tag{2-2}$$

$$\kappa = 2a/\lambda_0, \ \lambda_0 = \text{free space wavelength}$$

$$c = (\kappa^2 - 1)^{1/2}$$

Since $P = 0$ at $\theta = 0$ and $\theta = \pi$ we assume

$$P = \sum_{n=1}^{\infty} F_n \sin n\theta \tag{2-3}$$

where the F's are functions of v to be determined by the equations

$$F''_1 + [\kappa^2(1 + a_0 - a_2/2) - 1]F_1 + \frac{\kappa^2}{2} \sum_{n=2}^{\infty} (a_{n-1} - a_{n+1})F_n = 0 \tag{2-4}$$

$$\begin{aligned} F''_m + [\kappa^2(1 + a_0 - a_{2m}/2) - m^2]F_m + \frac{\kappa^2}{2} (a_{m-1} - a_{m+1})F_1 \\ + \frac{\kappa^2}{2} \sum_{n=2}^{\infty}{}' (a_{|m-n|} - a_{m+n})F_n = 0 \end{aligned} \tag{2-5}$$

in which $m = 2, 3, 4, \cdots$ and the primes on F_m and $\sum$ have the same significance as in (1-4) and (1-5).

The principal problem here is to solve equations (2-4) and (2-5) simultaneously subject to

$$\begin{aligned} F_1 &= e^{-icv} + R_H e^{icv}, & v &\to -\infty \\ F_1 &= T_H(v), & v &\to +\infty \end{aligned} \tag{2-6}$$

which again corresponds to a unit wave in the dominant mode incident from the left. $T_H(v)$ and the remaining F's correspond to outward traveling waves as before. $F_2, F_3, \cdots$ all approach zero as $v \to -\infty$.

3. *Remarks on Solving the Equations of Sections 1 and 2 for the Reflection Coefficient*

Suppose that we have a system in which the wave propagation is governed by the single differential equation

$$\frac{d^2 y}{dv^2} - h^2 y = 0 \tag{3-1}$$

where $h \equiv h(v)$ is a positive imaginary function of v, twice differentiable and such that $h \to ic$, c being a constant; as $v \to -\infty$. We desire the solution of (3-1) which, together with its first derivative, is continuous everywhere and at $\pm\infty$ satisfies the conditions

$$y = e^{-icv} + Re^{icv}, \qquad v \to -\infty \tag{3-2}$$

$$y' + (h + h'/(2h))y \to 0, \; v \to \infty \tag{3-3}$$

The constant R (the reflection coefficient) is to be determined. Condition (3-3), in which the primes denote differentiation with respect to v, is suggested by the fact that we want y to represent a wave traveling in the positive v direction (the factor exp $(i\omega t)$ is suppressed). In writing (3-3) we have assumed that h is such that for large values of v the two solutions of (3-1) are asymptotically proportional to*

$$y = h^{-\frac{1}{2}} e^{\pm\xi}, \tag{3-4}$$

$$\xi \equiv \xi(v) = icv + \int_{-\infty}^{v} (h - ic)\, dv. \tag{3-5}$$

Physical considerations suggest that solutions satisfying (3-2) and (3-3) exist in most cases of practical importance. However, if the function h is picked arbitrarily the corresponding solutions may be incapable of satisfying

* S. A. Schelkunoff[2] mentions that this approximation, sometimes designated by "WKB", goes back to Liouville. The ideas we shall use are quite similar to those in Schelkunoff's paper.

the conditions. For example, if $h = ic/(1 + \exp v)$ then $h \rightarrow ic \exp(-v)$ as $v \rightarrow \infty$, and the solutions of (3-1) behave like Bessel functions of order zero and argument $c \exp(-v)$. It may be verified that these solutions do not satisfy (3-3). Again, condition (3-3) may be satisfied without y having much resemblance to an outgoing wave at $v = \infty$. Thus if $h \rightarrow ia/v$ as $v \rightarrow \infty$, y increases like v^n where $n^2 - n - a^2 = 0$. When $0 < a < 1/2$ both values of n lie between 0 and 1, and both solutions satisfy (3-3). Despite these shortcomings it still seems best to retain (3-3) to specify the behavior of y at $v = \infty$.

It should be mentioned that P. S. Epstein[3] has obtained the reflected wave by transforming the hypergeometric differential equation into the form (3-1). This method has been extended by K. Rawer[4] who gives a number of references in which the approximation (3-4) is used to study propagation in a medium having a variable dielectric "constant". An interesting paper on the general subject of reflection in non-uniform transmission lines has been written by L. R. Walker and N. Wax[5].

1. When most of the reflection occurs in a short interval, say near $v = 0$' R may be obtained by numerical integration of (3-1). One method is to start at $v = 0$ with the initial conditions $y = 1$, $y' = 0$ and work outwards in both directions. Let $Y_a(v)$ denote this solution and $Y_b(v)$ the solution obtained by starting with $y = 0$, $y' = 1$. The general solution is

$$y = C_1 Y_a(v) + C_2 Y_b(v). \tag{3-6}$$

C_1 and C_2 are to be determined by the conditions

$$y = (\text{constant})\, h^{-1/2} e^{-\xi} \quad , \quad v > v_2 \tag{3-7}$$

$$y = (ic/h)^{1/2}[e^{-\xi} + Re^{\xi}] \quad , \quad v < v_1 \tag{3-8}$$

where v_1 and v_2 are large negative and positive values, respectively, of v. These conditions lead to equations for C_1, C_2, R:

$$\begin{aligned} &[y' + \theta^+ y]_{v=v_2} = 0 \\ &[y' - \theta^- y + 2(ich)^{1/2} e^{-\xi}]_{v=v_1} = 0 \\ &[y' + \theta^+ y - 2(ich)^{1/2} Re^{\xi}]_{v=v_1} = 0 \end{aligned} \tag{3-9}$$

in which ξ is given by (3-5) and

$$\theta^{\pm} = h \pm h'/(2h). \tag{3-10}$$

The required value of R is obtained by letting $v_1 \rightarrow -\infty$, $v_2 \rightarrow \infty$ in the expressions, which follow from (3-9),

$$
\begin{aligned}
\gamma &= C_2/C_1 = -[(Y_a' + \theta^+ Y_a)/(Y_b' + \theta^+ Y_b)]_{v=v_2} \\
\Gamma &= [y'/y]_{v=v_1} = [(Y_a' + \gamma Y_b')/(Y_a + \gamma Y_b)]_{v=v_1} \qquad (3\text{-}11)
\end{aligned}
$$

$$
R = [(\theta^+ + \Gamma)/(\theta^- - \Gamma)]_{v=v_1} \exp\left[-2icv_1 - 2\int_{-\infty}^{v_1} (h - ic)\,dv\right]
$$

where the arguments of $Y_a(v)$ and $Y_b(v)$ have been omitted for brevity.

If h should change from a positive imaginary quantity to a positive real quantity in (v_1, v_2) and remain greater than some fixed positive number for $v > v_2$ it may be shown that $|R| = 1$ (γ and Γ are real and Im $\theta^+ =$ Im θ^-, Real $\theta^+ = -$Real θ^- at $v = v_1$). This complete reflection is to be expected from physical consideration.

2. An exact expression for the reflection coefficient which holds when h satisfies the conditions following (3-1) (in particular it must not pass through zero anywhere in $-\infty < v < \infty$) is

$$
R = \tfrac{1}{2}(ic)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-\xi} y(v) \frac{d^2}{dv^2} h^{-\frac{1}{2}}\,dv \qquad (3\text{-}12)
$$

where ξ is given by (3-5). Before this integral for R may be evaluated $y(v)$, and hence R itself, must be known. Nevertheless, when R is small a useful approximation may be obtained by using the *WKB* approximation

$$
y(v) \approx (ic/h)^{1/2} e^{-\xi} \qquad (3\text{-}13)
$$

Thus

$$
\begin{aligned}
R &\approx \frac{1}{2}\int_{-\infty}^{\infty} e^{-2\xi} h^{-\frac{1}{2}} \frac{d^2}{dv^2} h^{-\frac{1}{2}}\,dv \\
&= \frac{1}{2i}\int_{-\infty}^{\infty} e^{-2\xi}\left[\frac{5}{16}K^{-5/2}\left(\frac{dK}{dv}\right)^2 - \frac{1}{4}K^{-3/2}\frac{d^2K}{dv^2}\right] dv
\end{aligned}
\qquad (3\text{-}14)
$$

in which $K = -h^2$.

The expression (3-12) for R is obtained by letting $v_0 \to -\infty$ in the integral equation

$$
\begin{aligned}
y(v_0) &= (ic/h_0)^{\frac{1}{2}} e^{-\xi_0} - \int_{-\infty}^{+\infty} G_a(v_0, v) y(v) h^{\frac{1}{2}} \frac{d^2}{dv^2} h^{-\frac{1}{2}}\,dv, \\
G_a(v_0, v) &= -\tfrac{1}{2} h_0^{-\frac{1}{2}} h^{-\frac{1}{2}} \begin{cases} e^{\xi - \xi_0}, & v < v_0 \\ e^{\xi_0 - \xi}, & v > v_0 \end{cases} \qquad (3\text{-}15) \\
\xi_0 - \xi &= \int_v^{v_0} h\,dv, \qquad h_0 \equiv h(v_0), \qquad \xi_0 \equiv \xi(v_0).
\end{aligned}
$$

$G_a(v_0, v)$ is the approximate Green's function suggested by (3-13). The integral equation may be obtained from the differential equation (3-1) and the boundary conditions (3-2) and (3-3) by the one-dimensional analogue of the method used in Section 3 of the companion paper[1]. If we multiply both sides of

$$\frac{d^2 y}{dv^2} - h^2 y = s(v) \tag{3-16}$$

(where $s(v)$ has been added for generality) by $G_a(v_0, v)$, integrate twice by parts over the intervals $(v_1, v_0 - \epsilon)$, $(v_0 + \epsilon, v_2)$ with $\epsilon > 0$ and $v_1 < v_0 < v_2$, and finally let $\epsilon \to 0$ we obtain

$$\begin{aligned} y(v_0) = \int_{v_1}^{v_2} G_a(v_0, v) \left[s(v) - y(v) h^{\frac{1}{2}} \frac{d^2}{dv^2} h^{-\frac{1}{2}} \right] dv \\ + G_a(v_0, v_1)[y' - \theta^- y]_{v=v_1} - G_a(v_0, v_2)[y' + \theta^+ y]_{v=v_2} . \end{aligned} \tag{3-17}$$

Equation (3-15) follows when we put $s(v) = 0$ and let $v_1 \to -\infty$, $v_2 \to \infty$. It will be recognized that (3-17) and (3-15) are closely related to integral equations occurring in the work of R. E. Langer[6] and E. C. Titchmarsh[9].

When h has, for example, one or more simple zeros in $-\infty < v < \infty$ the integral in (3-15) contains a factor which becomes infinite and the integral equation fails. However, we shall not concern ourselves with this case beyond remarking that it involves results obtained by H. Jeffreys[10], Langer[7], Furry[11] and others.

3. So far we have been considering the solution of only one equation whereas we really require the solution of a set of equations. If it is apparent that most of the disturbance is given by the first equation of the set it may be possible to proceed by successive approximations, each of the remaining equations being of the form (3-16) with $s(v)$ determined by the solution of the first equation.

Another method of dealing with a system of N equations is that of numerical integration. As a contribution towards obtaining the boundary conditions at large positive and negative values of v we shall state a generalized form of the *WKB* solution. Although this solution is related to the general results obtained by Birkhoff[12], Langer[8], and Newell[13] concerning the asymptotic forms assumed by the solutions of a system of ordinary linear differential equations of the first order, it is worth mentioning explicitly.

Let the mth equation of the set be

$$\ddot{y}_m = \sum_{n=1}^{N} A_{mn} y_n , \qquad m = 1, 2, \cdots, N \tag{3-18}$$

where the A_{mn}'s are relatively slowly varying functions of v (see equations

(3-22) for a more precise statement of the assumptions) and the dots denote differentiation with respect to v. We shall reserve primes to denote transposition of matrices. It is supposed that $A_{mn} = A_{nm}$ (equations(2-4) plus (2-5) satisfy this condition and (1-4) plus (1-5) may be made to do so by setting $\tilde{F}_0 = 2^{1/2}F_0$).

The solution of (3-18) is approximately

$$y_m \approx \sum_{\ell=1}^{N} S_{m\ell}[e^{\xi_\ell} d_\ell^- + e^{-\xi_\ell} d_\ell^+] \tag{3-19}$$

where the $d_\ell^\pm$ are the $2N$ constants of integration and

$$\varphi_\ell^2 S_{m\ell} = \sum_{n=1}^{N} A_{mn} S_{n\ell}$$

$$\varphi_\ell \sum_{n=1}^{N} S_{n\ell}^2 = 1 \tag{3-20}$$

$$\xi_\ell = \int_{v_{3\ell}}^{v} \varphi_\ell \, dv$$

serve to determine φ_ℓ, ξ_ℓ, and $S_{m\ell}$ (the last to within a plus or minus sign). We assume the N roots $\varphi_1^2, \varphi_2^2, \cdots \varphi_N^2$ of the determinantal equation arising from the first of equations (3-20) to be unequal, and denote by φ_ℓ that square root of φ_ℓ^2 which has a positive real part or, if the real part be zero, which has a positive imaginary part. $v_{3\ell}$ is any convenient constant.

The approximation (3-19) may be obtained by setting the assumed form

$$y_m = g_m e^{\pm\xi}, \qquad \xi = \int_{v_3}^{v} \varphi \, dv$$

in (3-18). The result is a set of N equations of which the mth is

$$\ddot{g}_m \pm 2\dot{g}_m \varphi \pm g_m \dot{\varphi} + g_m \varphi^2 = \sum_n A_{mn} g_n . \tag{3-21}$$

We also assume

$$|\dot{\varphi}| \ll |\varphi^2|, \; |\ddot{g}_m| \ll |\dot{g}_m\varphi| \ll |g_m\varphi^2| \tag{3-22}$$

$$g_m = g_{m0} + g_{m1} + g_{m2} + \cdots$$

where g_{mr} and its first two derivatives satisfy inequalities of the type

$$|g_{m0}| \gg |g_{m1}| \gg |g_{m2}| \cdots$$

The first and second order terms in (3-21) give, respectively,

$$g_{m0}\varphi^2 - \sum_n A_{mn} g_{n0} = 0 \tag{3-23}$$

Although the WKB approximation has the same form as (3-30) in the region where v is finite, we regard (3-30) and (3-31) as being the exact limiting forms of y. Hence, g^+ may differ from f^+.

Letting $v_0 \to -\infty$ in (3-29) and comparing the result with (3-30) gives the exact result

$$f^- = \tfrac{1}{2} \int_{-\infty}^{\infty} e^{-\Xi} [\ddot{S}' - 2\Phi\dot{S}' - \dot{\Phi}S'] y(v)\, dv \tag{3-32}$$

which leads to an approximation for the reflected wave when $y(v)$ is known approximately.

The integral equation (3-29) may be obtained by premultiplying both sides of $\ddot{y} = Ay$ by the transpose of the approximate Green's matrix

$$G_a(v_0, v) = \begin{cases} -\tfrac{1}{2} S e^{\Xi - \Xi_0} S_0', & v < v_0 \\ -\tfrac{1}{2} S e^{\Xi_0 - \Xi} S_0', & v > v_0. \end{cases}$$

and integrating by parts twice. It is seen that each column of $G_a(v_0, v)$ is an approximate solution of $\ddot{y} = Ay$, in which the columnar constants of integration are the columns of S_0', and represents a wave traveling away from v_0 in both directions. $G_a(v_0, v)$ is continuous at $v = v_0$ and

$$\left[\frac{\partial}{\partial v} G_a(v_0, v)\right]_{v=v_0+0} - \left[\frac{\partial}{\partial v} G_a(v_0, v)\right]_{v=v_0-0} = S_0 \Phi_0 S_0' = I$$

Thus the nth column of $G_a(v_0, v)$ gives the approximate values of $y_1(v)$, $y_2(v), \cdots, y_n(v)$, subject to the conditions that all these and all of their first derivatives are continuous at $v = v_0$ except $\dot{y}_n(v)$ which has the jump $\dot{y}_n(v_0 + 0) - \dot{y}_n(v_0 - 0) = 1$.

The presence of

$$\begin{aligned} 2\Phi\dot{S}' + \dot{\Phi}S' &= \Phi\dot{S}' - \Phi S'\dot{S}S^{-1} \\ &= \Phi(\dot{S}'S - S'\dot{S})S^{-1} \end{aligned}$$

in (3-29) and (3-32) makes the N variable case somewhat different from the case $N = 1$.

5. When Z_{mn} and Y_{mn} are slowly varying functions of v the approximate solution of the transmission line equations

$$\begin{aligned} \frac{dV_m}{dv} &= -\sum_{n=1}^{N} Z_{mn} J_n \\ \frac{dJ_m}{dv} &= -\sum_{n=1}^{N} Y_{mn} V_n \end{aligned} \tag{3-32}$$

where $Z_{mn} = Z_{nm}$ and $Y_{mn} = Y_{nm}$ is, as in (3-19),

$$V_m \approx \sum_{\ell=1}^{N} S_{m\ell} \, [e^{\xi_\ell} d_\ell^- + e^{-\xi_\ell} d_\ell^+]$$

$$J_m \approx \sum_{\ell=1}^{N} T_{m\ell} \, [e^{\xi_\ell} d_\ell^- - e^{-\xi_\ell} d_\ell^+]. \tag{3-33}$$

Here ξ_ℓ is the integral of φ_ℓ as given by (3-20), and φ_ℓ is determined by setting the determinant of the matrix $\varphi^2 I - ZY$ to zero. When φ_ℓ is known, $S_{m\ell}$ and $T_{m\ell}$ are determined (to within a plus or minus sign which may be absorbed by the constants $d_\ell^\mp$ of integration) by the relations

$$\varphi_\ell S_{m\ell} = -\sum_{n=1}^{N} Z_{mn} T_{n\ell}$$

$$\varphi_\ell T_{m\ell} = -\sum_{n=1}^{N} Y_{mn} S_{m\ell} \tag{3-34}$$

$$\sum_{n=1}^{N} S_{m\ell} T_{m\ell} = 1$$

The last condition, which arises from the condition that the equations for the second-order terms be consistent, may be regarded as a generalization of Slater's[15] result for the case $N = 1$.

4. *Transformation for Wave Guide Plus Horn*

The system to which we shall apply some of the preceding equations consists of a straight wave guide starting at $x = -\infty$ and running to $x = 0$ where it is connected to a sectoral horn. The horn is flared in the (x, y) plane only. The dimension of the system normal to the (x, y) plane is constant and equal to a or b according to whether the electric or magnetic vector is in the plane of the horn.

One might expect that the field in this system may also (in addition to our method) be determined by an alternating procedure of the type described by Poritsky and Blewett[16] using the equations obtained by Barrow and Chu[17] for transmission in the horn. However, we shall not investigate this possibility as we are primarily interested in using the system as an example to which we may apply the foregoing equations.

If the total angle of the horn is $2\alpha\pi$, and if the sides of the straight guide are at $y = 0$ and $y = b$, (assuming the electric vector to be in the plane of the horn), the equation of the lower side, i.e., the continuation of the side $y = 0$, of the horn is $y = -x \tan \alpha\pi$ and that of the upper side is $y = b + x \tan \alpha\pi$. If $z = x + iy$ and $w = v + i\theta$ then the Schwarz-Christoffel transformation

$z = f(w)$ which carries the guide plus horn in the z plane into the straight guide with walls at $\theta = 0$, $\theta = \pi$ in the w plane may be obtained from

$$\frac{dz}{dw} = (1 - e^{2w})^{\alpha}\, b/\pi \tag{4-1}$$

This gives, upon setting

$$\left|\frac{dz}{dw}\right|^2 = |f'(v + i\theta)|^2 = [1 - 2e^{2v}\cos 2\theta + e^{4v}]^{\alpha}\, b^2/\pi^2,$$

the relation (4-2)

$$1 + g(v, \theta) = [1 - 2e^{2v}\cos 2\theta + e^{4v}]^{\alpha}$$

from which the a_n's may be obtained in accordance with (1-2).

5. *Expressions for the a_n's for Horn*

The Fourier coefficients of $1 + g(v, \theta)$ appearing in (1-2) and (2-2) are the same. It may be shown from (4-2) that

$$1 + a_0 = \begin{cases} e^{4\alpha v}F(-\alpha, -\alpha; 1; e^{-4v}) & , \quad v > 0 \\ \Gamma(1 + 2\alpha)/\Gamma^2(1 + \alpha) & , \quad v = 0 \\ F(-\alpha, -\alpha; 1; e^{4v}) & , \quad v < 0 \end{cases} \tag{5-1}$$

and

$$a_{2r} = \begin{cases} 2e^{4\alpha v - 2rv}(-\alpha)_r F(-\alpha, r - \alpha; r + 1; e^{-4v})/r!\,, & v > 0 \\ 2(-\alpha)_r(1 + a_0)_{v=0}/(1 + \alpha)_r\,, & v = 0 \\ 2e^{2rv}(-\alpha)_r F(-\alpha, r - \alpha; r + 1; e^{4v})/r!\,, & v < 0 \end{cases} \tag{5-2}$$

where the F's denote hypergeometric functions, $r = 1, 2, \cdots$ and we have used the notation

$$(\beta)_0 = 1,\ (\beta)_r = \beta(\beta + 1) \cdots (\beta + r - 1) \tag{5-3}$$

When n is odd, $a_n = 0$ because of symmetry about $\theta = \pi/2$. The expressions for $v > 0$ in (5-1) and (5-2) may be verified by expanding the two factors in

$$1 + g(v, \theta) = e^{4\alpha v}(1 - e^{2i\theta - 2v})^{\alpha}(1 - e^{-2i\theta - 2v})^{\alpha}$$

by the binominal theorem and picking out the terms containing $e^{2ri\theta}$. When $v < 0$ we use the relation $1 + g(v, \theta) = e^{4\alpha v}[1 + g(-v, \theta)]$, and when $v = 0$ we may sum the hypergeometric series.

Differentiation of (5-1) and (5-2) leads to

$$\frac{d}{dv}(1 + a_0) = \begin{cases} 4\alpha e^{4\alpha v} F(-\alpha, 1 - \alpha; 1; e^{-4v})\,, & v > 0 \\ 2\alpha(1 + a_0)_{v=0}\,, & v = 0 \\ 4\alpha^2 e^{4v} F(1 - \alpha, 1 - \alpha; 2; e^{4v}), & v < 0 \end{cases} \tag{5-4}$$

$$\frac{d^2}{dv^2}(1 + a_0) = \begin{cases} 16\alpha^2 e^{4\alpha v}(1 - e^{-4v})^{2\alpha-1} F(\alpha, \alpha; 1; e^{-4v}), & v > 0 \\ 16\alpha^2 e^{4v}(1 - e^{4v})^{2\alpha-1} F(\alpha, \alpha; 1; e^{4v}), & v < 0 \end{cases} \tag{5-5}$$

where in obtaining (5-5) use was made of Euler's transformation

$$F(a, b; c; x) = (1 - x)^{c-a-b} F(c - a, c - b; c; x)$$

It is seen that $d(1 + a_0)/dv$ is continuous at $v = 0$ but the second derivative becomes infinite as $v^{2\alpha-1}$.

When $1 + a_0$ and a_2 are expressed as the customary integrals defining the Fourier coefficients it is seen that one of the coefficients occurring in equation (2-4) for F_1 is given by

$$\begin{aligned} 1 + a_0 - a_2/2 &= \frac{2}{\pi}\int_0^\pi (1 - 2e^{2v}\cos 2\theta + e^{4v})^\alpha \sin^2\theta \, d\theta \\ &= (e^{2v} + 1)^{2\alpha} F(-\alpha, \tfrac{1}{2}; 2; \operatorname{sech}^2 v) \end{aligned} \tag{5-6}$$

At $v = 0$, $1 + a_0 - a_2/2$ and its first and second derivatives are continuous, their values being

$$\begin{gathered} \frac{\Gamma(2 + 2\alpha)}{\Gamma(1 + \alpha)\Gamma(2 + \alpha)}, \qquad \frac{2\alpha\Gamma(2 + 2\alpha)}{\Gamma(1 + \alpha)\Gamma(2 + \alpha)}, \\ \frac{4\alpha(2\alpha^2 + 2\alpha + 1)\Gamma(1 + 2\alpha)}{\Gamma(1 + \alpha)\Gamma(2 + \alpha)}, \end{gathered} \tag{5-7}$$

respectively. These may be obtained by differentiating the integral in (5-6) and setting $v = 0$.

A second expression for $1 + a_0 - a_2/2$ follows from (5-1) and (5-2):

$$1 + a_0 - a_2/2 = \begin{cases} e^{4\alpha v}[F(-\alpha, -\alpha; 1; e^{-4v}) + \alpha e^{-2v}F(-\alpha, 1 - \alpha; 2; e^{-4v})], & v > 0 \\ F(-\alpha, -\alpha; 1; e^{4v}) + \alpha e^{2v}F(-\alpha, 1 - \alpha; 2; e^{4v}), & v < 0. \end{cases} \tag{5-8}$$

6. *Approximation to Reflection Coefficient of Horn, Electric Vector in* (x, y) *Plane*

When the flare angle $2\alpha\pi$ of the horn is very small the reflection coefficient may be shown to be

$$R_E = \frac{i\alpha}{2k} + 0(\alpha^2) \tag{6-1}$$

where $0(\alpha^2)$ denotes correction terms of the order α^2. This result is based upon the fact that when terms of order α^2 are neglected the set of differential equations (1-4) and (1-5) reduce to the single equation

$$F_0'' + (1 + a_0)k^2 F_0 = 0 \tag{6-2}$$

where, from (5-1, 4, 5),

	$v > 0$	$v < 0$
$1 + a_0$	$e^{4\alpha v}$	1
$\frac{d}{dv}(1 + a_0)$	$4\alpha e^{4\alpha v}$	0
$\frac{d^2}{dv^2}(1 + a_0)$	$16\alpha^2 e^{4\alpha v}(1 - e^{-4v})^{2\alpha-1}$	$16\alpha^2 e^{4v}(1 - e^{4v})^{2\alpha-1}$

The reflection coefficient (6-1) is the one corresponding to the differential equation (6-2) and may be computed by setting

$$(1 + a_0)k^2 = -h^2 = K \tag{6-3}$$

in the integrals (3-14).

The expression (6-1) for R_E may be obtained quickly (but the procedure is not trustworthy) by assuming that the principal contribution to the first integral in (3-14) comes from the region close to $v = 0$, say in $-\epsilon < v < \epsilon$, where the second derivative of $h^{-1/2}$ is infinite but integrable. When the integration is performed approximately by replacing the second derivative by the first, (3-14) gives

$$R_E \approx \frac{1}{2}\left[h^{-\frac{1}{2}} \frac{d}{dv} h^{-\frac{1}{2}}\right]_{-\epsilon}^{+\epsilon} \tag{6-4}$$

$$\approx \frac{1}{2ik}\left[\frac{d}{dv}(1 + a_0)^{-\frac{1}{4}}\right]_{-\epsilon}^{\epsilon} = \frac{i\alpha}{2k}$$

where ϵ is assumed to be so small that $1 + a_0$ is effectively unity and $d(1 + a_0)/dv$ changes from 0 at $-\epsilon$ to 4α at $+\epsilon$.

A more careful investigation based on the second integral in (3-14) also leads to the value (6-1) for R_E. It further suggests that possibly most of the correction term, denoted by $0(\alpha^2)$ in (6-1), is given by

$$\frac{\alpha^2}{2ik}\int_0^\infty e^{-2\xi-2\alpha v}\, dv = \frac{1}{4ix} + \frac{e^{ix}}{4i}[Si(x) - \pi/2 + iCi(x)] \tag{6-5}$$

with $x = k/\alpha$ and $2\xi = ix[\exp(2\alpha v) - 1]$. $Si(x)$ and $Ci(x)$ denote the integral sine and cosine functions. Incidentally, the rather curious result

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{mn(m + n)} = 2 \sum_{n=1}^{\infty} n^{-3}$$

turned up in the investigation of the orders of magnitude of the various terms.

7. *Approximation to Reflection Coefficient of Horn, Magnetic Vector in (x, y) Plane*

The work of this section is quite similar to that in Section 6 except that here we enter into more of the details. We shall show that when α is small the reflection coefficient appearing in equation (2-6) is

$$R_H = \frac{i\alpha}{2c^3} + 0(\alpha^2). \tag{7-1}$$

From (2-4) the analogue of the differential equation (3-1) is

$$F_1'' + [\kappa^2(1 + a_0 - a_2/2) - 1]F_1 = 0 \tag{7-2}$$

and the K appearing in the second of equations (3-14) is now

$$K = -h^2 = \kappa^2(1 + a_0 - a_2/2) - 1 \tag{7-3}$$

The largest terms in the expression (5-8) for $1 + a_0 - a_2/2$ yield, to within terms of $0(\alpha)$,

$$\begin{aligned} K &= \kappa^2(e^{4\alpha v} + \alpha e^{-2v}) - 1\,, \quad v > 0 \\ \dot{K} &= \kappa^2(4\alpha\, e^{4\alpha v} - 2\alpha e^{-2v}) \\ \ddot{K} &= \kappa^2(16\alpha^2 e^{4\alpha v} + 4\alpha e^{-2v}) \end{aligned} \tag{7-4}$$

$$\begin{aligned} K &= \kappa^2(1 + \alpha e^{2v}) - 1 = c^2 + \kappa^2\alpha e^{2v} \quad , \quad v < 0 \\ \dot{K} &= 2\alpha\kappa^2 e^{2v} \\ \ddot{K} &= 4\alpha\kappa^2 e^{2v} \end{aligned} \tag{7-5}$$

where the dots denote differentiation with respect to v and $c^2 = \kappa^2 - 1$. We have retained the α^2 in $\ddot{K}$ as given by (7-4) because at this stage we do not know whether it may be neglected or not.

When $v < 0$, the definition (3-5) of ξ and (7-5) yield

$$\begin{aligned} \xi &= icv + i \int_{-\infty}^{v} (K^{\frac{1}{2}} - c)\, dv \\ &= icv + ic \int_{-\infty}^{v} [(1 + \kappa^2 c^{-2} \alpha e^{2v})^{\frac{1}{2}} - 1]\, dv = icv + 0(\alpha) \end{aligned} \tag{7-6}$$

and we have

$$\int_{-\infty}^{0} e^{-2\xi} K^{-5/2} \dot{K}^2 \, dv = \int_{-\infty}^{0} e^{-2icv} c^{-5} 4\alpha^2 \kappa^4 e^{4v} \, dv \tag{7-7}$$
$$= 0(\alpha^2)$$

which may be neglected. The other integral suggested by (3-14) is

$$\int_{-\infty}^{0} e^{-2\xi} K^{-3/2} \ddot{K} \, dv = \int_{-\infty}^{0} e^{-2icv} c^{-3} 4\alpha\kappa^2 e^{2v} \, dv \tag{7-8}$$
$$= 2\alpha\kappa^2 c^{-3}/(1 - ic)$$

When $v > 0$,

$$\xi = \xi_{v=0} + i \int_0^v [\kappa^2(e^{4\alpha v} + \alpha e^{-2v}) - 1]^{\frac{1}{2}} \, dv$$
$$= i \int_0^v (\kappa^2 e^{4\alpha v} - 1)^{\frac{1}{2}} \, dv + 0(\alpha), \tag{7-9}$$
$$= \frac{i}{2\alpha} [x - \tan^{-1} x - c + \tan^{-1} c] + 0(\alpha),$$
$$x = (\kappa^2 e^{4\alpha v} - 1)^{\frac{1}{2}}, \qquad 2\alpha \, dv = x(1 + x^2)^{-1} \, dx$$

In the integrals containing exp $(-2v)$ as a factor, ξ may be taken to be icv since the integrand becomes negligibly small by the time icv differs significantly from (7-9). We have

$$\int_0^\infty e^{-2\xi} K^{-5/2} \dot{K}^2 \, dv = \int_0^\infty e^{-2\xi} (\kappa^2 e^{4\alpha v} - 1)^{-\frac{5}{2}} \kappa^4 \alpha^2 (4e^{4\alpha v} - 2e^{-2v})^2 \, dv$$
$$= \int_0^\infty e^{-2\xi} x^{-5} \kappa^4 \alpha^2 16 e^{8\alpha v} \, dv \tag{7-10}$$
$$= 8\alpha \int_c^\infty e^{-i[x - \tan^{-1} x - c + \tan^{-1} c]/\alpha} (x^{-4} + x^{-2}) \, dx$$

where the integrals containing e^{-2v} and e^{-4v} have been neglected since their contribution is $0(\alpha^2)$. When α becomes exceedingly small the exponential term oscillates rapidly and the last line of (7-10) is likewise $0(\alpha^2)$. This may be verified by integrating by parts, starting with

$$\exp Y \, dx = i\alpha \, x^{-2}(1 + x^2) d(\exp Y),$$
$$Y = -i(x - \tan^{-1} x) / \alpha$$

The last integral which must be considered is

$$\int_0^\infty e^{-2\xi} K^{-3/2} \ddot{K} \, dv = \int_0^\infty e^{-2\xi}(\kappa^2 e^{4\alpha v} - 1)^{-\frac{3}{2}} \kappa^2[16\alpha^2 e^{4\alpha v} + 4\alpha e^{-2v}] \, dv$$

$$= 16\kappa^2 \alpha^2 \int_0^\infty e^{-2\xi} x^{-3} e^{4\alpha v} \, dv$$

$$+ \int_0^\infty e^{-2icv-2v} c^{-3} \kappa^2 4\alpha \, dv \tag{7-11}$$

$$= 8\alpha \int_c^\infty e^{-i[x-\tan^{-1}x-c+\tan^{-1}c]/\alpha} x^{-2} \, dx + 2\alpha\kappa^2 c^{-3}/(1 + ic)$$

$$= 0(\alpha^2) + 2\alpha\kappa^2 c^{-3}/(1 + ic).$$

That the integral having x as the variable of integration is $0(\alpha^2)$ may be shown as in (7-10).

When we combine our results in accordance with (3-14) we obtain

$$R_H = \frac{1}{2i} \int_{-\infty}^\infty e^{-2\xi}[\tfrac{5}{16} K^{-5/2} \dot{K}^2 - \tfrac{1}{4} K^{-3/2} \ddot{K}] \, dv$$

$$= -\frac{\alpha\kappa^2 c^{-3}}{4i} \left[\frac{1}{1 - ic} + \frac{1}{1 + ic}\right] + 0(\alpha^2) \tag{7-12}$$

$$= i\alpha/(2c^3) + 0(\alpha^2)$$

which is (7-1).

If, instead of discarding (7-10) because it is $0(\alpha^2)$, we retain it and the corresponding integral in (7-11) (in the hope that they represent most of the difference between the approximate value (7-1) for R_H and the true value) we obtain the approximation

$$R_H = \frac{i\alpha}{2c^3} - \frac{i\alpha}{4} \int_c^\infty e^{-i(x-\tan^{-1}x-c+\tan^{-1}c)/\alpha}(5x^{-4} + x^{-2}) \, dx \tag{7-13}$$

in which the integral may be evaluated by numerical integration.

The approximations (6-1) and (7-1) for the reflection coefficients may also be obtained from an equation given by N. H. Frank.[18] However, care must be taken to suitably define the wave guide characteristic impedance which appears in his expression.

8. *Speculation on the Reflection Obtained from Horn Flared in Both Directions*

All the work from Section 4 onward applies only to a horn flared in one plane. Nevertheless, it is interesting to speculate on how close an estimate of the reflection from a three-dimensional horn may be obtained by superposing the two reflection coefficients (6-1) and (7-1). It must be kept in mind that the flare angles (the α's) may be different in the two directions, that k is given by (1-2) and c by (2-2), and finally the difference (not the sum) of R_E and R_H must be taken. In (6-1) R_E is the reflection coefficient of the component of the magnetic vector normal to the (x, y) plane (which is proportional to Q), while in (7-1) R_H is the reflection coefficient of the transverse electric vector (which is proportional to P) and there is a difference in sign just as in the case of voltage and current reflection coefficients. If $a > b$ and λ_0 is the wavelength in free space, the superposition gives the following expression for the reflection coefficient of the electric vector:

$$R = R_H - R_E$$
$$= \frac{i}{2}[(2a/\lambda_0)^2 - 1]^{-\frac{1}{2}}(\alpha_H/[(2a/\lambda_0)^2 - 1] - a\alpha_E/b) \quad (8\text{-}1)$$

where $2\pi\alpha_H$ and $2\pi\alpha_E$ are the total horn angles in the planes of H and E, respectively. Of course this approximation can be expected to hold only when α_H and α_E are small.

9. *Numerical Calculations—R_H for 60° Horn*

The value of R_H, the reflection coefficient when the magnetic vector lies in the plane of the flare, was computed on the assumption that only the dominant mode need be considered.* Thus, instead of the system of equations (2-4) and (2-5), only their simplified version, namely the single second order differential equation (7-2), was used. This equation may be written as

$$\frac{d^2 F_1}{dv^2} + KF_1 = 0 \quad (9\text{-}1)$$

where, according to (5-6),

$$K = -h^2 = \kappa^2(1 + a_0 - a_2/2) - 1 \quad (9\text{-}2)$$
$$1 + a_0 - a_2/2 = (e^{2v} + 1)^{2\alpha}F(-\alpha, 1/2; 2; \operatorname{sech}^2 v).$$

The problem was to obtain the R_H appearing in that solution F_1 of (9-1) which satisfies the boundary conditions (2-6).

No computations for R_E were made.

In the first method of calculation the integrals in the approximation (3-14), namely

$$R_H = \frac{1}{2i}\int_{-\infty}^{\infty} e^{-2\xi} K^{-1/4} \frac{d^2}{dv^2} K^{-1/4}\, dv \quad (9\text{-}3)$$
$$2\xi = 2icv + 2i\int_{-\infty}^{v} (K^{\frac{1}{2}} - c)\, dv,$$

* I am indebted to Miss M. Darville for carrying out the computations of this section.

were evaluated by Simpson's rule. The second derivative of $K^{-1/4}$ was computed from the even order central differences of $K^{-1/4}$.[19] For $\alpha = 1/6$, corresponding to an angle of $\pi/3$ between the two sides of the horn, calculations at two representative wave lengths led to the table

λ_0	c	$\kappa^2 = 1+c^2$	R_H
1.549a	.8173	1.6680	$-.0420 + i.0724$
1.610a	.7376	1.5441	$-.0551 + i.0878$

(9-4)

An idea of the variation of K may be obtained from its values at $-\infty$, $-.6$, 0, .6, 1.8, 3.6 which are approximately .67, .76, .98, 1.62, 4.56, 17.4, respectively. The range of integration was $-3 \leq v \leq 4.4$.

The second method of computation used the formulas (3-11) with F_1 playing the role of y. The differential equation (9-1) was integrated by the Kutte-Runge method, the interval between successive values of v being 0.2. For $c = .8173$ the values obtained were

v_1	v_2	γ	Γ	R_H
$-.6$	.6	$-.202 - i.981$	$-.142 - i.794$	$-.0167 + i.0658$
-1.2	1.2	$-.218 - i1.004$	$-.049 - i.696$	$-.0525 + i.0754$
-1.8	1.8	$-.225 - i.989$	$+.086 - i.716$	$-.0512 + i.0753$
-2.4	2.4	$-.220 - i1.000$	$.136 - i.842$	$-.0424 + i.0722$

(9-5)

In order to gain an idea of the meaning of these values of v it should be recalled that $w = v + i\theta$ and the walls of the guide are at $\theta = 0$ and $\theta = \pi$. An interval of length $\pi = 3.14 \cdots$ in the v direction therefore corresponds roughly to a distance equal to the width of the guide. The above table indicates that, loosely speaking, most of the reflection occurs close to the junction of the horn and wave guide.

The last value of R_H in (9-5) agrees quite well with the value $-.0420 + i.0724$ obtained from the approximate expression (9-3). It appears that the method leading to (9-5) is superior to the one based on (9-3) since, in theory, it may be made as accurate (insofar as the single equation (9-1) may replace the set of equations (2-4, 5)) as desired. Moreover, less actual work seems to be required.

The approximation (7-1) yields, for $c = .8173$,

$$R_H = \frac{i\alpha}{2c^3} = \frac{i(\frac{1}{6})}{2(.8173)^3} = i.153$$

which is considerably in error, as we might expect, since $\alpha = 1/6$ is not small. However, if we use the approximation (7-13) and evaluate the integral by Simpson's rule we obtain

$$R_H = i.153 - (.061 + i.077)$$
$$= -.061 + i.076$$

which is in better agreement with the earlier values of R_H.

No similar computations have been made to test the corresponding approximation for R_E obtained when the correction term (6-5) is added to the leading term in (6-1). However, it appears that for $\alpha = 1/6$ and the representative value $k = .38$, (6-5) is only about one sixth as large as $i\alpha/(2k)$ and hence is relatively unimportant.

References

1. S. O. Rice, Reflection from Corners in Rectangular Wave Guides—Conformal Transformation, this issue of the *B. S. T. J.*
2. S. A. Schelkunoff, Solution of Linear and Slightly Non-linear Differential Equations, *Quart. Jl. Appl. Math.*, *3*, 348–360 (1946).
3. P. S. Epstein, Reflection of Waves in an Inhomogeneous Absorbing Medium, *Proc. Nat'l. Acad. Sci.*, *16*, 627–637 (1930).
4. K. Rawer, Electrische Wellen in einem geschichteten Medium, *Ann. der Physik*, *35*, 385–416 (1939).
5. L. R. Walker and N. Wax, Non-Uniform Transmission Lines and Reflection Coefficients, *Jl. Appl. Phys.*, *17*, 1043–1045 (1946).
6. R. E. Langer, The Asymptotic Solutions of Ordinary Linear Differential Equations of the Second Order, with Special Reference to the Stokes Phenomenon, *Bull. Amer. Math. Soc.*, *40*, 545–582 (1934). Many references to the earlier work are given.
7. R. E. Langer, On the Connection Formulas and the Solutions of the Wave Equation, *Phys. Rev.*, *51*, 669–676 (1937).
8. R. E. Langer, The Boundary Problem of an Ordinary Linear Differential System in the Complex Domain, *Trans. Am. Math. Soc.*, *46*, 151–190 (1939).
9. E. C. Titchmarsh, Eigen-function Transformations, Oxford (1946).
10. H. Jeffreys, On Certain Approximate Solutions of Linear Differential Equations of the Second Order, *Proc. London Math. Soc.*, (2), *23*, 428–436 (1924).
11. W. H. Furry, Two Notes on Phase Integral Methods, *Phys. Rev.*, *71*, 360–371 (1947).
12. G. D. Birkhoff and R. E. Langer, The Boundary Problems and Developments Associated with a System of Ordinary Linear Differential Equations of the First Order, *Proc. Amer. Acad. Arts and Sci.*, *58*, 49–128 (1923).
13. H. E. Newell, The Asymptotic Forms of the Solutions of an Ordinary Linear Matric Differential Equation in the Complex Domain, *Duke Math. Jl.*, *9*, 245–258 (1942) and *10*, 705–709 (1943).
14. See, for example, Frank and von Mises, Differential und Integralgleichungen, Vol. 1, page 56, Braunschweig (1930).
15. J. C. Slater, Microwave Transmission, page 74, McGraw-Hill (1942).
16. Poritsky and Blewett, A Method of Solution of Field Problems by Means of Overlapping Regions, *Quart. Jl. Appl. Math. 3*, 336–347 (1946).
17. Barrow and Chu, Theory of the Electromagnetic Horn, *Proc. I.R.E.*, *27*, 51–64 (1939).
18. N. H. Frank, Reflections from Sections of Tapered Transmission Lines and Wave Guides, *Radiation Lab. Rep.* 43-17, Jan. 6, 1943.
19. Whittaker and Robinson, Calculus of Observations, page 64, Blackie (1926).

Abstracts of Technical Articles by Bell System Authors

Pulse Echo Measurements on Telephone and Television Facilities.[1] L. G. Abraham, A. W. Lebert, J. B. Maggio, and J. T. Schott. Pulse echo measurements have been used on telephone and television facilities since 1940 to locate impedance irregularities and control quality in manufacture and installation. These sets send a pulse into a line and observe on an oscilloscope the echoes returned from irregularities. The shape and width of the pulse, the rate at which it is repeated and the pulse magnitude are important in determining the accuracy of the results and the requirements of the measuring apparatus. The "coaxial pulse echo set" is used for factory and field testing of coaxial cables. The "Lookator" was developed for use on much narrower band systems such as spiral-four field cable and open wire lines.

Television Network Facilities.[2] L. G. Abraham and H. I. Romnes. This paper describes television network facilities which are needed to connect studios and other pickup points to transmitters in the same and in distant cities, and discusses their transmission characteristics. Short-haul television circuits may be by microwave radio or over wire circuits. Long-haul television connections may be by radio relay or over coaxial systems of the type originally developed for carrier telephone circuits. Transmission requirements include adequate frequency band, accurate gain and phase equalization, and freedom from interference resulting from excessive noise, crosstalk, or modulation. Radio and wire systems are under development to provide extensive high-quality television networks.

A Carrier Telephone System for Rural Service.[3] J. M. Barstow. The M1 carrier telephone system was designed for the purpose of extending telephone service into areas served by rural power lines, but not served by co-existing rural telephone lines. To the local office operator and to a carrier subscriber the service provided is the same, so far as procedures involved in establishing a connection are concerned, as a voice-frequency line.

At the office end of the system a telephone wire line extends from the office to a point near the power line. Here is located a converter (called common terminal) which converts the voice-frequency signal to be transmitted to the subscriber to an amplitude-modulated double-sideband carrier signal. This signal is coupled to the power line through a coupling unit

[1] *Trans., A. I. E. E.*, vol. 66, 1947 (pp. 541–548).
[2] *Transactions, A. I. E. E.*, vol. 66, 1947 (pp. 459–464).
[3] *Trans. A. I. E. E.*, vol. 66, 1947 (pp. 501–507).

and high-voltage capacitor. At the subscriber's location the signal is taken off by similar means and led by separate wires to the subscriber premises, where it is reconverted to voice frequency by means of a subscriber terminal. A signal transmitted from the subscriber to the central office goes through similar conversions.

The usual number of parties per two-way channel may be assigned according to local custom, and divided-code or full-code ringing is provided. Equipment is available for five two-way channels over a single-power line employing frequencies in the range 150 to 425 kilocycles. A sixth channel has been discontinued because of radio interference.

A description is given of the manner in which the power line should be treated in order to reduce reflection effects. The power line treatment does not affect its capabilities in regard to power transmission.

Application of Rural Carrier Telephone System.[4] E. H. B. Bartelink, L. E. Cook, F. A. Cowan,* and G. R. Messmer. This paper deals with the application of a carrier system developed primarily for providing rural telephone service over power distribution circuits in areas where this means of extending telephone service may be more attractive than other available methods. The modifications required in the power circuits to permit carrier frequency transmission are described, including the effect of these modifications on the operation of the power system. Construction features also are discussed. The use of the rural carrier telephone system over open wire telephone pairs is discussed briefly.

An Improved Cable Carrier System.[5] H. S. Black, F. A. Brooks, A. J. Wier and I. G. Wilson. A new 12-channel cable carrier system is described which is suitable for transcontinental communications. Important features are negative feedback amplifiers of improved design, new arrangements for accurate equalization of the cable loss, and automatic thermistor regulators which continuously control the transmission of each system.

Joint Use of Pole Lines for Rural Power and Telephone Services.[6] J. W. Campbell,* L. W. Hill, L. M. Moore, and H. J. Scholz. The use of poles to carry both power and communication circuits is not new, having been employed before 1890. There are today more than 6,000,000 poles used jointly by power and telephone organizations in the United States. The great bulk of these poles are located in urban areas where the voltages of the power circuits concerned are generally less than 5,000 volts and the span lengths between poles generally do not exceed about 150 feet.

As power and telephone lines were extended into rural territory, new

[4] *Trans. A. I. E. E.*, vol. 66, 1947 (pp. 511–517).
[5] *Trans. A. I. E. E.*, vol. 66, 1947 (pp. 741–746).
[6] *Trans. A. I. E. E.*, vol. 66, 1947 (pp. 519–524).
* Of Bell Tel. Labs.

problems were encountered in the application of joint construction because of the use of longer spans and higher voltages for the power circuits and the increased noise induction in the necessarily longer exposures. However, progress in the art through cooperation of the telephone industry with the Edison Electric Institute and the Rural Electrification Administration has brought about the developments reviewed in this paper which now make long span higher voltage rural joint use feasible where conditions are favorable.

Atomic Energy.[7] KARL K. DARROW. (*The 1947 Norman Wait Harris Lectures at Northwestern University.*) This little book, which reproduces four lectures substantially as they were given, is at once a very readable and a very accurate account of enough of the facts of nuclear physics to convey a good understanding of the atomic bomb and the possibilities of atomic power. The scientific accuracy of the presentation is instanced by the author's apologies for his title; he emphasizes that in reality his subject is *nuclear* energy, but that on the day of Hiroshima somebody wrote of an *atomic* bomb and the misusage spread like a chain reaction.

The role of electrons, protons and neutrons in atom building is told in a simple and entertaining style (but with a degree of ornamentation that may disturb some readers), and the discussion of rest mass and the Einstein relation between mass and energy is pointed up by well-chosen numerical illustrations beginning with the lightest composite nuclei. The role of fast-particle bombardment in increasing and decreasing the size of nuclei is also explained. The reader thus acquires a clear understanding of the basic phenomena for which nuclear fission is famed. The text is augmented by well-chosen cloud chamber photographs.

Though the author's treatment is accurate, his style and marshalling of facts are very readable. This is well illustrated by the closing paragraph of the third lecture, which follows immediately upon the author's development of the idea of the chain reaction:

> Here is the climax of my lectures, and here is where you should be frightened; and, if I had an orchestral accompaniment, here is where the orchestra would have mounted to a tumultuous fortissimo, with the drums rolling and the trumpets blaring and the tuba groaning and the strings in a frenzy, and whatever else a Richard Wagner could contrive to cause a sense of Götterdämmerung; for, let there be no doubt of it, this *is* something that could bring on the twilight of civilization. But at this crucial juncture I have only words to serve me, and all the words are spoiled. We speak of an awful headache, a dreadful cold, a frightful bore, and an appalling storm; and now when something comes along that is really awful and dreadful and frightful and appalling, all these words have been devaluated and have no terror in them. I have to fall back on the saying, of unknown origin and dubious value, that the strongest emphasis is understatement. Let then this picture, with its circles and its symbols and its numbers, be considered an emphatic understatement of the most terrific thing yet known to man.

---

[7] Publishedby John Wiley & Sons,Inc., New York, and Chapman & Hall, Ltd., London. 80 pages. $2.00.

The book will be welcomed particularly by those who at one time or another have had a general acquaintance with radio-activity, cosmic rays and the results of cloud chamber research, but whose vocational activities have forced such special knowledge well into the nebulous regions of their memories.

Network Theory Comes of Age.[8] R. L. Dietzold. The third decade in the growth of modern network theory, the decade of maturity, is considered in this review of the advances in network theory evolved over the past ten years. New types of networks developed during the war are included.

Thermistors as Components Open Product Design Horizons.[9] K. P. Dowell. These thermally sensitive resistors with high negative temperature coefficients have come a long way since they were laboratory curiosities and are now available in a wide range of types with diverse and stable characteristics. You may be able to transfer to your own problems some of the unusual design ideas described here.

Gas Pressure for Telephone Cables.[10] R. C. Giese. Communication cables consist of a number of electric conductors insulated from one another and encased in a metal sheath. This encasement is subjected to numerous hazards, such as those caused by electrolysis, crystallization, various kinds of mechanical damage, and lightning burns. Any damage to the sheath which will permit water to enter the cable will decrease the effectiveness of the insulation material and thus cause an impairment or an interruption to the service. The entrance of moisture through small openings in the sheath can be materially retarded when the space inside the cable, not occupied by the conductors or insulation, is filled with a gas maintained under controlled conditions. Nitrogen is the gas usually used for this purpose because it is inert and does not combine chemically with the conductors or insulation. In addition the use of the gas provides a method of locating openings in the sheath by means of a pressure gradient, which is a material aid in cable maintenance.

Rural Radiotelephone Experiment at Cheyenne Wells, Colo.[11] J. Harold Moore, Paul K. Seyler and S. B. Wright. The first rural party-line telephone service by radio installations operating on the subscribers' premises was inaugurated August 20, 1946. This paper describes the equipment used, how it operates, and the results obtained during the preliminary testing and the initial period of regular operation. Radio is one of several new methods which the Bell System is exploring in its program for extension of telephone service in rural areas. It is expected that experience gained in

[8] *Electrical Engineering*, September 1948 (pp. 895–899).
[9] *Elec'l. Mfg.*, August 1948 (pp. 84–91).
[10] *Transactions, A. I. E. E.*, vol. 66, 1947 (pp. 471–478).
[11] *Trans. A. I. E. E.*, vol. 66, 1947 (pp. 525–528).

this experiment will aid in developing a standard rural radiotelephone system.

Effect of Passive Modes in Traveling-Wave Tubes.[12] J. R. PIERCE. As the beam current in a traveling-wave tube is increased, the local fields due to the bunched beam become appreciable compared with the fields propagating along the circuit. The effect is to reduce gain, to increase the electron speed for optimum gain, to introduce a lower limit to the range of electron speeds for which gain is obtained, and to change the initial loss.

New Test Equipment and Testing Methods for Cable Carrier Systems.[13] W. H. TIDD, S. ROSEN and H. A. WENK. Three portable test sets developed for the improved cable carrier telephone system are described: A high sensitivity selective transmission measuring set covering 10 to 150 kc, a decade oscillator for frequencies from 2 to 79 kc, and a tube test set.

A New Microwave Television System.[14] J. F. WENTZ and K. D. SMITH. A microwave point-to-point radio system is described which is designed for the transmission of television programs. This system is intended to supplement wire facilities for local distribution of television signals from pickup points to studios or from studios to broadcast transmitters and to long distance network terminals. The circuits and equipment are described in detail. Performance obtained in tests during 1946 is given.

[12] *Proc. I. R. E.*, August 1948 (pp. 993–997).
[13] *Trans. A. I. E. E.*, vol. 66, 1947 (pp. 726–730).
[14] *Transactions, A. I. E. E.*, vol. 66, 1947 (pp. 465–470).

Contributors to this Issue

W. J. ALBERSHEIM, Technical Colleges, Aachen and Munich, Germany; B.Sc., Aachen, 1920; E.E., Aachen, 1922; Doctor of Engineering, Aachen, 1924; Professional Engineer, University of State of New York, 1937. Electrical Research Products, Inc., 1929–41; Bell Telephone Laboratories, 1941–. Dr. Albersheim was concerned with radar and jamming work during the war; he is now engaged in broadband FM transmission problems.

W. B. HEBENSTREIT, A.A., University of Chicago, 1935; B.S., California Institute of Technology, 1941. Bell Telephone Laboratories, 1941–47. While at the Bell Telephone Laboratories Mr. Hebenstreit worked on various microwave vacuum tubes including magnetrons, traveling wave tubes and the new double stream amplifier. He is now with the Hughes Aircraft Company.

A. V. HOLLENBERG, A.B., Willamette University, 1931; M.S., 1933, Ph.D., 1938, New York University; Instructor in Physics, Queens College, 1938–42. Columbia Radiation Laboratory, 1942–45; Bell Telephone Laboratories, 1946–. Dr. Hollenberg was engaged in research and development work on microwave magnetrons at the Columbia Radiation Laboratory. At Bell Telephone Laboratories he has been concerned with traveling wave amplifiers.

J. R. PIERCE, B.S. in Electrical Engineering, California Institute of Technology, 1933; Ph.D., 1936. Bell Telephone Laboratories, 1936–. Engaged in study of vacuum tubes.

S. O. RICE, B.S. in Electrical Engineering, Oregon State College, 1929; California Institute of Technology, 1929–30, 1934–35. Bell Telephone Laboratories, 1930–. Mr. Rice has been concerned with various theoretical investigations relating to telephone transmission theory.

SLOAN D. ROBERTSON, B.E.E., University of Dayton, 1936; M.Sc., Ohio State University, 1938, Ph.D., 1941; Instructor of Electrical Engineering, University of Dayton, 1940. Bell Telephone Laboratories, 1940–. Dr. Robertson was engaged in microwave radar work in the Radio Research Department during the war. He is now engaged in fundamental microwave radio research.

CLAUDE E. SHANNON, B.S. in Electrical Engineering, University of Michigan, 1936; S.M. in Electrical Engineering and Ph.D. in Mathematics, M.I.T., 1940. National Research Fellow, 1940. Bell Telephone Laboratories, 1941–. Dr. Shannon has been engaged in mathematical research principally in the use of Boolean Algebra in switching, the theory of communication, and cryptography.

Contributors to this Issue

W. J. ALBERSHEIM, Technical Colleges, Aachen and Munich, Germany; B.Sc., Aachen, 1920; E.E., Aachen, 1922; Doctor of Engineering, Aachen, 1924; Professional Engineer, University of State of New York, 1937. Electrical Research Products, Inc., 1929–41; Bell Telephone Laboratories, 1941–. Dr. Albersheim was concerned with radar and jamming work during the war; he is now engaged in broadband FM transmission problems.

W. B. HEBENSTREIT, A.A., University of Chicago, 1935; B.S., California Institute of Technology, 1941. Bell Telephone Laboratories, 1941–47. While at the Bell Telephone Laboratories Mr. Hebenstreit worked on various microwave vacuum tubes including magnetrons, traveling wave tubes and the new double stream amplifier. He is now with the Hughes Aircraft Company.

A. V. HOLLENBERG, A.B., Willamette University, 1931; M.S., 1933, Ph.D., 1938, New York University; Instructor in Physics, Queens College, 1938–42. Columbia Radiation Laboratory, 1942–45; Bell Telephone Laboratories, 1946–. Dr. Hollenberg was engaged in research and development work on microwave magnetrons at the Columbia Radiation Laboratory. At Bell Telephone Laboratories he has been concerned with traveling wave amplifiers.

J. R. PIERCE, B.S. in Electrical Engineering, California Institute of Technology, 1933; Ph.D., 1936. Bell Telephone Laboratories, 1936–. Engaged in study of vacuum tubes.

S. O. RICE, B.S. in Electrical Engineering, Oregon State College, 1929; California Institute of Technology, 1929–30, 1934–35. Bell Telephone Laboratories, 1930–. Mr. Rice has been concerned with various theoretical investigations relating to telephone transmission theory.

SLOAN D. ROBERTSON, B.E.E., University of Dayton, 1936; M.Sc., Ohio State University, 1938, Ph.D., 1941; Instructor of Electrical Engineering, University of Dayton, 1940. Bell Telephone Laboratories, 1940–. Dr. Robertson was engaged in microwave radar work in the Radio Research Department during the war. He is now engaged in fundamental microwave radio research.

CLAUDE E. SHANNON, B.S. in Electrical Engineering, University of Michigan, 1936; S.M. in Electrical Engineering and Ph.D. in Mathematics, M.I.T., 1940. National Research Fellow, 1940. Bell Telephone Laboratories, 1941–. Dr. Shannon has been engaged in mathematical research principally in the use of Boolean Algebra in switching, the theory of communication, and cryptography.