

# The Bell System Technical Journal

Vol. XXX

July, 1951

No. 3

---

Copyright, 1951, American Telephone and Telegraph Company

---

## Reduction of Skin Effect Losses by the Use of Laminated Conductors

By A. M. CLOGSTON

It has recently been discovered that it is possible to reduce skin effect losses in transmission lines by properly laminating the conductors and adjusting the velocity of transmission of the waves. The theory for such laminated transmission lines is presented in the case of planar systems for both infinitesimally thin laminae and laminae of finite thickness. A transmission line completely filled with laminated material is discussed. An analysis is given of the modes of transmission in a laminated line, and of the problem of terminating such a line.

### I. INTRODUCTION

It has long been recognized that an electromagnetic wave propagating in the vicinity of an electrical conductor can penetrate only a limited distance into the interior of the material. This phenomenon is known as "skin effect" and is usually measured by a so-called "skin depth"  $\delta$ . If  $y$  is measured from the surface of a conductor into its depth, the amplitude of the electromagnetic wave and the accompanying current density decreases as  $e^{-y/\delta}$ , provided the conductor is several times  $\delta$  in thickness, so that for  $y = \delta$  the amplitude has fallen to  $1/e = 0.367$  times its value at the surface. The skin depth  $\delta$  is given by

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} \quad (\text{I-1})$$

where  $\sigma$  is the conductivity of the material,  $\mu$  is its permeability and  $\omega$  is  $2\pi$  times the frequency  $f$  under consideration. Throughout this paper rationalized MKS units are used.

From one point of view, skin effect serves a most useful purpose; for instance, in shielding electrical equipment or reducing crosstalk between communication circuits. On the other hand, the effect severely limits the high frequency performance of many types of electrical apparatus, including in particular the various kinds of transmission lines.

Surprisingly enough, it has been discovered that it is possible, within limits, to increase the distance to which an electromagnetic wave penetrates

into a conducting material. This is done essentially by fabricating the conductor of many insulated laminae or filaments of conducting material arranged parallel to the direction of current flow. If the transverse dimensions of the laminae or filaments are small compared to the skin depth  $\delta$  at the frequency under consideration, and if the velocity of the electromagnetic wave along the conductor is close to a certain critical value, the wave will penetrate into the composite conductor a distance great enough to include a thickness of conducting material many skin depths deep. Physically speaking, the lateral change of the wave through the conducting regions is very nearly cancelled by the change through the insulating regions.

In Fig. 1 there is shown a cross-section view of a coaxial cable with a

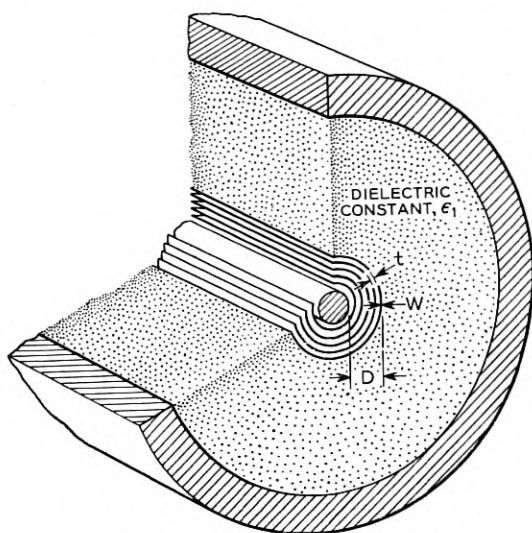


Fig. 1—Laminated transmission line.

laminated center conductor. The center conductor is formed of a non-conducting core surrounded by alternate layers of a conductor of thickness  $W$  and conductivity  $\sigma$ , and an insulator of thickness  $t$  and dielectric constant  $\epsilon$ . The center conductor is embedded in an insulator of dielectric constant  $\epsilon_1$  which is in turn encased in the outer conductor. We will assume all the conductors and insulators to have the permeability  $\mu_0$  of free space.

We will associate with the inner laminated conductor an average dielectric constant<sup>1</sup> for transverse electric fields given by

$$\bar{\epsilon} = \epsilon \left( 1 + \frac{W}{t} \right) \quad (\text{I-2})$$

<sup>1</sup> A similar average dielectric constant has been considered by Tokio Sakurai, *Journal of Physical Society of Japan*, Vol. 5, No. 6, pp. 394-398, Nov.-Dec. 1950.



It will be shown in the following sections that the electromagnetic wave and the accompanying currents will penetrate most deeply into the center conductor if the wave travels through the line with a velocity

$$v = \frac{1}{\sqrt{\epsilon\mu_0}} \quad (\text{I-3})$$

One way to make the wave assume this velocity is to let the dielectric constant  $\epsilon_1$  have the value

$$\epsilon_1 = \bar{\epsilon} = \epsilon \left( 1 + \frac{W}{t} \right) \quad (\text{I-4})$$

If the depth of the stack of laminations  $D$  is small compared to the distance between the stack and the outer conductor, and if the wave travels with the velocity given in equation (I-3), it will be shown that the wave decreases with distance into the center conductor as  $e^{-x/\delta_w}$  where  $\delta_w$  is given by

$$\delta_w = \sqrt{3} (1 + t/W)(\delta/W)\delta; \quad W \ll \delta \quad (\text{I-5})$$

Here  $\delta = \frac{1}{\sqrt{\pi f \mu_0 \sigma}}$  is the skin depth appropriate to the material of the conducting laminae and the frequency  $f$  under consideration. Let us now also associate with the center conductor an average longitudinal conductivity given by

$$\bar{\sigma} = \sigma \frac{W}{W + t} \quad (\text{I-6})$$

We will suppose for the present case that most of the attenuation of the transmission line results from the currents flowing in the inner conductor. It is easy to see that the attenuation of the line for very low frequencies will be  $A/\bar{\sigma}D$  where  $A$  is a constant depending on the impedance of the line. As the frequency increases,  $\delta_w$  decreases, and when  $\delta_w$  becomes several times smaller than  $D$  it will be shown that the attenuation becomes  $A/\bar{\sigma}\delta_w$ . At still higher frequencies  $\delta$  will similarly become several times smaller than  $W$ , and the attenuation then becomes  $A/\sigma\delta$ . From these considerations, a qualitative picture of the attenuation of the laminated line can be sketched as in Fig. 2.

For comparison, we have also sketched in Fig. 2 the attenuation that would be obtained if the laminations in Fig. 1 were replaced with solid metal. At low frequencies, the attenuation of this line would clearly be  $A/\sigma D$ . When the frequency becomes high enough for  $\delta$  to be several times smaller than  $D$  the attenuation will be shown to become  $A/\sigma\delta$ .

It will be observed how the attenuation of the unlaminated line remains

constant over a low range of frequencies and then rises at a rate proportional to the square root of the frequency. The laminated line has a higher initial attenuation, but remains constant to higher frequencies. At high enough frequencies the attenuation of the laminated line rises at a rate directly proportional to frequency for a while, and then eventually approaches the attenuation of the unlaminated line.

The frequency at which the attenuation of the laminated line begins to increase is greater than the corresponding frequency for the conventional line by a factor

$$\sqrt{3} \left( \frac{\sigma}{\bar{\sigma}} \right) \left( \frac{D}{W} \right)$$

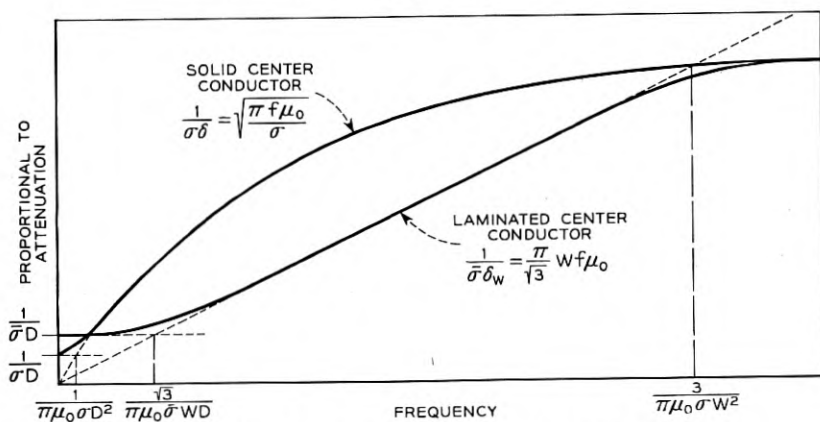


Fig. 2—Comparison of conventional and laminated transmission lines.

This is accomplished with an increase in initial attenuation by a factor

$$\sigma/\bar{\sigma} = \left( 1 + \frac{l}{W} \right)$$

which we will see later will be about 3/2 in a typical case. We might make a corresponding increase in the flat range of the conventional line by decreasing  $\sigma$  to a new value  $\sigma_1$ . In that case the attenuation would be increased by a factor

$$\sqrt{3} \left( \frac{\sigma}{\bar{\sigma}} \right) \left( \frac{D}{W} \right)$$

which may be very large since  $\left( \frac{\sigma}{\bar{\sigma}} \right) \left( \frac{D}{W} \right)$  is just the number of laminations of conductor or dielectric used on the center conductor.

The flat range of the conventional line might be alternatively increased to equal that of the laminated line by decreasing  $D$  to a new value  $D_1$ . In this case the attenuation would be increased by a factor

$$\sqrt{\sqrt{3} \left( \frac{\sigma}{\bar{\sigma}} \right) \left( \frac{D}{W} \right)}$$

just the square root of the factor achieved by changing  $\sigma$ , but still a large number.

In the frequency range in which the attenuation of the laminated line is governed by the skin depth  $\delta_w$ , and is therefore increasing linearly with frequency, this attenuation is less than the attenuation of the conventional line by a factor

$$\frac{\sigma \delta}{\bar{\sigma} \delta_w} = \frac{1}{\sqrt{3}} \left( \frac{W}{\delta} \right) \quad (\text{I-7})$$

It is interesting to note that the position of this region is governed by the conductivity of the conducting laminations, but that the attenuation is independent of the conductivity.

Considerable theoretical and experimental work has been carried out on laminated transmission lines by the author's colleagues. The following report therefore will be limited to bringing out some of the fundamental ideas in a simple way. We will for instance consider only planar systems so that the results will be only approximately applicable to real transmission lines. Other papers will more fully develop the formal theory, particularly for cylindrical systems, and discuss the practical and experimental aspects of the problem.

## II. SKIN EFFECT

We shall begin the discussion with this section by considering skin effect in various kinds of conducting media. We will first derive the skin depth equation (I-1) for an ordinary conductor like copper, and then discuss the behavior of a composite conductor made up of many thin, insulated conducting laminae. This second discussion will be first carried out for the case of infinitesimally thin laminae, and then in Section III the effects of the finite thickness of the conducting sheets will be considered.

Let us first set down and integrate Maxwell's equations in a form that will be useful in all our following discussions. Referring to the orthogonal coordinate system in Fig. 3 we shall be concerned with fields that have no variation along the  $z$ -axis and for which the  $z$ -component of electric field is zero. The only component of magnetic field is then  $H_z$  and the field equations become, in rationalized MKS units,

$$\frac{\partial H_z}{\partial y} = i\omega D_x + J_x, \quad (\text{II-1})$$

$$-\frac{\partial H_x}{\partial x} = i\omega D_y + J_y, \quad (\text{II-2})$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -i\omega B_z, \quad (\text{II-3})$$

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} = \rho \quad (\text{II-4})$$

In these equations  $H$ ,  $B$ ,  $D$ ,  $E$ ,  $J$  and  $\rho$  all have their usual meanings. A positive time factor  $e^{i\omega t}$  has been introduced.

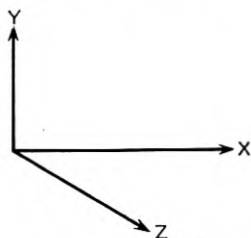


Fig. 3—Rectangular coordinate system.

Let us for the moment suppose that we are dealing with an anisotropic medium such that the following relations exist:

$$J_x = \sigma_x E_x; \quad J_y = \sigma_y E_y \quad (\text{II-5})$$

$$D_x = \epsilon_x E_x; \quad D_y = \epsilon_y E_y \quad (\text{II-6})$$

$$B_z = \mu_0 H_z \quad (\text{II-7})$$

Here the  $\sigma$ 's are conductivities, the  $\epsilon$ 's dielectric constants, and  $\mu_0$  is the permeability of free space. Suppose also now that the fields all vary with  $x$  according to a factor  $e^{-ikx}$ . If  $k$  has a positive real part we will be dealing with a wave moving along the  $x$ -axis in a positive direction, and a negative imaginary part will indicate that this wave is attenuated.

Using the above relations, one can easily find the following equations:

$$\frac{\partial^2 H_z}{\partial y^2} = \frac{i\omega\epsilon_x + \sigma_x}{i\omega\epsilon_y + \sigma_y} [i\omega\mu_0\sigma_y - \omega^2\mu_0\epsilon_y + k^2] H_z \quad (\text{II-8})$$

$$E_x = \frac{1}{i\omega\epsilon_x + \sigma_x} \frac{\partial H_z}{\partial y} \quad (\text{II-9})$$

$$E_y = \frac{ik}{i\omega\epsilon_y + \sigma_y} H_z \quad (\text{II-10})$$

Let us imagine that we have a semi-infinite volume of material arranged as shown in Fig. 4 where the  $z$ -axis is pointing out of the paper. If  $H_{z0}$  is the value of  $H_z$  at  $y = 0$ , it is clear from equation (16) that  $H_z$  must depend upon  $y$  according to

$$H_z = H_{z0} e^{-\alpha y} \quad (\text{II-11})$$

where

$$\alpha = \pm \sqrt{\frac{i\omega\epsilon_x + \sigma_x}{i\omega\epsilon_y + \sigma_y} [i\omega\mu_0\sigma_y - \omega^2\mu_0\epsilon_y + k^2]} \quad (\text{II-12})$$

and the sign is chosen so that the real part of  $\alpha$  is positive.

We can now consider the case when the material under consideration is an ordinary conductor such as copper or silver. In this case we must let

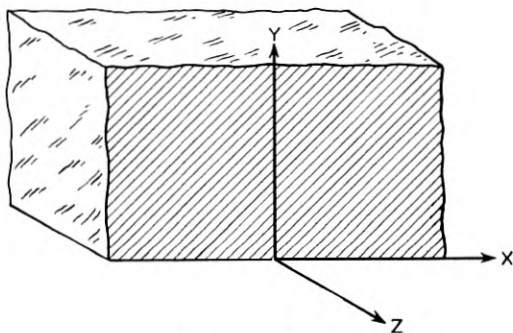


Fig. 4—Orientation of solid conductor.

$\sigma_x = \sigma_y = \sigma$  and  $\epsilon_x = \epsilon_y = \epsilon$ . Then  $\alpha$  becomes (the subscript  $S$  stands for solid)

$$\alpha_S = \pm \sqrt{i\omega\mu_0\sigma - \omega^2\mu_0\epsilon + k^2} \quad (\text{II-13})$$

Now, under any practical circumstances the propagation constant  $k$  will certainly not be more than a factor 100 larger than the propagation constant of free space  $k = \sqrt{\omega^2\mu_0\epsilon_0}$ . This applies also to the factor  $\sqrt{\omega^2\mu_0\epsilon}$ .

Consider then the ratio  $\frac{\omega\mu_0\sigma}{\omega^2\mu_0\epsilon_0} = \frac{\sigma}{\omega\epsilon_0}$ . For the metal copper, for instance,

$\sigma = 5.80 \times 10^7$  mhos/meter and the dielectric constant of free space  $\epsilon_0 = .885 \times 10^{-11}$ . If we consider frequencies as high as 10,000 megacycles the ratio is still as great as  $10^8$ . Thus, the second two factors under the square root sign in equation (II-13) are entirely negligible and we have

$$\alpha_S = \pm \sqrt{i\omega\mu_0\sigma} \quad (\text{II-14})$$

$$= (1 + i) \sqrt{\frac{\omega \mu_0 \sigma}{2}} \quad (\text{II-15})$$

Finally we have

$$\text{Real part } (\alpha_s) = \frac{1}{\delta} = \sqrt{\frac{\omega \mu_0 \sigma}{2}} \quad (\text{II-16})$$

We can now turn our attention to the stack of laminations shown in Fig. 5. There are shown a series of conducting sheets of conductivity  $\sigma$  and thickness  $W$ , separated by a series of insulating sheets of thickness  $t$  and dielectric constant  $\epsilon$ . Suppose we let  $W$  and  $t$  approach zero while maintaining a constant ratio to obtain a homogeneous but anisotropic material

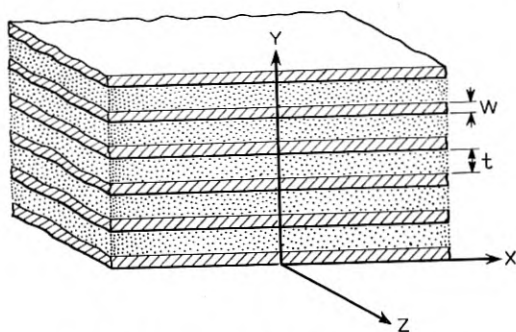


Fig. 5—Orientation of laminated conductor.

to which we can apply equation (II-12). In order to obtain  $\epsilon_y$ ,  $\sigma_y$ ,  $\epsilon_x$  and  $\sigma_x$  we can write

$$\frac{1}{\sigma_y + i\omega\epsilon_y} = \frac{1}{W + t} \left[ \frac{W}{\sigma} + \frac{t}{i\omega\epsilon} \right] \quad (\text{II-17})$$

and

$$\sigma_x + i\omega\epsilon_x = \frac{1}{W + t} [W\sigma + t i\omega\epsilon] \quad (\text{II-18})$$

One then has approximately, letting  $\epsilon_y = \bar{\epsilon}$  and  $\sigma_x = \bar{\sigma}$

$$\bar{\epsilon} = \epsilon \left( 1 + \frac{W}{t} \right) \quad (\text{II-19})$$

$$\bar{\sigma} = \sigma \left( \frac{W}{W + t} \right) \quad (\text{II-20})$$

$$\epsilon_x = \epsilon \left( \frac{t}{t + W} \right) \quad (\text{II-21})$$

$$\sigma_y = \sigma \left( \frac{\omega \epsilon}{\sigma} \right)^2 \frac{W}{l} \left( \frac{W}{l} + 1 \right) \quad (\text{II-22})$$

As before,  $\omega \epsilon_x$  is completely negligible compared to  $\bar{\sigma}$ , and furthermore  $\sigma_y$  is negligible compared to  $\omega \bar{\epsilon}$ . We have therefore for  $\alpha$ , (the subscript 0 stands for zero thickness)

$$\alpha_0 = \pm \sqrt{\frac{\bar{\sigma}}{i\omega\bar{\epsilon}} \left[ k^2 - \omega^2 \mu_0 \bar{\epsilon} + i\omega \mu_0 \sigma \left( \frac{\omega \epsilon}{\sigma} \right)^2 \frac{W}{l} \left( \frac{W}{l} + 1 \right) \right]} \quad (\text{II-23})$$

This situation is rather surprising. Let us suppose conditions are such that most of the energy of the wave is flowing in the region outside the stack of laminations. If this region is filled with an insulator of dielectric constant  $\epsilon_1$ ,  $k^2$  will be very nearly equal to  $\omega^2 \mu_0 \epsilon_1$ . Then  $\alpha_0$  will be given by

$$\alpha_0 = \pm \sqrt{\frac{1}{i} \bar{\sigma} \omega \mu_0 \left[ \left( \frac{\epsilon_1}{\bar{\epsilon}} - 1 \right) + i \left( \frac{\omega \epsilon}{\sigma} \right) \left( \frac{W}{l} \right) \right]} \quad (\text{II-24})$$

If  $\epsilon_1$  is made equal to  $\bar{\epsilon}$ ,  $\alpha_0$  becomes

$$\alpha_0 = \frac{W}{W + l} k \quad (\text{II-25})$$

$$= \frac{1}{1 + l/W} \frac{2\pi}{\lambda} \quad (\text{II-26})$$

where  $\lambda$  is the longitudinal wavelength. Thus, under these conditions the wave will penetrate into the laminations to a depth  $\frac{\lambda}{2\pi} \left( 1 + \frac{l}{W} \right)$  before it has decreased by a factor  $1/e$ . This distance is of course enormous compared to ordinary skin depth.

We will see in the next section that the finite thickness of the laminae limits the penetration of the waves for  $\epsilon_1 = \bar{\epsilon}$  to a distance much smaller than that implied in equation (II-26) but still large compared to conventional skin depths. In any case, we see in this simple way the suggestion of a method for obtaining great penetrations and consequently considerably reducing the attenuation of a transmission line.

The analysis of this section, carried out by assuming the medium to be anisotropic but homogeneous, can be given more physical meaning by examining a little more closely how the fields vary through the laminations shown in Fig. 5. From equations (II-8) and (II-9) one finds for a general case

$$\frac{\partial E_x}{\partial y} = \frac{1}{i\omega \epsilon_y + \sigma_y} [i\omega \mu_0 \sigma_y - \omega^2 \mu_0 \epsilon_y + k^2] H_z \quad (\text{II-27})$$



Let the value of  $H_z$  at the interface of a conducting lamina and a dielectric lamina be  $(H_z)_0$ . From equation (II-27), one finds just within the conducting lamina

$$\frac{\partial E_x}{\partial y} = i\omega\mu_0(H_z)_0, \quad (\text{II-28})$$

while just with the adjacent dielectric lamina

$$\frac{\partial E_x}{\partial y} = i\omega\mu_0 \left[ 1 - \frac{\epsilon_1}{\epsilon} \right] (H_z)_0. \quad (\text{II-29})$$

Thus, if the laminae are very thin, the change in  $E_x$  across the conducting lamina is

$$(\Delta E_x)_m = i\omega\mu_0(H_z)_0 W \quad (\text{II-30})$$

and the change in  $E_x$  across the dielectric lamina is

$$(\Delta E_x)_d = i\omega\mu_0(H_z)_0 \left( 1 - \frac{\epsilon_1}{\epsilon} \right) t \quad (\text{II-31})$$

Therefore, when  $\epsilon_1 = \epsilon \left( 1 + \frac{W}{t} \right)$ , the change in  $E_x$  across the conducting lamina is just balanced by the change in  $E_x$  across the dielectric lamina. This is the basic reason for the deep penetration of the fields into the laminated structure. When  $\epsilon_1 = \epsilon$ , there is no change in  $E_x$  across the dielectric lamina. In this case we note from equation (II-24) that  $\alpha_0 = \frac{W}{W+t} \alpha_s$ , and we see that the waves will penetrate into the laminae an increased distance that is just accounted for by the spacing of the laminae. Thus, for  $\epsilon_1 = \epsilon$ , the attenuation of a laminated line will be unchanged if the laminae are replaced by solid metal.

### III. LAMINATIONS OF FINITE THICKNESS

Let us refer again to Fig. 5 where a stack of conducting laminae of thickness  $W$  and conductivity  $\sigma$  are shown separated by insulating laminae of thickness  $t$  and dielectric constant  $\epsilon$ . First we shall inquire as to how the fields change across the conducting laminations. According to equations (II-8) and (II-9) one has

$$\frac{\partial^2 H_z}{\partial y^2} = i\omega\mu_0 \sigma H_z \quad (\text{III-1})$$

$$E_x = \frac{1}{\sigma} \frac{\partial H_z}{\partial y} \quad (\text{III-2})$$

where  $\omega\epsilon$  has been neglected against  $\sigma$  as before. Now, letting  $\eta = \sqrt{i\omega\mu_0\sigma}$ , we can write within any conducting lamina

$$H_z = Ae^{\eta y} + Be^{-\eta y} \quad (\text{III-3})$$

$$E_x = \frac{\eta}{\sigma} [Ae^{\eta y} - Be^{-\eta y}] \quad (\text{III-4})$$

If  $H_0$  and  $E_0$  are the values of  $H_z$  and  $E_x$  at the lower surface of a particular lamination, and if  $H_1$  and  $E_1$  are their values at the upper surface of the lamination, one can find from equations (III-3) and (III-4)

$$H_1 = H_0 \cosh \eta W + E_0 \frac{\sigma}{\eta} \sinh \eta W \quad (\text{III-5})$$

$$E_1 = H_0 \frac{\eta}{\sigma} \sinh \eta W + E_0 \cosh \eta W \quad (\text{III-6})$$

If we wish, this can be expressed as a matrix equation

$$\begin{Bmatrix} H_1 \\ E_1 \end{Bmatrix} = \begin{Bmatrix} \cosh \eta W & \frac{\sigma}{\eta} \sinh \eta W \\ \frac{\eta}{\sigma} \sinh \eta W & \cosh \eta W \end{Bmatrix} \begin{Bmatrix} H_0 \\ E_0 \end{Bmatrix} \quad (\text{III-7})$$

For the dielectric laminae, equations (II-8) and (II-9) become

$$\frac{\partial^2 H_z}{\partial y^2} = (k^2 - \omega^2 \mu_0 \epsilon) H_z \quad (\text{III-8})$$

$$E_x = \frac{1}{i\omega\epsilon} \frac{\partial H_z}{\partial y} \quad (\text{III-9})$$

Just as for the conducting laminae, let  $H_1$  and  $E_1$  be the values of  $H_z$  and  $E_x$  at the lower surface of a dielectric lamination and let  $H_2$  and  $E_2$  be these values at the top surface. Then, if  $\xi = \sqrt{k^2 - \omega^2 \mu_0 \epsilon}$ , one has

$$\begin{Bmatrix} H_2 \\ E_2 \end{Bmatrix} = \begin{Bmatrix} \cosh \xi l & \frac{i\omega\epsilon}{\xi} \sinh \xi l \\ \frac{\xi}{i\omega\epsilon} \sinh \xi l & \cosh \xi l \end{Bmatrix} \begin{Bmatrix} H_1 \\ E_1 \end{Bmatrix} \quad (\text{III-10})$$

From equations (III-7) and (III-10), we can find the variation of  $H_z$  and  $E_x$  from the bottom surface of a conducting lamination designated as point zero, to the top surface of the adjacent dielectric lamination designated as point two. Thus,

$$\begin{Bmatrix} H_2 \\ E_2 \end{Bmatrix} = \begin{Bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{Bmatrix} \begin{Bmatrix} H_0 \\ E_0 \end{Bmatrix} \quad (\text{III-11})$$

where the  $T$ 's are given below

$$T_{11} = \cosh \eta W \cosh \xi l + \frac{i\omega\epsilon}{\sigma} \frac{\eta}{\xi} \sinh \eta W \sinh \xi l \quad (\text{III-12})$$

$$T_{12} = \frac{\sigma}{\eta} \sinh \eta W \cosh \xi l + \frac{i\omega\epsilon}{\xi} \cosh \eta W \sinh \xi l \quad (\text{III-13})$$

$$T_{21} = \frac{\xi}{i\omega\epsilon} \cosh \eta W \sinh \xi l + \frac{\eta}{\sigma} \sinh \eta W \cosh \xi l \quad (\text{III-14})$$

$$T_{22} = \frac{\sigma}{i\omega\epsilon} \frac{\xi}{\eta} \sinh \eta W \sinh \xi l + \cosh \eta W \cosh \xi l \quad (\text{III-15})$$

It is easy to verify from the above that

$$T_{11}T_{22} - T_{12}T_{21} = 1 \quad (\text{III-16})$$

If we now designate the lower surface of each conducting lamination successively as points 0, 1, 2, 3,  $\dots$ , we can write down the following simultaneous difference equations

$$H_{n+1} = T_{11}H_n + T_{12}E_n \quad (\text{III-17})$$

$$E_{n+1} = T_{21}H_n + T_{22}E_n \quad (\text{III-18})$$

The solutions of these difference equations are

$$H_n = A\beta^n + B\beta^{-n} \quad (\text{III-19})$$

$$E_n = A \frac{\beta - T_{11}}{T_{12}} \beta^n + B \frac{1/\beta - T_{11}}{T_{12}} \beta^{-n} \quad (\text{III-20})$$

where

$$\beta = \left( \frac{T_{11} + T_{22}}{2} \right) + \sqrt{\left( \frac{T_{11} + T_{22}}{2} \right)^2 - 1} \quad (\text{III-21})$$

Let us now proceed to determine the skin depth to be associated with the stack of laminae in Fig. 5. Since we have assumed the stack to be very deep,  $A$  must be taken zero in equations (III-19) and (III-20), and the fields vary into the stack according to a factor  $\beta^{-n}$ , so that

$$H_n = H_0\beta^{-n} \quad (\text{III-22})$$

If we now define

$$y_n = (W + l)n \quad (\text{III-23})$$

one has

$$\begin{aligned} H_n &= H_0 \beta^{-(y_n/(W+t))} \\ &= H_0 e^{-(1n\beta/(W+t))y_n} \\ &= H_0 e^{-\alpha_w y_n} \end{aligned} \quad (\text{III-24})$$

where (the subscript  $w$  indicates a thickness  $W$  for the conducting laminae)

$$\alpha_w = \frac{\cosh^{-1} \left( \frac{T_{11} + T_{22}}{2} \right)}{(W + t)} \quad (\text{III-25})$$

From equations (III-12) and (III-15) one has

$$\begin{aligned} &\left( \frac{T_{11} + T_{22}}{2} \right) \\ &= \cosh \eta W \cosh \xi t + \frac{1}{2} \left( \frac{i\omega\epsilon}{\sigma} \frac{\eta}{\xi} + \frac{\sigma}{i\omega\epsilon} \frac{\xi}{\eta} \right) \sinh \eta W \sinh \xi t \end{aligned} \quad (\text{III-26})$$

As a practical matter, only rarely will  $k^2$  be greater than ten times  $\omega^2\mu_0\epsilon$  and  $\epsilon$  greater than  $10\epsilon_0$ . Hence,  $\xi$  will be no larger than  $10\sqrt{\omega^2\mu_0\epsilon_0}$ . Furthermore, we will see that  $t$  should not be very different in magnitude from  $W$ , which must be smaller than  $\sqrt{\frac{2}{\omega\mu_0\sigma}}$ . Thus, we can be sure that  $\xi t$  is smaller than  $100\sqrt{\frac{2\omega\epsilon_0}{\sigma}}$ , a quantity which is, as before, much smaller than unity. Under these conditions, equation (III-26) becomes approximately

$$\left( \frac{T_{11} + T_{22}}{2} \right) = \cosh \eta W + \frac{\xi t}{2} \left( \frac{i\omega\epsilon}{\sigma} \frac{\eta}{\xi} + \frac{\sigma}{i\omega\epsilon} \frac{\xi}{\eta} \right) \sinh \eta W \quad (\text{III-27})$$

$$= \cosh \eta W + \frac{1}{2} \left( \frac{t}{W} \right) \left[ \frac{i\omega\epsilon}{\sigma} + \left( 1 - \frac{\epsilon_1}{\epsilon} \right) \right] \eta W \sinh \eta W \quad (\text{III-28})$$

where we have again let  $k^2 = \omega^2\mu_0\epsilon_1$  as in equation (II-24).

Let us set

$$P = \frac{1}{2} \left( \frac{t}{W} \right) \left( 1 - \frac{\epsilon_1}{\epsilon} \right) \quad (\text{III-29})$$

$$= \frac{1}{2} \left[ -1 + \left( 1 + \frac{t}{W} \right) \left( 1 - \frac{\epsilon_1}{\epsilon} \right) \right] \quad (\text{III-30})$$

Then, again neglecting  $\frac{\omega\epsilon}{\sigma}$ , we have for  $\alpha_w$

$$\alpha_w = \frac{1}{W + t} \cosh^{-1} [\cosh \eta W + P(\eta W) \sinh \eta W] \quad (\text{III-31})$$

By definition,  $\eta W = (1 + i) \frac{W}{\delta}$ . We can therefore write approximately,

$$\cosh \eta W \simeq \left[ 1 - \frac{1}{6} \left( \frac{W}{\delta} \right)^4 \right] + i \left( \frac{W}{\delta} \right)^2 \quad (\text{III-32})$$

$$(\eta W) \sinh \eta W \simeq -\frac{2}{3} \left( \frac{W}{\delta} \right)^4 + i2 \left( \frac{W}{\delta} \right)^2 \quad (\text{III-33})$$

Using expressions (III-32) and (III-33), equation (III-31) becomes

$$\alpha_w = \frac{1}{W + i} \cosh^{-1} \cdot \left( \left[ 1 - \frac{(1 + 4P)}{6} \left( \frac{W}{\delta} \right)^4 \right] + i(1 + 2P) \left( \frac{W}{\delta} \right)^2 \right) \quad (\text{III-34})$$

Provided that  $(1 + 2P) \left( \frac{W}{\delta} \right)^2 \ll 1$ , equation (III-34) can be expanded in orders of magnitude. Thus, for

$$\left| 1 - \frac{\epsilon_1}{\epsilon} \right| \ll \frac{W}{W + i} \left( \frac{\delta}{W} \right)^2 \quad (\text{III-35})$$

we find approximately

$$\alpha_w = \frac{1}{(W + i)} \left( \frac{W}{\delta} \right) \cdot \left( \sqrt{-f + \sqrt{f^2 + g^2}} \pm i \sqrt{f + \sqrt{f^2 + g^2}} \right) \quad (\text{III-36})$$

where

$$f = \frac{1 + 4P}{6} \left( \frac{W}{\delta} \right)^2 \quad (\text{III-37})$$

and

$$g = 1 + 2P \quad (\text{III-38})$$

The plus sign is to be used when  $g$  is positive and the minus sign when  $g$  is negative.

Equations (III-36) for  $\alpha_w$  and (II-24) for  $\alpha_0$  are very similar. With a little manipulation we can rewrite equation (II-24) as

$$\alpha_0 = \frac{1}{\delta} \sqrt{\frac{\delta}{\sigma}} \left( \sqrt{\left( \frac{\omega \epsilon}{\sigma} \right) \left( \frac{W}{l} \right)} + \sqrt{\left( \frac{\omega \epsilon}{\sigma} \right)^2 \left( \frac{W}{l} \right)^2 + \left( 1 - \frac{\epsilon_1}{\epsilon} \right)^2} \right) \pm i \sqrt{-\left( \frac{\omega \epsilon}{\sigma} \right) \left( \frac{W}{l} \right) + \sqrt{\left( \frac{\omega \epsilon}{\sigma} \right)^2 \left( \frac{W}{l} \right)^2 + \left( 1 - \frac{\epsilon_1}{\epsilon} \right)^2}} \quad (\text{III-39})$$

Also equation (III-36) can be written

$$\alpha_w = \frac{1}{\delta} \sqrt{\frac{\bar{\sigma}}{\sigma}} \left( \sqrt{-\left(\frac{W}{W+t} f\right)} + \sqrt{\left(\frac{W}{W+t} f\right)^2 + \left(1 - \frac{\epsilon_1}{\bar{\epsilon}}\right)^2} \right) \pm i \sqrt{\left(\frac{W}{W+t} f\right)} + \sqrt{\left(\frac{W}{W+t} f\right)^2 + \left(1 - \frac{\epsilon_1}{\bar{\epsilon}}\right)^2} \quad (\text{III-40})$$

Equation (III-39) is a very good approximation for  $\alpha$  for a stack of infinitesimally thin laminae, but is inadequate for the discussion of situations where the finite value of  $W/\delta$  is important. Equation (III-40) on the other hand is a good approximation to  $\alpha$  when  $W/\delta$  is appreciable, but the assumptions made in deriving (III-40) do not allow us to go correctly to the limit  $W/\delta = 0$ .

To estimate how small  $W$  must be before equation (III-40) fails, let us set  $\left(-\frac{W}{W+t} f\right)$  equal to  $\left(\frac{\omega\epsilon}{\sigma}\right)\left(\frac{W}{t}\right)$ . We will also set  $\epsilon_1 = \bar{\epsilon}$ , since it is only then that these terms are important. In this case,

$$W = \frac{1}{\sigma} \sqrt{12 \frac{\bar{\epsilon}}{\mu_0}} \quad (\text{III-41})$$

If we take  $\bar{\epsilon}$  to be 5 times the dielectric constant  $\epsilon_0$  of free space, and take  $\sigma = 5.8 \times 10^7$  mhos/meter for copper, we obtain  $W = 3.5 \times 10^{-8}$  cm. Thus, for any practical purpose we can ignore the failure of equation (III-40) to have the proper limit for  $W/\delta = 0$ .

By definition, the fields decrease into the stack according to  $e^{-\alpha_w y}$ . Let us define the distance at which the fields have decreased by  $1/e$  to be the effective skin depth  $\delta_w$ . Then we have

$$\frac{1}{\delta_w} = \text{Real part } (\alpha_w) \\ = \frac{1}{\delta} \sqrt{\frac{\bar{\sigma}}{\sigma}} \sqrt{-\left(\frac{W}{W+t} f\right)} + \sqrt{\left(\frac{W}{W+t} f\right)^2 + \left(1 - \frac{\epsilon_1}{\bar{\epsilon}}\right)^2} \quad (\text{III-42})$$

From equations (III-30) and (III-37) we find

$$-\left(\frac{W}{W+t} f\right) = \frac{1}{\delta} \left[ \frac{W}{W+t} + 2\left(\frac{\epsilon_1}{\bar{\epsilon}} - 1\right) \right] \left(\frac{W}{\delta}\right)^2 \quad (\text{III-43})$$

For  $\epsilon_1 = \bar{\epsilon}$  equations (III-43) and (III-42) give us finally

$$\delta_w(\epsilon_1 = \bar{\epsilon}) = \sqrt{3} \left(1 + \frac{t}{W}\right) \left(\frac{\delta}{W}\right) \delta \quad (\text{III-44})$$

With the results obtained in Sections II and III, we can compare the curves of attenuation as a function of frequency for conventional and laminated lines. Let us consider a transmission line such as that shown in Fig. 1, where we may imagine the center conductor to be either laminated as shown or made of solid metal. Let us suppose, as in the introduction, that most of the power loss is in the center conductor and that the distance between the stack and outer conductor is large compared to the depth of the stack. Clearly, the attenuation of the line will be proportional to the power per unit area flowing into the center conductor for a given power flow in the line. If  $H_z$  is the transverse magnetic field and  $E_x$  is the longitudinal electric field, the power flow into the center conductor per unit area will be given by

$$\frac{1}{2} |H_z| \cdot |E_x| \cos \Phi = \frac{1}{2} \text{Real part } (H_z \cdot E_x^*) \quad (\text{III-45})$$

where  $\Phi$  is the phase angle between  $H_z$  and  $E_x$  and (\*) indicates the conjugate of a complex quantity. If  $C$  is the circumference of the center conductor, and  $Z$  is the characteristic impedance of the line, the attenuation of the line will be given by

$$\gamma = \frac{1}{2CZ |H_z|^2} \text{Real part } (H_z \cdot E_x^*) \quad (\text{III-46})$$

First, let us suppose that the inner conductor is solid and that the frequency is very low. In that case, the uniform current density in the metal will be  $H_z/D$ , and therefore  $E_x$  will be  $H_z/\sigma D$ . Hence, for this case the attenuation will be

$$\gamma_s(f \text{ small}) = \frac{1}{2CZ\sigma D} \quad (\text{III-47})$$

In a similar manner, the attenuation of the line when the center conductor is laminated will be for very low frequencies

$$\gamma_w(f \text{ small}) = \frac{1}{2CZ\bar{\sigma}D} \quad (\text{III-48})$$

Next, let us consider the solid conductor again but for frequencies where  $\delta \ll D$ . Then we have from equations (II-9) and (II-15)

$$E_x = - \frac{1 + i}{\sigma \delta} H_z \quad (\text{III-49})$$

Hence,

$$\gamma_s = \frac{1}{2CZ\sigma\delta} \quad (\text{III-50})$$



Finally, we desire the attenuation  $\gamma_w$  of the laminated line at elevated frequencies. Therefore we need the relation between  $E_x$  and  $H_x$  at the first surface of the stack which, for definiteness, we can take to be a metal lamina. For a sufficiently deep stack, this ratio is the same at each successive corresponding point. Referring to equations (III-7) and (III-10) we wish to find

$$R = \frac{E_0}{H_0} = \frac{E_2}{H_2} \quad (\text{III-51})$$

By eliminating among these equations and using the same approximations made previously, we obtain

$$R = \frac{1}{\sigma W} \frac{\eta W}{\sinh \eta W} \frac{1 - \beta \cosh \eta W}{\beta} \quad (\text{III-52})$$

If use is now made of the relation  $\beta = e^{\alpha_w(W+t)}$ , one finds to first order

$$R = -\frac{\alpha_w}{\bar{\sigma}} \quad (\text{III-53})$$

Therefore, for the attenuation one has,

$$\begin{aligned} \gamma_w &= \frac{1}{2CZ\bar{\sigma}} \text{Real part } (\alpha_w^*) \\ &= \frac{1}{2CZ\bar{\sigma}\delta_w} \end{aligned} \quad (\text{III-54})$$

which is equivalent to an expression used in the introduction with  $1/2CZ$  being the value of the constant A.

We have compared the attenuation of conventional and laminated lines as a function of frequency for an insulator between inner and outer conductors having the critical dielectric constant  $\epsilon_1 = \bar{\epsilon}$ . It is also of interest to draw the comparison as a function of dielectric constant  $\epsilon_1$  at a fixed frequency. The ratio of the two attenuations will be

$$\frac{\gamma_w}{\gamma_s} = \frac{\sigma\delta}{\bar{\sigma}\delta_w} \quad (\text{III-55})$$

This curve is drawn in Fig. 6 for  $W/\delta = 1/3$  and  $t/W = 1$ .

For  $\epsilon_1 = \bar{\epsilon}$  we have,

$$\frac{\gamma_w}{\gamma_s} = \frac{1}{\sqrt{3}} \left( \frac{W}{\delta} \right) \quad (\text{III-56})$$

It will also be observed that for  $\epsilon_1 = \epsilon$

$$\frac{\gamma_w}{\gamma_s} = 1 \tag{III-57}$$

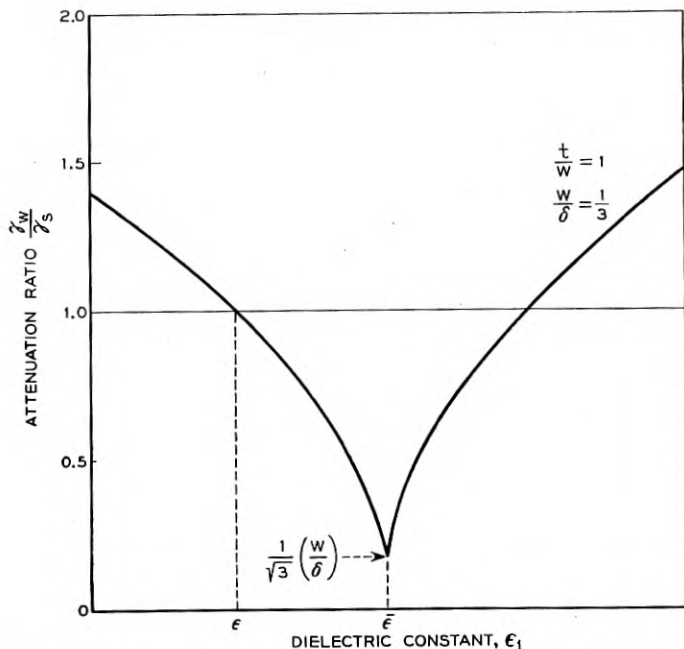


Fig. 6—Relative attenuation as a function of dielectric constant of material between stack and outer conductor.

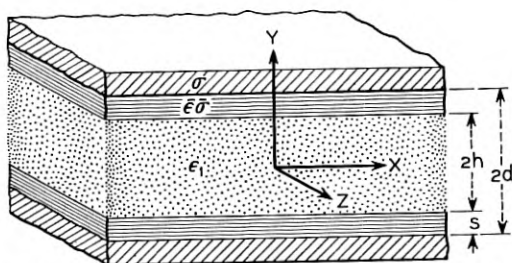


Fig. 7—Plane parallel transmission line with laminated conductors.

#### IV. TRANSMISSION LINE WITH LAMINATED CONDUCTORS

In the preceding sections, we have considered the case of a transmission line with a depth of laminations small compared to the spacing of the con-

ductors yet large compared to the effective skin depth  $\delta_w$ . In the following sections, we will deal with several situations in which the stacks are much smaller in depth than  $\delta_w$ . The size of the stack, then, reflected in the imaginary part of the propagation constant  $k$  has more effect on  $\alpha$  than anything else and we may consider  $W/\delta$  and, all the more,  $\sigma_y$  to be zero.

Under these conditions, we shall calculate the attenuation of the parallel plane transmission line shown in Fig. 7. In that figure we have two parallel plates or shields of conductivity  $\sigma$  separated a distance  $2d$ . Inside each plate there is a thickness  $s$  of laminated conductor of average conductivity  $\bar{\sigma}$  and average transverse dielectric constant  $\bar{\epsilon}$ . The interior of the line is filled with a dielectric of thickness  $2h = 2(d - s)$  and having a dielectric constant  $\epsilon_1$ . The calculations to be made will be valid down to some low frequency at which the skin depth in the outer shields becomes equal to their thickness.

With reference to equations (II-8) and (II-9), we can write down the following expressions for the fields in the various parts of the line:

$$\text{In shield:} \quad H_z = A e^{-\eta(y-d)} \quad (\text{IV-1})$$

$$E_x = -A \frac{\eta}{\sigma} e^{-\eta(y-d)} \quad (\text{IV-2})$$

$$\eta = \sqrt{i\omega\mu_0\sigma} \quad (\text{IV-3})$$

$$\text{In laminae:} \quad H_z = B \cosh \zeta(y - d) + C \sinh \zeta(y - d) \quad (\text{IV-4})$$

$$E_x = \frac{\zeta}{\bar{\sigma}} [B \sinh \zeta(y - d) + C \cosh \zeta(y - d)] \quad (\text{IV-5})$$

$$\zeta = \sqrt{\frac{\bar{\sigma}}{i\omega\bar{\epsilon}} (k^2 - \omega^2\mu_0\bar{\epsilon})} \quad (\text{IV-6})$$

$$\text{In dielectric:} \quad H_z = \cosh \xi y \quad (\text{IV-7})$$

$$E_x = \frac{\xi}{i\omega\epsilon_1} \sinh \xi y \quad (\text{IV-8})$$

$$\xi = \sqrt{k^2 - \omega^2\mu_0\epsilon_1} \quad (\text{IV-9})$$

where  $A$ ,  $B$  and  $C$  are constants.

The fields  $H_z$  and  $E_x$  must match at the boundaries  $y = h$  and  $y = d$ . Imposing these conditions, we find the characteristic equation for determining  $k$  to be

$$\tanh \zeta s = - \frac{1 + \left(\frac{\sigma}{i\omega\epsilon_1}\right) \frac{\xi}{\eta} \tanh \xi h}{\left(\frac{\sigma}{i\omega\epsilon_1}\right) \left(\frac{\bar{\sigma}}{\sigma}\right) \frac{\xi}{\zeta} \tanh \xi h + \left(\frac{\sigma}{\bar{\sigma}}\right) \frac{\zeta}{\eta}} \quad (\text{IV-10})$$

The constants are also determined to be

$$B = \frac{\cosh \xi h}{\cosh \zeta s + \frac{\bar{\sigma}}{\sigma} \frac{\eta}{\zeta} \sinh \zeta s} \quad (\text{IV-11})$$

$$A = B$$

$$C = -\frac{\bar{\sigma}}{\sigma} \frac{\eta}{\zeta} B \quad (\text{IV-12})$$

Just as before, it is obvious from equation (IV-6) that  $k$  must be nearly equal to  $\omega^2 \mu_0 \bar{\epsilon}$  if the fields are to penetrate deeply into the laminations. Let us guess on physical grounds that this can be accomplished by setting  $\epsilon_1 = \bar{\epsilon}$ . Under these conditions from equation (IV-9) and (IV-6)

$$\xi = \sqrt{\frac{i\omega\bar{\epsilon}}{\bar{\sigma}}} \zeta \quad (\text{IV-13})$$

and

$$\xi h = \sqrt{\frac{i\omega\bar{\epsilon}}{\bar{\sigma}}} \left(\frac{h}{s}\right) (\zeta s)$$

Thus, if  $(\zeta s)$  is not to be very large and  $\left(\frac{h}{s}\right)$  is not greater than 100,  $\xi h$  will be a very small number. We can therefore set  $\tanh \xi h = \xi h$  and rewrite equation (IV-10) as

$$(\zeta s) \tanh(\zeta s) + \frac{\left(\frac{h}{s}\right) (\zeta s)^2 + \left(\frac{\bar{\sigma}}{\sigma}\right) \eta s}{1 + \left(\frac{\bar{\sigma}}{\sigma}\right) \eta h} = 0 \quad (\text{IV-14})$$

From equation (IV-3),  $\eta h = (1 + i) \frac{h}{\delta}$ . We shall imagine that  $h$  is many times greater than the skin depth in the outer shield and may therefore reduce equation (IV-14) to

$$\theta \tanh \theta + \frac{1 - i}{2} \left(\frac{\delta}{s}\right) \left(\frac{\sigma}{\bar{\sigma}}\right) \theta^2 + \frac{s}{h} = 0 \quad (\text{IV-15})$$

where  $(\zeta s)$  has been set equal to  $\theta$ . Now, if  $s$  is also many times the skin depth in the outer conductor, the second term in equation (IV-15) may be

neglected compared to the first term. We have then, finally, for the characteristic equation

$$\theta \tanh \theta + \frac{s}{h} = 0 \quad (\text{IV-16})$$

For  $s$  much smaller than  $h$ , approximate solutions of equation (IV-16) may be written

$$\theta_0^2 = -\frac{s}{h} \quad (\text{IV-17})$$

and

$$\theta_n = in\pi \left[ 1 + \frac{s}{h} \frac{1}{(n\pi)^2} \right] \quad n = 1, 2, 3, \dots \quad (\text{IV-18})$$

The fundamental solution (IV-17) obviously agrees with our assumption that  $(\delta/s)$  is not large. Similarly, the higher order solutions (IV-18) are consistent with that assumption if  $n$  is not taken too large.

We may now return to equation (IV-6) and obtain approximate expressions for  $k$ .

$$k_0 = \sqrt{\omega^2 \mu_0 \bar{\epsilon}} \left[ 1 + \frac{1}{i\omega \mu_0 \bar{\sigma}} \frac{1}{2sh} \right] \quad (\text{IV-19})$$

$$k_n = \sqrt{\omega^2 \mu_0 \bar{\epsilon}} \left( 1 + \frac{1}{i\omega \mu_0 \bar{\sigma}} \left[ \frac{1}{sh} + \frac{1}{2} \left( \frac{n\pi}{s} \right)^2 \right] \right) \quad n = 1, 2, 3, \dots \quad (\text{IV-20})$$

We see that  $k^2$  is indeed approximately equal to  $\omega^2 \mu_0 \bar{\epsilon}$ . The imaginary parts of (IV-19) and (IV-20) are negative and give us the desired attenuation for the fundamental and higher modes of the line in nepers per meter.

$$I(k_0) = -\frac{1}{2sh} \frac{1}{\bar{\sigma}} \sqrt{\frac{\bar{\epsilon}}{\mu_0}} \quad (\text{IV-21})$$

$$I(k_n) = -\frac{1}{2sh} \left[ 2 + \frac{h}{s} (n\pi)^2 \right] \frac{1}{\bar{\sigma}} \sqrt{\frac{\bar{\epsilon}}{\mu_0}} \quad n = 1, 2, 3, \dots \quad (\text{IV-22})$$

where we have assumed  $\delta \ll s \ll h$  and  $\epsilon_1 = \bar{\epsilon}$ .

Let us first comment on the fact that there exist several modes of transmission in this line. The fundamental mode with propagation constant  $k_0$  corresponds to the ordinary mode of transmission that would exist between a pair of parallel plates such as shown in Fig. 8. The higher modes are waves that are confined almost entirely to the laminations and are not encountered in an ordinary transmission line. These modes will be more fully discussed in Section VI.

For comparison with equations (IV-21) and (IV-22), we shall next calculate the attenuation of the parallel plate transmission line shown in Fig. 8.

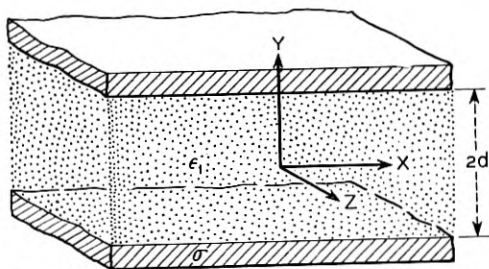


Fig. 8—Plane parallel transmission line with solid conductors.

For the fields we have again

In shield:

$$H_z = A e^{-\eta(y-d)} \quad (\text{IV-23})$$

$$E_x = -A \frac{\eta}{\sigma} e^{-\eta(y-d)} \quad (\text{IV-24})$$

$$\eta = \sqrt{i\omega\mu_0\sigma} \quad (\text{IV-25})$$

In dielectric:

$$H_z = \cosh \xi y \quad (\text{IV-26})$$

$$E_x = \frac{\xi}{i\omega\epsilon_1} \sinh \xi y \quad (\text{IV-27})$$

$$\xi = \sqrt{k^2 - \omega^2\mu_0\epsilon_1} \quad (\text{IV-28})$$

By matching the fields at  $y = d$  we readily obtain the characteristic equation

$$(\xi d) \tanh(\xi d) + \frac{i\omega\epsilon_1}{\sigma} \eta d = 0 \quad (\text{IV-29})$$

which gives approximately

$$(\xi d)^2 = -\frac{i\omega\epsilon_1}{\sigma} \eta d \quad (\text{IV-30})$$

We also determine the constant  $A$  to be unity. Proceeding as before, we find for the propagation constant

$$k = \sqrt{\omega^2\mu_0\epsilon_1} \left[ 1 + \frac{1}{2d} \frac{1}{\sqrt{i\omega\mu_0\sigma}} \right] \quad (\text{IV-31})$$

and for the attenuation

$$I(k) = -\frac{1}{d} \sqrt{\frac{1}{8} \frac{\omega \epsilon_1}{\sigma}} \quad (\text{IV-32})$$

It is observed at once that the attenuation expressed in equation (IV-21) is independent of frequency, while that given in equation (IV-32) increases

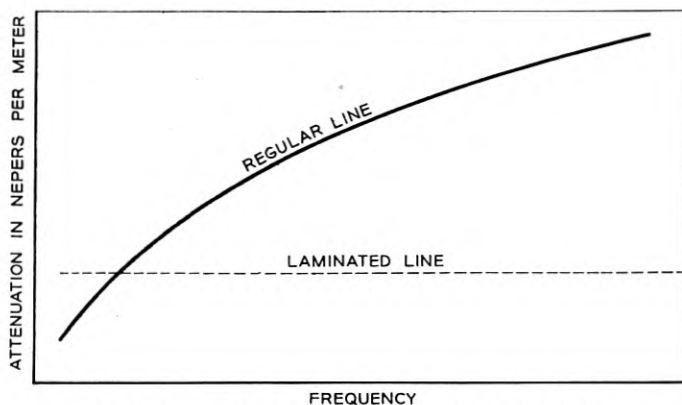


Fig. 9—Comparison of conventional and laminated lines.

as the square root of the frequency. If we take the ratio of equation (IV-32) to equation (IV-21) we obtain

$$\frac{\text{attenuation of regular line}}{\text{attenuation of laminated line}} = \left(\frac{h}{d}\right) \left(\frac{\bar{\sigma}}{\sigma}\right) s \sqrt{\frac{\omega \mu_0 \sigma}{2}} \quad (\text{IV-33})$$

$$= \left(\frac{h}{d}\right) \left(\frac{\bar{\sigma}}{\sigma}\right) \left(\frac{s}{\delta}\right) \quad (\text{IV-34})$$

which is of course a thoroughly reasonable result. A sketch of the attenuation curves for the two lines is shown in Fig. 9.

As a final point, we observe that the attenuation given in equation (IV-21) is proportional to  $\frac{1}{\bar{\sigma}} \sqrt{\bar{\epsilon}}$ . Therefore, for stacks in which the laminations have the dimensions shown in Fig. 5, the attenuation is seen from equations (II-19) and (II-20) to be proportional to

$$\left(1 + \frac{t}{W}\right) \sqrt{1 + \frac{W}{t}}$$

This expression has a minimum for  $\frac{W}{t} = 2$ . Thus, the optimum arrange-



ment is for the conductor to be twice the thickness of the dielectric. This rule holds, of course, only as long as the effective skin depth defined in equation (III-42) is greater than  $s$ . If use of the ratio  $t/W = 1/2$  finds  $\delta_w$  smaller than  $s$ , the best thing to do is to increase  $t/W$  until the effective skin depth  $\delta_w$  becomes equal to  $s$ .

### V. TRANSMISSION LINE FILLED WITH LAMINATIONS

In the last section, we have calculated the attenuation of the transmission line shown in Fig. 7. By reference to equation (IV-21), it is seen that this attenuation decreases as  $s$  increases. Since we have assumed in deducing equation (IV-21) that  $s \ll h$ , we cannot use that equation to find the attenuation for  $s = d$ . Nevertheless, the suggestion is evident that this case may be particularly interesting.

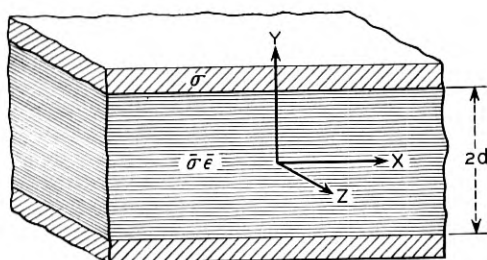


Fig. 10—Plane parallel transmission line filled with laminations.

Accordingly, let us consider the transmission line shown in Fig. 10, where the space between the outer shields has been completely filled with laminations. As before, we can write down the following fields:

In shield:

$$H_x = A e^{-\eta(y-d)} \quad (V-1)$$

$$E_x = -A \frac{\eta}{\sigma} e^{-\eta(y-d)} \quad (V-2)$$

$$\eta = \sqrt{i\omega\mu_0\sigma} \quad (V-3)$$

In laminae:

$$H_x = \cosh \zeta y \quad (V-4)$$

$$E_x = \frac{\zeta}{\sigma} \sinh \zeta y \quad (V-5)$$

where  $A$  is some constant. Matching fields at  $y = d$ , we obtain the characteristic equation

$$(\zeta d) \tanh(\zeta d) = -\frac{\bar{\sigma}}{\sigma} \eta d \quad (\text{V-6})$$

We might also have obtained this equation by placing  $h = 0$  and  $s = d$  in equation (IV-10).

We can verify that an approximate solution of equation (V-6) is

$$(\zeta d) = \frac{n\pi i}{2} \left[ 1 - \frac{\sigma}{\bar{\sigma}} \frac{1}{\eta d} \right] \quad n = 1, 3, 5, \dots \quad (\text{V-7})$$

Proceeding as before, we find for the propagation constant

$$k_n = \sqrt{\omega^2 \mu_0 \bar{\epsilon}} \left[ 1 + \frac{1}{2i\omega\mu_0\bar{\sigma}} \left( \frac{n\pi}{2d} \right)^2 \right], \quad (\text{V-8})$$

and for the attenuation

$$I(k_n) = -\frac{1}{2\bar{\sigma}} \sqrt{\frac{\bar{\epsilon}}{\mu_0}} \left( \frac{n\pi}{2d} \right)^2 \quad (\text{V-9})$$

If we place  $n = 1$  in equation (V-9) and compare the result with equation (IV-21), we see that the attenuation of the transmission line has been indeed decreased by completely filling it with laminations, and without sacrifice of the frequency independent characteristic. Furthermore, it is no longer necessary to supply a dielectric with dielectric constant equal to  $\bar{\epsilon}$ . This case clearly represents something unfamiliar in the way of transmission lines. We might in fact consider the laminated material as a new kind of transmission medium.

In order to visualize this situation more completely, let us study the distribution of fields and currents inside the transmission line. We will be interested in  $H_x$ ,  $E_y$  which can be obtained from equations (II-10) and (V-8), the current  $J = \bar{\sigma} E_x$ , the Poynting vector  $P = 1/2$  Real part  $(E_y H_x^*)$ , the total current  $I = \left| \int_0^d J dy \right|$ , and the total power  $W = \int_{-d}^d P dy$ . From equation (V-7), we can take  $\zeta$  equal approximately to  $\frac{n\pi i}{2d}$  and obtain

$$H_x = \cos \frac{n\pi y}{2d} \quad (\text{V-10})$$

$$E_y = \sqrt{\mu_0/\bar{\epsilon}} \cos \frac{n\pi y}{2d} \quad (\text{V-11})$$

$$J = - \left( \frac{n\pi}{2d} \right) \sin \frac{n\pi y}{2d} \quad (\text{V-12})$$

$$P = 1/2 \sqrt{\mu_0/\bar{\epsilon}} \cos \frac{n\pi y}{2d} \quad (\text{V-13})$$

$$I = 1 \quad (\text{V-14})$$

$$W = 1/2 \sqrt{\mu_0/\bar{\epsilon}} d \quad (\text{V-15})$$

For comparison let us write down these same quantities for the transmission line of Fig. 8. This can be done in an obvious way from equations (IV-23) to (IV-28) and by use of the characteristic equation (IV-30). We have then approximately,

$$H_s = \frac{1}{\sqrt{2}} \quad (\text{V-16})$$

$$E_v = \frac{1}{\sqrt{2}} \sqrt{\frac{\mu_0}{\epsilon_1}} \quad (\text{V-17})$$

$$P = \frac{1}{2\sqrt{2}} \sqrt{\frac{\mu_0}{\epsilon_1}} \quad (\text{V-18})$$

$$W = \frac{1}{2} \sqrt{\frac{\mu_0}{\epsilon_1}} d \quad (\text{V-19})$$

and in the shield

$$J = - \frac{1}{\sqrt{2}} \eta e^{-\eta(v-d)} \quad (\text{V-20})$$

$$I = \frac{1}{\sqrt{2}} \quad (\text{V-21})$$

If we let  $\epsilon_1 = \bar{\epsilon}$ , the second set of equations has been normalized to the same transmitted power as the first set. A comparison of these equations is shown in Fig. 11. The decreased attenuation of the laminated line, is, of course, accounted for by its much smaller current density, even though its total current is bigger by a factor  $\sqrt{2}$ . Only the fundamental mode of the laminated line is considered in Fig. 11. The higher modes will be discussed in the next section.

## VI. MODES OF TRANSMISSION

We have seen that both the transmission line partly filled with laminations as in Section IV, and the completely filled line described in Section V,

have fundamental and higher modes of transmission. Let us now examine this situation more closely.

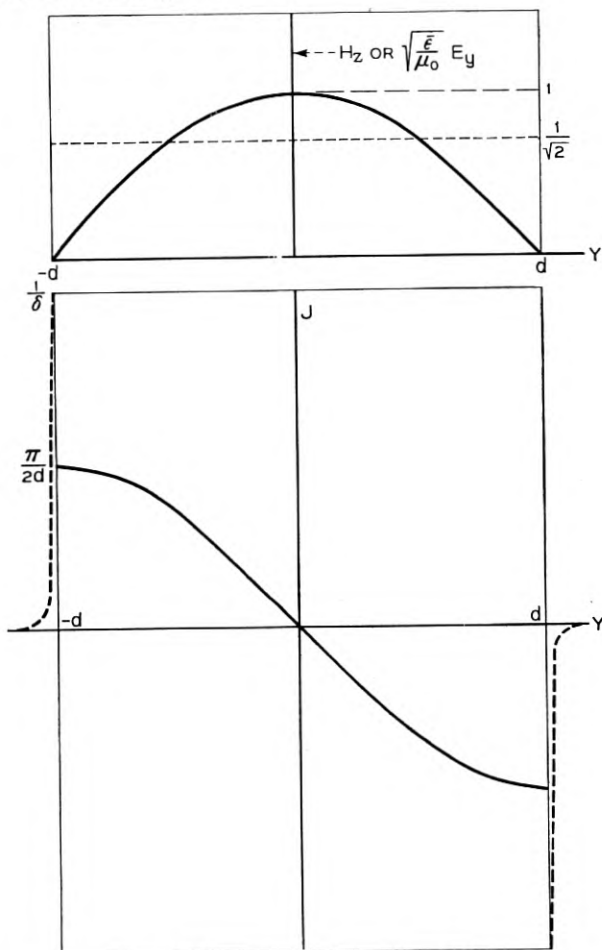


Fig. 11—Distribution of fields and current in transmission line filled with laminations.

First, in the case of the partially filled line, we can use the results of Section IV to find the following approximate expressions for the current density and magnetic field in the lamina:

$$J_0 = -\frac{1}{s} \quad (\text{VI-1})$$

$$J_n = (-1)^{n+1} \left(\frac{h}{s}\right) (n\pi)^2 \frac{1}{s} \cos \frac{n\pi}{s} (y - d) \quad (\text{VI-2})$$

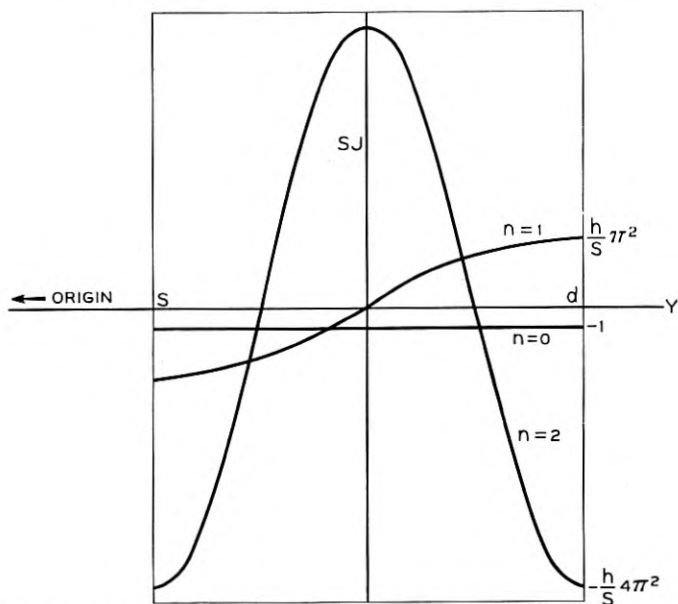


Fig. 12—Distribution of current in laminae of partially filled line for various modes.

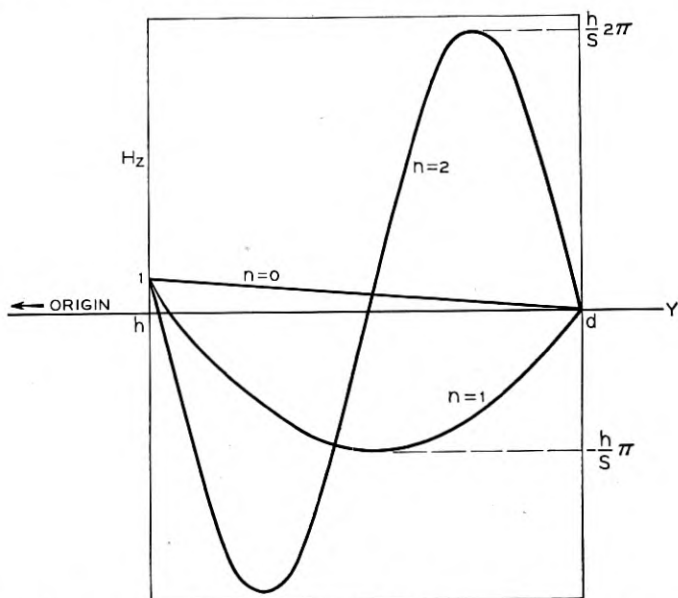


Fig. 13—Distribution of magnetic field in laminae of partially filled line for various modes.

$$(H_z)_0 = -\frac{1}{s}(y-d) \quad (\text{VI-3})$$

$$(H_z)_n = (-1)^{n+1} \left(\frac{h}{s}\right) n\pi \sin \left[ \left( n\pi + \frac{s}{h} \frac{1}{n\pi} \right) \frac{y-d}{s} \right] \quad (\text{VI-4})$$

The current densities for  $n = 0, 1, 2$  are shown in Fig. 12. We see that the current density for  $n = 0$  is all of one phase, while for the higher modes the current density has one or more reversals of phase. For these higher

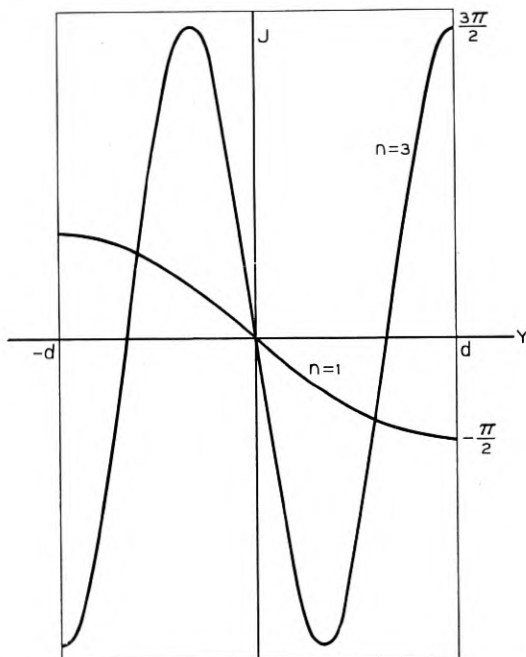


Fig. 14—Current distribution in completely filled line for various modes.

modes, the net current in the laminae is essentially zero and consequently we should expect only small fields in the interior of the line. This supposition is confirmed in Fig. 13, where the magnetic fields in the laminae are drawn for  $n = 0, 1, 2$ . The fields for all the modes have a common value unity at  $y = s$ , but for the higher modes the fields in the interior of the laminae are much greater than unity.

For the completely filled line, we have from equation (V-12)

$$J_n = -\frac{n\pi}{2d} \sin \frac{n\pi y}{2d} \quad (\text{VI-5})$$

These current densities are drawn in Fig. 14 for  $n = 1$  and  $n = 3$ . It will be observed that in each case a net current flows in one sense on one side of the center line and in the opposite sense on the other side of the line.

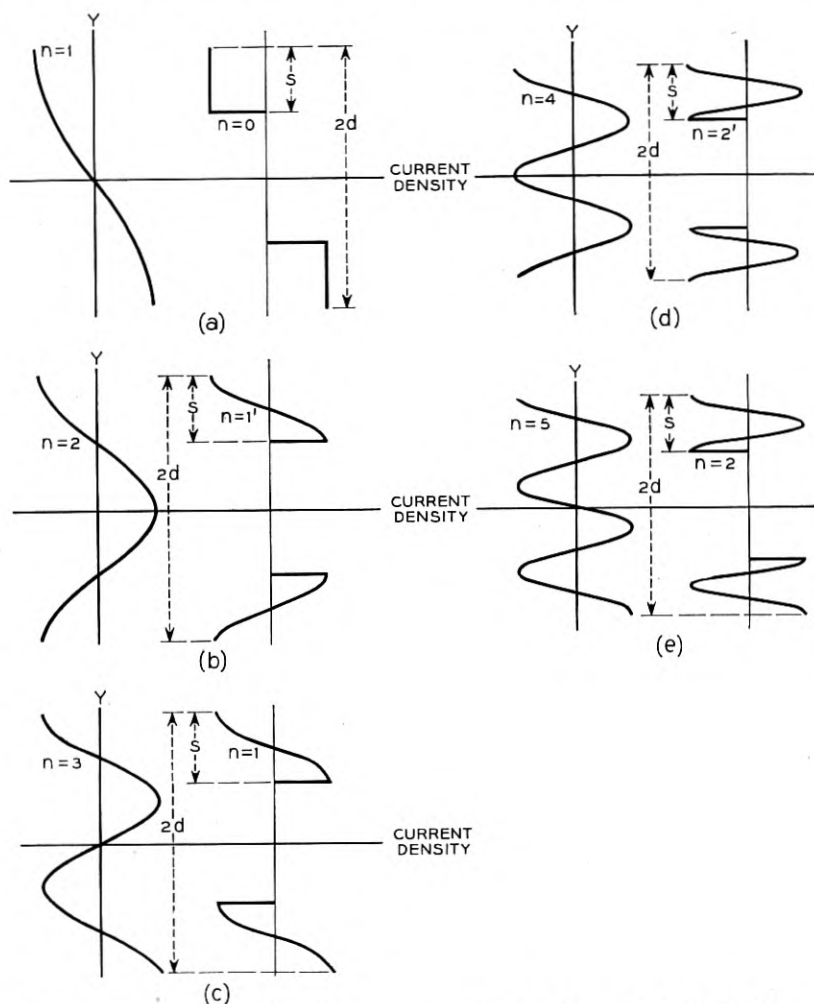


Fig. 15—Correspondence of modes between partially filled line and wholly filled line.

We can now profitably compare the modes for the two types of lines. Clearly there should be a one to one correspondence of the modes since the partially filled line can be made to approach the completely filled line continuously by adding more laminated material. This correspondence is shown



schematically in Fig. 15. The modes we have discussed to this point have all been antisymmetric in current density about the center plane. It will be clear from Fig. 15 that there are in addition another set of modes symmetric in the current density. For the completely filled case these are the modes  $n = 2, 4, 6 \dots$ , and for the partially filled case the modes  $n = 1', 2', 3', \dots$ .

An important point can now be made. For the completely filled case, there are higher modes such as  $n = 3$  where a net current flows on one side of the center plane. The corresponding mode, however, for the partially filled case with  $s \ll h$  has nearly zero net current on either side of the center plane. Thus, for the partially filled case with  $s$  much smaller than  $h$  there are no modes except the fundamental with large fields in the interior of the line.

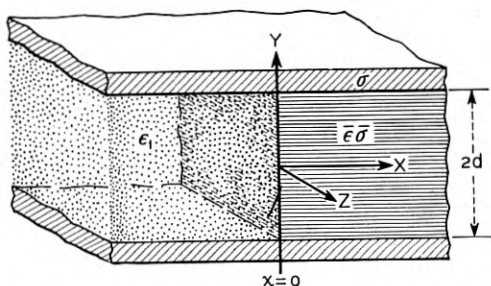


Fig. 16—Junction between two plane parallel transmission lines, one of which is filled with laminations.

## VII. TERMINATION OF A LAMINATED LINE

The discussion of modes of transmission in the last two sections enables us now to consider what occurs at the junction of an un laminated transmission line and one partially or completely filled with laminated material. We will, for simplicity, consider mainly the case of the completely filled line as shown in Fig. 16. To the left of  $x = 0$  there is an un laminated line such as shown in Fig. 8 filled with a dielectric of dielectric constant  $\epsilon_1$ . To the right of  $x = 0$  there is a line of the type considered in Section V filled with laminated material of average transverse dielectric constant  $\bar{\epsilon}$  and average longitudinal conductivity  $\bar{\sigma}$ . We shall consider separately what happens to a wave incident upon the boundary from left or right.

When expressing the fields in the un laminated line, we shall have to include certain unpropagated modes which have not yet been discussed. These modes must attenuate to the left, and can be written

$$H_z = \cos \frac{n\pi y}{d} e^{n\pi x/d} \quad (\text{VII-1})$$

$$E_x = -\frac{1}{i\omega\epsilon_1} \left( \frac{n\pi}{d} \right) \sin \frac{n\pi y}{d} e^{n\pi x/d} \quad n = 1, 2, 3 \dots \quad (\text{VII-2})$$

$$E_y = -\frac{1}{i\omega\epsilon_1} \left( \frac{n\pi}{d} \right) \cos \frac{n\pi y}{d} e^{n\pi x/d} \quad (\text{VII-3})$$

It has been assumed that the wavelength in the dielectric of the frequency under consideration is much greater than  $d$ .

We shall first consider a wave incident upon the boundary from the left. From equations (V-16), (V-17), (VII-1) and (VII-3) we have for  $x < 0$

$$H_z = A e^{-ikx} + B e^{ikx} + \sum_m C_m \cos \frac{m\pi y}{d} e^{m\pi x/d} \quad (\text{VII-4})$$

$$E_y = \sqrt{\frac{\mu_0}{\epsilon_1}} [A e^{-ikx} - B e^{ikx}] - \frac{1}{i\omega\epsilon_1} \sum_m C_m \left( \frac{m\pi}{d} \right) \cos \frac{m\pi y}{d} e^{m\pi x/d} \quad (\text{VII-5})$$

where  $m = 1, 2, 3 \dots$  and  $k$  is given by (IV-31). For  $x > 0$  we have from equations (V-10) and (V-11)

$$H_z = \sum_n D_n \cos \frac{n\pi y}{2d} e^{-ik_n x} \quad (\text{VII-6})$$

$$E_y = \sqrt{\frac{\mu_0}{\bar{\epsilon}}} \sum_n D_n \cos \frac{n\pi y}{2d} e^{-ik_n x} \quad (\text{VII-7})$$

where  $n = 1, 3, 5 \dots$  and  $k_n$  is given by (V-8). At  $x = 0$  the boundary conditions give

$$A + B + \sum_m C_m \cos \frac{m\pi y}{d} = \sum_n D_n \cos \frac{n\pi y}{2d} \quad (\text{VII-8})$$

$$\begin{aligned} \sqrt{\frac{\mu_0}{\epsilon_1}} (A - B) - \frac{1}{i\omega\epsilon_1} \sum_m C_m \left( \frac{m\pi}{d} \right) \cos \frac{m\pi y}{d} \\ = \sqrt{\frac{\mu_0}{\bar{\epsilon}}} \sum_n D_n \cos \frac{n\pi y}{2d} \end{aligned} \quad (\text{VII-9})$$

If  $\epsilon_1 = \bar{\epsilon}$ , it can be seen from (VII-8) and (VII-9) that  $B = C_m = 0$ . Thus there is no reflected wave and no unpropagated waves are needed. Let us consider only this case. The coefficients  $D_n$  are determined by

$$A = \sum_n D_n \cos \frac{n\pi y}{2d} \quad (\text{VII-10})$$

which yields

$$D_n = \frac{4}{n\pi} (-1)^{(n-1)/2} A \quad (\text{VII-11})$$

Referring to equation (V-15) we have for the power flow in the transmitted wave,

$$W = \frac{d}{2} \sqrt{\frac{\mu_0}{\epsilon}} \sum_n D_n^2 \quad (\text{VII-12})$$

$$= A^2 \sqrt{\frac{\mu_0}{\epsilon}} \frac{8d}{\pi^2} \sum_n \frac{1}{n^2} \quad (\text{VII-13})$$

Let us now find the ratio of the power transmitted in the fundamental mode  $n = 1$  to the total power transmitted, which is also the total incident power as can be checked from equation (VII-12). We have

$$\frac{\text{power in fundamental}}{\text{total power}} = \frac{8}{\pi^2} \quad (\text{VII-14})$$

Thus, in exciting the fundamental mode of the laminated line we have a power loss which can be expressed in db as

$$\begin{aligned} \text{db loss} &= 10 \log \frac{\pi^2}{8} \\ &= 0.912 \end{aligned} \quad (\text{VII-15})$$

Let us next consider a wave composed of the fundamental mode of the laminated line incident upon the boundary from the right. For  $x < 0$  we have in this case

$$H_z = B e^{ikx} + \sum_m C_m \cos \frac{m\pi y}{d} e^{m\pi x/d} \quad (\text{VII-16})$$

$$E_y = -B \sqrt{\frac{\mu_0}{\epsilon_1}} e^{ikx} - \frac{1}{i\omega\epsilon_1} \sum_m C_m \left( \frac{m\pi}{d} \right) \cos \frac{m\pi y}{d} e^{m\pi x/d} \quad (\text{VII-17})$$

where again  $m = 1, 2, 3, \dots$  and  $k$  is given by (IV-31). For  $x > 0$

$$H_z = M e^{ik_1 x} \cos \frac{\pi y}{2d} + \sum_n N_n \cos \frac{n\pi y}{2d} e^{-ik_n x} \quad (\text{VII-18})$$

$$E_y = \sqrt{\frac{\mu_0}{\epsilon}} \left[ -M e^{ik_1 x} \cos \frac{\pi y}{2d} + \sum_n N_n \cos \frac{n\pi y}{2d} e^{-ik_n x} \right] \quad (\text{VII-19})$$

with  $k_n$  given by (V-8) and  $n = 1, 3, 5, \dots$ . At  $x = 0$ ,

$$B + \sum_m C_m \cos \frac{m\pi y}{d} = M \cos \frac{\pi y}{2d} + \sum_n N_n \cos \frac{n\pi y}{2d} \quad (\text{VII-20})$$

$$\begin{aligned} -B - \frac{1}{ik} \sum_m C_m \left( \frac{m\pi}{d} \right) \cos \frac{m\pi y}{d} \\ = \sqrt{\frac{\epsilon_1}{\epsilon}} \left[ -M \cos \frac{\pi y}{2d} + \sum_n N_n \cos \frac{n\pi y}{2d} \right] \end{aligned} \quad (\text{VII-21})$$

Let us again let  $\epsilon_1 = \bar{\epsilon}$ . By adding and subtracting we find

$$-\frac{1}{ik} \sum_m C_m \frac{m\pi}{d} \cos \frac{m\pi y}{d} = 2 \sum_n N_n \cos \frac{n\pi y}{2d} \quad (\text{VII-22})$$

$$2B + \frac{1}{ik} \sum_m C_m \frac{m\pi}{d} \cos \frac{m\pi y}{d} = 2M \cos \frac{\pi y}{2d} \quad (\text{VII-23})$$

From (VII-23) it is clear that

$$2B = \frac{1}{2d} \int_{-d}^d 2M \cos \frac{\pi y}{2d} dy \quad (\text{VII-24})$$

or

$$B = \frac{2}{\pi} M \quad (\text{VII-25})$$

Thus, we have for the transmitted power

$$\begin{aligned} W_t &= d \sqrt{\frac{\mu_0}{\epsilon_1}} B^2 \\ &= d \sqrt{\frac{\mu_0}{\epsilon_1}} \left( \frac{2}{\pi} \right)^2 M^2 \end{aligned} \quad (\text{VII-26})$$

and for the incident power

$$W_i = \frac{d}{2} \sqrt{\frac{\mu_0}{\bar{\epsilon}}} M^2$$

The ratio of transmitted power to incident power is therefore

$$\frac{W_t}{W_i} = \frac{8}{\pi^2} \quad (\text{VII-27})$$

just as in equation (VII-14). There is thus the same power loss in crossing the boundary in either direction. It is interesting to note, however, that in the second case the non-propagating modes in the unlaminated line are excited.

From equations (VII-22), (VII-23) and (VII-25) we can find

$$\sum_n N_n \cos \frac{n\pi y}{2d} = M \left( \frac{2}{\pi} - \cos \frac{\pi y}{2d} \right) \quad (\text{VII-28})$$

and consequently

$$N_1 = M \left( \frac{8}{\pi^2} - 1 \right), \quad (\text{VII-29})$$

and

$$N_n = M \frac{8}{\pi^2} \frac{1}{n} (-1)^{(n-1)/2}, \quad n \neq 1. \quad (\text{VII-30})$$

The reflected power is found as before to be for the fundamental mode of the laminated line

$$W_{r1} = \frac{d}{2} \sqrt{\frac{\mu_0}{\epsilon}} M^2 \left( \frac{8}{\pi^2} - 1 \right)^2 \quad (\text{VII-31})$$

and for the higher modes

$$\begin{aligned} \sum_{n \neq 1} W_{rn} &= \frac{d}{2} \sqrt{\frac{\mu_0}{\epsilon}} M^2 \sum_{n \neq 1} \left( \frac{8}{n\pi^2} \right)^2, \\ &= \frac{d}{2} \sqrt{\frac{\mu_0}{\epsilon}} M^2 \frac{8}{\pi^2} \left( 1 - \frac{8}{\pi^2} \right). \end{aligned} \quad (\text{VII-32})$$

We can now easily check that

$$\begin{aligned} \frac{W_t}{W_i} + \frac{W_{r1}}{W_i} + \frac{1}{W_i} \sum_{n \neq 1} W_{rn} &= \frac{8}{\pi^2} \\ &+ \left( \frac{8}{\pi^2} - 1 \right)^2 + \frac{8}{\pi^2} \left( 1 - \frac{8}{\pi^2} \right) = 1 \end{aligned} \quad (\text{VII-33})$$

The case of the partially filled line can be studied in a manner similar to the above discussion, and will show smaller power losses for waves transmitted through the boundary. The problem is further complicated by the presence of unpropagated modes in the partially filled line similar to those in the unlaminate line.

## APPENDIX A

### PLANE WAVES

It is interesting to inquire about the waves that exist in a laminated medium of infinite extent. Let us return to equation (II-23). It is easy to

show that  $\alpha_0$  is zero for  $k$  given approximately by

$$k = \sqrt{\omega^2 \mu_0 \bar{\epsilon}} \left[ 1 - i \frac{1}{2} \frac{W}{t} \left( \frac{\omega \epsilon}{\sigma} \right) \right] \quad (\text{A-1})$$

For  $\alpha_0 = 0$ ,  $E_x$  is of course zero. We thus have a plane wave propagating through the medium with a wavelength appropriate to the average dielectric constant  $\bar{\epsilon}$  and with a very small attenuation proportional to the square of the frequency. Equation (A-1) can be obtained from equation (45) in Sakurai's paper.<sup>2</sup>

If we wish next to observe the effect of finite thickness laminations, we can require  $\alpha_w$  to be zero in equation (III-34). In this case we obtain

$$k = \sqrt{\omega^2 \mu_0 \bar{\epsilon}} \left[ 1 - i \frac{1}{12} \frac{W}{W+t} \left( \frac{W}{\delta} \right)^2 \right] \quad (\text{A-2})$$

Again the attenuation is proportional to the square of the frequency. The attenuation given by equation (A-2) is equal to that obtained from equation (A-1) if

$$W = \frac{1}{\sigma} \sqrt{12 \left( 1 + \frac{W}{t} \right) \frac{\epsilon}{\mu_0}} \quad (\text{A-3})$$

For copper, equation (A-3) requires  $W$  to be of the order of  $10^{-8}$  cm. Under ordinary circumstances, therefore, the attenuation given in equation (A-1) is much smaller than that obtained when consideration is given to the finite thickness of the laminations.

## APPENDIX B

### TRANSMISSION LINE FILLED WITH LAMINATIONS OF FINITE WIDTH

Let us consider the transmission line shown in Fig. 17. As before we have a set of metal laminae of width  $W$  and conductivity  $\sigma$  separated by insulating laminae of width  $t$  and dielectric constant  $\epsilon$ . The laminae will be numbered as shown in the figure from 1 to  $N$ . Let us define  $r$  and  $p$  by the relations

$$r = \frac{\beta - T_{11}}{T_{12}}; \quad p = \frac{1/\beta - T_{11}}{T_{12}} \quad (\text{B-1})$$

Then, from equations (III-19) and (III-20), we can write for the  $z$ -component of magnetic field and the  $x$ -component of electric field,

$$H_n = A\beta^n + B\beta^{-n} \quad (\text{B-2})$$

$$E_n = rA\beta^n + pB\beta^{-n} \quad (\text{B-3})$$

<sup>2</sup> Tokio Sakurai, loc. cit., page 398.

From equations (II-9) and (II-15) we clearly have

$$\frac{E_N}{H_N} = -R; \quad \frac{E_0}{H_0} = R \quad (\text{B-4})$$

where  $R = \frac{\alpha_s}{\sigma}$  and  $\alpha_s = \frac{1+i}{\delta}$ . It is easy now to obtain the characteristic equation

$$\begin{vmatrix} R\beta^N & R\beta^{-N} & \beta^N & \beta^{-N} \\ R & R & -1 & -1 \\ r & 0 & -1 & 0 \\ 0 & p & 0 & -1 \end{vmatrix} = 0 \quad (\text{B-5})$$

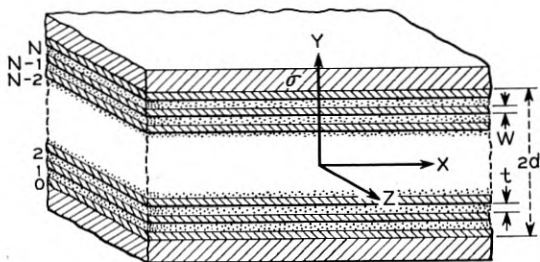


Fig. 17—Transmission line completely filled with laminations.

After expansion and use of expressions (B-1), the characteristic equation is found to take the form

$$\coth(N \ln \beta) = \frac{R^2 T_{12} + T_{21}}{R \left( \frac{1}{\beta} - \beta \right)} \quad (\text{B-6})$$

which can be written, using the relation  $\ln \beta = \alpha_w(W + t)$ ,

$$2 \sinh \alpha_w(W + t) \coth(2d\alpha_w) = RT_{12} + \frac{1}{R} T_{21} \quad (\text{B-7})$$

Let us now assume, as in Section III, that  $|\xi t|$  and  $|\eta W|$  are much smaller than unity. Equation (B-7) can then be written

$$(2\alpha_w d) \coth(2\alpha_w d) = - \left( \frac{W}{W + t} \right) (\alpha_s d) \cdot \left( 2(P + 1) + i \frac{2}{3} (3P + 1) \left( \frac{W}{\delta} \right)^2 \right) \quad (\text{B-8})$$

The quantity on the right of equation (B-8) is of the order of  $d/\delta$ . We write, therefore,

$$(2\alpha_w d) \coth (2\alpha_w d) = 0 \left( \frac{d}{\delta} \right) \quad (\text{B-9})$$

and

$$(2\alpha_w d) = \pi i \left[ 1 + 0 \left( \frac{\delta}{d} \right) \right] \quad (\text{B-10})$$

Thus we have approximately

$$\cosh \alpha_w (W + t) = 1 - \frac{1}{2} \left( \frac{\pi(W + t)}{2d} \left[ 1 + 0 \left( \frac{\delta}{d} \right) \right] \right)^2 \quad (\text{B-11})$$

Furthermore, from equation (III-34)

$$\cosh \alpha_w (W + t) = 1 - \frac{1 + 4P}{6} \left( \frac{W}{\delta} \right)^4 + i(1 + 2P) \left( \frac{W}{\delta} \right)^2 \quad (\text{B-12})$$

We can now equate equations (B-11) and (B-12) and after suitable manipulation obtain

$$k = \sqrt{\omega^2 \mu_0 \bar{\epsilon}} \left( 1 - \frac{i}{2 \left( 1 + \frac{t}{W} \right)} \left[ \frac{\pi^2}{8} \left( 1 + \frac{t}{W} \right)^2 \left( \frac{\delta}{d} \right)^2 + \frac{1}{6} \left( \frac{W}{\delta} \right)^2 \right] \right) \quad (\text{B-13})$$

where we have now dropped a term of order  $\left( \frac{\delta}{d} \right)$  compared to unity.

Equation (B-13) is of considerable interest. It is observed that for  $\left( \frac{W}{\delta} \right) \ll \left( \frac{\delta}{d} \right)$  the attenuation becomes that given in equation (V-8) for infinitely thin laminae. For  $\left( \frac{W}{\delta} \right) \gg \left( \frac{\delta}{d} \right)$ , on the other hand, the attenuation approaches that given in equation (A-2) for an unbounded array of finite laminae. This can be considered in two ways. Let us ask for the condition that the two terms contributing to the attenuation in (B-13) are equal. We find for this to be true that

$$d = \frac{\pi}{2} \sqrt{3} \left( 1 + \frac{t}{W} \right) \left( \frac{\delta}{W} \right) \delta \quad (\text{B-14})$$

In other words, at a given frequency, the attenuation of the line can be little reduced by making  $d$  larger than the value given in equation (B-14)



which will be recognized as approximately the skin depth given in equation (III-44).

Alternatively, we see that the attenuation as a function of frequency remains constant from low frequencies up to the point where  $\delta$  satisfies equation (B-14). At higher frequencies, the attenuation increases parabolically. At frequencies where  $\delta \ll W$ , the attenuation will clearly correspond to propagation in a parallel plate transmission line of width  $l$  and will therefore increase with the square root of the frequency. This behavior is similar to that discussed in Section I for a line with a thin stack of laminations on its inner conductor.

## Some Circuit Properties and Applications of *n-p-n* Transistors

By R. L. WALLACE, JR. and W. J. PIETENPOL

Shockley, Sparks, and Teal have recently described the physical properties of a new kind of transistor. Preliminary studies of circuit performance show that it is a stable, high gain, quiet amplifier of considerable practical interest. This paper analyzes the performance of a few early experimental units.

### INTRODUCTION

Almost two years ago, W. Shockley<sup>1, 2</sup> first published the theory of a transistor made from a single piece of germanium in which the conductivity type varies in such a way as to produce two rectifying junctions. Since that time, M. Sparks, G. K. Teal and others at the Bell Telephone Laboratories<sup>3, 4</sup> have contributed notably to the physical realization of this device.

Recently Sparks has produced a number of *n-p-n* transistors and has found their behavior to be closely in accord with Shockley's theory.<sup>4</sup> Preliminary circuit studies on these devices have shown that in several respects their performance is remarkable. In view of this, our transistor development group has undertaken to produce small quantities of *n-p-n* transistors in a form suitable for incorporation in working circuits.

This paper will deal principally with the circuit aspects of the *n-p-n* transistor by presenting and analyzing performance data on a small number of experimental units. For a discussion of the solid state physics of its design and operation the reader is referred to the previously mentioned works of Shockley, Sparks, and Teal.

### OUTSTANDING PROPERTIES

Before getting lost in a maze of detail, it seems worthwhile to list and mention briefly the salient features of this new transistor. They are:

1. *Relatively low noise figure.* Most of the units measured so far have a noise figure between 10 and 20 db at 1000 cps.
2. *Complete freedom from short-circuit instability.* The input and output impedances are always positive whether the transistor is connected ground-emitter, grounded-base, or grounded collector. This permits a great deal of freedom in circuit design and makes it possible, by choosing the appropriate connection, to obtain a considerable variety of input and output impedances.
3. *High gain.* Power gains of the order of 40 to 50 db per stage have been obtained

4. *Power handling capacity and efficiency.* The design can readily be varied to permit the required amount of power dissipation up to at least two watts. Furthermore the static characteristics are so nearly ideal that Class A efficiencies of 48 or 49 out of a possible 50% can be realized. The efficiencies for Class B and Class C operation are correspondingly high.

5. *Ruggedness and small size.* The germanium part of the transistor is enclosed in a hard plastic bead about  $\frac{3}{16}$  inch in diameter. Inside the bead three connections are mechanically as well as electrically fastened to the germanium and are brought out as pigtailed through the bead. This gives a very sturdy unit.

6. *Freedom from microphonics.* Vibration tests in the audio frequency range indicate that these devices are relatively free from microphonic noise.

7. *Limited frequency response.* Collector capacitance limits the frequency response at full gain to a few kilocycles. By using a suitable impedance

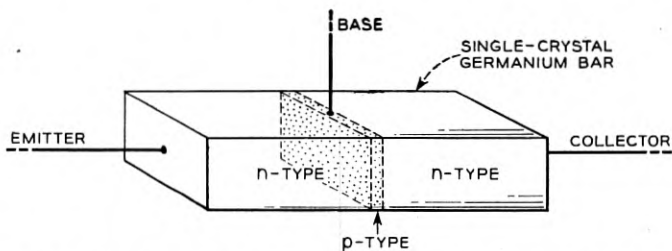


Fig. 1—The heart of an *n-p-n* transistor is a tiny bar of germanium to which three mechanically strong electrical connections are made.

mismatch it is possible to maintain the frequency response flat to at least one megacycle while still obtaining a useful amount of gain.

8. *Operation with exceedingly small power consumption.* Perhaps the most remarkable feature of these transistors is their ability to operate with exceedingly small power consumption. The best example of this to date is an audio oscillator which requires for a power supply only 6 microamperes at 0.1 volts. This represents 0.6 microwatts of power which contrasts sharply with the million or more microwatts required to heat the cathode of an ordinary receiving-type vacuum tube.

#### PHYSICAL APPEARANCE AND CONSTRUCTION

Figure 1 shows schematically the configuration of an *n-p-n* transistor. The small bar of single crystal germanium contains a thin layer of *p*-type interposed between regions of *n*-type. Mechanically strong ohmic connections are made to the three regions as indicated and brought out through a

hard plastic bead. A finished transistor is shown in the photograph of Fig. 2. It should be pointed out that Fig. 1 is not drawn to scale and that the  $p$ -layer may be less than a thousandth of an inch thick.

### STATIC CHARACTERISTICS

A great deal of information about the low frequency performance of a transistor can be obtained from a set of static characteristics such as those shown in Fig. 4. Curves of this sort are obtained simply by connecting suit-

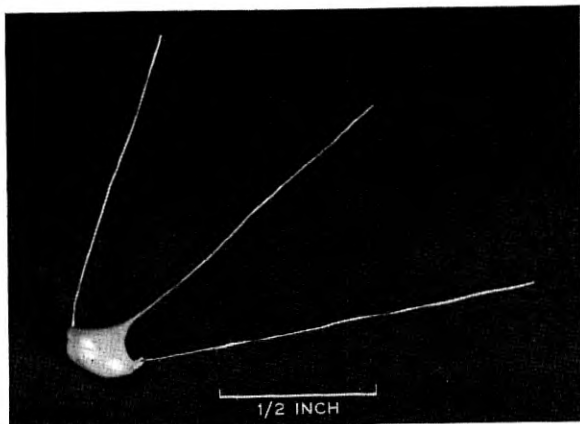


Fig. 2—A beaded  $n-p-n$  transistor.

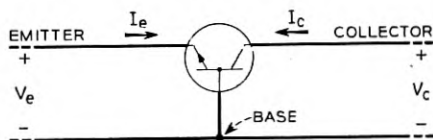


Fig. 3—The symbol for a  $p$ -type transistor on which the convention of signs for currents and voltages is indicated.

able current sources to the emitter and collector circuits of the transistor and measuring the resulting voltages. The currents are called positive when they flow into the emitter and collector as shown and the voltages are called positive when they have the signs shown in Fig. 3.

Let us first examine these curves with an eye to finding out what kind of voltage and current supplies are needed to bias the transistor into the range in which it can amplify. To make this easy, that part of the characteristics which lies within the normal operating range has been shown as solid lines and that part of the characteristics corresponding to cutoff has been shown as dotted lines.

Note from the upper set of curves that  $V_c$  is positive in the operating range. This means that the collector must be biased positive with respect to the

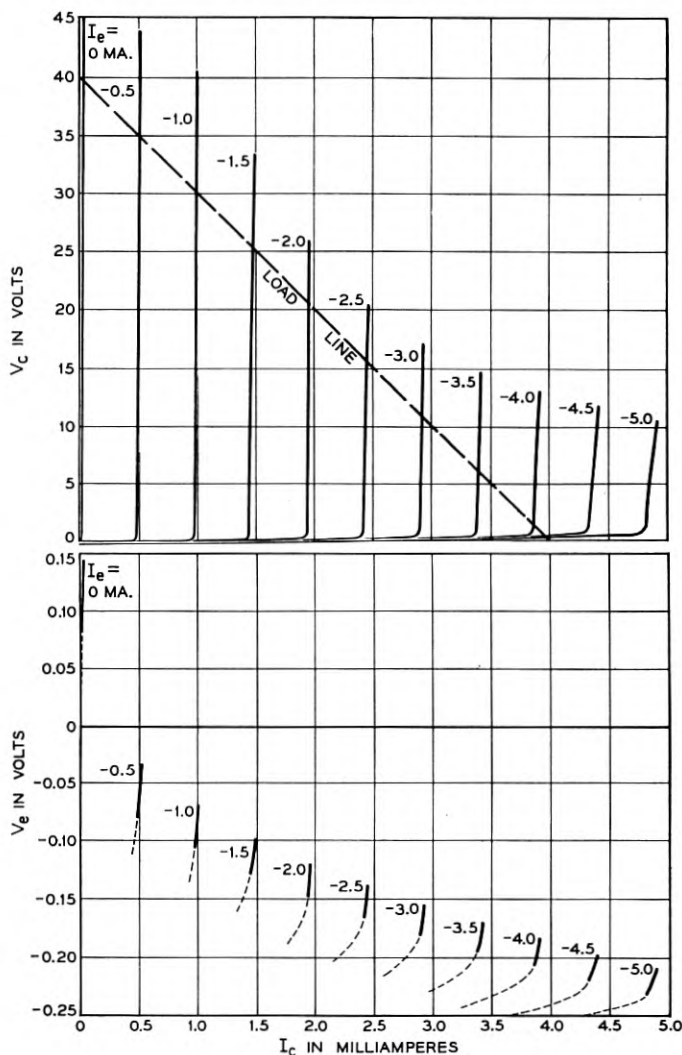


Fig. 4—Static characteristics of an *n-p-n* transistor.

base. For this particular transistor a bias voltage anywhere between about 0.1 volts and 35 volts is suitable. Note also that all the curves on this plot correspond to negative emitter currents. This means that the emitter must

be biased in such a way that current flows out of the emitter into a suitable current supply. Furthermore, the collector current corresponding to any given emitter current can be seen to be almost equal in magnitude to the emitter current. Since these two currents are opposite in sign, this means that most of the current which flows into the collector leaves by way of the emitter with the result that the current in the base circuit is very small.

Suppose that the collector is held at a constant positive voltage as, for example, by connecting a battery between collector and base (with a transformer winding in series, perhaps). Now if a negative current is forced into the emitter by a battery and resistance connected in series between emitter and base, the collector current can be controlled by varying the emitter current and will always be approximately equal in magnitude to the emitter current. Suitable collector currents for this particular transistor range from about 20 microamperes to about five milliamperes.

The exact choice of collector current and voltage within the ranges mentioned above will be dictated largely by the amount of power output required. The more power output required, the more current and voltage will be needed from the power supply. Since the collector circuit efficiency cannot exceed the theoretical limit of 50% in Class A operation, the signal power output cannot exceed half the power supplied by the battery. This means, for example, that if the collector is worked at 20 volts and 2 milliamperes the Class A power output cannot exceed 20 milliwatts.

From the lower plot of Fig. 4 it is possible to obtain information about the bias voltage required for the emitter. Note, first, that the entire emitter voltage plot corresponds to a very small range of emitter voltages near zero and, furthermore, that the part of the characteristics corresponding to the operating range covers only a few thousandths of a volt. This means that if the collector voltage is held constant very small changes in emitter voltage will produce fairly large changes in collector current, or if the collector current is held constant very small changes in emitter voltage will produce relatively enormous changes in collector voltage. This at once suggests the use of this transistor as a d-c. amplifier between a low impedance source and a high impedance load. In this application, voltage stepup of the order of 10,000 times is possible.

The very great sensitivity of the collector circuit to emitter voltage suggests, however, that for a-c. amplifiers one should use a current source as an emitter bias supply. This can be obtained from a battery and a large resistance in series. Furthermore, since the emitter voltage is always nearly zero, the emitter current can be calculated in advance by dividing the battery voltage by the value of the series resistance (provided, of course, that

the supply voltage is large compared to the few hundredths of a volt drop across the emitter circuit).

One can also draw some interesting conclusions from the static characteristics about the large signal operation of the transistor. If the load is resistive, the instantaneous operating point will swing up and down along a straight line such as the load line shown in the upper plot of Fig. 4. This particular load line corresponds to an a-c. load resistance of 10,000 ohms. Suppose that the steady collector biases are 20 volts and 2 milliamperes so that the drain from the power supply is 40 milliwatts. Now consider the permissible swings of collector voltage and current. Since the collector characteristics are quite straight and evenly spaced over a wide range of current and voltage values, the output signal can swing nearly down to zero collector volts and nearly up to zero collector current without distortion. The limit on the lower end is imposed by the fact that the collector characteristics begin to be curved when  $V_c$  is less than about 0.1 volts; and the limit on the upper end is imposed by the fact that the collector current does not drop completely to zero when  $I_e$  drops to zero. The lower limit of collector current is, in this case, about 50 microamperes and, since this amount of current in 10,000 ohms corresponds to 0.5 volts, this means that the instantaneous collector voltage is limited to swings between 39.5 volts and 0.1 volts. Starting from a quiescent value of 20 volts, the permissible positive swing is then 19.5 volts and the permissible negative swing is 19.9 volts. Reducing the quiescent voltage to 19.8 volts (and keeping the same load line) makes it possible to obtain a peak swing of 19.7 volts which corresponds to 19.45 milliwatts of signal delivered to the load. This gives a collector circuit efficiency of 48.5% out of a possible 50%. Some transistors take even less collector current when the emitter current is zero and hence permit even higher efficiencies.

These computations of efficiency have all been based on the assumption of sinusoidal *current* applied to the emitter. It will be shown in a later section that the emitter resistance varies with emitter current, however, and this means that to realize high efficiency with low distortion it is necessary to drive the emitter from a high impedance source.

#### OPERATION WITH SMALL POWER CONSUMPTION

For small signal applications the transistor represented by the characteristics of Fig. 4 can deliver useful gain at very much lower voltages and currents than those used in the example above. In order to show this, the characteristics of Fig. 5 have been plotted for a range of collector voltage extending up to only 2 volts and for a range of collector currents extending

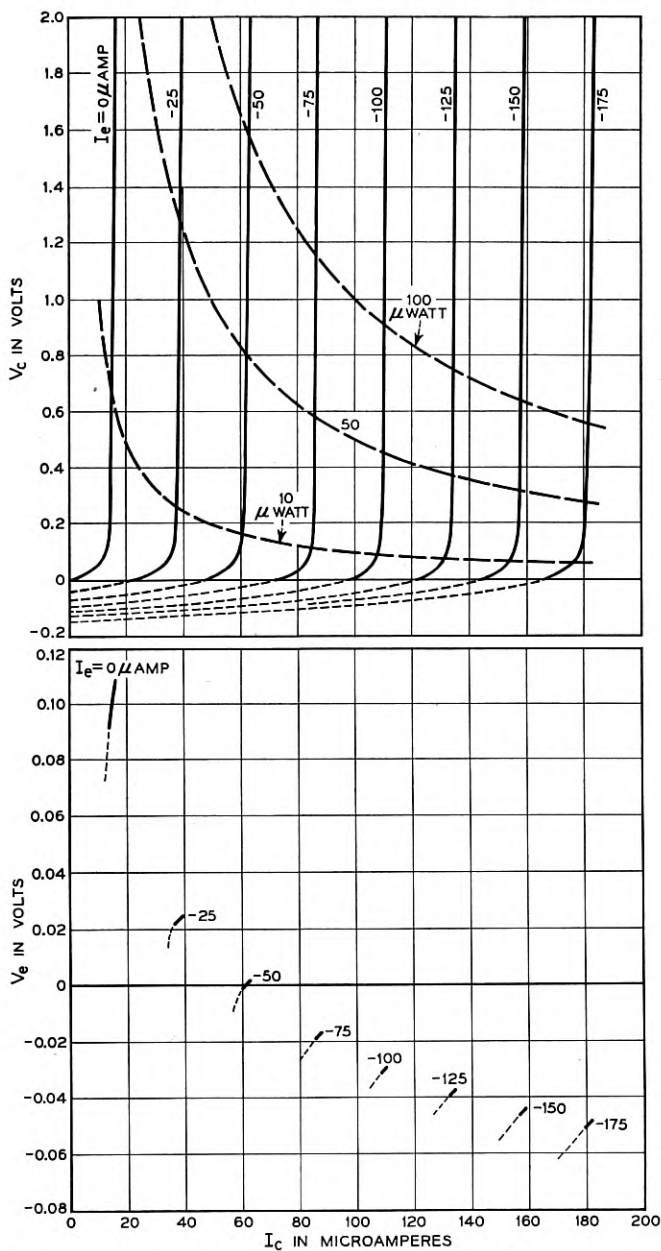


Fig. 5—Static characteristics showing behavior at very low applied voltages and currents.



up to only 200 microamperes. It can be seen from the upper plot that the collector circuit characteristics are still quite usefully straight and evenly spaced in this micro-power range. In fact, for small signal operation it is sufficient to use a collector voltage only a little in excess of 0.1 volts and a collector current a little in excess of 20 microamperes. This means that the power required to bias the collector into the operating range amounts to only a few microwatts. Contours are shown for 10, 50, and 100 microwatts of power supply.

This ability of the transistor to work with extremely small power consumption is one of its most striking and perhaps most important features

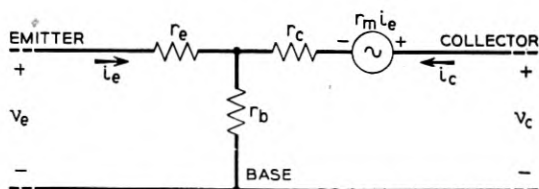


Fig. 6—The low-frequency equivalent circuit of a transistor.

When one considers that the total power consumption of a single transistor stage can be smaller by many thousands of times than the power required to heat the cathode in a vacuum tube, it is obvious that the advent of this device will make possible many new kinds of application.

#### VARIATION OF TRANSISTOR PROPERTIES WITH OPERATING POINT

Ryder and Kircher<sup>5</sup> have shown that it is convenient to analyze the small signal properties of a transistor at low frequencies in terms of the equivalent circuit of Fig. 6 where  $r_e$  is called the emitter resistance,  $r_b$  is called the base resistance, and  $r_c$  is called the collector resistance. The internal generator,  $r_m i_e$ , is the active part of the circuit and in this respect corresponds to the familiar  $\mu e_0$  of vacuum tube circuit theory. It is the purpose of this section to show what values these quantities have for a particular *n-p-n* transistor and to show how they vary with the applied biases. This will form a basis for the next section in which these quantities will be used to compute such things as the input and output impedances and the gains of various transistor connections.

Ryder and Kircher have shown that these four  $r$ 's can be obtained directly from static characteristics such as those shown in Fig. 4 and Fig. 5. In the case of *n-p-n* transistors, however, the magnitudes of these quantities are such that it is difficult to obtain satisfactory accuracy in this way and it has been more convenient to measure the 4-pole  $r$ 's by a-c. methods.

These measurements have shown that all of the  $r$ 's are, to a first approximation, independent of collector voltage so long as the collector voltage is above a few tenths of a volt and so long as the total dissipation is small enough to prevent appreciable heating of the transistor.

In view of this fact it is perhaps sufficient to show how these quantities vary with emitter current for a moderate fixed value of collector voltage. Figures 7 and 8 show that  $r_e$  and  $r_m$  are very nearly equal and that they tend to decrease as  $I_e$  increases. Theoretically  $r_m$  and  $r_e$  should both be infinite. The fact that they reach values as low as 10 megohms in this case is a meas-

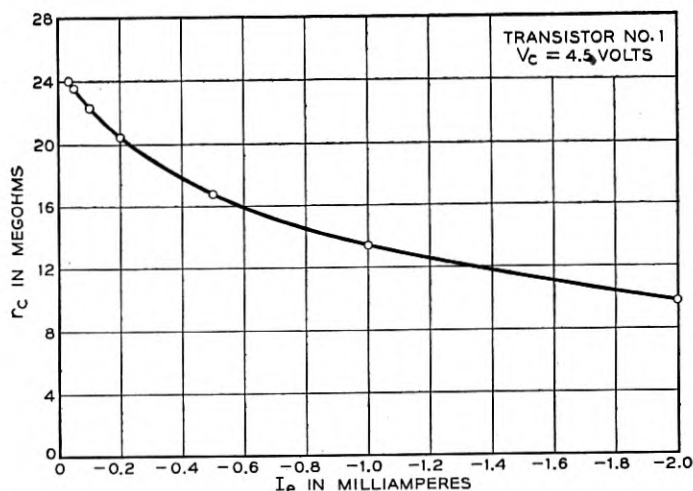


Fig. 7—The variation of collector resistance with emitter current at a fixed value of collector voltage.

ure of the imperfection in technique of fabricating the transistor. Values as high as 60 megohms have been achieved in the laboratory.

Figure 9 shows that  $r_b$  in this transistor is approximately 240 ohms and is independent of  $I_e$ .

Figure 10 shows that  $r_e$  decreases with increasing emitter current, ranging from about 500 ohms at 50 microamperes down to about 5 ohms at 5 milliamperes. Shockley<sup>4</sup> has shown that  $r_e$  should be given by

$$r_e = \frac{kT}{qI_e} \quad (1)$$

where  $q$  is the charge on an electron,  $k$  is Boltzman's constant,  $T$  is the Kelvin temperature and  $I_e$  is the emitter current. When the temperature

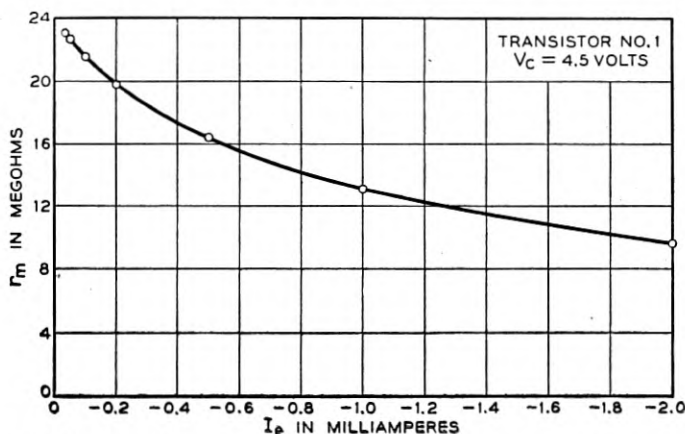


Fig. 8—Variation of  $r_m$  with emitter current at fixed collector voltage.

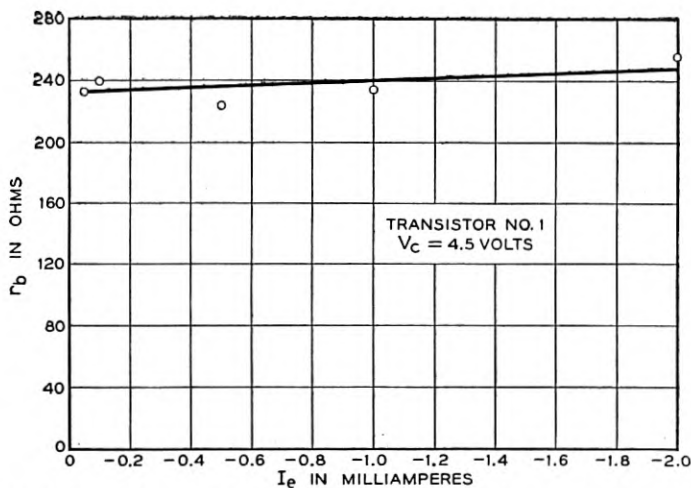


Fig. 9—Variation of base resistance with emitter current. Scatter of the data indicates that the measurements were not accurate.

is about  $80^\circ$  F., this reduces to

$$r_s = \frac{25.9}{I_e} \quad (2)$$

where  $I_e$  is measured in milliamperes. Within experiment error, values of  $r_s$  computed from this relation agree perfectly with the measured curve shown in Fig. 10.

Figure 11 introduces a new quantity,  $\alpha$ , the current amplification factor of the transistor. This quantity is defined by the equation

$$\alpha = \frac{r_m + r_b}{r_c + r_b} \quad (3)$$

Since  $r_m$  and  $r_c$  are both very large compared to  $r_b$ ,  $\alpha$  is approximately equal to the ratio of  $r_m$  to  $r_c$ . It will be shown in a later section that this quantity is important in determining some of the circuit properties of the transistor and that many of the circuit properties become more desirable as  $\alpha$  approaches unity.

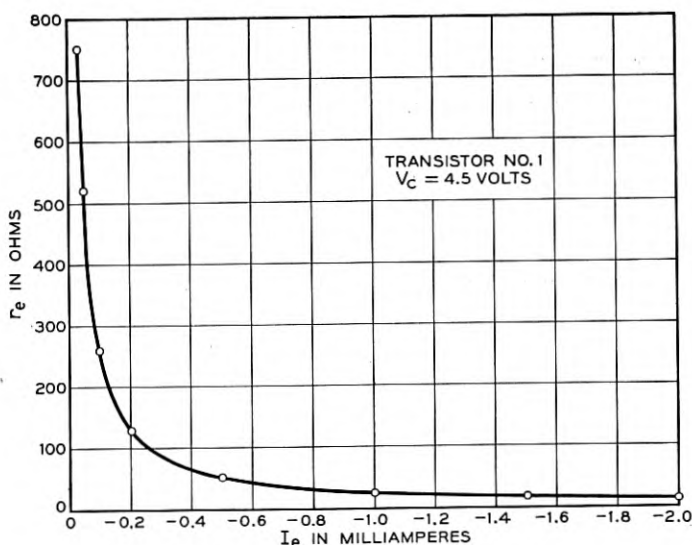


Fig. 10—The emitter resistance is inversely proportional to emitter current.

It can be seen from Fig. 11 that in this transistor  $\alpha$  is approximately equal to 0.98 and that it increases slightly with increasing emitter current. The highest value of  $\alpha$  so far achieved is 0.9965.

Those units which have been made in the laboratory so far show considerable variation in some of the properties, but this is partly due to the fact that changes have been made deliberately to test one aspect or another of Shockley's theory. The data in table I are presented to indicate what properties have been achieved to date. The collector capacitance  $C_c$  will be discussed in a later section.

#### GENERAL CONSIDERATIONS AND FORMULAE

It is a consequence of the fact that  $\alpha$  is always less than unity in this structure that these transistors are unconditionally stable with all termina-

tions. This means that stability considerations do not prevent working with matched terminations. Furthermore it is possible to obtain a variety of input and output impedances by connecting the transistor as a grounded-emitter, grounded-base, or grounded-collector stage. It is the purpose of this section to give some idea of the characteristics of these various stages and to show in each case at least one way of supplying the required biases and couplings to the stage.

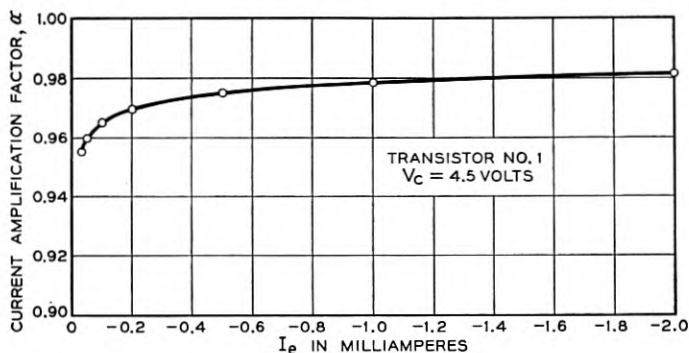


Fig. 11—The current amplification factor,  $\alpha$ , increases slightly with increasing emitter current. Note the expanded scale for  $\alpha$ .

TABLE I  
CONSTANTS FOR VARIOUS TRANSISTORS MEASURED AT  $V_c = 4.5$  v.,  $I_e = 1.0$  ma.

Transistor No.	I	II	III	IV	V
$r_e$ (ohms)	25.9	31.6	33.1	30.2	38.8
$r_b$ (ohms)	240	44	300	3070	180
$r_c$ (megohms)	13.4	0.626	1.11	1.21	2.00
$r_c - r_m$ (megohms)	0.288	0.00387	0.0168	0.00422	0.0439
$\alpha$	0.9785	0.9936	0.9848	0.9965	0.9780
$C_c$ ( $\mu\mu\text{f.}$ )	7	7.7	18.9	27.9	21.2

It will be convenient to begin by writing down general relationships which will apply to all the possible connections. To this end let the transistor be represented by the box in Fig. 12. At low frequencies, the signal currents and voltages are related through the equations:

$$\begin{aligned} R_{11}i_1 + R_{12}i_2 &= v_1 \\ R_{21}i_1 + R_{22}i_2 &= v_2 \end{aligned} \quad (4)$$

If a generator of open circuit voltage  $v_g$  and internal resistance  $R_g$  is connected to the input terminals of the device as shown in Fig. 13, then

$$v_1 = v_g - i_1 R_g \quad (5)$$

and if a load of resistance  $R_L$  is connected to the output terminals

$$v_2 = -R_L i_2. \quad (6)$$

The equations for the circuit of Fig. 13 are, therefore,

$$(R_{11} + R_g) i_1 + R_{12} i_2 = v_g \quad (7)$$

$$R_{21} i_1 + (R_{22} + R_L) i_2 = 0$$

Solving for the voltage developed across the load ( $= -R_L i_2$ ) gives

$$v_2 = \frac{R_L R_{21}}{(R_{11} + R_g)(R_{22} + R_L) - R_{12} R_{21}} v_g \quad (8)$$

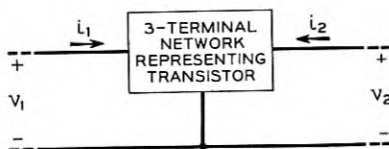


Fig. 12—A three-terminal network representing either grounded emitter, grounded base, or grounded collector connection of a transistor. Note the convention of signs.

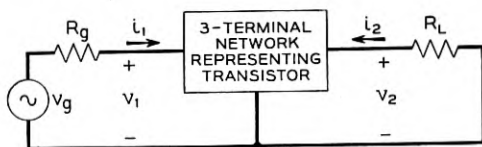


Fig. 13—The three-terminal network of Fig. 12 connected between a generator and a load.

The power gain in the circuit is the power delivered to the load ( $v_2^2/R_L$ ) divided by the power available from the generator ( $v_g^2/4R_g$ ). From equation (8), this gives

$$G = \frac{4R_g R_L R_{21}^2}{[(R_{11} + R_g)(R_{22} + R_L) - R_{12} R_{21}]^2} \quad (9)$$

The gain depends on  $R_g$  and  $R_L$  and will be maximum when these are chosen to match the input and output impedances of the transistor stage. But the input impedance depends on  $R_L$  and the output impedance depends on  $R_g$  in the following way:

$$\text{Input impedance} = R_i = R_{11} - \frac{R_{12} R_{21}}{R_{22} + R_L} \quad \text{and} \quad (10)$$

$$\text{Output impedance} = R_o = R_{22} - \frac{R_{12} R_{21}}{R_{11} + R_g} \quad (11)$$

If  $R_i = R_o$  and  $R_o = R_L$  then impedances are matched at the input and output terminals and the gain is a maximum. The conditions are:

Matched input impedance =

$$R_{im} = R_{11} \sqrt{1 - R_{12} R_{21} / R_{11} R_{22}}, \quad (12)$$

Matched output impedance =

$$R_{om} = R_{22} \sqrt{1 - R_{12} R_{21} / R_{11} R_{22}}, \quad (13)$$

Maximum available gain =

$$\text{M.A.G.} = \frac{R_{21}^2}{R_{11} R_{22}} \frac{1}{[1 + \sqrt{1 - R_{12} R_{21} / R_{11} R_{22}}]^2} \quad (14)$$

#### THE GROUNDED BASE STAGE

In this and the following two sections we will put into equations (7) through (14) the appropriate 4-pole  $r$ 's to obtain expressions for impedances and gains. As a numerical example we will substitute into the resulting equations the measured values of these  $r$ 's for Transistor No. I working at  $V_e = 4.5\text{v.}$  and  $I_e = 1\text{ ma.}$  It must be understood that the numerical values may vary appreciably from transistor to transistor and that these numerical calculations are intended only for illustration and not as a basis for final circuit design. The numerical values to be used are

$$r_e = 25.9 \text{ ohms}$$

$$r_b = 240 \text{ ohms}$$

$$r_c = 13.4 (10)^6 \text{ ohms} \quad (15)$$

$$r_c - r_m = 0.288 (10)^6 \text{ ohms}$$

$$\alpha = 0.9785$$

In this section it will be shown that the grounded base connection is suitable for working between a low impedance source and a high impedance load. The input impedance may be of the order of a hundred ohms and the output impedance of the order of one or more megohms. In this connection the current amplification is always less than unity but the voltage amplification may be very large indeed. Power gains of the order of 40 to 50 db can be obtained between matched impedances, and appreciable gains can still be obtained if the load resistance is reduced to a few thousand ohms (because the current gain is then almost equal to unity). In this case the gain of the stage is almost completely independent of those transistor properties

which tend to vary from unit to unit. This sort of stage does not produce a phase reversal.

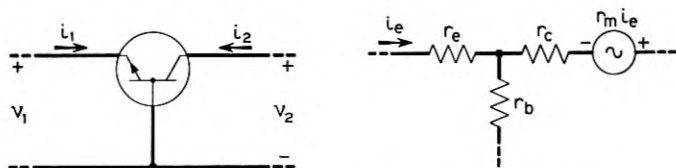


Fig. 14—The grounded base connection of a transistor.

For the grounded base stage shown in Fig. 14,

$$\begin{aligned}
 R_{11} &= r_e + r_b = 266 \text{ ohms} \\
 R_{12} &= r_b = 240 \text{ ohms} \\
 R_{21} &= r_m + r_b = 13.1 (10)^6 \text{ ohms} \\
 R_{22} &= r_c + r_b = 13.4 (10)^6 \text{ ohms} \quad (16) \\
 \alpha &= \frac{r_m + r_b}{r_c + r_b} \\
 &\doteq \frac{r_m}{r_c} \\
 &= 0.9785
 \end{aligned}$$

In this case if  $r_b$  is neglected by comparison with  $r_m$  and  $r_c$ , equation (8) leads to

$$v_2 = \frac{\alpha R_L v_g}{(r_e + r_b + R_g)(1 + R_L/r_c) - \alpha r_b} \quad (17)$$

Since for these transistors  $\frac{r_m}{r_c} (\doteq \alpha)$  is always less than unity, the output voltage is in phase with the input voltage. Furthermore, if  $R_L$  is very high, the output voltage is enormous by comparison with the input voltage. For example, if  $R_g = 0$  and  $R_L$  is infinite

$$v_2 = v_g \frac{r_m}{r_c + r_b} \quad (18)$$

and for the numerical example this is

$$v_2 = 4.93 (10)^4 v_g.$$



To achieve this step-up would require a load impedance very large compared to 13 megohms, but even with more modest values of load impedance the voltage step-up is large.

If  $R_L$  is small compared to  $r_c$ , the second of equations (7) leads to

$$\begin{aligned} i_2 &= -\frac{r_m}{r_c} i_1 \\ &\doteq -i_1 \end{aligned} \quad (19)$$

and the current delivered to the load is approximately equal to the current which the generator delivers to the transistor.

From equations (10) and (11), the input and output impedances are

$$R_i = r_e + r_b - \frac{r_b(r_m + r_b)}{r_c + R_L + r_b} \quad (20)$$

$$R_o = r_c + r_b - \frac{r_b(r_m + r_b)}{r_e + r_b + R_o} \quad (21)$$

As the load impedance varies from zero to infinity, the input impedance varies from

$$\begin{aligned} R_i &= r_e + r_b \left[ 1 - \frac{r_m + r_b}{r_c + r_b} \right] \quad \text{for } R_L = 0 \\ &\doteq r_e + r_b(1 - \alpha) \\ &= 31.1 \text{ ohm} \end{aligned} \quad (22)$$

to

$$R_i = r_e + r_b = 266 \text{ ohms} \quad \text{for } R_L = \infty. \quad (23)$$

When  $R_o = 0$ , the output impedance is

$$R_o = r_c - \frac{r_b}{r_e + r_b} (r_m - r_e) \quad (24)$$

$$\doteq r_c - \frac{r_b}{r_e + r_b} r_m \quad (25)$$

$$= 1.56 (10)^6 \text{ ohms.}$$

As  $R_o$  increases to infinity

$$R_o = r_c + r_b = 13.4 (10)^6 \text{ ohms.} \quad (26)$$

From equations (12) and (13), the matched input and output impedances are approximately

$$\begin{aligned} R_{im} &= (r_e + r_b) \sqrt{1 - \alpha r_b / (r_e + r_b)} \\ &= 91 \text{ ohms} \end{aligned} \quad (27)$$

$$\begin{aligned} R_{om} &= (r_c + r_b) \sqrt{1 - \alpha r_b / (r_e + r_b)} \\ &= 4.58(10)^6 \text{ ohms.} \end{aligned} \quad (28)$$

With matched impedances, the maximum available gain is

$$\begin{aligned} \text{M.A.G.} &= \frac{\alpha(r_m + r_b)}{r_e + r_b} [1 + \sqrt{1 - \alpha r_b / (r_e + r_b)}]^{-2} \\ &= 2.7 (10)^4 \text{ or } 44.3 \text{ db.} \end{aligned} \quad (29)$$

The matched output impedance of this stage is inconveniently high but a useful amount of gain can be maintained if  $R_L$  is reduced to a more rea-

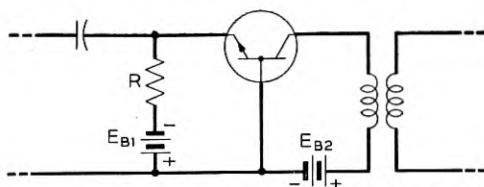


Fig. 15—One practical arrangement of a grounded base amplifier stage.

sonable value. For example, if  $R_L = 200,000$  and  $R_g = 25$ , equation (9) gives

$$G = 5.3 (10)^3 \text{ or } 37.2 \text{ db.}$$

If stages of this sort are to be cascaded, a step-down transformer must be used to couple each collector to the following emitter. Otherwise, since the current amplification factor of the transistor is slightly less than unity, the gain per stage will also be slightly less than unity.

One practical arrangement of a grounded base stage would be as shown in Fig. 15. The required value of  $R$  will be approximately

$$R = \frac{E_{B1}}{I_c} \quad (30)$$

where  $I_c$  is the desired collector current and  $E_{B1}$  is the voltage of the emitter-bias battery. For operating at  $I_c = 1 \text{ ma}$ , for example,  $E_{B1} = 1.5 \text{ v}$  and  $R = 1500 \text{ ohms}$  would be suitable.

## THE GROUNDED EMITTER STAGE

For many applications the grounded emitter connection is more desirable than either of the other two. The power gains which can be obtained are high—of the order of 50 db—and the interstage coupling problem is simplified by the fact that the input impedance is somewhat higher than that of the grounded base stage while the output impedance is very much lower. The input impedance may be of the order of a few hundred ohms and the output impedance of the order of a few hundred thousand ohms. Both voltage and current amplification are produced (with a phase reversal) and gains of the order of 30 db or more per stage can be obtained without the

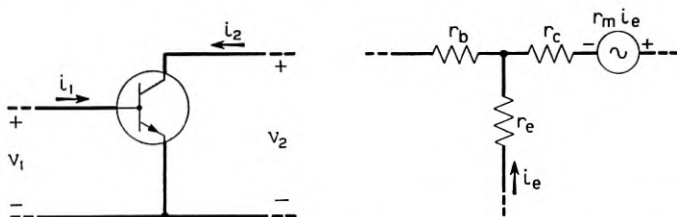


Fig. 16—The grounded emitter connection of a transistor and the equivalent circuit

use of interstage coupling transformers. The input and output impedance depend very critically on  $\alpha$  and may vary appreciably from unit to unit

For this connection, which is indicated schematically in Fig. 16,

$$\begin{aligned}
 R_{11} &= r_e + r_b = 266 \text{ ohms} \\
 R_{12} &= r_e = 25.9 \text{ ohms} \\
 R_{21} &= r_e - r_m = -13.1 (10)^6 \text{ ohms} \\
 R_{22} &= r_e + r_c - r_m = 0.288 (10)^6 \text{ ohms}
 \end{aligned}
 \tag{31}$$

Putting these values into equation (8) shows  $v_2$  is always opposite in sign compared with  $v_o$ , that is, that the grounded emitter stage produces a phase reversal as does the grounded cathode vacuum tube.

If  $R_L$  is infinite and  $R_o = 0$

$$\begin{aligned}
 v_2 &= v_o \frac{r_e - r_m}{r_e + r_b} \\
 &= -4.93 (10)^4 v_o
 \end{aligned}$$

which is the same as for the grounded base stage. But if  $R_L = 0$

$$i_2 = \frac{r_m - r_e}{r_e + r_c - r_m} i_1 \quad (32)$$

$$\doteq \frac{\alpha}{1 - \alpha} i_1 \quad (33)$$

$$= 45.5 i_1.$$

Thus it is seen that the grounded emitter amplifier can produce quite appreciable current amplification—particularly so when  $\alpha$  approaches unity.

The input impedance to the stage is

$$R_i = r_e + r_b + \frac{r_e(r_m - r_e)}{r_e + r_c - r_m + R_L}. \quad (34)$$

When  $R_L = 0$  this reduces to

$$R_i = r_b + r_e \frac{1}{\frac{r_e}{r_c} + 1 - \frac{r_m}{r_c}} \quad (35)$$

$$\doteq r_b + r_e \frac{1}{1 - \alpha} \quad (36)$$

$$= 1440 \text{ ohms.}$$

As  $R_L$  increases to infinity, the input impedance decreases to  $r_e + r_b$  which, for the numerical example, is 266 ohms.

The output impedance is

$$R_o = r_e + r_c - r_m + \frac{r_e(r_m - r_e)}{r_e + r_b + R_g} \quad (37)$$

When  $R_g = 0$ , this gives

$$R_o = r_c - \frac{r_b}{r_e + r_b} (r_m - r_e) \quad (38)$$

$$\doteq r_c \left[ 1 - \frac{r_b}{r_e + r_b} \alpha \right] \quad (39)$$

$$= 1.56 (10)^6 \text{ ohms}$$

As  $R_g$  increases to infinity,  $R_o$  decreases to

$$R_o = r_e + r_c - r_m \quad (40)$$

$$= 0.288 (10)^6 \text{ ohms.}$$

The matched input and output impedances are

$$R_{im} = (r_e + r_b) \sqrt{1 + r_e(r_m - r_e)/(r_e + r_b)(r_e + r_c - r_m)} \quad (41)$$

= 619 ohms and

$$R_{om} = (r_e + r_c - r_m) \sqrt{1 + r_c(r_m - r_e)/(r_e + r_b)(r_e + r_c - r_m)} \quad (42)$$

= 0.671 (10)<sup>6</sup> ohms

As  $\alpha$  increases toward unity the matched input impedance increases and the matched output impedance decreases. They approach the limits

$$R_{im} = \sqrt{(r_b + r_e)(r_e + r_b)} \quad (43)$$

$$R_{om} = r_e \sqrt{(r_b + r_c)/(r_e + r_b)} \quad (44)$$

as  $\alpha \rightarrow 1$ .

If  $r_m$  in the transistor of our numerical examples could be increased to exactly the value of  $r_c$  ( $\alpha = 1$ ) then the matched impedances would be

$$R_{im} = 59,700 \text{ ohms}$$

$$R_{om} = 5,800 \text{ ohms}$$

From this example, it is seen that the impedances vary rapidly with  $\alpha$  as  $\alpha$  approaches unity.

With matched impedances, the maximum available gain from the grounded emitter stage is

$$\begin{aligned} \text{M.A.G.} &= \frac{(r_e - r_m)^2}{r_e r_c} \left[ \sqrt{\left(1 + \frac{r_b}{r_e}\right) \left(\frac{r_e}{r_c} + 1 - \frac{r_m}{r_c}\right)} \right. \\ &\quad \left. + \sqrt{1 + \frac{r_b}{r_e} \left(\frac{r_e}{r_c} + 1 - \frac{r_m}{r_c}\right)} \right]^{-2} \quad (45) \\ &= 2.02 (10)^5 \text{ or } 53 \text{ db} \end{aligned}$$

When  $\alpha$  is exactly unity this expression reduces to  $r_c/r_e$  provided  $r_e$  and  $r_b$  are small compared with  $r_c$ . For values of  $\alpha$  which are enough smaller than unity so that

$$\frac{r_e}{r_c} \ll 1 - \alpha$$

the expression for maximum available gain reduces to the approximate expression

$$\text{M.A.G.} = \alpha(r_m/r_c) \left[ \sqrt{(1 - \alpha) \frac{r_b}{r_e}} + \sqrt{1 + (1 - \alpha) \frac{r_b}{r_e}} \right]^{-2} \quad (46)$$

base and ground as shown, for example, in Fig. 18. Since the base floats at a positive potential with respect to ground, this circuit produces a base current of the right sign to decrease the collector current. As the value of the series resistance is decreased to zero, the collector current decreases to

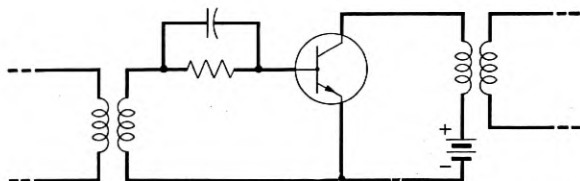


Fig. 18—Modification of Fig. 17 to obtain lower collector current.

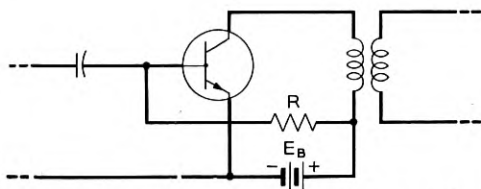


Fig. 19—Modification of Fig. 17 to obtain higher collector current.

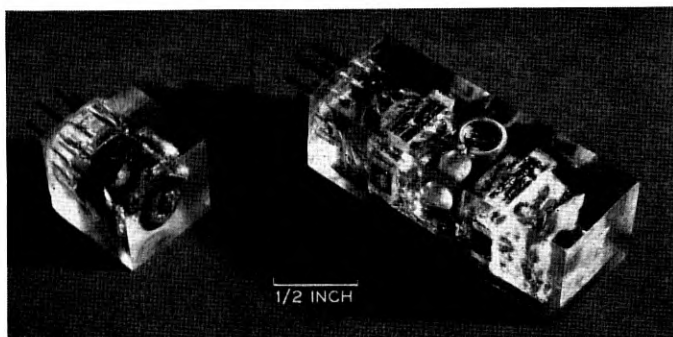


Fig. 20—A two-stage grounded emitter amplifier which produces approximately 90 db power gain is shown on the right and a micro-power audio oscillator is shown on the left.

a value corresponding to zero emitter voltage. A still further decrease in collector current can be obtained by inserting resistance between emitter and ground.

In order to increase the collector current to values higher than that corresponding to zero base current, a high resistance path between base and the positive supply voltage may be used as shown in Fig. 19. In this case the

collector current will increase by  $1/(1 - \alpha)$  microamperes for each microampere which flows through the bias resistor. Since the current in the bias resistor will be approximately  $E_B/R$ , it is a simple matter to compute the required value of bias resistor once the desired collector current is known.

Figure 20 shows a two-stage audio amplifier which gives approximately 90 db gain. The circuit is shown in Fig. 21.

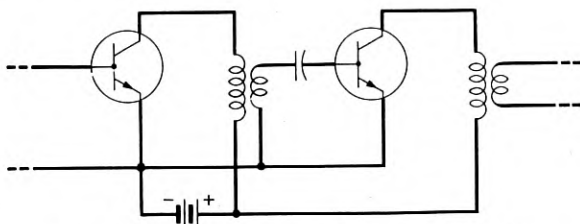


Fig. 21—Circuit of the amplifier shown in Fig. 20.

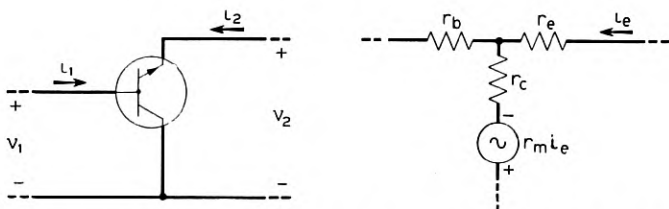


Fig. 22—The grounded collector connection of a transistor and the equivalent circuit.

### THE GROUNDED-COLLECTOR STAGE

Although the power gain obtainable from this connection is relatively low—of the order of 15 or 20 db—it has very interesting possibilities in producing very high input impedances or very low output impedances. If it is worked into a fairly high load impedance, the input impedance may be several megohms, or if it is worked from a source of moderately low impedance (a few thousand ohms), the output impedance may be of the order of 25 ohms or lower.

For this type of stage, which is shown schematically in Fig. 22,

$$R_{11} = r_b + r_c = 13.4 (10)^6 \text{ ohms}$$

$$R_{12} = r_c - r_m = 0.288 (10)^6 \text{ ohms}$$

$$R_{21} = r_c = 13.4 (10)^6 \text{ ohms}$$

$$R_{22} = r_e + r_c - r_m = 0.288 (10)^6 \text{ ohms}$$

(48)

If this stage is worked from a zero impedance generator into an infinite impedance load

$$v_2 = v_g \left( \frac{r_c}{r_b + r_c} \right) \quad (49)$$

$$\doteq v_g$$

and so, like a cathode follower, it gives an output voltage which is less than the input voltage, but in the same phase.

If the stage is operated into a short circuit

$$i_2 = -i_1 \frac{r_c}{r_e + r_c - r_m} \quad (50)$$

$$\doteq -i_1 \frac{1}{1 - \alpha} \quad (51)$$

$$= -46.5 i_1$$

which indicates that the stage can give an appreciable current gain.

The input impedance is

$$R_i = r_b + r_c - \frac{r_c(r_c - r_m)}{r_e + r_c - r_m + R_L} \quad (52)$$

When  $R_L = 0$ , this reduces to

$$R_i = r_b + r_e \frac{1}{\frac{r_e}{r_c} + 1 - \frac{r_m}{r_c}} \quad (53)$$

$$\doteq r_b + r_e \frac{1}{1 - \alpha} \quad (54)$$

$$= 1445 \text{ ohms.}$$

When  $R_L$  is infinite

$$R_i = r_b + r_c \quad (55)$$

$$= 13.4(10)^6 \text{ ohms.}$$

With respect to input impedance, the grounded collector stage is again seen to be like a cathode follower in that the input impedance is high when the load impedance is high.

The output impedance is

$$R_o = r_e + r_c - r_m - \frac{r_c(r_c - r_m)}{r_b + r_c + R_g} \quad (56)$$



For  $R_g = 0$ , this reduces to

$$R_o = r_e + r_b \frac{r_c - r_m}{r_b + r_c} \quad (57)$$

$$\begin{aligned} &\doteq r_e + r_b (1 - \alpha) \\ &= 31.1 \text{ ohms.} \end{aligned} \quad (58)$$

For  $R_g$  infinite

$$\begin{aligned} R_o &= r_e + r_c - r_m \\ &= 0.288(10)^6 \text{ ohms.} \end{aligned} \quad (59)$$

The matched input impedance is

$$R_{im} = (r_b + r_c) \sqrt{\frac{r_b}{r_b + r_c} + \frac{r_c r_e}{(r_b + r_c)(r_e + r_c - r_m)}} \quad (60)$$

$$\begin{aligned} &\doteq \sqrt{r_c[r_b + r_e/(1 - \alpha)]} \\ &= 139,000 \text{ ohms.} \end{aligned} \quad (61)$$

The matched output impedance is

$$R_{om} = (r_e + r_c - r_m) \sqrt{\frac{r_b}{r_b + r_c} + \frac{r_c r_e}{(r_b + r_c)(r_e + r_c - r_m)}} \quad (62)$$

$$\begin{aligned} &\doteq (1 - \alpha) \sqrt{r_c[r_b + r_e/(1 - \alpha)]} \\ &= 2990 \text{ ohms.} \end{aligned} \quad (63)$$

With matched impedance, the maximum available gain of the grounded collector stage is

$$\text{M.A.G.} = \frac{\frac{r_c^2}{(r_b + r_c)(r_e + r_c - r_m)}}{\left[1 + \sqrt{\frac{r_b}{r_b + r_c} + \frac{r_c r_e}{(r_b + r_c)(r_e + r_c - r_m)}}\right]^2} \quad (64)$$

As  $\alpha$  approaches unity, this approaches approximately

$$\text{M.A.G.} = r_c/4r_e \quad (65)$$

but so long as  $r_e \ll r_c - r_m$ , a good approximation is

$$\begin{aligned} \text{M.A.G.} &= 1/(1 - \alpha) \\ &= 46.5 \text{ or } 16.7 \text{ db.} \end{aligned} \quad (66)$$

The considerations involved in supplying biases to a grounded collector stage are rather similar to those discussed already for the grounded emitter case. If the base is allowed to float, the collector current will be given approximately by equation (47) as discussed for the grounded emitter case. A resistance between base and the negative side of the supply battery in Fig. 24 will serve to decrease the collector current while a resistance between base and ground will serve to increase it. In applications where it is desired to make full use of the high input impedance which this stage can afford, it may be most desirable to let the base float as shown in Fig. 23.

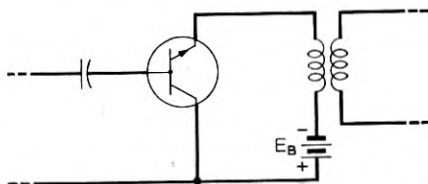


Fig. 23—One practical arrangement of a grounded collector stage.

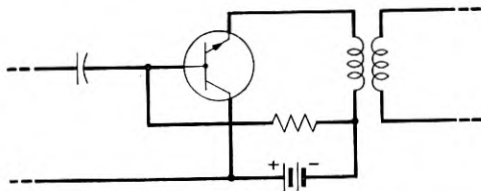


Fig. 24—Modification of Fig. 23 to obtain lower collector current. To raise collector current remove the resistance shown and connect a high resistance between base and ground.

#### FREQUENCY RESPONSE—GENERAL REMARKS

Shockley has shown that there are several different physical considerations which lead one to expect a high-frequency cutoff in the response of  $n-p-n$  transistors. The frequency at which cutoff occurs depends in a theoretically understandable way on such things as the geometry of the transistor and the physical properties of the germanium from which it is made. If these factors could all be controlled and varied at will, it would be possible to design a transistor to have a specified cutoff frequency.

One limitation comes about in the following way: In order to produce transistor action, the electrons which are injected into the  $p$  layer at the emitter junction must travel across this thin layer and arrive at the collector junction. They do this principally by a process of diffusion and require a finite (but small) amount of time to make the journey. If this time were

exactly the same for all electrons, the effect would be simply to delay the output signal with respect to the input and there would be no effect on frequency response. But there is a certain amount of dispersion in transit time which means that the electrons corresponding to a particular part of the input signal wave do not all arrive simultaneously at the collector. When this difference in time of arrival amounts to an appreciable part of a cycle there is a tendency for some of the electrons to cancel the effect of others so that the frequency response begins to fall off. As the signal frequency increases beyond this point, the effect becomes more and more pronounced and the response continues to fall with increasing frequency.

In terms of the equivalent circuit, this dispersion in transit time means that beyond a certain frequency,  $r_m$  (and hence  $\alpha$ ) begins to decrease with increasing frequency and so the transistor may be said to have a certain  $\alpha$ -cutoff which we will call  $f_{c\alpha}$ .

Shockley has shown that  $f_{c\alpha}$  is inversely proportional to the square of the  $p$ -layer thickness and hence increases rapidly as the  $p$  layer is made thinner. For  $n$ - $p$ - $n$  transistors now available, this cutoff should occur at frequencies between five and twenty megacycles.

Another limitation on frequency response comes about from the fact that, at sufficiently high frequencies, the emitter junction fails to behave as a pure resistance and is, in effect, shunted by a capacitance. In terms of the equivalent circuit, this means that  $r_e$  is shunted by a capacitance.

The effect which this has on frequency response can be reduced by reducing the impedance of the source from which the emitter is driven. But so far as the emitter junction is concerned,  $r_b$  is always in series with the source impedance and so it is the value of  $r_b$  which ultimately determines the emitter cutoff frequency.

This capacitative reactance should begin to become appreciable with respect to emitter resistance at a frequency which may be of the same order as  $f_{c\alpha}$ . If  $r_b$  is high, the emitter cutoff frequency  $f_{ce}$  will then be of the same order of magnitude as  $f_{c\alpha}$  and will increase as  $r_b$  is decreased.

A third cause for limited frequency response is the capacitance of the collector junction. The  $n$ -type germanium on one side of the junction behaves as one plate of a parallel-plate condenser and the  $p$ -type germanium on the other side behaves as the other plate. Since the transition from  $n$  to  $p$  type germanium may be made in an exceedingly small fraction of an inch, the plates of the condenser are very closely spaced and the capacitance may be appreciable.

Collector capacitance also depends on collector voltage, decreasing with increasing voltage. Theoretically, the capacitance should be in proportion to the negative one-third power of  $V_c$ .

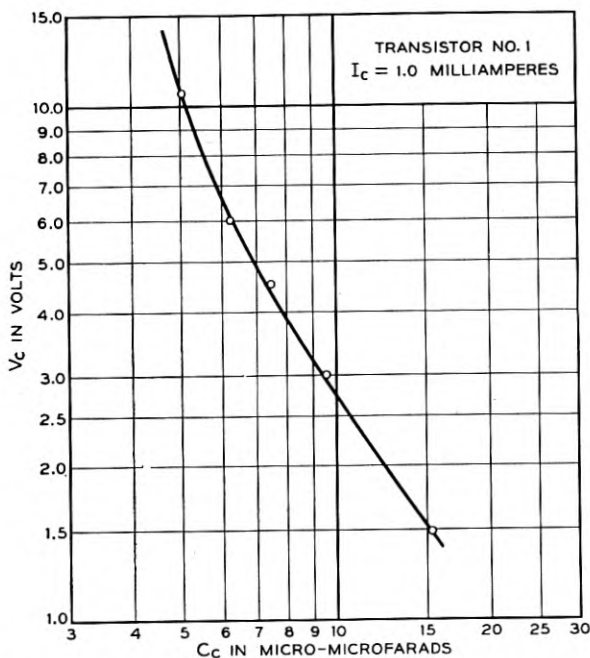


Fig. 25—Collector capacitance decreases as collector voltage is increased.

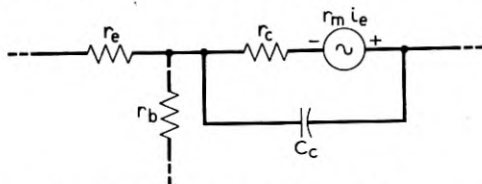


Fig. 26—The equivalent circuit of a transistor with collector capacitance shown.

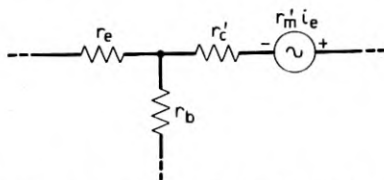


Fig. 27—The effect of collector capacitance is to change  $r_m$  and  $r_c$  to  $r'_m$  and  $r'_c$ . See equations (67) and (68).

Figure 25 shows measured values of  $C_c$  as a function of collector voltage. For reasons which are not understood at present, these data show a departure from the usual inverse one-third power variation. At  $V_c = 4.5$  volts

the capacitance is seen to be approximately 7 micro-microfarads. In terms of the equivalent circuit, this capacitance is in shunt with the series combination of  $r_c$  and the generator,  $r_m i_e$ , as shown in Fig. 26. This can be shown to be equivalent to the circuit of Fig. 27 in which  $r_c$  has been replaced by

$$r_c' = r_c / (1 + jC_c r_c \omega) \quad (67)$$

and  $r_m$  has been replaced by

$$r_m' = r_m / (1 + jC_c r_c \omega). \quad (68)$$

The effect of collector capacitance can be computed by substituting  $r_c'$  and  $r_m'$  for the values of  $r_m$  and  $r_c$  (implicitly contained) in equation (8). In the sections which follow, this will be done for each of the three transistor connections and the resulting collector cutoff frequencies  $f_{cc}$  will be computed. It will be shown that at least for the transistor on which data are presented collector capacitance tends to produce a cutoff frequency well below those to be expected from emitter cutoff or alpha cutoff. For this reason, only collector cutoff will be considered.

#### COLLECTOR CUTOFF IN THE GROUNDED BASE STAGE

If the values of  $r_c'$  and  $r_m'$  from equations (67) and (68) are substituted for  $r_c$  and  $r_m$  in equations (16) and the resulting values of the  $R$ 's are substituted into equation (8), the result is

$$v_2/v_g = \frac{\alpha R_L}{(r_e + r_b + R_g)[1 + R_L(1 + j\omega C_c r_c)/r_c]} \quad (69)$$

The cutoff frequency  $f_{cc}$  is defined as the frequency at which the voltage across the load has dropped 3 db compared to its low-frequency value. This is the frequency at which the imaginary part of the denominator of (69) is equal to the real part. Solving for  $f_{cc}$  gives

$$f_{cc} = \frac{1}{2\pi C_c} \left[ \frac{1}{R_L} + \frac{1}{r_c} - \frac{\alpha r_b}{R_L(r_e + r_b + R_g)} \right] \quad (70)$$

Substituting into this equation  $C_c = 7(10)^{-12}$  farad, numerical values of the  $r$ 's from (16), and the values  $R_g = 91$  and  $R_L = 4.58(10)^6$  ohms, corresponding to maximum available gain gives

$$f_{cc} = 3390 \text{ cps.}$$

With these terminations, the low-frequency gain is 44.3 db. If  $R_g$  and  $R_L$  are reduced to 25 and 200,000 ohms, respectively,  $f_{cc}$  is raised to 23,500 cps and the gain is lowered to 37.2 db. A further reduction of  $R_L$  to 20,000 ohms increases  $f_{cc}$  to 0.22 megacycles and reduces the gain to 27.8 db. This corre-

sponds to a gain-bandwidth product of  $1.2(10)^8$  cps and shows that useful gain could be obtained at frequencies well above a megacycle, *provided* alpha and emitter cutoffs did not interfere.

#### COLLECTOR CUTOFF IN THE GROUNDED-EMITTER STAGE

The procedure described in the last section leads, in this case, to

$$v_2/v_1 = \frac{-R_L r_m/r_c + (R_L r_c/r_c)(1 + jr_c C_c \omega)}{r_e + (r_b + R_\theta)(1 - r_m/r_c) + [r_e R_L/r_c + (r_b + R_\theta)(r_e + R_L)/r_c](1 + jr_c C_c \omega)} \quad (71)$$

For the transistor of our numerical example, the imaginary term in the numerator is completely negligible at frequencies below  $(10)^9$  cps. Neglecting it leads to

$$f_{cc} = \frac{1}{2\pi C_c} \frac{1 + R_L/r_c + [(r_b + R_\theta)/r_c][1 - (r_m - r_e - R_L)/r_c]}{R_L + [(r_b + R_\theta)/r_c](r_e + R_L)} \quad (72)$$

In this case the values of  $R_\theta$  and  $R_L$  (619 ohms and 671,000 ohms respectively) which correspond to maximum available gain give

$$f_{cc} = 3740 \text{ cps and}$$

$$\text{M.A.G.} = 53 \text{ db.}$$

Reducing  $R_L$  to 100,000 and increasing  $R_\theta$  to 1000 ohms gives

$$f_{cc} = 11,120 \text{ cps}$$

$$G = 50 \text{ db.}$$

For  $R_\theta = 1000$  and  $R_L = 10,000$ ,

$$f_{cc} = 97,900 \text{ cps}$$

$$G = 41.3 \text{ db.}$$

and for  $R_\theta = R_L = 1000$  ohms,

$$f_{cc} = 943,000 \text{ cps}$$

$$G = 31.4 \text{ db.}$$

The gain-bandwidth product for this stage is  $1.3(10)^9$  cps as compared to  $1.2(10)^8$  cps for the same transistor connected as a grounded base amplifier. It should be pointed out, however, that this stage is particularly sensitive to change in  $\alpha$  and on this account alpha cutoff may influence the response at fairly low frequencies. For example, when the terminating resistances are both 1000 ohms, reducing  $\alpha$  from 0.9785 to 0.900 reduces the gain from 31.4 db to 0.2 db.

## COLLECTOR CUTOFF IN THE GROUNDED COLLECTOR STAGE

In this case

$$v_2/v_0 = \frac{R_L}{[r_e + R_L + (r_b + R_0)(1 - r_m/r_c)] + (1/r_c)(r_b + R_0)(r_e + R_L)(1 + j\omega C_c r_c)} \quad (73)$$

and

$$f_{cc} = \frac{1}{2\pi C_c} \left[ \frac{1}{r_c} + \frac{1}{r_b + R_0} + \frac{1 - r_m/r_c}{r_e + R_L} \right] \quad (74)$$

For matched impedances ( $R_0 = 139,000$  ohms and  $R_L = 2990$  ohms),

$$f_{cc} = 320,000 \text{ cps}$$

$$G = 16.7 \text{ db.}$$

The cutoff frequency can be raised by decreasing either  $R_0$  or  $R_L$ . With  $R_0 = 139,000$  and  $R_L = 25$  ohms

$$f_{cc} = 9.77 \text{ megacycles}$$

$$G = 1.8 \text{ db}$$

The gain-bandwidth product in this case is  $1.5(10)^7$ .

### NOISE

The data now available on noise are insufficient to give an adequate picture of the performance of *n-p-n* transistors in this respect. Such measurements as have been made, however, make it clear that these devices are very much quieter than early point-contact transistors reported on by Ryder and Kircher.

Transistor noise seems still to decrease with increasing frequency at a rate of something like 11 db per decade. It also decreases as the thickness of the *p* layer is decreased.

Of the order of half a dozen units of various dimensions have been measured at 1000 cps and have shown noise figures as low as 8 db and as high as 25 db.

The dependence of noise figure on operating point has been measured for only one transistor. As indicated in Fig. 28 and Fig. 29, these data show that the noise figure improves as  $V_c$  is reduced and that it may be roughly independent of collector current. These data were taken on a grounded emitter stage with impedance match at the input terminals. Noise figure for this connection varies slightly with source impedance and has been found

to be a minimum when the source impedance is roughly equal to the input impedance of the stage.

It must be emphasized that this functional dependence of noise figure on operating point and source impedance has been measured for only one transistor. Further measurements may show that these results are not typical.

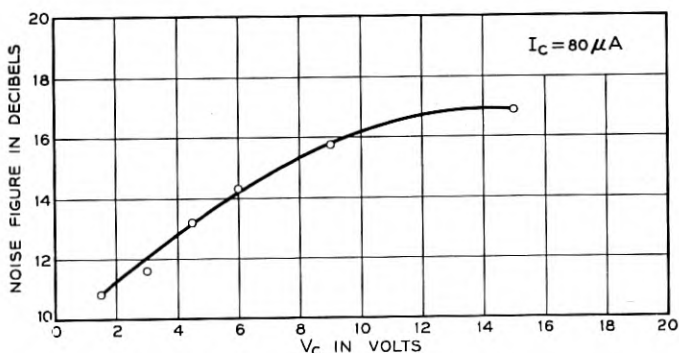


Fig. 28—Noise figure increases with increasing collector voltage.

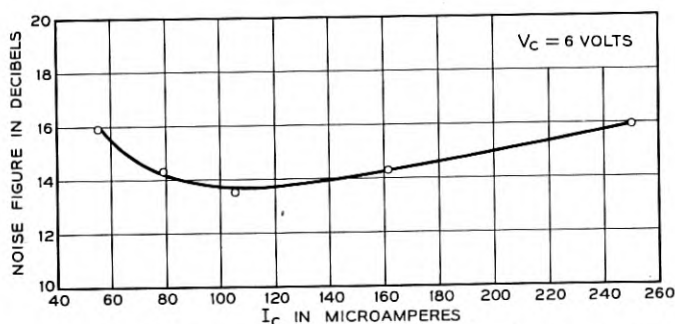


Fig. 29—Noise figure does not vary much with collector current.

#### FINAL COMMENTS

In this paper we have attempted to present what is known about the circuit performance of *n-p-n* transistors. Since these devices are still undergoing exploratory development and since only a limited number has been produced, it is obviously impossible to give statistical data on reproducibility or on such reliability factors as the effect of ambient temperature.

It is much too soon to know what properties may be achieved after further development, but the results obtained to date seem encouraging and worth reporting.



## ACKNOWLEDGMENT

The authors are happy to acknowledge their indebtedness to W. Shockley who was first to conceive the  $n$ - $p$ - $n$  transistor and has provided much of the inspiration and guidance which has made its physical realization possible. His comments have been of great help in the preparation of this paper.

We are also much indebted to J. A. Morton for his encouragement and helpful guidance and to M. Sparks for providing most of the transistors which have been studied. We wish to thank L. O. Schott, L. C. Geiger, and K. D. Smith for taking some of the data presented and to thank G. Raisbeck and L. G. Schimpf for proofreading and correcting the manuscript.

## REFERENCES

1. W. Shockley, "The Theory of  $p$ - $n$  Junctions in Semiconductors and  $p$ - $n$  Junction Transistors," *B. S. T. J.*, XXVIII, 435 (1949).
2. W. Shockley, "Electrons and Holes in Semiconductors," Van Nostrand (1950).
3. F. S. Goucher, G. L. Pearson, M. Sparks, G. K. Teal, W. Shockley, "Theory and Experiment for a Germanium  $p$ - $n$  Junction," *Phys. Rev.*, 81, 637 (1951).
4. W. Shockley, M. Sparks, G. K. Teal, " $p$ - $n$  Transistors," *Phys. Rev.*, 83, 151 (1951).
5. R. M. Ryder, R. J. Kircher, "Some Circuit Aspects of the Transistor," *B. S. T. J.*, XXVIII, 367 (1949).

# A Photographic Method for Displaying Sound Wave and Microwave Space Patterns

By W. E. KOCK and F. K. HARVEY

(Manuscript Received Oct. 27, 1950)

A photographic method using mechanical scanning for displaying the space patterns of sound and microwaves is described. A probe pick-up scans the sound or microwave field and the amplified probe output controls the brilliance of a small lamp affixed to the probe. A camera set at time exposure records the light intensity variations of the lamp as it moves across the scanned field, forming a pattern on the film of the amplitude distribution. Phase fronts can be delineated by adding a constant amplitude signal to the probe output. Photographs are included which show: sound and microwave patterns of lenses, diffraction at a straight edge and disk, refraction by a prism, diffusion of sound by a divergent lens, and radiation from loud speakers. Also, by transposition of source and receiver, directional patterns of transducers acting as microphones are obtained which (by reciprocity) appear identical with their radiation patterns. This provides a means for examining the directional characteristics of non-reversible transducers such as a carbon microphone. A calibration method is described which allows the relative value of the field intensities to be determined.

## INTRODUCTION

In analyzing the performance of an acoustic or microwave radiator it is helpful to know the way in which the waves proceed as they emerge from the source. It is desirable in some cases to have a photographic record of the distribution of intensity in the field generated by the radiator. This paper describes a simple mechanical scanning method for accomplishing this result. For acoustic analysis, a probe microphone is moved back and forth through the sound field to be explored and a small lamp is affixed to the microphone. When the output of the microphone is connected to the lamp through an amplifier, the intensity of the light varies in accordance with the sound level encountered by the probe microphone. A camera set at time exposure records the variations of light intensity. In this way the desired picture is built up by scanning the sound field somewhat after the manner in which a television image is formed.<sup>1</sup> For analyzing microwave fields the microphone is replaced with a microwave pickup probe.

## EXPERIMENTAL

The scanning device is shown in Fig. 1.<sup>2</sup> The microphone is located at the end of the rocking arm. Attached to it (at the left) is a small neon lamp

<sup>1</sup> A similar procedure in which the probe output, instead of lighting a lamp and being photographically recorded, is traced on spark paper (Teledeltos paper) has been employed for microwave presentations by H. Iams, "A Phase Front Plotter for Centimeter Waves," *R.C.A. Review*, 8, 270 (1947); also *Proc. I.R.E.*, 38, 543 (1950).

<sup>2</sup> This device was designed by Mr. T. Aamodt of these Laboratories.

which is energized by the amplified output of the microphone. As the rocker arm moves through its vertical motion the entire assembly including the motor and gear reducer is caused to move slowly towards the reader by the

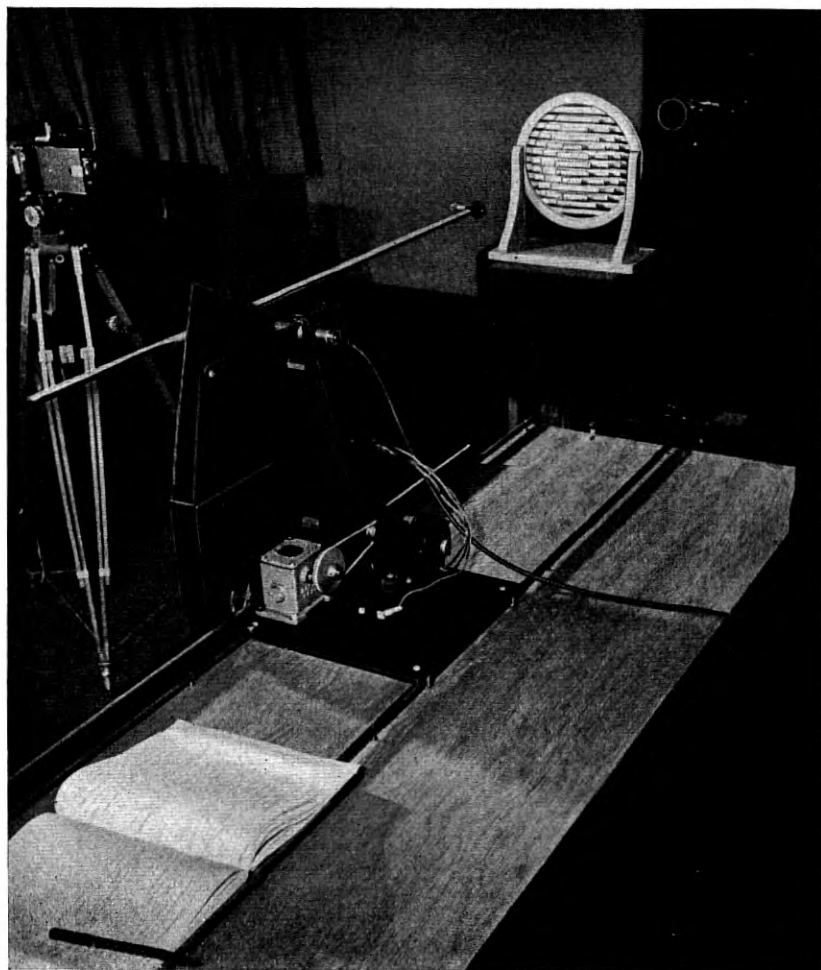


Fig. 1—The scanning mechanism set up for photographing the sound field in front of an acoustic lens.

rack and pinion shown at the left of the photograph. An acoustic lens<sup>3</sup> is seen in its test position between the end of the rocking arm and a tweeter

<sup>3</sup> W. E. Kock and F. K. Harvey, "Refracting Sound Waves," *Jour. Acous. Soc. Am.*, 21, 471 (1949).

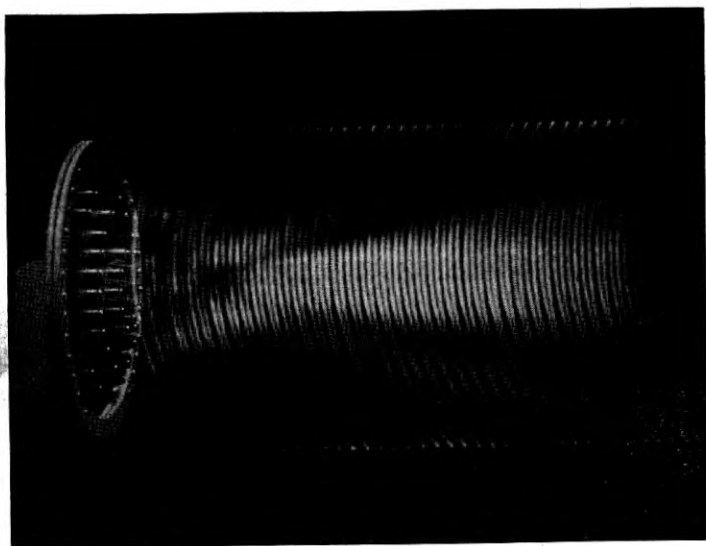


Fig. 2—An early photo of a sound field in which the scanning strokes were too coarse.

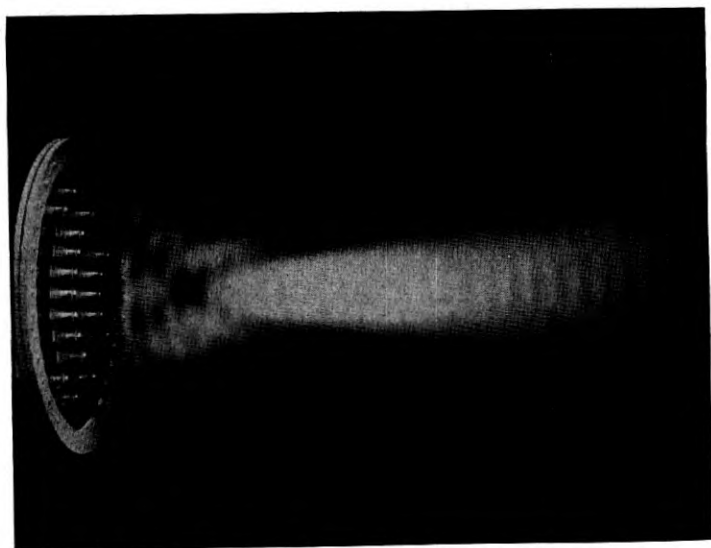


Fig. 3—Finer grained scanning produces a smooth pattern of the 10" acoustic lens of Fig. 2.  $f = 9$  KC ( $\lambda = 1.51''$ ).

loud speaker which supplies a single frequency sound field. The Polaroid-Land camera in the background has been favored because its short develop-

ment time of 1 minute allows the photographic record to be inspected almost immediately after the scanning run has been completed.

#### AMPLITUDE PATTERNS

The first pictures were taken to determine the amplitude distribution in the focal region of a 10" diam. acoustic lens. This acoustic lens is made of rigid metal strips arranged in an open construction.<sup>2</sup> Figure 2 shows an early photograph in which the horizontal motion of the scanning device was too rapid. The line structure of the scanner is therefore very coarse but

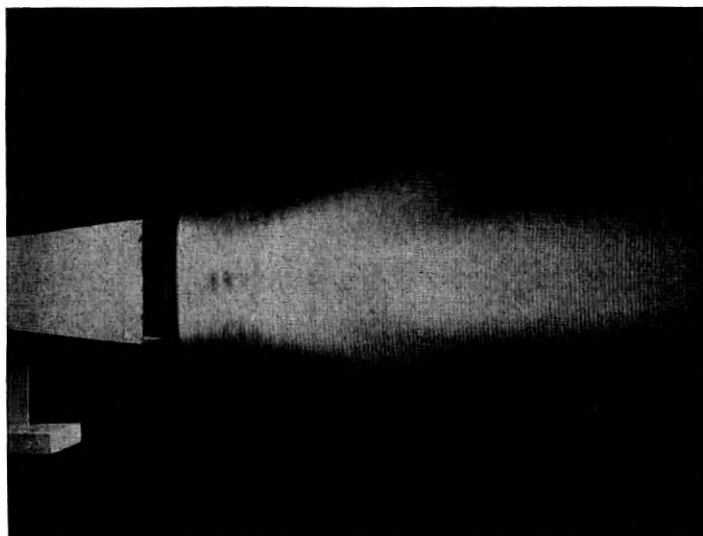


Fig. 4—Radiation pattern of a 6" square aperture horn.  $f = 9$  KC.

the lobe structure and focusing effect of the lens are evident. Figure 3 shows a later photograph in which a finer gradation scan with a longer stroke provides a smooth appearing pattern.

The radiation pattern of a long horn with a 6" square aperture is shown in Fig. 4, taken at a frequency of 9 KC. As in the lens pictures, minor lobes can be seen forming at the sides of the major lobe, while several minima appear faintly in the central region near the aperture (the close-in or Fresnel field).

The refracting property of a prism made of rigid strips is illustrated by the amplitude pattern in Fig. 5. In combination with the strip lens of Fig. 3, it bends the sound beam downward away from the axis. The lens itself is an example of a refractor, of course, but the prism is usually chosen to demon-

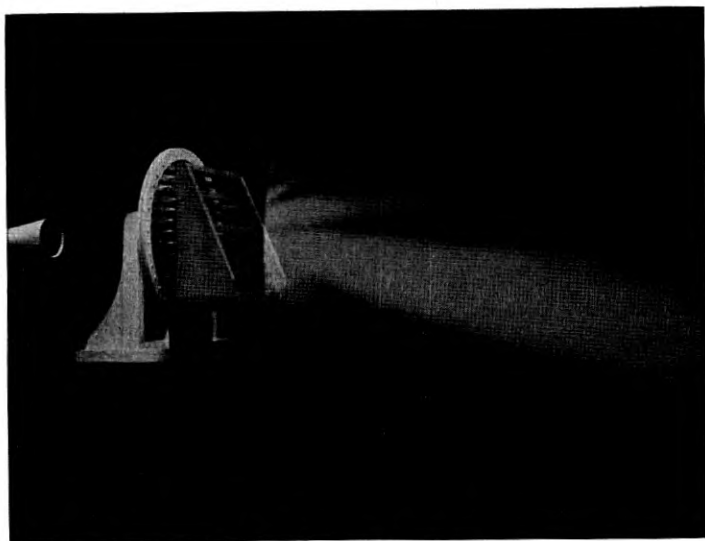


Fig. 5—A strip prism placed before the lens of Fig. 3 tilts the focussed beam downward.  
 $f = 9$  KC.

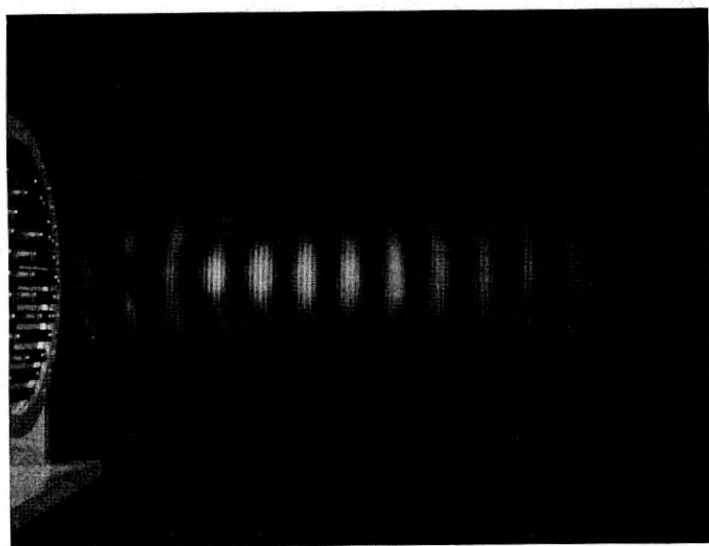


Fig. 6—Adding a signal of constant phase to the signal picked up by the scanning microphone delineates the position of the wave crests as they progress outward toward the right. (These are not standing waves.)  $f = 9$  KC.

strate refraction in its simplest form. One sees, along with the downward beam tilt, a dissymmetry of the minor lobe structure, the upper lobe being quite pronounced.

#### ADDITION OF PHASE

From this point on, the portrayal of phase will be added to most of the photographs. This is accomplished by combining, with the signal picked up by the probe microphone, a constant-phase constant-amplitude signal from the oscillator feeding the source loud speaker. As the probe moves away

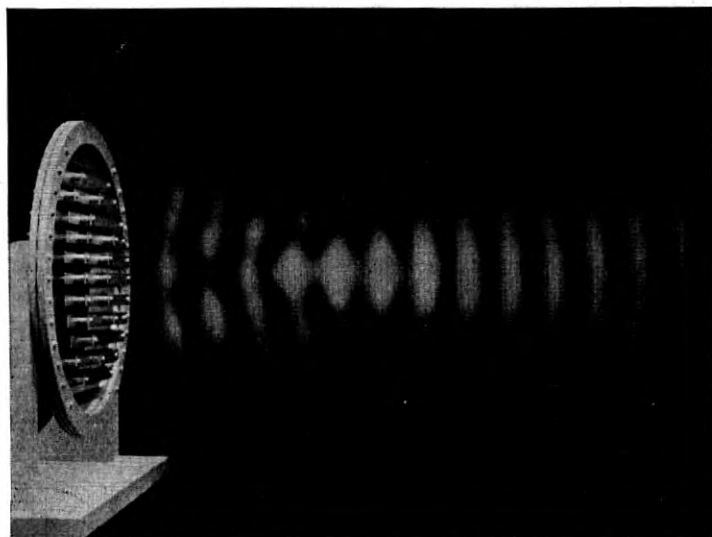


Fig. 7—Increasing the sound intensity in Fig. 6 brings out the details of the minor lobes.  $f = 9$  KC.

from the loud speaker and lens combination, the phase of the microphone signal varies with respect to the constant signal from the oscillator. Constructive and destructive interference results and pictures such as shown in Fig. 6 are obtained. This is a pattern for a progressive sound wave (not a standing wave) emerging from the strip lens; it is, in fact, the same pattern as that of Fig. 3 except that phase has been added and the signal intensity reduced somewhat. The curved wave fronts converging towards the brighter focal spot are evident and as the sound energy progresses through the focal point the wave fronts become concave outward. Figure 7 shows a similar pattern but with the intensity of the signal increased. This procedure brings up the intensity of the weaker field and shows more clearly the minor lobe

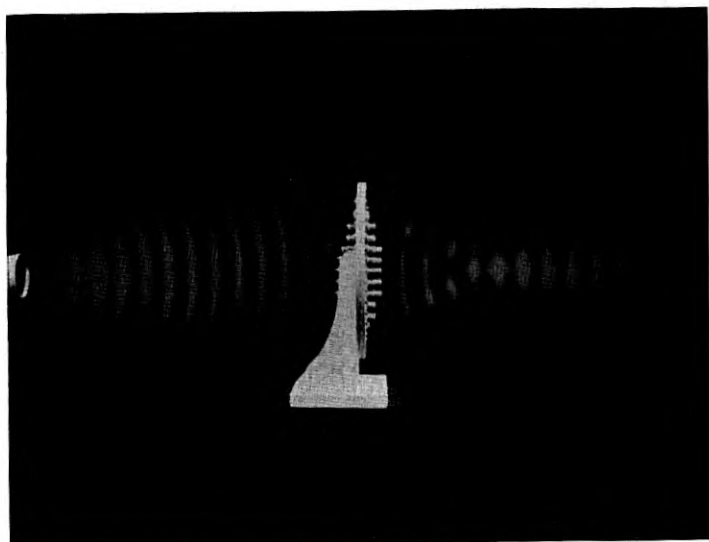


Fig. 8—A composite pattern of horn and lens which shows that the lens delays the wave fronts by approximately one wavelength.  $f = 9$  KC.

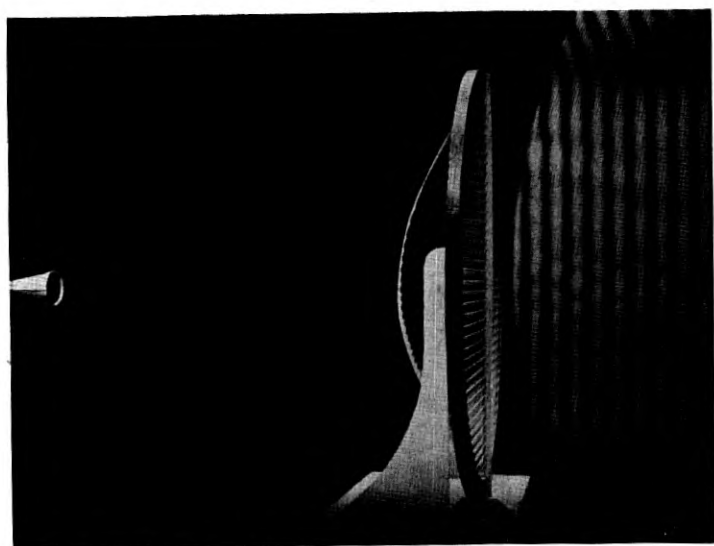


Fig. 9—Flat phase fronts are obtained with a 30" diameter slant plate lens "illuminated" by the 3" aperture feed horn of Fig. 8 placed at the focus.  $f = 9$  KC.



structure of the lens. The phase reversal at each successive minor lobe is readily seen from the fact that the bright areas in the minor lobes line up with the dark areas of the adjacent lobes.

The addition of phase makes it possible to obtain motion pictures of progressive wave motion. By taking successive movie "stills", in which the phase front pattern is advanced one-eighth of a wavelength, a complete cycle is obtained. This series of eight pictures is then repeated until a reasonable length or loop of film is obtained.

An example of the retarding effect caused by a delay lens is shown in Fig. 8. This wave pattern is a composite (two exposure) picture and was obtained in the following way: the left side of the photo was scanned with only the feed horn active (lens removed). The lens was then put in place and the probe continued to scan the area to the right. At the top of the photo, the circular wave fronts are seen to be continuous from the horn out, but in the shadow region behind the lens the wave fronts are retarded. One sees that the insertion of the lens has caused a delay equal to about one wavelength along the axis. This results in a curved wave front emerging from the lens and a consequent focusing action.

An acoustic (or microwave) lens in which the delay elements are slanted guides similar in construction to a venetian blind<sup>3, 4</sup> is shown in Fig. 9. The feed horn is the same as that in Fig. 8 and is set at the focus of this lens so as to produce flat rather than converging phase fronts in the emerging wave. As the directional pattern of the feed horn would indicate (Fig. 8), the center portion of this lens is rather strongly "illuminated," but the resulting energy concentration, in passing through the lens, is shifted upwards by the guiding action of the slanted plates. The resulting dissymmetry of the vertical amplitude distribution at the aperture of this lens (indicated by the increased thickness of the phase lines in the upper section of the photograph) is in a large part responsible for the unsymmetrical minor lobe structure as reported.<sup>4</sup> The straightness of the phase lines, however, indicates that the lens has converted the circular wave fronts from the horn into the desired plane wave fronts.

The undesirable concentration of energy at the center portion of the lens can be materially reduced by substituting, for the small feed horn of Fig. 8, a full conical horn shield having its throat at the focal point of the lens. The energy distribution entering the lens is then fairly uniform and the slant plates cause only a slight dissymmetry in the amplitude distribution of the emerging wave as shown in Fig. 10. The distribution of such a "shielded" lens in the horizontal plane, however, is even more uniform since it is not skewed by the slant plates in this plane (Fig. 11).

<sup>4</sup> W. E. Kock, "Path Length Microwave Lenses," *Proc. I.R.E.*, 37, 852 (1949).

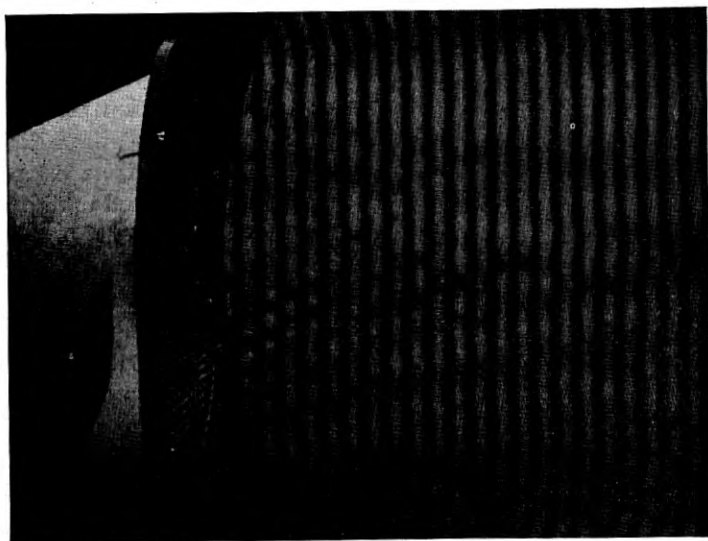


Fig. 10—The pattern of the slant plate lens of Fig. 9 when enclosed in a horn (vertical plane).  $f = 9$  KC.

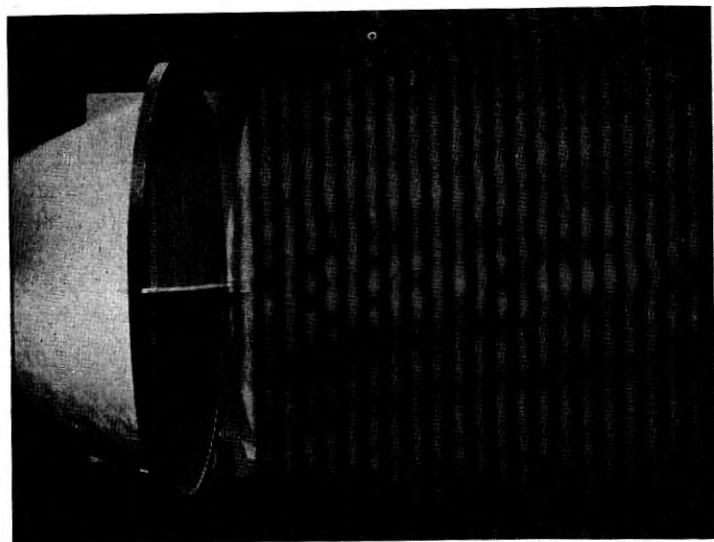


Fig. 11—The horizontal plane pattern of the horn and lens of Fig. 10.  $f = 9$  KC.

In some acoustic investigations it is desirable to have a source of plane waves of extended area. Figures 10 and 11 show that a shielded acoustic lens provides a simple way of achieving this result. The broad band nature of

the slant plate lens ensures that plane wave fronts will be produced from 14 KC down to the lowest frequency contemplated for use, and almost any size can be employed since microwave lenses of 10 and 20 foot apertures<sup>5</sup> have been built and found quite satisfactory.

In the three preceding photographs, variations in the intensity of the emerging waves (variations in the thickness of the phase lines) can be observed. To show this effect more clearly and to indicate the symmetry of these amplitude variations, Fig. 11 was retaken with the phasing signal

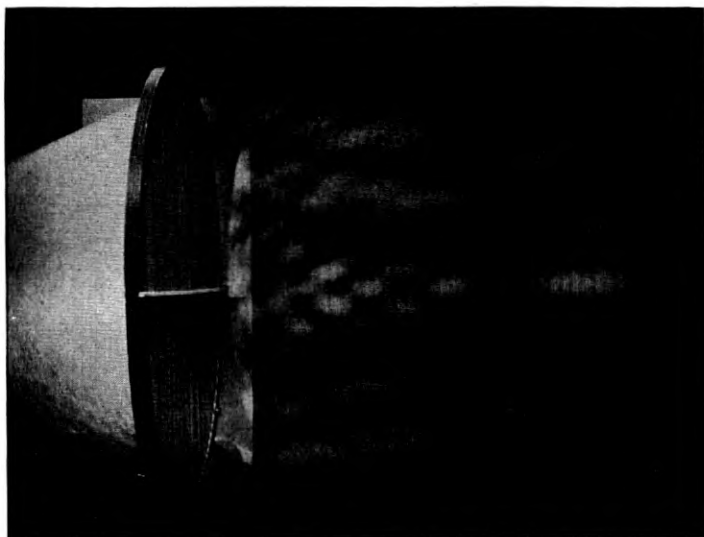


Fig. 12—Removal of the phase signal in Fig. 11 shows more clearly the amplitude variations present in this horn-lens pattern.  $f = 9$  KC.

removed so as to obtain a pure amplitude pattern (Fig. 12). As mentioned in connection with Fig. 4, these patterns simply show the intensity variations which are always present in the close-in or Fresnel field of any plane wave source having a finite cross-sectional extent. For a given wavelength and cross-sectional dimension, these variations can be calculated from diffraction theory and evaluated with the aid of Cornu's spiral. It is interesting to compare Fig. 12 with Fig. 13 which is a Schlieren photo of the sound field in front of an ultrasonic quartz radiator of 14 wavelengths aperture dimension.<sup>6</sup> Although the aperture dimensions in wavelengths in the two cases are different and the actual wavelengths are quite different, there is a

<sup>5</sup> W. E. Kock, "Metal Lens Antennas," *Proc. I.R.E.*, 34, 828 (1946).

<sup>6</sup> K. Osterhammel, "Optische Untersuchung des Schallfeldes Kolbenförmig schwingender Quarze," *Akustische Zeits.* 6, 82. (1941).

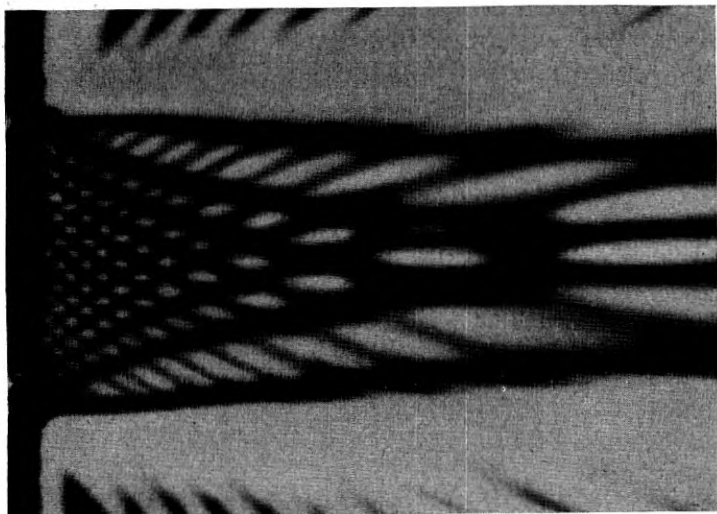


Fig. 13—A Schlieren photograph of an ultrasonic quartz radiator shows the presence of amplitude variations similar to those of Fig. 12. (After Osterhammel.)

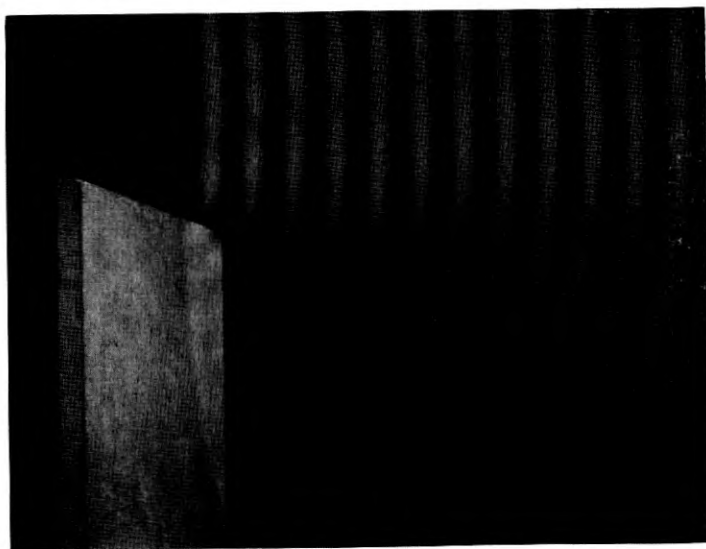


Fig. 14—Plane waves diffracted by a knife edge become cylindrical in the shadow region.  
 $f = 9 \text{ KC.}$

remarkable similarity in the general appearance of the patterns. Immediately along the axis of the radiators, one observes quite large amplitude

fluctuations which become more rapid as the radiating source is approached; similar fluctuations are observed in front of large-area microwave antennas.

#### DIFFRACTED WAVES

Figure 14 shows the diffraction of waves over a straight edge. The sound waves in the upper half of the pattern are seen to progress with flat phase fronts, but in the shadow region of the  $\frac{3}{4}$ " wooden board circular wave fronts are evident, indicating that the edge is acting as a new Huyghens source. Figure 15 is a similar pattern showing diffraction around a wooden disk.

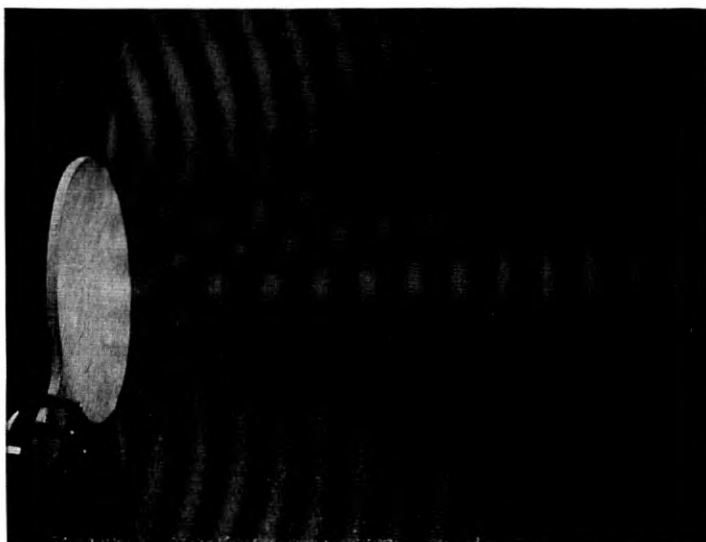


Fig. 15—The circular wave fronts in the shadow of a 10" diameter "opaque" disk combine to produce a lobe structure.

The circular wave fronts emanating from the top and bottom edges are evident. Similar wave fronts are re-radiated from all around the circular edge of the disk and these combine to produce a concentration of energy along the axis corresponding to the "bright spot" of optics in the shadow of an opaque disk. Figure 16 is a repeat of 15 with the phase signal removed; this amplitude pattern shows the bright spot more clearly.

Another diffraction effect of optics, the pattern produced by two small slits, can be duplicated by two non-directional sound sources separated several wavelengths and having equal amplitudes and phase. This produces the multi-lobed pattern of Fig. 17. Destructive interference occurs in those directions for which the two sources are out of phase, and constructive inter-

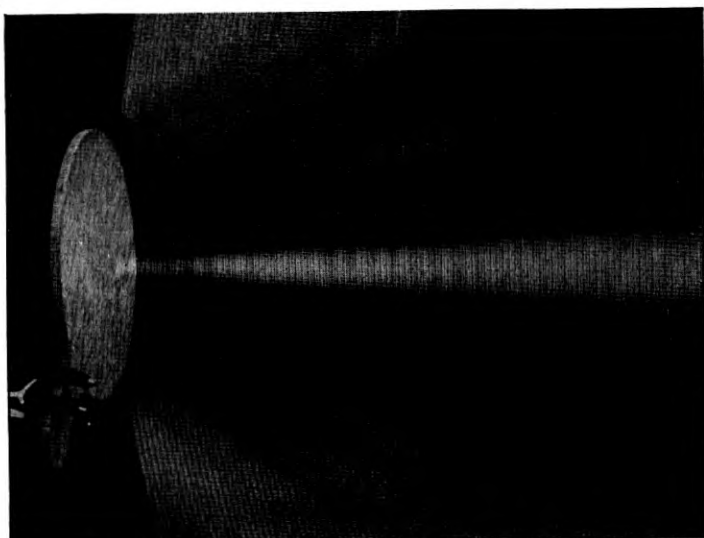


Fig. 16—Removal of the phase signal of Fig. 15 shows more clearly the major lobe of radiation in the shadow of a disc, the "bright spot" of optics.  $f = 9$  KC.

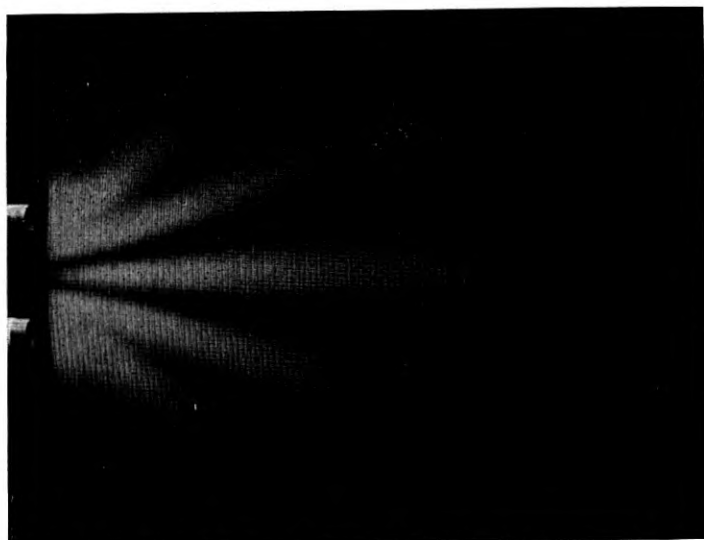


Fig. 17—Radiation pattern of two equal sources of low directivity separated 3 wavelengths center to center.  $f = 9$  KC.

ference in whose directions for which the sources are in phase (such as straight ahead). The interference minima become filled in near the radiators be-

cause the amplitudes of the two signals are unequal when their distances to the point of interference are unequal. Perfect cancellation can then not occur. It is interesting to compare this photograph with the contour curves for this case as calculated from diffraction theory<sup>7</sup> and shown in Fig. 18. Al-

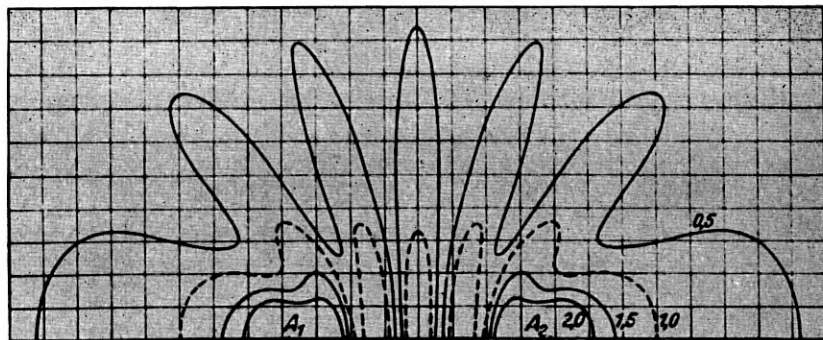


Fig. 18—Calculated radiation pattern for two non-directional sources separated 3 wavelengths. (After Stenzel.)

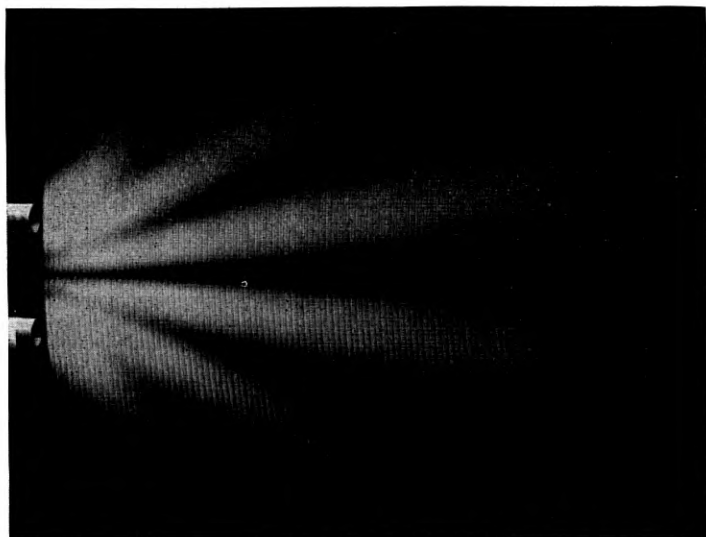


Fig. 19—The pattern of Fig. 17 except that the sound sources have opposite phase.

though the contour plot gives more quantitative information, the photograph allows one to grasp the qualitative effects more quickly. Figure 19 is similar

<sup>7</sup> Heinrich Stenzel, "Leitfaden zur Berechnung von Schallvorgängen," (Julius Springer, Berlin, 1939), p. 59.

to Fig. 17 except that the connections to one of the sound sources were reversed. With the two sources out of phase, cancellation now occurs straight ahead.

The diffraction pattern of one wide slit would be similar to that of the 4 wavelength radiator of Fig. 4.

To show the diffraction rings produced at the focal point of a lens, the sound field of a strip lens was scanned in a plane perpendicular to that of the previous photos. Figure 20 shows such a scan made in a plane passing through the focal point and perpendicular to the lens axis. The usual optical formulae

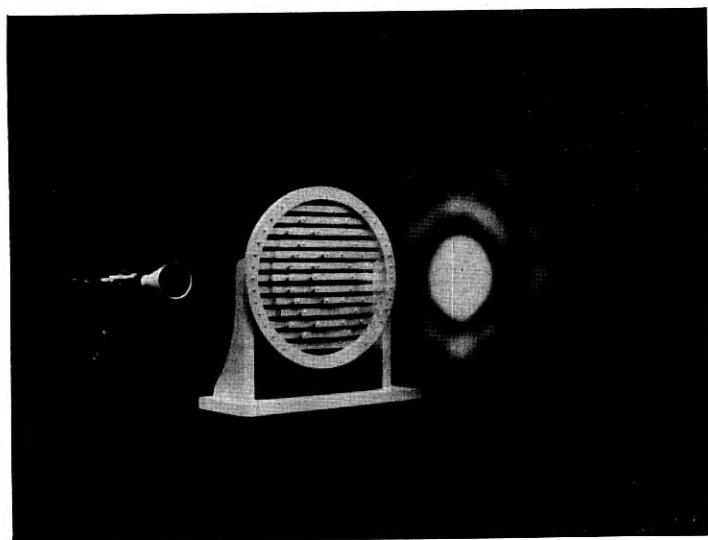


Fig. 20—By scanning a plane perpendicular to the axis of radiation, the diffraction rings around the focal spot of the lens of Fig. 3 are portrayed.  $f = 9$  KC.

determine the size of the focal spot and the position of the surrounding diffraction rings. They are functions of the focal distance, the aperture, and the wavelength. In the previous lens patterns of Figs. 3 and 7, these diffraction rings show up as minor lobes. For a perfectly symmetrical lens construction, the rings would be more perfect, and the minor "lobes" would in reality be cones of energy surrounding the major lobe.

#### DIFFUSION OF SOUND

The previous lenses have been of the convergent type for focusing or beaming energy. In the next pair of pictures is shown the effect of a divergent lens diffusing energy. The phase pattern of a 6" square aperture horn



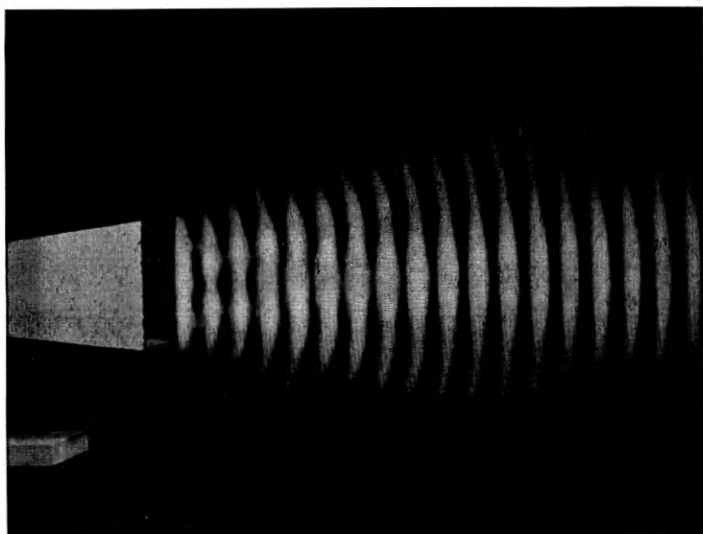


Fig. 21—The beam from the 6" aperture horn loud speaker of Fig. 4 has fairly flat wave fronts and a narrow angular coverage.  $f = 9$  KC.

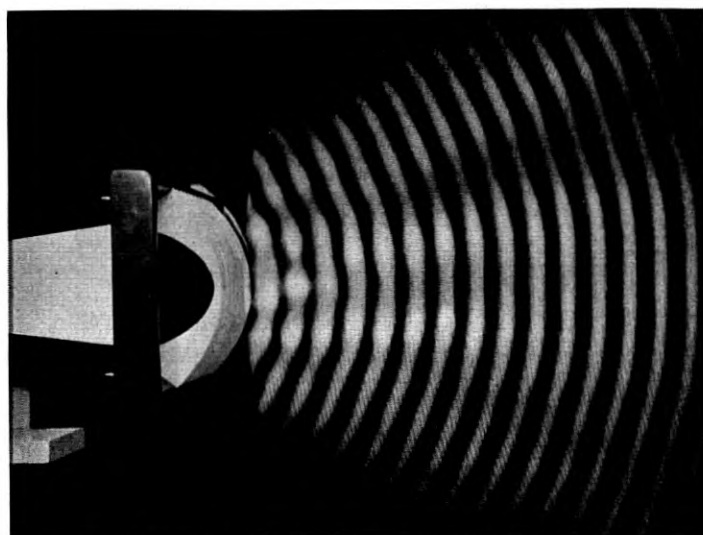


Fig. 22—A diverging acoustic lens in the aperture of the horn in Fig. 21 converts the straight line waves into circular waves with their greater angular coverage.  $f = 9$  KC.

appears in Fig. 21. As in the pure amplitude pattern of this horn (Fig. 4), the directivity of this aperture is seen to be fairly high and the sound energy

is collimated into a fairly sharp beam. High frequencies emanating from this horn would be projected through a rather small angular region and such a horn by itself would not be a very desirable general purpose loud speaker. Figure 22 shows the same horn with a slant plate divergent lens (described in reference 2) placed in front of its aperture. The transformation of the flat phase fronts into circular fronts and the wider angle of coverage can be readily seen. A frequency of 9 KC was used for these tests.

When a large cone type loud speaker is operated at the higher frequencies it too becomes quite directive. Figure 23 is a radiation pattern of a 12-inch

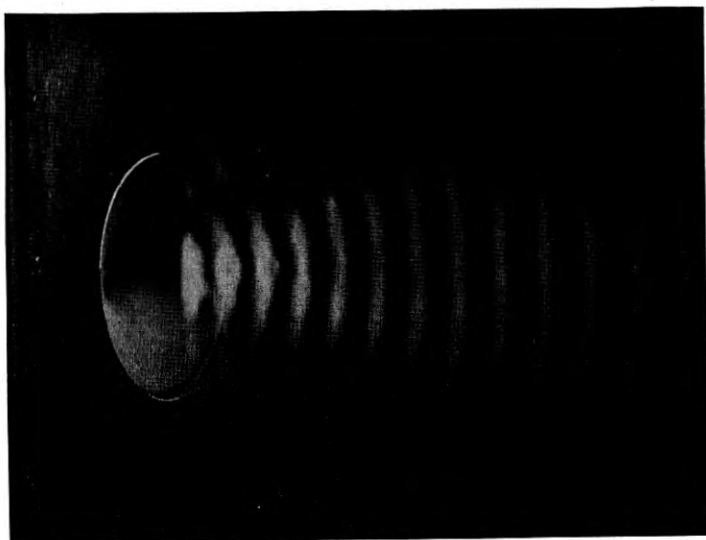


Fig. 23—A large aperture loud speaker (12 inch) also becomes directive at high frequencies as evidenced by the flat wave fronts in the beam.  $f = 8.5$  KC.

loud speaker at 8.5 KC. The minor lobe formation is noticeable, indicating the presence of a central lobe surrounded by a region of low intensity.

#### MICROWAVES AND SOUND

The lenses employed in the preceding photographs were originally conceived and constructed for use at the very short radio wavelengths known as microwaves.<sup>8</sup> The strip lens is, in fact, a small scale model of the type used in the antenna systems of the New York-Chicago microwave relay circuits of the Bell System for telephone and network television. A similar type of dual-purpose delay lens using disks instead of strips is shown in the

<sup>8</sup> W. E. Kock, "Metallic Delay Lenses," *Bell Sys. Tech. Jour.*, 27, 58 (1948).

next pair of pictures. In Fig. 24 the disks are arranged in an open construction so that sound waves as well as microwaves will pass through. An acoustic amplitude pattern is shown taken at a frequency of 12 KC ( $\lambda = 1.13''$ ). In Fig. 25 the disks are copper foil and are supported on polystyrene foam which is transparent for radio waves but opaque for sound waves. The pattern now shows the intensity distribution of the microwave field being focused by the lens. The frequency was 9100 megacycles ( $\lambda = 3.3$  cm. or  $1.3''$ ).

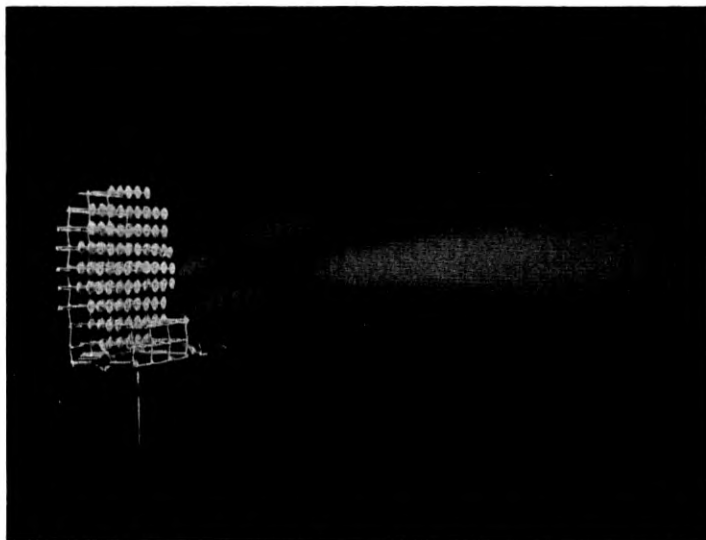


Fig. 24—A sound field pattern of a disk array lens originally designed for 3 cm. radio waves.  $f = 12$  KC. ( $\lambda = 1.13''$ ).

To obtain this microwave picture, the loud speaker was replaced by a microwave radiator and the pickup microphone replaced by a tiny dipole and crystal detector. The microwaves are modulated with 120- $\mu$  pulses so that the same audio frequency amplifiers are used as before with sound waves. However, with this low frequency supplied to it, the neon lamp trace appears as a series of dots.

In the next pair of pictures is shown a phase advance lens which likewise is effective only for microwaves. It uses parallel conducting plates in a waveguide construction.<sup>4</sup> Because the phase velocity is increased in passing through this medium, a concave lens is required for focusing (see Fig. 26). When the waveguide source (off to the left of Fig. 26) is brought in nearer so as to be at the focal point of the lens, the wavefronts straighten out and

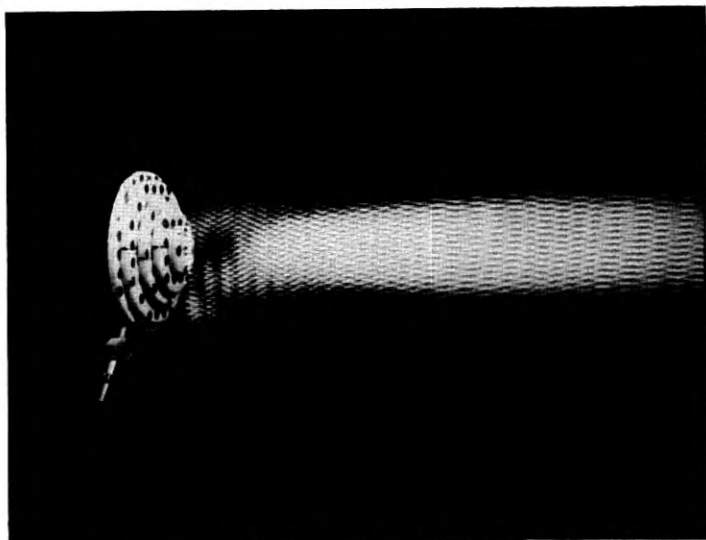


Fig. 25—An electromagnetic field pattern of a disk array lens in which the copper foil disks are mounted on polystyrene foam sheets.  $f = 9,100,000$  KC ( $\lambda = 1.3''$ ).

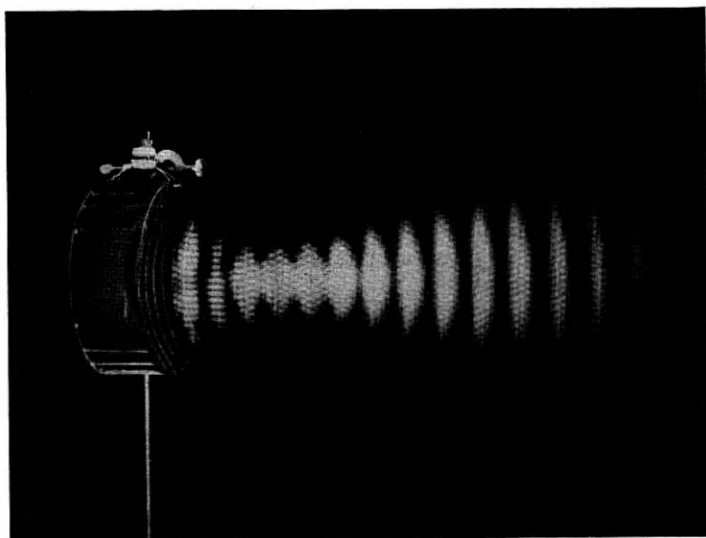


Fig. 26—A microwave field pattern of a waveguide type metal lens. Phase has been added, showing the curved wave fronts approaching and leaving the focal point.  $f = 9.1$  KMC.

the energy is collimated into a beam (Fig. 27). The reference signal for these phase patterns was obtained by the method illustrated in Fig. 28. A secondary microwave source was employed which was fed from the main source

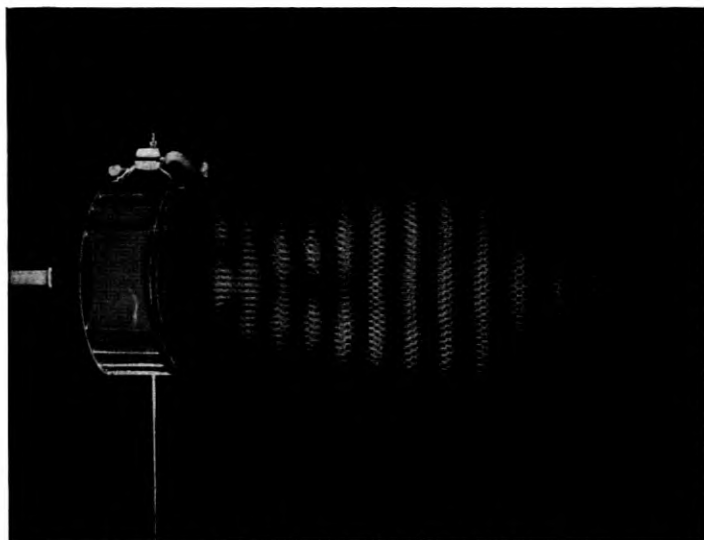


Fig. 27—When the waveguide feed of the lens of Fig. 26 is brought closer to the lens, a beam of flat wavefronts is produced.  $f = 9.1$  KMC.

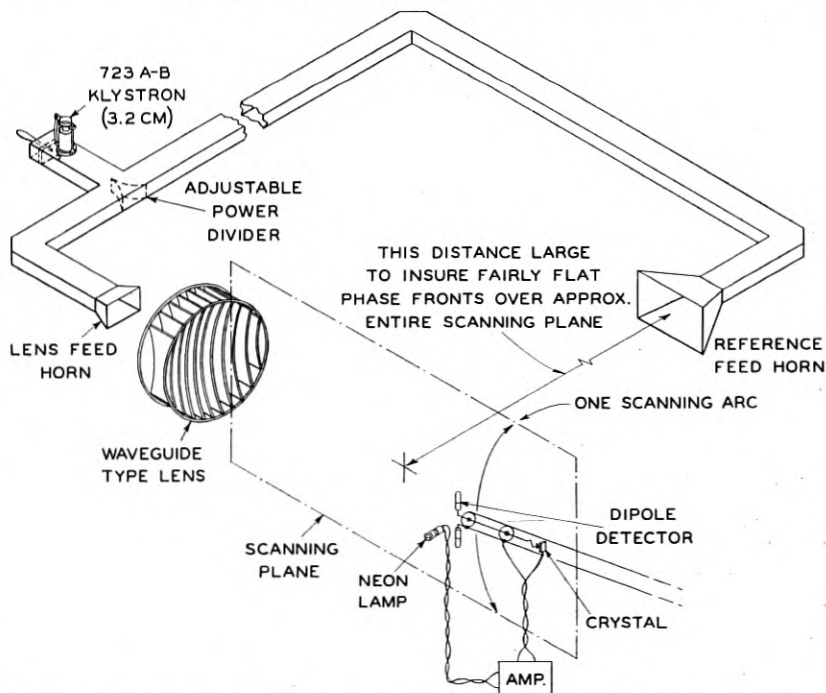


Fig. 28—The method of adding a constant-phase constant-amplitude microwave signal to the lens signals of Figs. 26 and 27 so as to portray the wave fronts.

and located on a line perpendicular to the scanning area and passing through its center. It was sufficiently far away to ensure that its wave front was approximately plane, i.e., of uniform phase and amplitude, over the scanned area. The scanning microwave probe thus picked up a nearly constant phase, constant amplitude signal from this secondary source in all positions of scan, and in addition sampled the variable intensity, variable phase microwave field coming from the lens.

#### MICROPHONE PATTERNS USING TRANSPOSITION TECHNIQUES

In the analysis of directional properties of acoustic or microwave radiators such as lenses, reciprocity can be employed, which is the property that equivalent directional characteristics will be exhibited whether the transducer is used as a transmitter or receiver. Some electro-acoustic transducers are not reversible, however, and the directive properties of a carbon microphone, for example, cannot be easily ascertained unless some means is employed which measures its characteristics while it is receiving acoustic energy, i.e., in the microphone condition. In all of the preceding photographs the scanning device probed an actual sound field. In the analysis of a microphone, however, we are interested in its ability to pick up sound coming from various directions in space. We can therefore replace the sound source in the preceding photographs with the microphone under test, and replace the probe microphone with a scanning sound source. As the source scans the space in front of the microphone, the signal in the microphone will vary depending upon the ability of the microphone to receive sounds from a particular spot in the scanned area. If this microphone signal is used to control the brilliance of the lamp (still affixed to the scanner), the resulting photo will indicate the directional characteristics of the microphone by itself or in combination with a directional device such as a lens. As in the preceding photos, phase can be added by combining a constant amplitude signal with the microphone signal. The end result of all this is simply to interchange the connections of source and sink.

Figure 29 shows a microphone pick-up pattern in which the strip lens has been placed in front of the microphone (not shown in the photo). Although this picture is now a representation of the microphone response for sound emanating from various points in the scanned space, it is seen to be nearly identical with the pattern of Fig. 7 which is an analysis of an actual sound field. This fact is simply a consequence of the principle of reciprocity. An interpretation which applies equally well to either of these two phase pictures is that the lines in each pattern are contours of equal phase length between the fixed and scanning transducers.

With this transposition technique, however, we are now able to examine

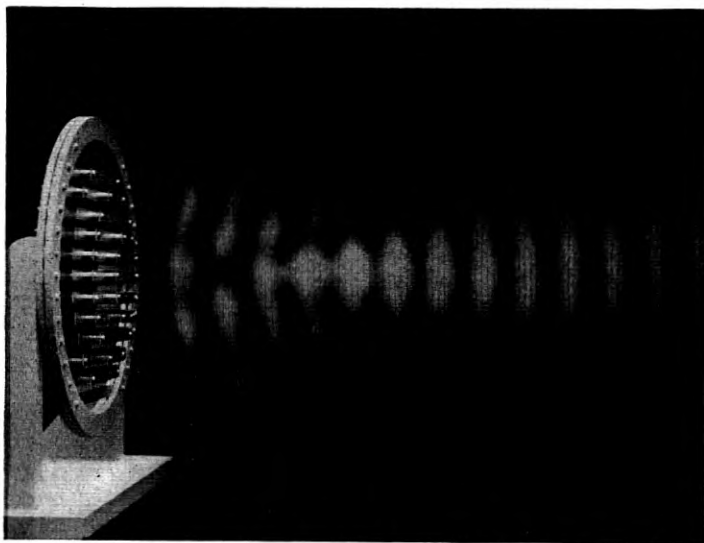


Fig. 29—A pattern taken with the microphone and loud speaker transposed. The sound source now scans and the pickup microphone is fixed (behind the lens). The phase signal is still employed indicating contours of equal phase "length" between the fixed and scanning transducers. The similarity to Fig. 7 can be observed.  $f = 9$  KC.



Fig. 30—A pattern of a carbon telephone transmitter obtained by the transposition method of Fig. 29. The phase fronts are those which would be observed if the microphone were radiating sound.  $f = 4$  KC ( $\lambda = 3.4''$ ).

the spatial response of a non-reversible transducer such as a carbon microphone. This is shown in Fig. 30 for an F-1 carbon telephone transmitter. It is also the directional characteristic the unit would have if it were capable of radiating.

The addition of phase to a microphone directional pattern may seem superfluous but a knowledge of phase can often be useful. For example, in highly directional microphone arrays, the wave fronts in the close-in field

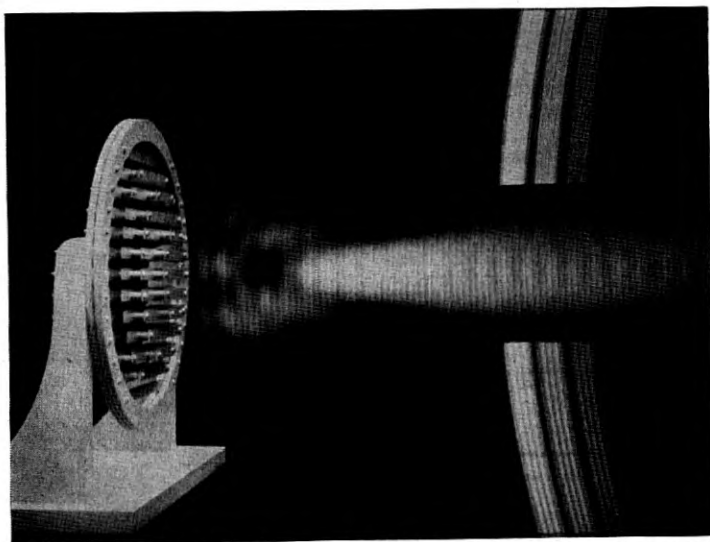


Fig. 31—A procedure for calibrating the photographic sound patterns. Part of the photo is scanned a second time with a constant signal on the lamp. Successive reductions of the signal (by 3 db steps in this case) produce the calibration arcs. (A mask between the camera and the scanner prevents the calibration arcs from registering in the pattern area.)

give information on the directional properties in the distant (far-field) region.

#### ADDITIONAL DETAILS

This concluding section will describe in more detail the scanning mechanism, the photographic procedure and methods for calibrating the photograph to provide a measure of the relative field intensities.

The scanning arm rocks up and down over an angle of about  $60^\circ$ , making one stroke every two seconds. The vertical travel of the lamp in the course of a stroke is adjustable up to 40" while the horizontal travel is fixed at  $\frac{1}{10}$  of an inch per stroke to provide a fine grain picture. The average picture is



built up with about 300 lines or strokes corresponding to a 10-minute scanning time. However, pictures up to twice this length are occasionally taken.

To make a picture (see Fig. 1) the loud speaker is turned on to provide a steady sound field. A filter in the microphone amplifier is tuned to the signal frequency to reduce the interference from external noises. Absorbing blankets are sometimes desirable on nearly reflecting surfaces. The scanner is first moved manually to sample the sound level at various points so as to determine the proper gain settings and then placed in a starting position close to the acoustic lens whose pattern is to be taken. The room is darkened, the camera shutter opened, and the scanner started. Because the scanning process is relatively slow, the observer sees only the individual strokes of the flickering lamp. However, all the strokes are recorded on the camera film and form the desired pattern. When the scanning is completed, the shutter is closed and the room lights turned on. The film is then re-exposed for a few seconds to add the image of the acoustic lens. A dark background is provided so that the sound pattern will not be obliterated.

The miniature glow lamps used for these pictures have been neon and argon types having no base resistances, e.g., the NE17, NE51, or AR4 and AR7. Neon seems to produce smoother gradations in intensity but argon is sometimes desirable because the film is blue sensitive. The lamps operate only over a voltage range from 70 to 120 volts, so that compression must be used in the amplifier circuit if the pattern is to show the maximum amplitude variations encountered which are of the order of 10 to 30 db. Many of the patterns were taken by just connecting the lamp to a high impedance circuit in the output of a power amplifier, the compression being obtained by the decreasing resistance characteristic of the lamp with applied voltage.

In this as in any photographic process, the operator can control the effects pictured, intensifying or subduing images as desired by adjusting the film exposure and the circuit compression. When directivity is to be displayed and the major lobe is of main interest, the minor lobes may not even be exposed; when the phase fronts are desired in weak regions, the maximum range of the film and largest circuit compression may be used.

When quantitative information is desired, the relative intensity of the sound field in the amplitude patterns can be ascertained by a calibration test performed before or after each run. This is illustrated in Fig. 31. A signal equal to the maximum signal the lamp receives during the scanning run is fed at constant level to the lamp while it is scanning the unused part of the photograph at the right. Successive reductions of the signal by a known amount of attenuation (3 db steps in this case) give a series of arcs of decreasing brightness. These can be compared to the various scanned portions of the photograph to get a measure of the amplitude.

## Some Basic Concepts of Translators and Identifiers Used in Telephone Switching Systems

By H. H. SCHNECKLOTH

(Manuscript Received April 26, 1951)

The functions and typical designs of translators as applied to automatic telephone switching systems are first reviewed. The fundamental similarity of some existing translation schemes is noted and a discussion given of the factors which can be juggled to obtain economic application of such schemes.

Identifiers and their uses are next described and some processes of identification are shown to have much in common with those of translation. Complications encountered in commercial application are discussed. Some needed improvements in general designs, possibly by new approaches, are indicated.

Finally, the author points out the frequent occurrence of translation and identification processes in switching elements which are not labeled "translators" or "identifiers" and suggests that future improvements in translation and identification methods may consequently be useful in switching circuit networks in general.

### INTRODUCTION

Those concerned with the technical details of automatic telephone switching circuits usually regard a switching system as made up of a number of types of blocks or elements named in accordance with their main accomplishment in the train of events in handling a telephone call.

Each of these elements is important as a useful cog in making the system work. Some of them, however, have an added distinction, if not a glamour, because their introduction to the switching world made possible fundamental changes in switching technique or in the service which could be offered to subscribers.

Two such elements are translators and identifiers. They will not be defined or explained until later, but a few words concerning their importance may be in order here.

These elements were not used at all in early types of automatic switching systems. The invention of the translator in 1905, by Mr. E. C. Molina,<sup>1</sup> and the philosophy that accompanied it are now generally credited with having laid the groundwork for the Bell System's adoption of systems of the common control type. In these, the paths necessary to reach a called number are not selected by the calling dial but by equipment which is common to many switching elements. The dial merely furnishes the customer's orders to the common equipment which, through the translation process, can set up a more suitable series of paths than is possible within the limitations of direct dial control.

<sup>1</sup> "Historic Firsts": Translation, p. 445 of Nov. 1948 *Bell Laboratories Record*.

The advent of the process which made the dial an order-passing device rather than a direct-control device, of course changed the entire conception of what could be used in the way of automatic switches, how they were controlled and the number and arrangement of switching paths. But more important than the effect on technical methods in the equipment was the fact that the principle of translation reduced the limitations which subscriber numbering arrangements formerly had on the economy of giving subscribers direct access, by dialing, to large numbers of central offices in complicated networks. First the benefit of this was in facilitating dial service in large metropolitan areas, later in making available suitable methods for dialing of toll calls by subscribers. In other words, the translator was a practical means of pushing back the horizon for automatic telephone service.

Systems based on the use of the translation principle are now found in many countries, and the Bell System's most modern crossbar arrangements depend on it more than ever. It is difficult to conceive of a nationwide automatic toll switching plan for a country as large as the United States without the use of translation.

Identifiers are not so old as translators and have, so far, influenced the general design of switching systems in only a minor way. Their importance lies in the fact that they were key elements in introducing to subscribers, both here and in Europe, a new kind of toll service. With this service, whenever a subscriber dials a toll call, equipment in the central office automatically determines his own number and prints or otherwise makes a record of it along with other details of the call so that eventually he can receive a complete statement of toll calls which have been dialed and the detailed charges for each.

Identifiers were essential in the first offices in which this service was provided in order to determine the number of the calling party, and are necessary in plans now envisioned for giving this service in the large number of old-type offices still in service in the Bell System.

Our newer type offices arranged for this service do not employ devices named "identifiers" as the identification function is spread over numerous elements. Nevertheless, the identification process is there and, named or not, it is found in many other situations.

Translators and identifiers, while used for functions which seem offhand to be quite different, are often similar in so far as the general operation and problems of design are concerned and, in fact, have much in common with numerous circuit elements called by different names. Because of their fundamental importance they have been the subject of much invention, directed toward use in specific conditions, the use of different types of apparatus,

improvements in reliability, increase in speed, and, most important of all, reduction in cost. Operable arrangements to meet all sorts of conditions of use can generally be selected from the present fund of knowledge but the choice in practice is much more limited than in theory. However, translators and identifiers, particularly those used on a large scale, can be varied in many ways to obtain maximum usefulness and to minimize the costs, not only of these devices themselves, but also of other parts of the associated system.

In the following discussion a number of devices of these types will be illustrated to provide a general review of some of the methods available, and the factors encountered in practice in trying to obtain a proper balance between costs and usefulness will be examined.

### TRANSLATORS

#### WHAT ARE TRANSLATORS?

The translators with which we are concerned in this discussion deal only with information in the form of electrically coded numbers. The "languages" carrying this information are the coding systems made up of the numbering bases and signalling methods. Unfortunately, the analogy between the switching system translator and an interpreter of languages, implied by our use of the word "translator," is not very consistent.

The term "translator," as used in this paper, means broadly a switching system element which, in response to an inquiry in the form of an input code, supplies an answer in the form of an output code to the element presenting the input code or to some other element. Each code may represent one or more numbers.

The translator may be a device which serves a number of other circuit elements in common, or it may be associated with a single such element or even built into it and unnamed.

Now in practice we will find that in some applications translators are used so that the input codes and corresponding output codes represent the same numbers in different bases or with different signalling methods, that is, the same information in different languages. Here we have the nearest approach to our implied analogy. However, quite commonly, the switching translator is required to do things which would be decidedly out of order in the case of language translation, such as changing the information instead of the language or changing both at the same time.

Some of the variations in switching translator applications which should be noted at this point are:

- (1) In practice, translators are arranged with a multiplicity of input and output possibilities. The inputs may be permanently associated with

corresponding outputs or the association may be changeable through movable jumpers or other means. Thus we have two major types of translators—*fixed* and *changeable*.

- (2) In most applications the input and output codes have no natural correspondence but require *arbitrary* translation determined by the designers of the system or those who operate it. Translation is of the *systematic* type where there is a definite relationship between the input and output codes. The relationship may be purely mathematical or may follow from some of the peculiarities of the switching system.
- (3) The input and output codes may be of the same numbering base with the same or a different number of places. Often the two codes are of different numbering systems and one or both may consist of mixed base numbers.

As discussed later, many devices we do not call translators do in fact have a translating function, but they have been designed with a different point of view and have accordingly been given different but suitable functional names. However, the vagaries of telephone switching nomenclature have in some cases resulted in giving other names to elements clearly having the same functions as elements earlier or elsewhere called translators. The more important variants will be pointed out as we go along. One of these, the term "code converter," seems more appropriate than the original term.

#### EXAMPLES OF USE

It may now be in order to give a few examples of the kinds of uses made of translators in automatic switching systems.

(1) Let us assume we have a common control system in a multi-office city. When a subscriber dials the digits of a local number which we call the "office code", indicating the central office unit in which the called subscriber is located, this number is received by the switching equipment of the originating office as a decimal number with three digits or less. The switching equipment, in extending the call to the central office indicated, may have to set up connections at numerous stages of switches, some in distant offices, which are not indicated by the dialed code. The switching operations which must be performed by the control equipment to reach the desired office are indeed represented by a numerical code, but it may be a number quite different from the dialed number and have no natural relation to it. For this purpose a translator is used to convert the dialed number to the required new code comprising the instructions for the switching operations which must be performed. In British practice a translator used in this way is sometimes called a "route table."

Since the routing of calls corresponding to any particular office code

through the switching equipment may have to be changed from time to time, provision must be made in the translator so that the correspondence between dialed codes and switching codes can be changed economically when necessary by the operating personnel by making changes in the wiring of the translators or by other means. At present this is generally done by re-arranging jumpers.

Here then we have an example of a translator of the *changeable* type with *arbitrary* correspondence between the input and output codes.

This type of translation is used with almost all common control systems in the world. With some of these systems it is required because of the non-decimal nature of the switching arrangement. With other types, having decimal switches, it is not required but it is nevertheless used in order to make the trunking arrangements more flexible and efficient.

Translators of this type are not large in some cases because of the small number of translations needed. A complete translator is sometimes permanently associated with each circuit element having need for translation service and the translator may be separately mounted or built into the associated circuit. In other cases one or more translators may be arranged for common use by numerous circuits requiring them.

(2) In some switching systems, for instance the panel type, additional use is made of translation in the switching operations involving the further extension of a call after it has reached the desired central office. The numerical code, usually the thousands, hundreds, tens, and units digits dialed by the calling subscriber would, with some types of offices, directly serve to control the switching operation for the final selections; but in panel-type offices the terminations for the subscriber numbers, while arranged in an orderly manner, are not grouped on a decimal basis and a switching control code corresponding to the actual location of the called number must be determined. Here again a translator is used, but the input codes in this case have a definite relation to the output codes which is invariable; so the translator used is of the *fixed* type with *systematic* correspondence between the codes.

This particular application followed from the first American invention relating to translators and marked an important step in the development of switching arrangements not requiring a direct correspondence between dialed numbers and switching operations. This application is now found in many systems.

When we examine the block diagrams of the switching arrangements for the panel and some other systems with this type of translation we do not find any block for this particular translator, as the systematic correspondence



between the input and output codes permitted the translating arrangement to be made a part of other equipment.

(3) In our latest crossbar system the idea of not using the numerical digits of the called number as the switching control code has been carried still further. There are no terminals numbered as subscribers' telephone numbers. The switches obtain access to the called subscribers by connecting to terminals for "line equipment numbers". The line equipments correspond to subscribers' lines and are associated with subscribers' numbers on an entirely arbitrary basis. The equipment numbers are not four-digit numbers but each is a series of five 1- or 2-digit numbers (a mixed base number) which indicate the locations of the equipments on the frames.

Here again use is made of translation to convert the dialed decimal number to the non-decimal number forming the switching instructions the common equipment must have in order to reach a called subscriber. In this case, however, the simple, fixed, systematic type of translator used with the panel system cannot be employed, as the associations of the input and output codes are entirely arbitrary and may be changed from time to time to make changes in number assignments, and so the type of translator employed is of the changeable type with arbitrary correspondence. This type of translator, especially when used for large offices having 10,000 or more numbers (input codes) is decidedly a large scale affair and too costly for permanent association with each circuit unit making use of it. The translating equipment is, therefore, made common and is sectionalized in groups called *number groups* each handling several hundred numbers and operating independently.

(4) In automatic equipments arranged for handling toll calls dialed by operators or subscribers, translation is an extremely important feature especially where the networks are as large and complicated as those of the Bell System. Our latest crossbar toll switching system has many important features and economic advantages made possible by the ingenious use of translation.

In automatic toll switching practice a numerical code of three to six decimal digits is sent to the switching equipment in the originating toll office to indicate the particular geographical area in which the called number is located. The area may be nearby or far away, trunking may involve only a few paths in series or a number of intermediate offices and many interswitch paths. The switching equipment must determine the necessary course of action from the 3- to 6-digit code which has been received, and this is done through the use of a translator.

In the case of our newest toll crossbar office, the output code of the

translator actually consists of many different numbers in various bases, each indicating some different item of information required in order to complete the call. The information includes switching instructions for interswitch paths and intertoll routes of the preferred combination and also indicates where alternate combinations may be found if the preferred choices are busy or out of order.

Because of the large number of possible codes, translators suitable for this application present special design problems.

### TYPICAL TRANSLATORS

Before undertaking an analysis of the general concepts used in translation, it may be well to review some of the typical translation methods which have been used in the field. In the following descriptions, details not necessary in illustrating the general method are omitted.

#### *Fixed Translators*

Figure 1 shows the principles of a fixed, systematic translation scheme employed in early panel and other systems for deriving from decimal numbers the switching instructions for controlling some of the non-decimal selections.

This is one of the solutions for case (2) of the examples of use just mentioned. The selection of a called subscriber's number by the terminating equipment in the called central office unit is governed by the last four decimal digits of the number, but the process required with the panel arrangement is in part non-decimal. The numbers are grouped in banks of 100, which means that, once switching has proceeded this far, the wanted number can be selected in the bank on a decimal basis as indicated directly by the tens and units digits of the number. Other groupings of final and preceding equipment involved are not decimal, so the preceding selections must be made on the basis of non-decimal switching control codes obtained by the common control equipment through the translation process.

The switching code wanted in this case consists of three numbers for controlling selections called incoming brush (IB), incoming group (IG) and final brush (FB), and must be derived from the combination of the thousands and hundreds digits of the called subscriber's number as recorded on the register switches by the calling subscriber's dial. This can be done because the non-decimal arrangement of the switching equipment is orderly and there is, therefore, a systematic relationship between the input and output codes.<sup>2</sup>

It will be noted that the input code consists of a ground on one lead in

<sup>2</sup> Oscar Myers, Codes and Translations, *A.I.E.E. Transactions*, Vol. 68, 1949.



each of four groups of leads from the "thousands" and "hundreds" switches. This results in an output code of one marked lead in each of the three groups of non-decimal output code marking leads.

The operation is simple. The IB code is determined directly from the pairing of consecutive terminals on the thousands register. The IG code is

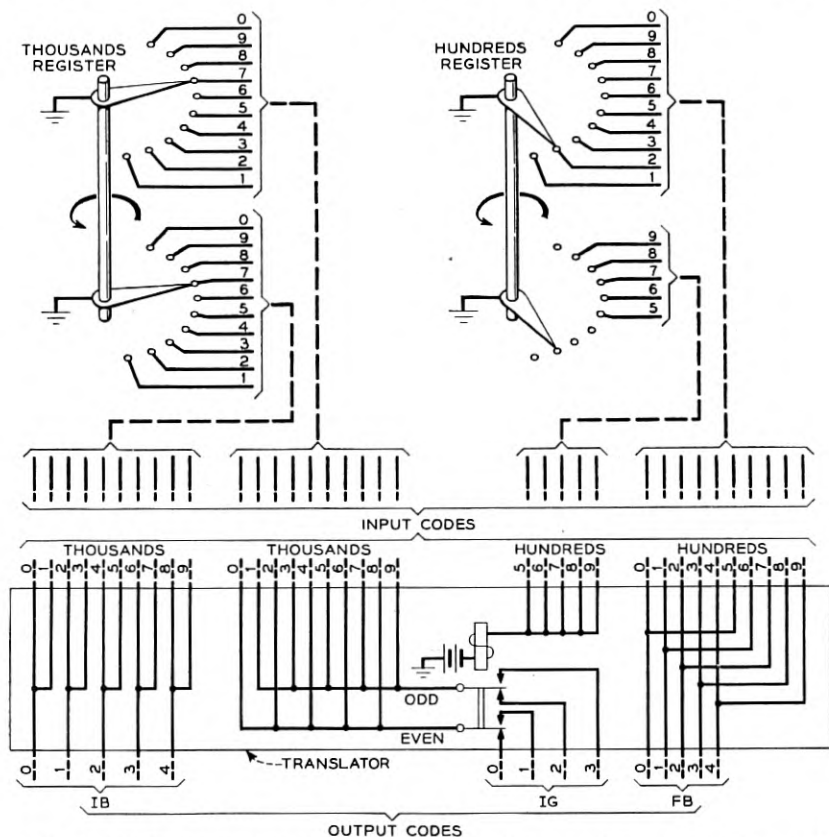


Fig. 1—Fixed, systematic translator used in panel system for converting thousands and hundreds digits of called numbers to 3-part code for control of non-decimal selections. (Changes information)

derived from a combination of the settings of the thousands and hundreds switches, which with the combining relay gives one of four possible code marks depending on whether the thousands number is even or odd and whether the hundreds number is in the first five or last five series. The FB code, one out of five, is derived from the hundreds switch by pairing the numbers in the first five series with those in the last five.

The simplicity of this type of translator makes it economical to furnish one for each switching element (sender) requiring its use rather than providing for use in common. In practice the translator is not even as discrete an element as shown here but is contained in and wired as part of each sender.

It will be obvious that this form of translator may also be constructed with other types of apparatus.

Figure 2 shows another form of fixed, systematic translator used as an element in changing the code for a number in a two-out-of-five system to a one-out-of-ten system. The operation is simply that a mark on two of the five input code leads will cause the operation of the two associated relays.

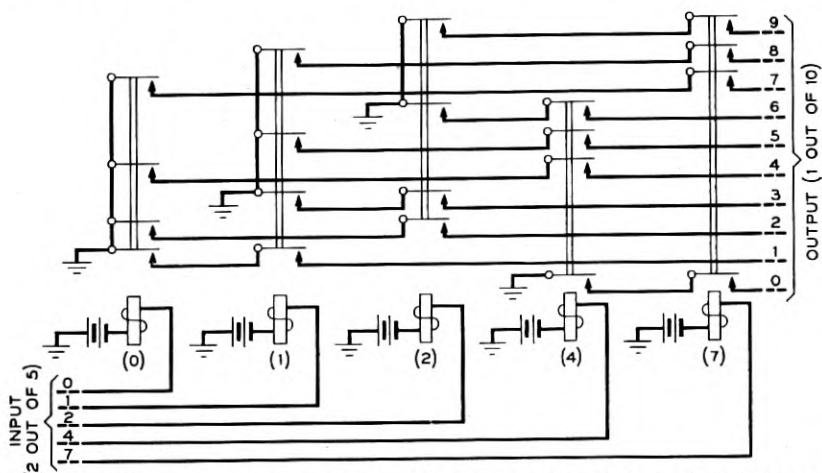


Fig. 2—Fixed, systematic translator for changing the coding for a single-digit decimal number from two out of five to one out of ten. (Changes "language")

which will place a mark on one out of the ten output leads. This element is repeated for each place in the decimal number involved or the same element may be used, by the addition of suitable controls, for translating numerous digits on a sequential basis. Note that the two numbers representing each input code add up to the number being translated, except that seven and four are assumed to add to zero.

Figure 3 shows a systematic translator for changing the code for a number from the decimal to the centesimal system. Here the input code consists of a ground on one lead in each of the two input groups and the output code consists of a ground on one of the hundred output code marking leads. The operation of one of the "tens" relays by the grounded tens lead connects

the units leads to ten output code leads one of which will then be grounded. By adding suitable relays this translator can be expanded to a system in which the output code is a mark on one of a thousand or ten thousand leads.

The arrangement of Fig. 3 occurs frequently as a selection device and when so used in the Bell System is now called a "relay tree".<sup>3</sup> It is often an element of other types of translators.

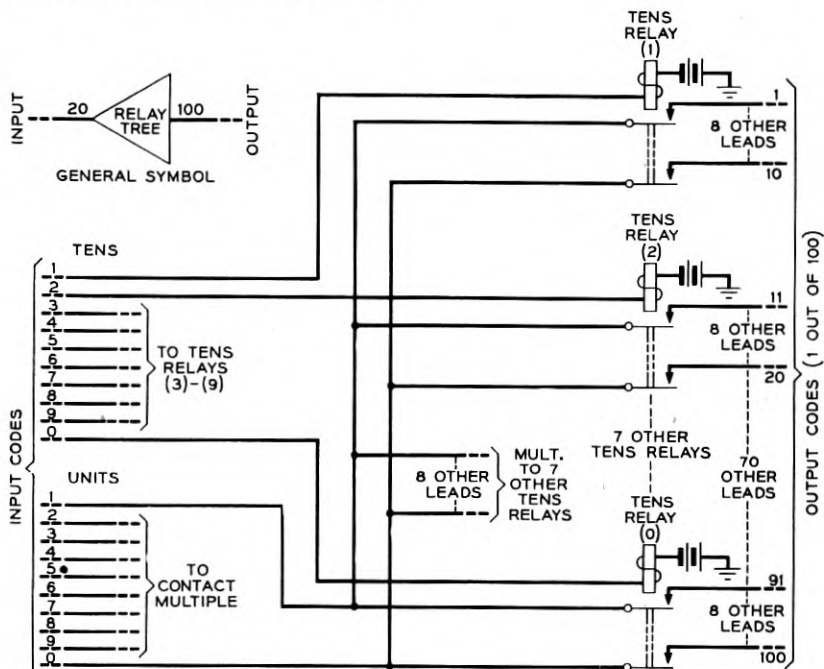


Fig. 3—Fixed, systematic translator for changing a 2-digit decimal number to a centesimal base. (Changes "language")

A 100-point step-by-step switch can, of course, be used as the equivalent of the foregoing type of translator, and is frequently so used. In this case the input code is two sets of decimal pulses to drive the selector to the required point, and the output code is again a mark on one out of one hundred output code marking leads connected to the bank terminals.<sup>4</sup> Such a switch is also used in place of a relay tree as a selecting element in other types of translators.

<sup>3</sup> S. H. Washburn, "Relay Trees" and Symmetric Circuits, *A.I.E.E. Transactions*, Vol. 68, 1949.

<sup>4</sup> This is an elementary sample of a translator with sequential input and combinational output.

*Changeable Translators*

The Western Electric Company's first full-automatic panel office made use of a rotary-type switch in order to obtain within the senders a trans-

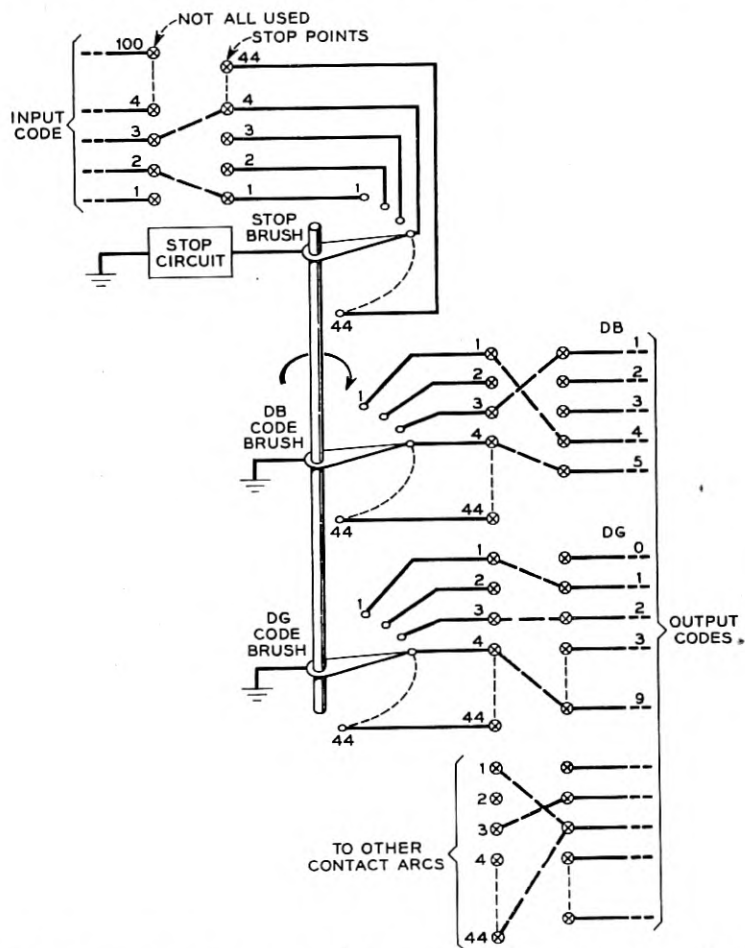


Fig. 4—Changeable office-code translator used in first full-automatic panel office.

lation from the 2-digit office codes to the switching control codes necessary to make the selections to reach the office containing the called number. The fundamental arrangement of this translator is shown in Fig. 4.

In this case the setting of the registers for the first two dialed digits causes a mark to be placed on a terminal in a one-out-of-one hundred

system. This terminal is cross-connected to a "stop-point" or code-point location terminal to which the rotary translator switch is driven. The brushes of the switch make contact with terminals in this position which are arbitrarily cross-connected to various groups of code marking leads for controlling the required selections.

In the first use mentioned above it was economical to provide one of these translators for each sender. In other cases, particularly those involving a larger number of translations, the device was sufficiently expensive to warrant a group of translators for use in common.

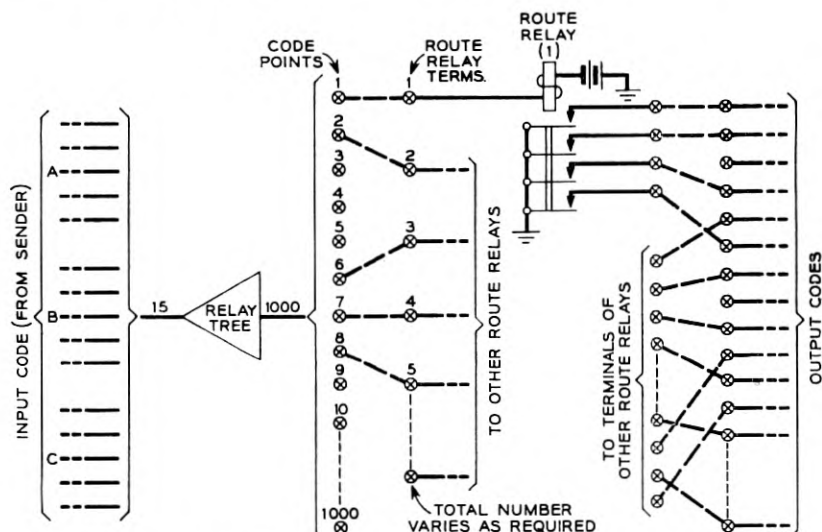


Fig. 5—Relay translator for 3-digit office-code.

It should be noted that this translation arrangement in reality consists of an initial decimal to centesimal translator such as shown in Fig. 3, formed by part of the register switches. This is followed by a cross connection which changes the systematic translation which has already been achieved to an arbitrary translation in the form of a mark on one of the stop points. This is followed by the translation achieved by the rotary switch which converts the one-out-of-one-hundred code to one consisting of marks on numerous code marking leads.

We shall now see how this method is in principle followed by translators used in modern systems.

Figure 5 shows a translator used in a modern crossbar office and, with variations, in other modern offices for translating the dialed office-code to

a switching control code. Note the similarity of the operation to that of the preceding translator.

The 3-digit decimal input code is first translated by the systematic relay tree to a one-out-of-one-thousand code. The marked code point is cross-connected to an arbitrary route relay, the contacts of which are in turn cross-connected as necessary to arbitrary code marking leads to provide the required information to the control equipment.

Depending on the economic factors determined by the holding times of the translators and elements served by them and the size of the translator, this type of translator may be individual to such elements or be arranged in common groups.

#### PRELIMINARY RESUMÉ

##### *Systematic Translators*

On the basis of the illustrations given so far, it is possible to make a few observations which will help in the discussion of the more advanced types of translators to follow.

In the first place, it will be evident that the systematic translators used as examples follow no common pattern. Each design has been carefully tailored to minimize the cost by taking full advantage of the particular relationship between the input and output codes.

This applies not only to the three varieties previously illustrated but also to numerous other forms used in practice. Among these may be mentioned systematic translators having the general functions of the one shown in Fig. 1, but dealing with the different numbering arrangements found in some common control systems outside of the Bell System.

Others are those used for translating code marking lead systems from a base of 1-out-of-10 to 2-out-of-5, 3-out-of-8 to 1-out-of-40, combinations of 4 to 1-out-of-10 or vice versa, etc. Translations between binary and decimal numeration also occur. These changes in the method of indicating a number within the same switching equipment are, of course, not due to engineers merely changing their minds, but are necessitated by the fact that, no matter what system of numeration and coding may be used within a switching system for the sake of efficiency, telephone numbers and charges for calls must be presented in generally understood terms in so far as the subscribers are concerned.

All of the above fixed translators involve rather simple relationships between the input and output codes. The rules for determining the output codes are in reality instructions for arithmetical treatment of the input code, but the translator does this in a special way without recourse to true computing processes.

There are some cases dealing with the charges for telephone calls where use is made of devices in which the input codes consist of two or more factors, the arithmetical combination of which determines the output code. Small scale electrical or clock-like mechanical computers *are* sometimes used here; but in other cases the devices, because of the limited number of possible outputs, are in reality systematic translators furnishing a limited reference table, although there is a tendency to call them *computers*.

Automatic computers could, of course, theoretically be used in many cases to provide the translation afforded by systematic translators; and it would also be possible to use one of the arbitrary translators mentioned in the foregoing section. We would then be in a position to say that the processes of systematic translation would follow a pattern for various types of applications. However, such arrangements could conceivably be economical only in cases requiring a large variety of fixed translations or in cases where the relationships between input and output codes were very complicated. It seems to the writer most likely that, for the simple types of systematic translation required in switching, special designs, each based on well-known principles of efficient switching network circuitry arranged to fit the special function, will continue to be the most economical.

#### *Arbitrary Changeable Translators*

Where there is no uniform relationship between the various possible input and output codes of a translator, and in particular where the output code associated with any given input code is subject to change, we can, in general, resort to one of two things. We can provide a separate translator for each input code, arranged so that it can be modified as the code associations are changed. This is seldom done in practice, and is economical only where there are very few codes and the changes are infrequent. The usual procedure is to provide a translator which can handle all or a substantial portion of the required codes in a uniform manner without regard in its detailed operation to the arbitrary and changeable relationship between the input and output codes, but providing facilities for the changes required.

The two changeable translators shown in Figs. 4 and 5, not only illustrate methods of doing this with switches and relays, but also illustrate the general principles used in all working translators of this type now in use. The main variations are improvements to reduce costs, either of the translator or elsewhere in the system, or to provide better tractability for changes.

The general principle is that the over-all translation operation consists in causing the input code to select a coding element which is capable of producing the required output code. Changes in the output code associated with the input code are made by wiring or other changes causing different



coding elements to be selected or causing the same coding element to produce different output codes.

For instance, in Fig. 4, the input code, consisting of a mark on one of 44 leads, causes the translator switch to select one of the 44 possible positions and the various terminals contacted in this position and their associated changeable cross connections constitute the coding elements to mark the required output code leads. The output code for any input code can be changed by changing the cross connections in the coding element. In this case they can also be changed by moving the "stop-point" cross connection to a new stop point causing a new coding element to be selected. (This permits numerous inputs to have the same output.)

Now, in the case of Fig. 5, the same general situation exists, each coding element consisting of a route relay with changeable output connections and with provision for changing the relay selected by the input code.

We can now go on to examine some of the variations of this general scheme in practice and theory.

#### NUMBER GROUP TRANSLATOR FOR #5 CROSSBAR OFFICE

Figure 6 shows the changeable translator used in the #5 Crossbar System to determine the equipment location number when the called directory number is known. This translator is arranged to handle 1000 directory numbers, but is not limited to this capacity.

The input code is a 3-place decimal number. The output is a mixed base number coded by one mark in each of six groups of code marking leads.

The operation here is preceded by the selection of the correct translator according to the thousands digit of the called number. Then the input code causes the relay tree to select (for three wires) a 3-point directory number terminal each point of which is connected to a coding element furnishing two items of output information.

The economy of this arrangement lies in the use made of resistances for the output coding elements. This is a sample of what has come to be known as passive-element coding, which will be further illustrated later.

The difficulty the designers faced in this case was that each translator, while having only 1000 possible inputs, must provide for translation to a very large number of possible outputs, each different and each consisting of a mark on six different leads. If the scheme of Fig. 4 had been employed, then the translator switch would have required provision for 1000 positions and 6000 wires. If the scheme of Fig. 5 had been used, then a total of 1000 coding relays would have been required and each coding relay would have had six contact sets and cross connections.

By employing a relay selecting tree carrying three wires and using three



separate cross connections for each directory number, each connected to a 2-code passive coding element permanently associated with one marking lead in each of two groups of output code bus bars, it was possible to effect considerable economy over at least schemes like those shown in Figs. 4 and 5.

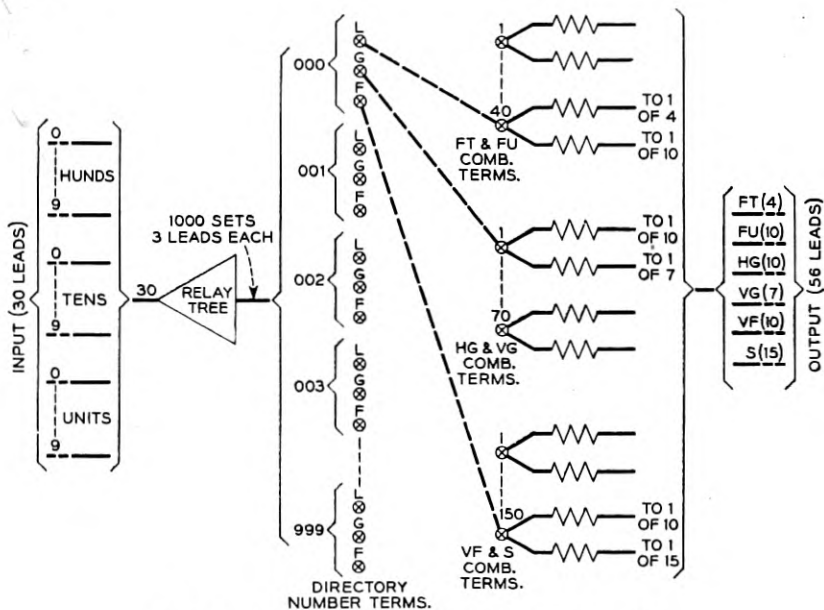


Fig. 6—Number group translator for #5 crossbar office.

### RING TYPE TRANSLATOR

An example of a translator with still further simplification and improvement of the coding equipment is that shown in Fig. 7, which is also used in the #5 Crossbar office, in this case for determining the directory number of a calling station when the equipment location number is known and it is desired to make a record of the calling number for charging for the call. This is the reverse translation of the case covered by Fig. 6.

As used in practice, the translator of Fig. 7 is limited to capacity for 1000 mixed base input codes, any of which may be translated to any of 40,000 directory number output codes in a 5-place decimal system.

Here the relay selection tree, under control of the input code, causes the selection of a one-wire circuit to one out of 1000 equipment number terminals each of which has a cross-connection wire which serves directly as the coding element for translating the associated equipment number. This is

due to the fact that each jumper is threaded through a common array of ring-type induction coils with one coil for each digit in each place of the output numbering scheme, that is, 44, as shown in Fig. 7.

After the equipment number selection has been made, a surge of current is passed through the jumper. Since the jumper acts as a single turn primary for all the coils through which it is threaded, a high voltage is induced in

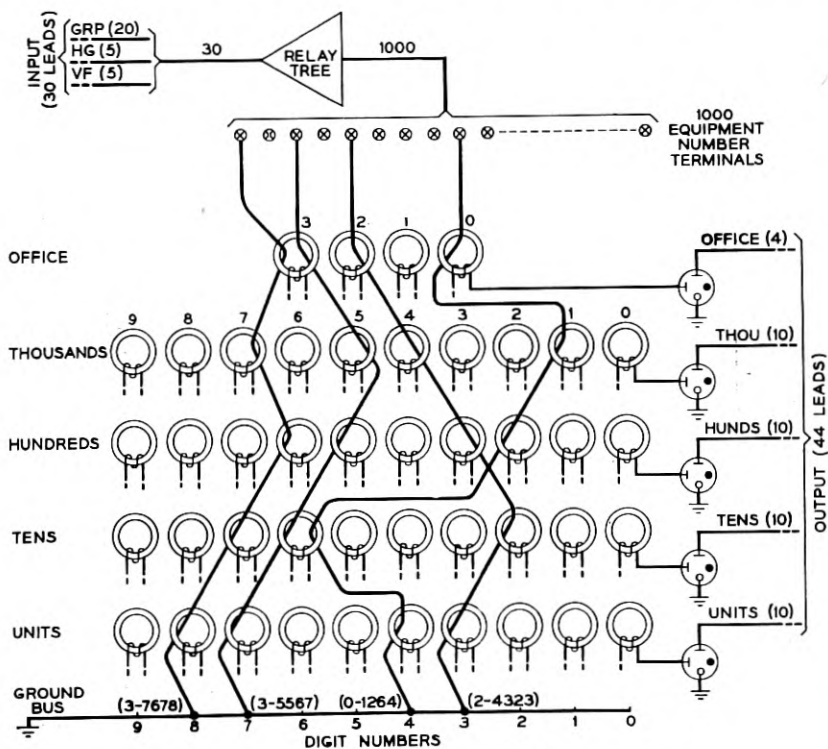


Fig. 7—Ring-type translator.

the high turn secondaries of these coils which causes the cold cathode tube associated with each of these coils to "trigger" and place a mark on the corresponding output code marking lead. The translation for any equipment number is changed by simply threading the associated jumper through different coils.

Many jumpers may be threaded through the same coils. The jumpers thus act as individual coding devices and the coils as a reading equipment common to all coding elements.

The improved coding element arrangements in this translator<sup>5</sup> have eliminated the output bus-bar multiple entirely (existing in Fig. 6 for instance) as each output code lead terminates in the translator only on the cathode of the reading tube. This improves the safety of the device and helps reduce the cost. But the main benefit is in the improved tractability for changes as they involve only one jumper per input code as compared to three separate jumpers in the previous case.

#### POSSIBLE METHODS OF TRANSLATOR COST REDUCTION

Let us consider again the translators of Figs. 6 and 7, which are modern types for local central offices and used on a sufficiently large scale so that the costs are important, and let us examine where there are possibilities of cost reductions.

Now these translators are designed for use in buildings with as many as four central office units, that is, 40,000 directory numbers and somewhat less equipment location numbers. Yet each translator can handle only 1000 input codes, so if both of these types of translation are involved, 40 translators of one type and somewhat less of the other will be required. Two things may immediately be considered as possibilities for reducing costs:

(1) Since the two types of translators effect translations between the same sets of equipment numbers and directory numbers but in reverse directions, it might appear that savings could be made by providing a single type of translator for both functions and arranged for 2-way operation without requiring duplication of the equipment.

(2) Since a considerable amount of equipment is involved in the large number of identical selection devices for coding elements in each translator and for connecting devices in the equipment selecting translators, savings might be made by reducing the number of translators required by increasing the speed of operation.

#### ELECTRONIC CROSS-REFERENCE SYSTEM (2-WAY TRANSLATOR)

The writer, as a result of considering the foregoing possibilities a number of years ago, proposed a 2-way translator intended to effect savings over existing schemes of using two different types of translators operating in reverse directions, as in the  $\#5$  Crossbar case discussed above. As it was

<sup>5</sup> It is interesting to note that the coding schemes of Fig. 6 and 7 were both proposed by Mr. T. L. Dimond of the Bell Telephone Laboratories. The arrangement of Fig. 7 is often referred to as the "Dimond ring" translator. This pun is one of the rare exceptions to dullness in switching nomenclature.

For an earlier version of the ring type translator see *U. S. Patent*  $\#2,265,884$  issued to Mr. F. A. Korn, of the Bell Telephone Laboratories. Note that this patent applies to an *identifier*.

desired to reduce the number of translators required to the ultimate of one, the great speed needed in order to obtain the required capacity indicated the need for full-electronic methods. Because, figuratively, the equipment could refer to the translation system in terms of one number and obtain the associated number as an answer, in either direction, much as in a cross reference card file, the system was called an electronic cross-reference system. It is promising, but not yet practical.

In this scheme the general concepts of other translators were not avoided but were applied in a different way. For instance the idea of having the input code select a changeable coding element was not avoided. However, the selecting element itself was eliminated, making possible important savings. The select function was combined with the coding function. Each coding element acts as its own selector under control of the input code, say code A, and then supplies the wanted output code, say code B. The operation may be reversed by supplying code B as an input, which results in auto-selection of the same coding element, but this time code A is supplied as an output.

The general arrangement is shown in Fig. 8.

The auto-selection and coding element for each associated pair of codes is a special cold cathode tube of the multi-anode type, having ten point anodes and a common cathode. The anodes are symmetrically spaced around the cathode so that each anode to cathode gap has the same characteristics. Only eight anodes are used in this illustration.

Since there is one tube for each pair of associated numbers there will be a total of 10,000 tubes in this case. Four groups of ten numerical bus bars for the A numbers and the same for the B numbers, passing the entire tube field and cross-connectable to the anodes of the tubes, complete the translator proper. The extraneous equipment consists of "inquiry" and "answer" sections, each of which may automatically be connected to either the side A or side B bus bar system. The "run-down" element of the "inquiry" section contains electronic pulsing and sequence-control equipment for causing the auto-selection and coding operations.

The operation is very fast, the entire auto-selection and output code reading process involved in each translation being completed in less than a millisecond if all the connecting switching gear is electronic.

Assuming that an "A" number, say 1925, has been recorded in the register of an inquiry section, that this section has been connected to the bus bar system of the "A" side and the corresponding answer section has been connected to B bus bar system, we are ready to make a translation. The electronic rundown equipment pulses into the bus bar system of the A

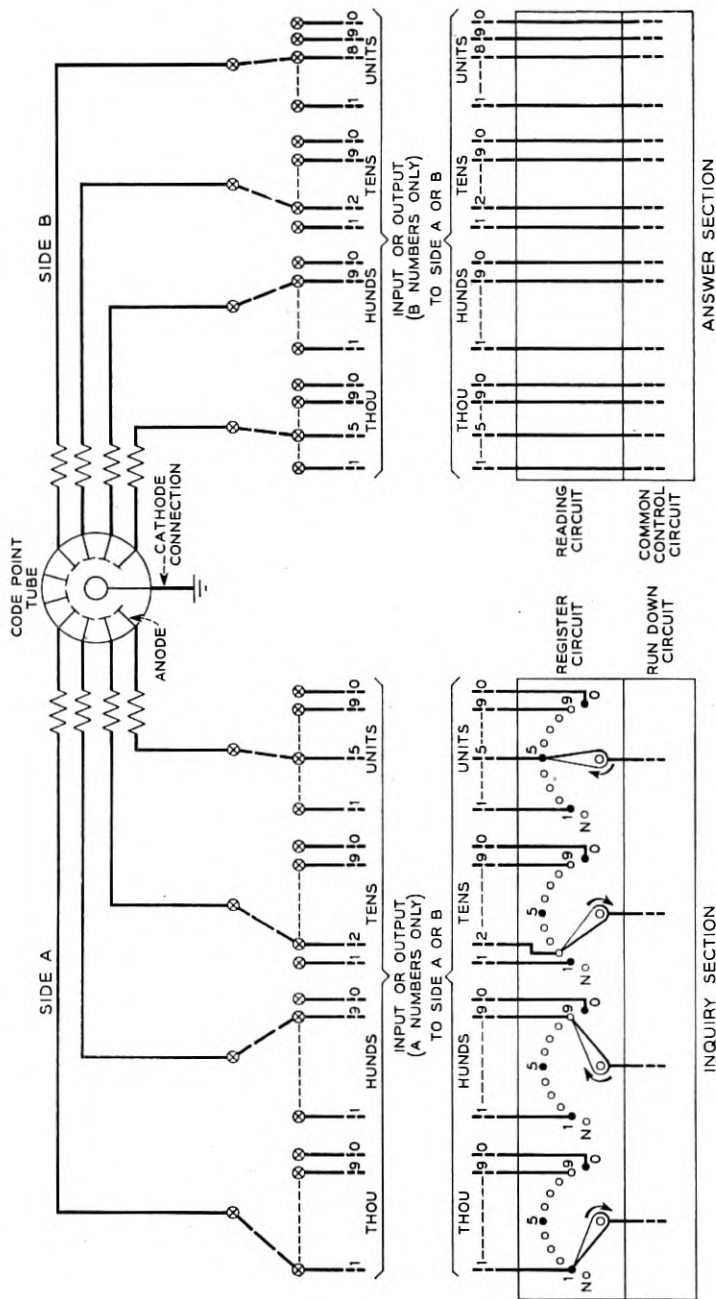


Fig. 8—Electronic cross-reference system (2-way translator).

side to cause the auto selection of the coding tube for (A) number 1925 as follows:

- (1) Breakdown voltage is applied through the thousands register to the \*1 bus in the thousands group of the A side. This causes breakdown, or firing, at the thousands anode of all tubes that have *one* as the thousands A digit. In a fully equipped system this would be  $\frac{1}{10}$  of all the tubes, or 1000.
- (2) "Sustain" or "hold" voltage (lower than breakdown voltage) is applied to the \*9 hundreds bus, followed by removal of the voltage from the thousands bus. This causes all previously fired tubes which have \*9 as the hundreds digit to be held ionized by the hundreds anode and the others dropped out, that is  $\frac{9}{10}$  of the total tubes dropped out and  $\frac{1}{10}$  or 100 held.
- (3) This drop-out process is continued through the tens and units steps so that first ten and finally only one tube (number 1925) is held by the units anode. Auto-selection of the code tube required is now complete and it remains only to read the coding of the B side.
- (4) Reading is accomplished by applying "hold" or "sustain" voltage to *all* the bus bars on the B side. This causes anodes connected to this bus system to fire by transfer in only the one tube which is ionized. Thus all B anodes in tube 1925 will be fired and the resultant currents in the associated bus bars (one in each of the 4 "B" groups) are marks which can be read in the electronic reading circuit to register the B output number 5928.
- (5) The translator is dropped by removing all voltages, causing the selected tube to deionize, and the translator is ready for another job. All of this requires only the time of the various breakdown and transfer steps and intermediate deionizing times, totaling less than 100 microseconds.

The advantages of possible cost reduction are obvious especially for large scale applications. The disadvantages, in the form here shown, are:

- (1) The possible hazards resulting from the fact that this is a single common unit with one-at-a-time operation.
- (2) The tractability for general use needs to be improved as it is necessary to change all the 4 or 5 required jumpers on one side of a coding element in order to make a translation change.
- (3) The special tubes need to be made in very large quantities in order to obtain the required low price.

Further work now being done may resolve all these difficulties.

This translator illustrates some of the variations the designer can consider

when dealing with a translation problem with a new approach, and also some which *must* be considered. These and other possible variations will be reviewed later.

#### SLIDE BAR TRANSLATOR

This proposed device was the forerunner of one of the Bell System's most important projected translators and is worth examining for the new principles.<sup>6</sup>

In this case, as in the cross-reference system just described, one of the objectives was reduction in cost by eliminating the separate equipment for selecting the coding elements and combining the selecting equipment as far as possible with the coding equipment. A further purpose was to provide non-electrical coding elements with improved tractability for making changes.

The construction of this device is shown in Fig. 9. The coding elements consist of thin slide bars shown in Fig. 9(a), each notched in accordance with the input code on one end and the output code on the other. These are stacked as shown in Fig. 9(d). The selecting equipment is a combination of code bars, Fig. 9(b), which work in combination with the wide and narrow notches of the slide bars, so that when a combination of select code bars is operated according to an input code all slide bars except the one carrying the input code are restrained from sliding when the common operating magnet is actuated. The one slide bar slides to the right, this representing the selection of the coding element.

The reading code bars are now all operated and only the set corresponding to the output code of the displaced slide bar can operate fully. Contacts on the output code bars then mark the output code leads.

Because this device is slow relative to relay or electronic translators owing to its mechanical elements, it has limited traffic capacity and would have to be duplicated many times in each office. Its advantage, however, lies in the fact that, when changes in translation are to be made, new coding elements (slide bars) can be prepared in advance and the changes made simply by substituting the new bars for the old, without changing cross connections.

#### CARD TRANSLATOR

While the slide bar translator is sometimes spoken of as a "card translator" it is similar to but by no means the same as the card translator which the Bell System proposes to adopt for toll crossbar offices.

<sup>6</sup> See *U. S. Patent #2,361,246* issued to Mr. George K. Stibitz.



This card translator will not be illustrated at this time, but the following general notes are in order:

The card translator is especially designed for toll application where a 3 to 6-digit decimal input must be translated to an output containing 30

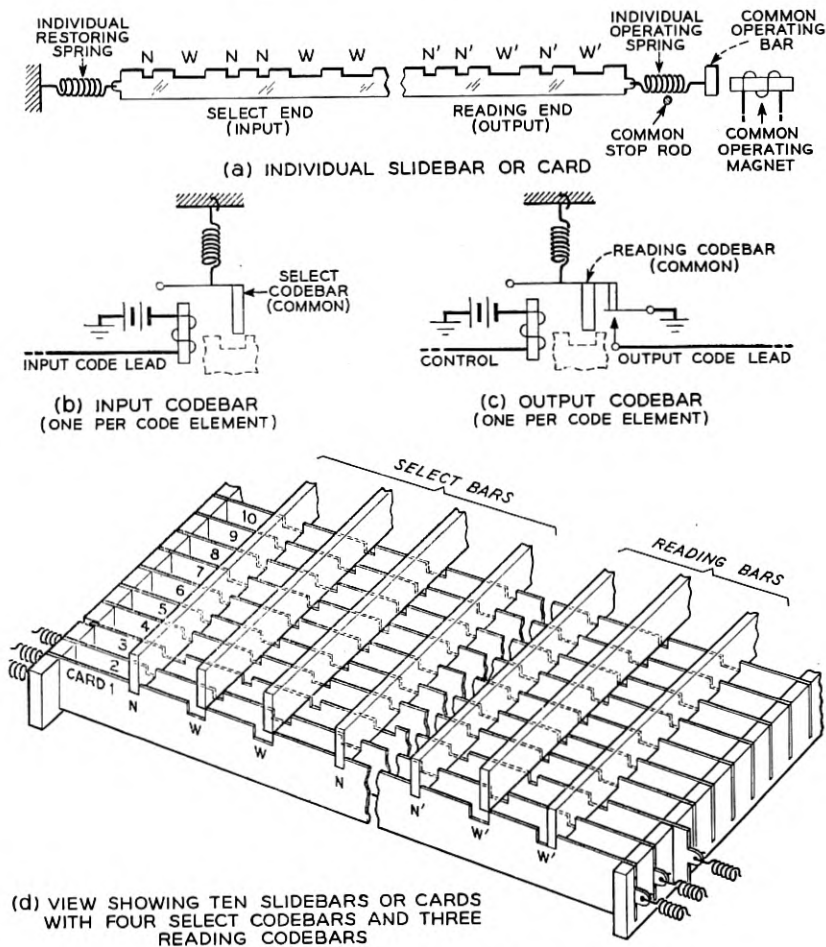


Fig. 9—Slide bar translator.

items of information with a very large number of possible combinations. The large number of possible outputs and the necessity of making changes quickly and frequently make the previously illustrated translators impracticable. For such application the card translator is well suited. The



coding elements consist of steel cards, notched at the lower edge according to the input code. The output code is carried on the card in the form of small and large holes.

The coding elements (cards) are selected, as in the slide bar translator, by selecting code bars that will drop the required card. Reading is done by photo cells (actually photo-transistors) which detect the presence or absence of light through the tunnels formed by the holes in the stacked cards.

#### PRE-TRANSLATOR FOR NO. 5 CROSSBAR SYSTEM

This all-relay translator is used for obtaining one of three possible sets of switching instructions at a time when it is impractical to consult the more complete office code translator associated with the marker in this system.

Provision is made for 576 possible input codes on a 3-place decimal basis each translatable to one of only three possible output codes each consisting of a mark on one of three output leads.

Because of these restricted capabilities it will not be illustrated here,<sup>7</sup> but its general principles are worth noting because it is an example of how close tailoring to the requirements can effect economy. This translator is one of the few examples of changeable translators which do not follow the general principle used in the previously described translators of this type. That principle, it will be recalled, is that each input code causes the selection of an individual coding element which determines the output code and changes are made by causing the selection of different coding elements or by changing the output of the coding elements.

In this pre-translator the input codes are teamed in groups of three, each of the group causing the selection of the same code point terminal. Hence only one third as many code point terminals are required as there are inputs. Each of these code point terminals is cross connected to one of 27 terminals, each of which represents a different permutation of each of the three possible outputs. At this terminal three possible answers are represented, and three relays beyond this point select the correct answer, depending on whether or not the input code is the first, second or third of the group of three.

Changes for any input code are made by changing the jumper affecting this code and its two associated codes to a new terminal representing the new permutation.

This arrangement reduces the required number of relay contacts, cross connections and cross-connecting terminals as compared to the number required by more conventional all-relay translators. It would become im-

<sup>7</sup> For a full description, see Pre-translation in No. 5 Crossbar, R. C. Avery, *Bell Laboratories Record*, April 1950.

practical if only a few more output codes were required because of the great increase in the number of permutations.

## GENERAL THOUGHTS ON TRANSLATORS

### *Broad Considerations in Choosing Translation System*

If a switching system designer were confronted with the problem of providing a large scale translator and he were given full latitude, there would be many factors he would have to consider and a wide variety of choices he could make in order to reach the most economical and useful design, even if he were uninventive and restricted to the combinations of the present art. Actually he would not have full latitude, for the translator design must always be coordinated with the design of other elements and often is subordinated or greatly limited by the importance of the more intricate or costly elements with which the translator must work. The best he can do is to arrange his design to help provide the most economical and serviceable system from an overall standpoint, and in this the translator design might not be ideal.

What are some of the more important factors that the designer can juggle and what choices can be made within the known possibilities?

Let us assume that the problem specifies, for some new type of common control office, large scale translation between equipment and directory numbers in both directions. Then the following decisions are certainly important:

(1) One-way or Two-way Operation?

This is determined by economic study if the available two-way translators are satisfactory.

(2) Should translators be provided for each circuit requiring their use or should they be provided for access to circuits in common?

This can be determined only by economic studies.

(3) If the translation system is to be common, should it be based on the use of a single full-capacity translator or numerous smaller translators involving more translator connecting devices?

This involves economic studies and questions of the speed or traffic capacity of the translator and the question of relative service hazards in the two arrangements.

(4) What type of translator shall be used?

This involves economics, speed, reliability, types of apparatus available and tractability for making changes.

### *Detailed Considerations*

It has already been brought out that large scale changeable translators known to the present art follow the same basic concept and that there are, as a result, certain problems common to all of them.

Let us examine Fig. 10 to see how some elements can be varied to change costs. This figure shows four general arrangements for coding, all working into the same output code marking lead bus-bar system. In practice all of the coding elements of the same translator would be of the same form and a large number would appear before the bus system or sections of the bus system. In the general arrangement covered here the coding elements are arranged in numerical order each permanently connected to its associated output bus bars. Types of systems having no output bus-bar multiple, such as the Dimond ring, slide bar or card translators are not illustrated, but covered in the discussion.

Starting at the left with the input code leads, we have theoretically much choice as to the types of signaling (various combinational or sequential types) and the system of numeration making up the input code. In spite of this freedom most translators in use employ decimal inputs with signaling generally on a code marking lead basis or sometimes on a decimal pulsing basis. Where code marking lead signaling is used the marks are almost always simple off-on marks and, for decimal notation, each place is represented by a 1 out of 10, 2 out of 5 or combinations of 4 group. The practical choice is limited by the fact that it has usually been uneconomical to change the coding system of the translator input to other than that existing at the output end of the relay or switch devices making use of the translator, as this would require a change of language by intermediate translation.

If it were not for these limitations the number of input leads could be reduced by use of binary numeration for marking leads, or by signaling over a single pair of wires with any of the other well known methods of signaling. However, the reduction of leads could, in any case, effect only minor savings as the leads are short. The largest savings possible would be in the reduction of the amount of translator selecting equipment required.

The language of the translator input, of course, also affects the design of the coding element selector which could be any type of selecting equipment such as switches, relays, tubes or code-bar mechanisms or, in one proposal, self-selecting coding elements. Relay trees are frequently used and optimum designs for the common types of inputs are well established. For different types of inputs the design of such trees profits by mathematical analysis.

Now we come to the coding elements themselves.

Because of the large number of these, optimum design is important. In the illustrations and in most applications the output codes are formed by the coding devices placing marks on the output code marking leads. The output language could be different, of course, and possibly with economy in the translator itself, but the philosophy of using code marking leads in the output end is the same as that just mentioned for the input system.

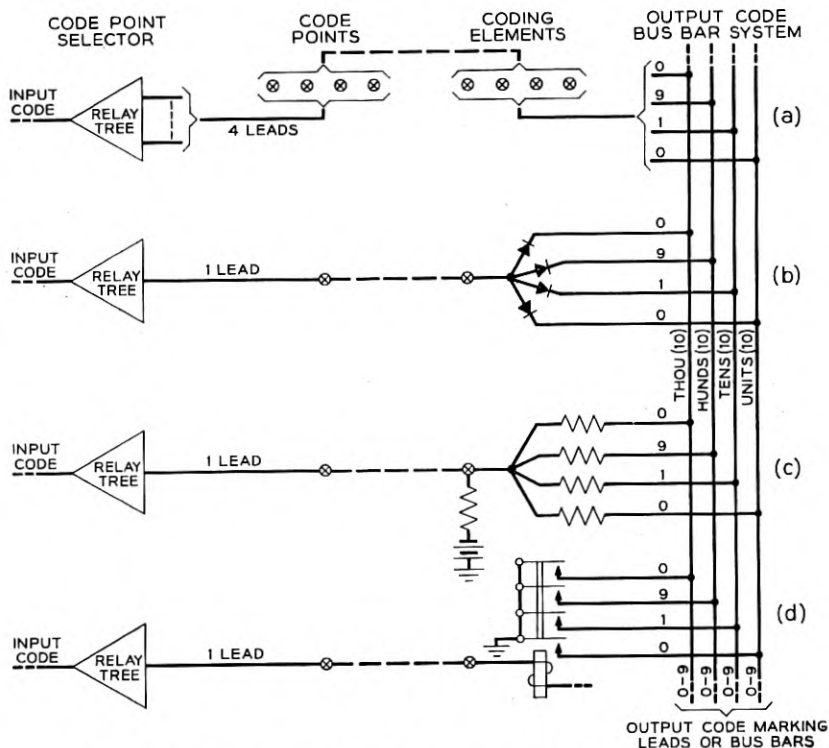


Fig. 10—Four general methods of output coding in changeable translators.

The coding methods shown require a multiple of all or part of the output code leads before the coding elements, that is, a bus system. It is the problem of the coding elements to mark the required buses in the various output groups without causing false marks on buses not involved by "back-up" through the connecting network or at least keeping the back-up below levels providing adequate discrimination between wanted and unwanted signals.

This back-up problem is solved in translators of the Dimond ring, slide

bar and card types by the fact that there is no bus-bar system, each output lead having only one connection on the translator. Each electrical bus-bar is replaced respectively by a coil, a code bar channel or a light tunnel.

In systems using bus-bar output multiples the coding elements solve the coding and back-up problems in four general ways, as shown in Fig. 10, assuming that the output consists of four code groups:

- (1) By having an individual lead through the code-point selector and cross-connection field for each code group as shown in Fig. 10(a). Each lead is directly connected to the required bus. All are open in the selector except those involved in any individual translation, thus avoiding back-up. No apparatus other than the wiring is involved in the coding element. This saves apparatus, but the cost of the cross connections and the wiring apparatus in the coding element selector are higher than for the other cases because of the larger number of wires.
- (2) Figure 10(b) shows a one-wire arrangement with back-up prevented by a unilateral or non-linear element in each lead to the bus-bars.
- (3) Figure 10(c) shows a one-wire arrangement which is connected to each of the four required output buses through a terminating and coding resistance network which reduces back-up through the unilateral effect provided.
- (4) Figure 10(d), again one-wire through the selection and cross-connect field uses a relay to effect coding, back-up of course being prevented by the fact that the leads to all code buses are open at the relay contacts except those involved in a particular translation. Figure 8 operates on similar principles for avoiding back-up.

Schemes (b), (c) and (d) reduce the wiring and the selector costs as compared to (a) through the use of only one wire but this is done at the expense of the additional apparatus in the coding element.

Figure 6 shows a compromise between (a) and (c) of Fig. 10.

Different conditions for one-way translators may warrant a careful choice between one of the four general methods of coding shown in Fig. 10 and the methods involving no bus-bars mentioned above.

In the case of the arrangements of Fig. 10 some cost changes can be effected by juggling, in the design, with the coding elements and the bus-bar grouping.

For instance, if the number base of the output were changed from decimal to binary, the number of output bus-bars would be reduced from 40 to 14, but the increased number of places in each output would require 14 leads from each coding element instead of the present four. This complication of the coding elements would probably prove-out this change.

On the other hand, if the output number base were changed to centesimal the number of output bus-bars would be increased to 200 but the *reduction* in the number of places would *reduce* the number of output leads from each coding element to two. This might be of possible economic advantage. If it were not permissible to end up with a centesimal output, it would be necessary to provide a secondary translator to change the output from a centesimal to the decimal or other system required, and this would reduce the economy of the centesimal output.

#### WHAT ABOUT NEW CONCEPTS FOR TRANSLATORS?

There are in the "proposed" state numerous interesting variations of the basic translator principle which has been outlined. These have as objectives lower system first costs and improved tractability for changes.

The general concept for changeable translators of the non-systematic type which has been repeatedly stated above appears in so many forms in practice and in inventions and suggestions ranging through varieties of mechanical, electromechanical, electro-optical and electronic types that one might wonder: (1) Is it not possible to design a large scale changeable translator with a different basic concept? (2) If it is possible, what might be gained?

We have noted, in the "pre-translator," a departure from the general concept of changeable translators, and there are others. However, they are all limited scale types applicable to special conditions.

A recently published article<sup>8</sup> indicates a new line of attack on the problem of obtaining a changeable translator with greater speed and reliability so that it could be used to carry a greater load than now customary. Not enough details are given to indicate whether the electronic arrangements outlined depart from the basic concept of existing translators, but, if they do not, they at least present interesting variations.

What could be gained by entirely new concepts for translators can not be answered in advance except in terms of what would be welcome. The present general designs give good performance and there is little need for improvement in this respect. What is always welcome is lower costs, particularly the cost of making changes.

#### IDENTIFIERS

##### WHAT IS MEANT BY "IDENTIFIER"

It was stated at the beginning of this paper that identifiers are relatively new in the switching art. This applies to identifiers which are sufficiently

<sup>8</sup>T. H. Flowers, "Introduction to Electronic Automatic Telephone Exchanges—Register—Translators," *Post Office Electrical Engineers' Journal*, January 1951.

limited in function or distinct as a switching unit to be so labeled. Un-named identifiers and identification processes have existed since the early days of the switching art. Only patent attorneys recognized these early arrangements and called them by their proper names.

Let us confine ourselves, for the moment, to the type of device generally named as an identifier. This is a device for indicating in code form the designation of a line, station, trunk, frame or other unit to which the device has a connection. The connection is generally electrical, but could conceivably be physical, optical or electro-magnetic.

#### EXAMPLES OF USE

The term identifier first came into general prominence in connection with the introduction of "automatic ticketing" in the United States and Europe. These are systems used in connection with subscriber dialed toll calls for printing automatically a ticket carrying calling and called numbers and other details necessary to charge for the call. The identifier is the automatic device for determining the calling number, a function ordinarily performed in manual service by the operator asking for the number. With the identifier it is also possible, on those calls requiring the service of an operator, to display the calling number automatically so that the operator's request can be avoided.

Another example is found in Bell Crossbar Offices of the toll type in which it is necessary for certain equipment to determine the number of the frame on which a calling trunk is located. For this purpose, what is known as a "frame" identifier, is used.

Of course, no arrangement for fully automatic completion of long haul toll calls can be successful without an automatic system for making a record of the details necessary to charge for the calls, including automatic identification of the calling number. Identification processes will, therefore, become more and more important.

#### TYPICAL IDENTIFIERS

##### *General*

Existing identifiers follow a number of basic concepts but many variations of these fundamental notions are possible. A few examples illustrating the different concepts with some of their variations will be given. The task is simplified because some of these concepts have a strong resemblance to translator principles which have already been discussed.



### *Calling Number Identifiers—Searching Type*

Figure 11 shows an identifier used in the Bell System's first application of automatic ticketing<sup>9</sup> and with variations in similar applications elsewhere.

The principle here is that the identifier, through the outgoing trunk to which the calling line has been extended, applies a tone to the sleeve terminal of the calling line, utilizing the sleeve through all the switching stages. This tone finds its way through an equipment-to-directory number translating jumper to one terminal common for each 1000 numbers, one common for each 100 and one per number in each one hundred block. The numbers of the terminals with tones correspond to the various decimal digits of the calling numbers.

Relays connect tone detection equipment sequentially to the various thousands, hundreds, etc. terminals and each time the tone is found the corresponding digit is registered on relays in the identifier. These relays mark the output code leads.

The variations in other identifiers of the searching type consist of the use of switches or tubes instead of relays, searching for special d-c. voltages instead of tone, and in transmitting the digits of the identified number back to the source of the identifying signal by pulses over the sleeve instead of transmitting them to code marking output leads.

This type of identifier is obviously rather slow because of the sequence of operations and it is, therefore, necessary to provide a plurality of identifiers for each office. The different identifiers are prevented from interfering with each other by preference lock-out circuits, by discriminating signals such as different tones or by other special means.

### *All-Electronic Calling Number Identifier*

This proposed identifier works much like a translator in that a coding element individual to each number is selected and this places marks on a decimal bus-bar output system. Referring to Fig. 12, it will be noted that there is a directory number field with a multi-anode tube of the type used in Fig. 8 for each number plus an RC filter to discriminate against surges. This identifier provides for party lines and for class of service indication requiring the two extra cross connections shown.

The operation consists in the application by the control unit through the trunk and the switch sleeves to the line equipment sleeve terminal of a 10-millisecond pulse of  $\pm 135$  volts. This finds its way through the normal

<sup>9</sup> O. A. Friend, "Automatic Ticketing of Telephone Calls," *Electrical Engineering*, Vol. 63, Transactions, pp. 81-88, March 1944.



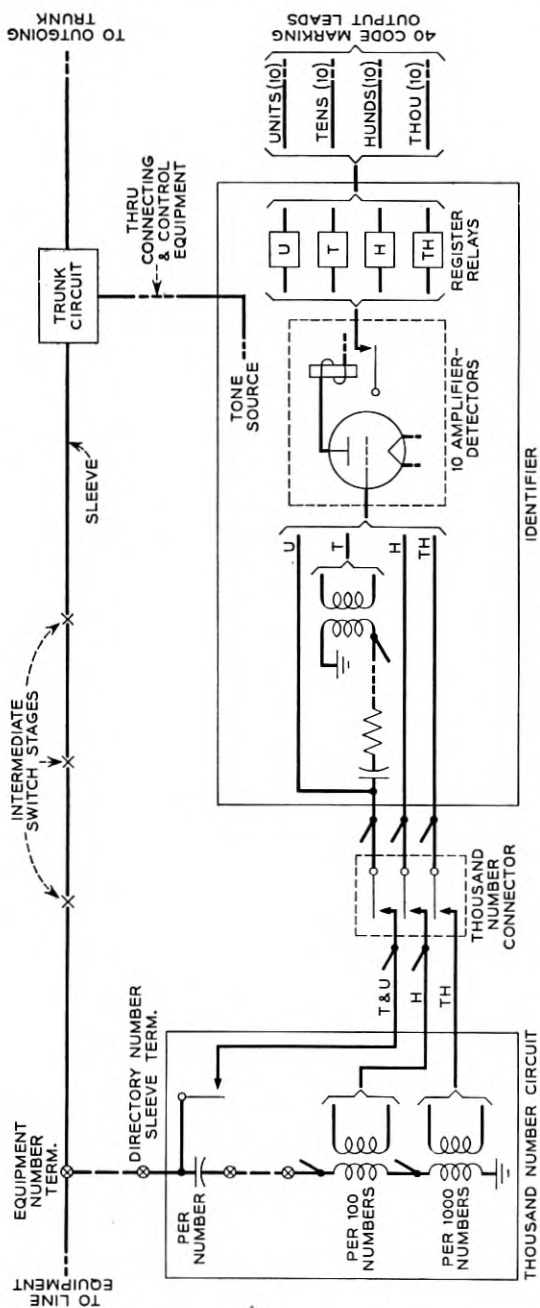


Fig. 11—Calling number identifier of tone-searching type.

cross connections (acting as translators) to the sleeve terminals of all the directory numbers associated with the line and the start anodes of the corresponding directory number tubes, firing each of these tubes at the start anodes. Continuous hold voltage (+75) is also placed on one of the hold or "party" buses through the party station register (station identity previously determined).

When the operating pulse ends, the cessation of current through the start anodes causes all the operated tubes to de-ionize except the one for the

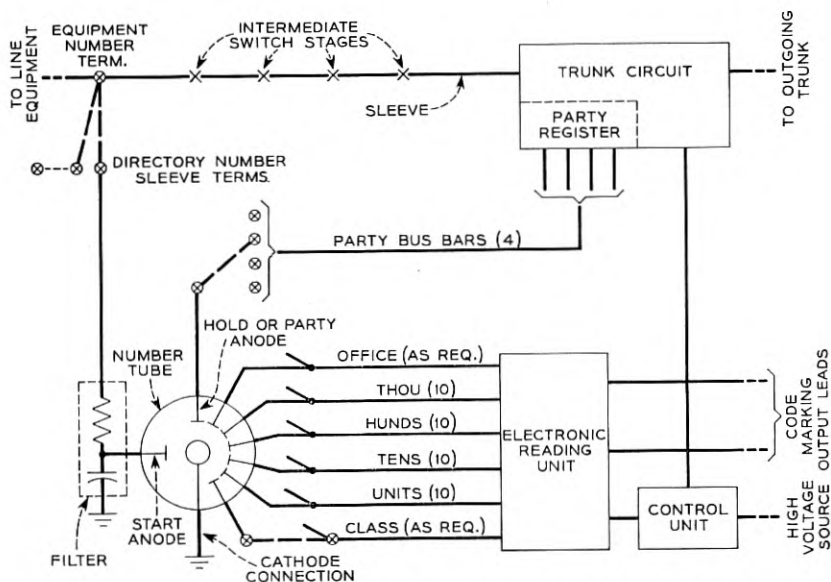


Fig. 12—All-electronic calling number identifier with multianode tubes as coding elements.

wanted number which is held, through ionization transfer, by the party anode. The tube for the wanted number has now been selected and the output code is read by the reading circuit by applying hold voltage to each bus of the system, causing all the coding anodes in the wanted number tube, and in only that one to conduct. Current then flows in one bus in each group, including the class of service group, which indicates the digits for the number and class of service, which are registered electronically in the read unit. The register reading is transferred to the output code marking lead system.

This operation requires less than 30 milliseconds, which means that a

single identifier has capacity to handle up to 50,000 numbers on a one-at-a-time basis.

This identifier has not been developed beyond the laboratory stage.<sup>10</sup>

#### *Passive Element Identifier*

Figure 13(a) shows a scheme for the identification of calling numbers by the use of an identifier containing passive element output coding devices consisting of condensers and resistors. The general operation is simple. A tone is applied, again through the trunk and the switching sleeve and the normal equipment to directory number cross connection to the directory number terminal. This tone passes through the termination and coding elements to the bus-bar system, the code for the associated number being formed by a tone mark on one bus in each decimal output group. Common detectors and discriminators detect these tones and register the identified number on an associated electronic register. From the register it is transferred to the output system by suitable marks on the code marking leads.

This identifier operates in about the same time as that of Fig. 12, that is, less than 30 milliseconds, and is therefore also capable of serving 50,000 lines without duplicating the translator equipment. This device is not in use but is testing satisfactorily in the laboratory. So far it seems the most economical of all the various types suitable for calling number identification.

#### *Identifiers Arranged for Transmitting Identity Code to Input*

Figure 13, (b) and 13(c), shows two variations of identifiers operating on the basis that the identification code formed by the coding elements is not transmitted to a common output but back to the input end.

Figure 13(b), used for calling number identification, uses non-linear coding elements which are non-conductive normally. When a start voltage is applied to one of the coding elements from a trunk or a control element through the switching stages on one of the switching conductors, the coding element becomes conductive and connects the input end to the coding bus-bar system which causes the identifying code to be sent back to the inquiring source. The identifying codes in this case are combinations of frequencies or multiplex pulses. This device is in commercial operation.

Figure 13(c) shows a similar arrangement used in the Bell System for identifying frame numbers in a crossbar system.<sup>11</sup>

<sup>10</sup> There is no published material on this identifier. A patent has been allowed Mr. R. P. Murphy of the B. T. L. An earlier version of this identifier is described in *U. S. Patent* 2,319,424 issued to Mr. M. E. Maloney of the B. T. L.

<sup>11</sup> O. Myers, "Multifrequency Frame Identification in Crossbar Toll," *Bell Laboratories Record*, September 1944.

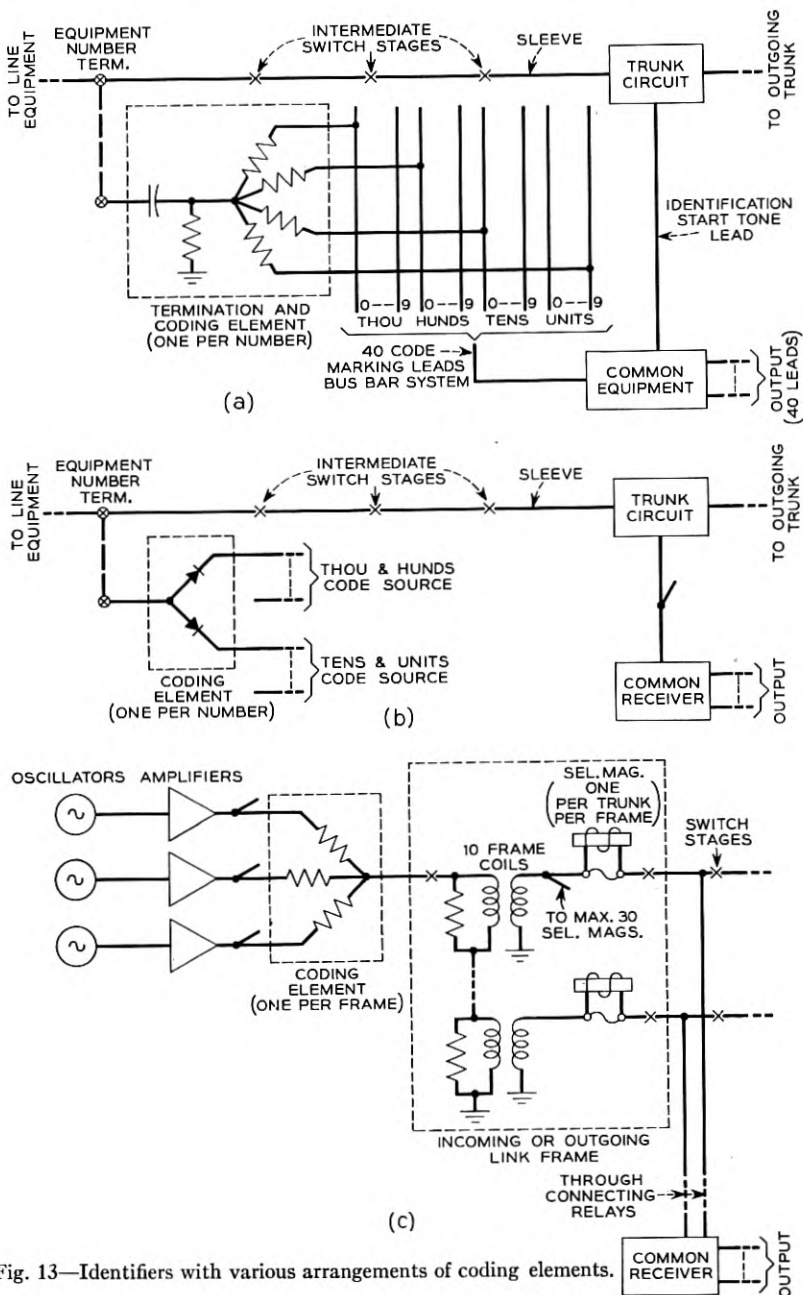


Fig. 13—Identifiers with various arrangements of coding elements.

Here all the trunks in the same frame have one lead permanently connected to a coding element which produces a 3-frequency code. This lead is connected to common detection equipment when the frame identification is wanted and this registers the frequencies and converts them to an output on a code marking lead system.

#### *Identification by Reading Positions of Normal Selectors*

There are numerous patents and a few commercial systems involving identification of calling subscribers, calling trunks, selected trunks, selected senders or registers, etc., in which the identification process consists in reading the position of the line finder, switch or relay unit which has been operated to select the line, trunk or other unit involved.

One of the oldest patents on calling number identification involves this principle.<sup>12</sup> The number of the switch group plus the number of the switch setting may in certain cases correspond directly to the wanted code. In other cases this indication must be translated, sometimes to a new arbitrary code and sometimes to obtain a code in a different numeration or signaling system.

Two methods of reading the switch positions are used: (1) counting of the steps or checking and registering other action taken by the switch *during* the time the involved line or other circuit is being selected, and (2) checking and registering the switch position *after* selection.

One of the interesting applications of this method for calling number identification where the line finder group and position numbers indicate the calling number is illustrated in the article covered by footnote No. 2.<sup>13</sup>

#### *Comments on Identifiers*

The identifiers we have discussed are divided into the following general types:

- (1) Searching types
- (2) Coding element types
  - Type A—with transmission to common output
  - Type B—With transmission to source over input lead
- (3) Switch position reading types

The first two types depend on the use of a considerable amount of equipment comprising the identifier and a large number of leads, and the problem of economy is generally solved by using one of the regular conductors of each connection (usually the sleeve) as part of the identification circuit.

<sup>12</sup> Mr. W. W. Carpenter et al, U. S. Pat. 2,112,951.

<sup>13</sup> R. F. Stehlik, "La Louviere Automatic Network of the Belgian Telephone System," *The Automatic Electric Company Technical Journal*, January 1951.

This procedure, in practice, requires that the identification equipment be arranged to discriminate against the surges directly introduced on these leads by other equipment connected to them and against crosstalk resulting from capacitive or magnetic coupling to other leads. This is done in part by adjustment of impedances and by discriminators either at the coding elements or in the common circuits. The problem of feedback has been discussed under Section II. There is no special technical difficulty in this but the economics is important.

The third general type has no special problems due to large numbers of wires or coding elements or feedback or crosstalk. The economic considerations involve the additional equipment in numerous other switching elements to read the switch positions and the fact that, if the switch position does not directly give the wanted number in the desired code form, resort must be made to additional translation.

Future possibilities lie in new methods avoiding these various problems economically, probably methods with entirely new basic concepts.

#### CONCLUSION

#### GENERAL REMARKS

The similarity of the problem of identification to that of translation is obvious. In identification the general problem is to construct an output code for information on an item to which the identifier has a connection. In translation the problem is to construct an output code for information on an item for which the translator has a previously registered code to use as input information.

Now, if we stretch the point, we could very well say that the identifier, because of its connection to the object being identified, also has an input code when a signal to start identification is applied to this connection, the input code simply being a mark on a one-out-of-X basis.

What is probably of more importance than the similarity of identifiers and translators is the frequent occurrence in switching networks of elements that functionally or operationally or both are essentially translators or identifiers although they are not so named.

The ordinary line relay, responding to the subscriber when he starts a call, can be considered as a coding element in a translator in some simple dial systems, as it translates an input code consisting of a mark on one out of 100 leads to marks in a two-place decimal system used to direct the line finder to the line. In the case of the No. 5 crossbar system, the line relays and their associated group equipments, although not so named, can certainly be considered as a identification system as the end result of their operation is the registration of the calling equipment number in a common

unit and in some patent literature the term "calling line identifier" is actually used. In practice this number is later used as an input to a translator (Fig. 7) to obtain the calling directory number. Any relay not serving merely as a means to renew or register a signal acts as the coding element of a translator or identifier.

Finally it looks to the writer as if any of the units made up of numerous relays, or other devices, as used in common control systems, act like translators with a vast number of possible input and output combinations with the action resulting from the output codes often fed back as part of a new input code.

There might then be considerable possibility that any fundamental improvement in general switching network theory or in the theory of translators and identifiers would be of mutual advantage.

#### ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Messrs. O. A. Friend, O. Myers and R. Marino, of the Bell Telephone Laboratories, for their assistance in the work on this paper.

#### BIBLIOGRAPHY

- "Historic Firsts: Translation," p. 445 of *Bell Laboratories Record*, November 1948.
- E. B. Craft, L. F. Morehouse, H. P. Charlesworth, "The Machine Switching Telephone System," *A. I. E. E. Journal*, April 1923.
- F. J. Scudder and J. N. Reynolds, "Crossbar Dial Telephone Switching System," *A.I.E.E. Transactions*, V. 58, 1939.
- Oscar Myers, "Codes and Translations," *A.I.E.E. Transactions*, Vol. 68, 1949.
- T. L. Dimond, "No. 5 Crossbar AMA Translator," *Bell Laboratories Record*, January 1951.
- O. J. Morzenti, "Number Group Frame for No. 5 Crossbar," *Bell Laboratories Record*, July 1950.
- R. C. Avery, "Pre-Translation in No. 5 Crossbar," *Bell Laboratories Record*, April 1950.
- J. A. Lawrence, "Contemporary Telephone Mechanization Abroad and Possible Future Trends," *Post Office Telecommunications Journal*, August 1950.
- T. H. Flowers, "Introduction to Electronic Automatic Telephone Exchanges: Register-Translators," *The Post Office Electrical Engineers' Journal*, Jan. 1951.
- O. A. Friend, "Automatic Ticketing of Telephone Calls," *A.I.E.E. Transactions*, V. 63, 1944.
- William Hatton, "Automatic Ticketing of Long Distance Connections," *Electrical Communication*, V. 18, 1940.
- J. E. Ostline, "The Strowger Automatic Toll Ticketing System," *Strowger Technical Journal*, June 1940.
- R. F. Stehlik, "La Louviere Automatic Network of the Belgian Telephone System," *The Automatic Electric Technical Journal*, Jan. 1951.
- G. T. Baker, "Calling Line Identification in Automatic Telephone Exchanges," *I. E. E. Journal*, Vol. 94, Part III, No. 28. March 1947.
- R. Taylor and J. McGavin, "The A. T. & E. System of Calling Line Identification," *The Strowger Journal*, April, 1949.

# Waves in Electron Streams and Circuits

By J. R. PIERCE

(Manuscript Received Jan. 9, 1951)

This paper reviews some of the assumptions made and some of the general problems involved in analyzing the behavior of electron streams coupled to circuits. It explains why a wave approach is used. The propagation constant of the wave is obtained in terms of the properties of the electron stream and the impedance of the circuit. Some general properties of waves are discussed. The importance of fitting boundary conditions in the solution of an actual problem is discussed, and examples, including that of "backward-gaining" waves, are discussed.

## INTRODUCTION

Of recent years, a good deal of work has appeared concerning small linear perturbations of uniform clouds of electrons and ions.\*<sup>1-4</sup> A number of questions can be raised concerning the physical interpretations of such mathematical labors.

First of all, for there to be a very direct physical interpretation, the unperturbed state must exist at some time or place and then be modified in the manner described by the perturbation. This condition is satisfied, for instance, in the case of an electron stream of moderate current shot into a long metal tube and confined by a longitudinal magnetic field. However, if the current is made large enough, the uniform flow becomes unstable<sup>5, 6</sup> and the method of perturbations can be used only to establish such instability and not to determine what form the flow will assume. I feel some misgivings about drawing physical interpretations from perturbations of uniform d-c. plasmas and infinitely extending clouds of charge unless these unperturbed states can be shown to exist physically, or unless the results can be shown

\* A few late references only are given; others are quoted in those cited.

<sup>1</sup> D. Bohm and E. P. Gross, *Theory of Plasma Oscillations: A. Origin of Medium-Like Behavior*, *Phys. Rev.*, Vol. 75, pp. 1851-1864 (1949); B. Excitation and Damping of Oscillations, *Phys. Rev.*, Vol. 75, pp. 1864-1876 (1949). Effects of Plasma Boundaries in Plasma Oscillations, *Phys. Rev.*, Vol. 79, pp. 992-1001 (1950).

<sup>2</sup> J. A. Roberts, "Wave Amplification by Interaction with a Stream of Electrons," *Phys. Rev.*, Vol. 76, pp. 340-344 (1949).

<sup>3</sup> V. A. Bailey, "The Growth of Circularly Polarized Waves in the Sun's Atmosphere and Their Escape into Space," *Phys. Rev.*, Vol. 78, pp. 428-443 (1950).

<sup>4</sup> "Traveling Wave Tubes," J. R. Pierce, Van Nostrand, 1950.

<sup>5</sup> A. V. Haeff, "Space-Charge Effects in Electron Beams," *Proc. I.R.E.*, Vol. 27, pp. 586-602 (1939).

<sup>6</sup> J. R. Pierce, "Limiting Stable Current in Electron Beams in the Presence of Ions," *Jour. App. Phys.*, Vol. 15, pp. 721-726 (1944); and "Note on Stability of Electron Flow in the Presence of Positive Ions," *Jour. App. Phys.*, Vol. 21, p. 1063, Oct. 1950.



to be approximations to those which would be obtained for more realistic but mathematically more refractory situations.

Other misinterpretations have arisen through combining non-relativistic equations of motion with Maxwell's equations and then attaching significance to terms of the order  $(v/c)^2$ .<sup>7</sup>

Finally, granting that all else is well, it is unsafe to draw conclusions from the examination of particular solutions of differential equations. In a very simple example, it is impossible to determine the gain of an amplifier tube which uses an electron stream simply by examining various "waves" which can travel on the stream. In solving a physical problem, one must not only solve the differential equation involved but he must satisfy the appropriate boundary conditions as well.

In all, such confusion as there has been concerning waves in clouds of electrons and ions seems to have arisen not through lack of mathematical ambitiousness but rather through simple errors in physical interpretation.

The following material concerns itself with some particular types of "waves" and with the importance and consequences of fitting boundary conditions. The work treats a very easy case, simplified and abstracted from a physically realizable system. The case was made so simple in order to avoid painful mathematics which might obscure the actual points to be made. The purpose is to explore this simple case thoroughly, avoiding basic misunderstandings. If it is objected that matters so simple should not be treated at such length, because no one could misunderstand them anyway, I can only reply that I did misunderstand some of the matters recounted herein.

### I. WHY ARE WAVES INTRODUCED?

We will consider the case of a narrow or thin beam of electrons across which we can assume that the electric field is constant.† In our calculations we assume that all electrons in a given very small region have the same velocity, thus neglecting the thermal velocity distribution.‡ We assume that the flow is a smoothed-out jelly of charge,‡ with the charge per unit mass characteristic of electrons; thus, we neglect individual interactions between electrons, and consider only a sort of average effect.

We will write the quantities involved in the following forms

$$\text{velocity} = v + u_0$$

<sup>7</sup> L. R. Walker, "Note on Wave Amplification by Interaction with a Stream of Electrons," *Phys. Rev.*, Vol. 76, pp. 1721-1722 (1949).

† This is in itself a drastic abstraction. No attempt will be made to justify it here, beyond saying that it is useful in considering the problems that follow.

‡ Other drastic approximations for which no justification will be given.

Here  $u_0$  is a constant component and  $v$  is a small fluctuating or a-c. component

$$\text{charge density} = \rho + \rho_0$$

where again  $\rho_0$  is the average or d-c. component, which will of course be negative, and  $\rho$  is the a-c. component

convection current density

$$i - I_0$$

Here  $I_0$  is the average or a c. current density and, as the electrons are assumed to move in the  $+z$  direction, the current density in the  $+z$  direction is taken as  $-I_0$ . In other work I have used  $i$  and  $I_0$  as current rather than as current density; I hope that this will cause no confusion.

It is assumed that there is no average field. It is assumed that there is an a-c. field in the  $z$  direction only, and this is called  $E$ .

We have two equations to work with. One is

$$\frac{d(v + u_0)}{dt} = -\frac{e}{m} E$$

Here  $e/m$ , the charge-to-mass ratio of the electron, is taken as a positive quantity. The time derivative is that moving with an electron. We can instead take derivatives at a fixed point

$$\frac{d(v + u_0)}{dt} = \frac{\partial(v + u_0)}{\partial t} + (v + u_0) \frac{\partial(v + u_0)}{\partial z}$$

which gives

$$\begin{aligned} & \frac{\partial v}{\partial t} + u_0 \frac{\partial v}{\partial z} \\ & + \frac{\partial u_0}{\partial t} + (v + u_0) \frac{\partial u_0}{\partial z} \qquad (1.1) \\ & + v \frac{\partial v}{\partial z} = -\frac{e}{m} E \end{aligned}$$

The terms on the second line are zero because  $\partial u_0 / \partial t = 0$ ,  $\partial u_0 / \partial z = 0$ . Further, let us consider a series of solutions of (1.1) for fields in which  $E$  has the same form in time and space, but varies in magnitude. As  $E$  is made smaller and smaller,  $v$  will become smaller and smaller, and the term  $v \partial v / \partial z$ , which is a product of two a-c. quantities, will become relatively smaller

than the other two terms involving  $v$ . In our small signal theory we neglect the term  $v\partial v/\partial z$ , and write

$$\frac{\partial v}{\partial t} + u_0 \frac{\partial v}{\partial z} = -\frac{e}{m} E \quad (1.2)$$

We note, then, that this approximates the true equation for small values of  $E$  and  $v$  only.

We have another equation

$$\frac{\partial}{\partial z} (i - I_0) = -\frac{\partial}{\partial t} (\rho + \rho_0) \quad (1.3)$$

This is the equation of continuity, or of conservation of charge. If we integrate it over a small distance  $\Delta z$  we obtain

$$(i - I_0)_{z+\Delta z} - (i - I_0)_z = -\frac{\partial}{\partial t} [(\rho + \rho_0)\Delta z]$$

The quantity  $(\rho + \rho_0)\Delta z$  is the charge per unit cross section in the distance  $\Delta z$ . Thus, the right-hand side is the rate at which charge in the distance  $\Delta z$  decreases. The quantity on the left is obviously the rate at which charge per unit cross section is flowing out of the space  $\Delta z$  long.

If we carry out the operations in (1.3) we obtain

$$\frac{\partial i}{\partial z} - \frac{\partial I_0}{\partial z} = -\frac{\partial \rho}{\partial t} - \frac{\partial \rho_0}{\partial t}$$

As  $\partial I_0/\partial z = 0$ ,  $\partial \rho_0/\partial t = 0$

$$\frac{\partial i}{\partial z} = -\frac{\partial \rho}{\partial t} \quad (1.4)$$

We need to add that the convection current is given by

$$\begin{aligned} i - I_0 &= (\rho + \rho_0)(v + u_0) \\ i - I_0 &= \rho v + \rho u_0 + \rho_0 u_0 + \rho v \end{aligned} \quad (1.5)$$

The term  $\rho_0 u_0$  is a constant term and is to be identified with  $-I_0$

$$-I_0 = \rho_0 u_0 \quad (1.6)$$

The term  $\rho v$  is a product of a-c. quantities. Suppose we solve all our equations neglecting  $\rho v$ . Then, the error caused by this approximation will be less as  $\rho$  and  $v$  are less, that is, at small signal levels. Thus, we write

$$i = \rho u_0 + \tau \rho_0 \quad (1.7)$$

We now have three approximate equations, which are good approximations at small signal levels

$$\frac{\partial v}{\partial t} + u_0 \frac{\partial v}{\partial z} = -\frac{e}{m} E \quad (1.2)$$

$$\frac{\partial i}{\partial z} = -\frac{\partial \rho}{\partial t} \quad (1.4)$$

$$i = \rho u_0 + v \rho_0 \quad (1.7)$$

We can eliminate  $\rho$  and  $v$  from these equations and obtain an equation relating  $i$  and  $E$ . To do this we solve (1.7) for  $v$

$$v = \frac{1}{\rho_0} i - \frac{u_0}{\rho_0} \rho$$

differentiate

$$\frac{\partial v}{\partial t} = \frac{1}{\rho_0} \frac{\partial i}{\partial t} - \frac{u_0}{\rho_0} \frac{\partial \rho}{\partial t}$$

use (1.4)

$$\frac{\partial v}{\partial t} = \frac{1}{\rho_0} \frac{\partial i}{\partial t} + \frac{u_0}{\rho_0} \frac{\partial i}{\partial z}$$

differentiate (1.2) with respect to  $t$ ,

$$\frac{\partial}{\partial t} \left( \frac{\partial v}{\partial t} \right) + u_0 \frac{\partial}{\partial z} \left( \frac{\partial v}{\partial t} \right) = -\frac{e}{m} \frac{\partial E}{\partial t}$$

and substitute for  $\partial v / \partial t$ , obtaining

$$\frac{\partial^2 i}{\partial t^2} + 2u_0 \frac{\partial^2 i}{\partial z \partial t} + u_0^2 \frac{\partial^2 i}{\partial z^2} = -\frac{e}{m} \rho_0 \frac{\partial E}{\partial t} \quad (1.8)$$

This is an equation relating  $i$  and its derivatives with  $E$ . It is a linear equation; that is,  $i$  and its derivatives, and  $E$  appear to the first power only. This is because we have neglected non-linear terms, saying that at low levels they are small compared with the linear terms.

Now, the electron flow interacts with surroundings of some sort, or, we shall say, with a circuit. Let us consider as an example of a circuit a transmission line with a distributed capacitance  $C$  per unit length and a distributed inductance  $L$  per unit length, which will transmit a slow wave. Suppose that the electron stream flows along very close to the line. Then

if the current  $\sigma i$  of the electron stream, where  $\sigma$  is the area of electron flow, changes with distance, a current  $J$  will flow into the line per unit length as shown in Fig. 1.1 where

$$J = -\sigma \frac{\partial i}{\partial z} \quad (1.9)$$

If  $V$  is the voltage on the line and  $I$  is the current in the line we write

$$\frac{\partial I}{\partial z} = -C \frac{\partial V}{\partial t} - \sigma \frac{\partial i}{\partial z} \quad (1.10)$$

$$\frac{\partial V}{\partial z} = -L \frac{\partial I}{\partial t} \quad (1.11)$$

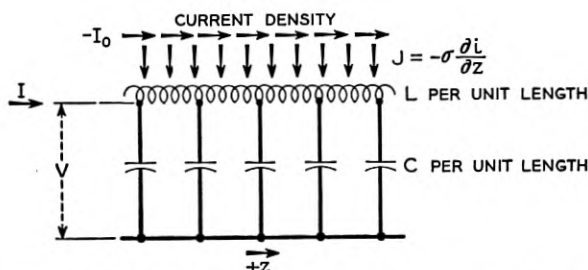


Fig. 1.1—A transmission line with an electron beam very close to it.

We can eliminate  $I$  by differentiating

$$\begin{aligned} \frac{\partial^2 I}{\partial z \partial t} &= -C \frac{\partial^2 V}{\partial t^2} - \sigma \frac{\partial^2 i}{\partial z \partial t} \\ \frac{\partial^2 V}{\partial z^2} &= -L \frac{\partial^2 I}{\partial z \partial t} \\ \sigma \frac{\partial^2 i}{\partial z \partial t} &= \frac{1}{L} \frac{\partial^2 V}{\partial z^2} - C \frac{\partial^2 V}{\partial t^2} \end{aligned} \quad (1.12)$$

We can further identify the field acting on the electrons as

$$E = -\frac{\partial V}{\partial z} \quad (1.13)$$

In (1.8), let us replace  $E$  by means of (1.13), and let us differentiate with respect to  $z$  and again with respect to  $t$ . We obtain

$$\frac{\partial^2}{\partial t^2} \left( \frac{\partial^2 i}{\partial z \partial t} \right) + 2u_0 \frac{\partial^2}{\partial z \partial t} \left( \frac{\partial^2 i}{\partial z \partial t} \right) + u_0^2 \frac{\partial^2}{\partial z^2} \left( \frac{\partial^2 i}{\partial z \partial t} \right) = \frac{e}{m} \rho_0 \frac{\partial^4 V}{\partial z^2 \partial t^2}$$

We can substitute for  $\partial^2 i / \partial z \partial t$  from (1.12) and obtain

$$u_0^2 \frac{\partial^4 V}{\partial z^4} + 2u_0 \frac{\partial^4 V}{\partial z^3 \partial t} + \left(1 - u_0^2 LC - \sigma \frac{e}{m} L \rho_0\right) \frac{\partial^4 V}{\partial z^2 \partial t^2} - 2u_0 LC \frac{\partial^4 V}{\partial z \partial t^3} - LC \frac{\partial^4 V}{\partial t^4} = 0 \quad (1.14)$$

Thus, we have obtained a linear partial differential equation in  $V$ ,  $z$  and  $t$ .

So far, nothing has been said about waves or wavelike behavior. We might solve (1.14) for any boundary conditions on  $V$  and its derivatives that we chose, by any means, as by using a differential analyzer or a digital computer. There is, however, a well-established technique for dealing with linear partial differential equations with constant coefficients, such as (1.14) is. It is known that they have solutions of the form

$$V = A e^{i\omega t} e^{-i\beta z} \quad (1.15)$$

As (1.14) is an entirely real equation, if (1.15) is a solution, the real part of (1.15) is also a solution, i.e.,

$$\text{Re} (A e^{i\omega t} e^{-i\beta z})$$

is a solution. Hence, we may regard the real part of the complex  $V$  as the true physical solution.

If we substitute (1.15) into (1.14) we obtain

$$u_0^2 \beta^4 - 2u_0 \omega \beta^3 + \left(\frac{1}{L} - u_0^2 LC - \sigma \frac{e}{m} L \beta\right) \omega^2 \beta^2 + 2u_0 LC \omega^3 \beta - LC \omega^4 = 0 \quad (1.16)$$

Now (1.16) is an algebraic equation in  $\omega$  and  $\beta$ . How are we to interpret it?

Suppose we are interested in devices driven from sinusoidal generators, such as amplifiers.\* This means that  $\omega$  is real, and that it is the radian frequency of the applied signal. We may then regard (1.16) as an equation in  $\beta$ , and, as it is a fourth degree equation, there will in general be four roots. We may regard these as pertaining to four waves, whose voltages vary as

$$V_1 = A_1 e^{j(\omega t - \beta_1 z)}$$

$$V_2 = A_2 e^{j(\omega t - \beta_2 z)}$$

$$V_3 = A_3 e^{j(\omega t - \beta_3 z)}$$

$$V_4 = A_4 e^{j(\omega t - \beta_4 z)}$$

\* We might, on the other hand, be interested in devices with an imposed spatial pattern, as in a magnetron oscillator. In this case we might assume  $\beta$  as a given, real quantity and solve for real or complex values of  $\omega$ .

Each of these four components is a solution of the *differential equation*. The solution of an actual *physical problem* will be the sum of the four components, or, if we like, the real part of that sum, and the amplitude factors  $A_1 - A_4$ , which are in general complex, will depend on the particular physical problem which is solved.

What has been the purpose of this argument? First of all, it is intended to indicate how the waves get into the picture. The differential equations for a long beam of constant average velocity  $u_0$  and charge density  $\rho_0$  were linearized by neglecting terms in which the products of a-c. quantities appeared. By this means a linear partial differential equation with constant coefficients which relates  $i$  and  $E$  was found. This was combined with the linear partial differential equation for a uniform transmission-line circuit, and an overall partial differential equation for  $V$  was obtained, linear and with constant coefficients. Such an equation could be solved by any means, but it is known to have wave-type solutions, and the solution of the original physical problem must be a sum of all such solutions.

In general, we will not expect so simple a relation between  $i$  and  $V$  or  $E$  as (1.12), that for a simple transmission line. Further, for broad electron streams the electronic behavior cannot be expressed so simply as it has been in (1.8). Nonetheless, we will find wave solutions in which all quantities vary with time and distance as

$$e^{j\omega t} e^{-j\beta z}$$

as long as

- (1) the d-c. beam properties (the undisturbed electron flow) and the circuit properties do not vary with  $z$ .
- (2) the signal amplitude is low enough so that terms involving products of a-c. quantities can be neglected.

When this is so, the solution of a physical problem can be expressed as the sum, or the real part of the sum, of such wave solutions, taken with the proper amplitudes.†

## II. THE COMPONENT WAVES

Once we are convinced that the solution of our problem can be expressed as the sum of a number of waves which are solutions of a linear partial differential equation, it is simplest to use this fact directly in finding certain properties of the waves of which the solution is to be made up.

Let us, for instance, let  $E$  in (1.8) contain the factor

$$e^{j\omega t} e^{-j\beta z}$$

† An additional overall condition is that the electron flow has no velocity distribution.

In other words, let  $E$  in (1.8) be one of the wave components of a solution. Then (1.8) becomes

$$(-\omega^2 + 2u_0\omega\beta - u_0^2\beta^2)i = -j\omega \frac{e}{m} \rho_0 E$$

$$i = \frac{-j\omega\epsilon \left( -\frac{e}{m} \frac{\rho_0}{\epsilon} \right)}{u_0^2 \left( \frac{\omega}{u_0} - \beta \right)^2} E \quad (2.1)$$

Here  $\epsilon$ , the dielectric constant of vacuum, has been introduced for reasons which will become apparent later. It is further of interest to introduce other simple parameters.

$$-\frac{e}{m} \frac{\rho_0}{\epsilon} = \omega_p^2 \quad (2.2)$$

$$\frac{\omega_p}{u_0} = \beta_p \quad (2.3)$$

$$\frac{\omega}{u_0} = \beta_0 \quad (2.4)$$

The quantity  $\omega_p$  is called the *plasma frequency* (a radian frequency).  $\omega_p^2$  is positive because  $\rho_0$  is negative.  $\beta_0$  would be the phase constant of a wave traveling with the electron velocity. While  $\beta_p$  would be the phase constant of a wave traveling with a phase velocity equal to the electron velocity, and having a frequency  $\omega_p$ , we may merely regard  $\beta_p$  as a convenient parameter which increases as the beam current is increased. In terms of  $\beta_p$  and  $\beta_0$

$$i = \frac{-\beta_p^2}{(\beta_0 - \beta)^2} (j\omega\epsilon E) \quad (2.5)$$

This may seem a strange form in which to write the equation. It will perhaps seem less strange, however, if we recall that the current density  $I$  in a dielectric medium is given by

$$I = j\omega\epsilon E$$

Thus, we see that for real values of  $\beta$  the electron convection current density  $i$  is that which would correspond to a negative dielectric constant or a negative capacitance. Its magnitude depends on  $\beta_p^2$ , which is proportional to the d-c. beam current density; and the magnitude becomes very large when the phase velocity of the wave approaches the velocity of the electrons, that is when  $\beta$  approaches  $\beta_0$ .



Suppose we consider a beam of area  $\sigma$ . We can write the total electron convection current  $I_e$  in the form

$$I_e = \sigma i = Y_e E \quad (2.6)$$

$$Y_e = \frac{-j\omega\epsilon\sigma\beta_p^2}{(\beta_0 - \beta)^2} \quad (2.7)$$

We will call  $Y_e$  the electronic admittance; it is measured in mho meters.

Later we will deal with waves in which the electron stream transfers power to the circuit, and it is interesting to see under what conditions this can take place. Let the amplitude of the wave under consideration vary with distance as

$$e^{(\alpha_1 - j\beta_1)z}$$

We may take the complex nature of the propagation constant into account by substituting in (2.7)

$$\begin{aligned} -j\beta &= \alpha_1 - j\beta_1 \\ \beta &= j\alpha_1 + \beta_1 \end{aligned} \quad (2.8)$$

This leads to

$$\begin{aligned} Y_e &= \frac{-j\omega\epsilon\sigma\beta_p^2}{(\beta_0 - \beta_1 - j\alpha_1)^2} \\ Y_e &= \frac{\omega\epsilon\sigma\beta_p^2[2\alpha_1(\beta_0 - \beta_1) - j(\beta_0 - \beta_1)^2]}{[(\beta_0 - \beta_1)^2 + \alpha_1^2]} \end{aligned} \quad (2.9)$$

The electron stream can transfer energy to the circuit only if the real part of  $Y_e$  is negative (a negative conductance). For a wave which increases in the direction of electron flow (the  $+z$  direction),  $\alpha_1$  is positive and the electronic conductance will be negative if  $\beta_1 > \beta_0$ ; that is, if the electron velocity is greater than the phase velocity of the wave.<sup>8</sup>

For a wave which decreases in the  $+z$  direction, the conductance will be negative if the electron velocity is smaller than the phase velocity of the wave.

Let us now consider the interaction of our thin electron stream with the circuit. Here there is some possibility of confusion. In (1.12) the field caused by impressing a current on a circuit was calculated. This may be likened to the voltage along an impedance  $Z$  caused by an impressed current  $I$ .

<sup>8</sup> This is indicated by very elementary arguments (J. R. Pierce and L. M. Field, "Traveling Wave Tubes," *Proc. I.R.E.*, Vol. 35, pp. 108-111, Feb. 1947). It is easy to forget, however, and was recently pointed out to me, to my consternation, by Dr. L. J. Chu.

Figure 2.1 will help to make this clear. Here the impressed current  $I$  flows to the right and back through the circuit of impedance  $Z$ . The voltage will increase to the right and hence the field will be directed to the left.

In general, for an impressed current  $I$  we will write the field produced as

$$E = -Z(\omega, \beta)I \quad (2.10)$$

Here  $Z(\omega, \beta)$  is a circuit impedance per unit length, which is usually a function of  $\omega$  and  $\beta$ . In terms of an admittance, the relation connecting impressed current and field is

$$I = -Y(\omega, \beta)E \quad (2.11)$$

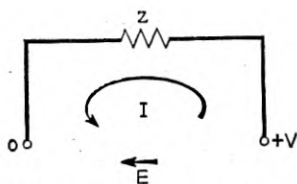


Fig. 2.1—The voltage and field produced by a current impressed on an impedance  $Z$ .

This can also be made clearer by means of an illustration. Suppose that the impressed current density in a very broad beam is  $i$  and the "circuit" is merely free space. Then

$$\rho = \frac{\beta}{\omega} i$$

and from Poisson's equation

$$\frac{\partial E}{\partial z} = -j\beta E = \frac{\rho}{\epsilon} = \frac{\beta i}{\omega \epsilon}$$

$$i = -j\omega\epsilon E$$

But, the admittance of a unit cube is just  $j\omega\epsilon$ , and the current through this admittance is  $j\omega\epsilon E$ .

Thus, when we have calculated the field caused by an impressed convection current, the admittance is the negative of the field divided by the convection current.

In (2.5),  $i$ , or rather,  $\sigma i$ , where  $\sigma$  is the area of the beam, may be regarded as the impressed current. If  $Y(\omega, \beta)$  is the circuit admittance, one way of writing the condition for a natural mode of propagation of stream and circuit is

$$\sigma i = -Y(\omega, \beta)E \quad (2.12)$$

Another way of putting this is to say

$$Y_e + Y(\omega, \beta) = 0 \quad (2.13)$$

From (2.13) and (2.7) we obtain

$$\beta = \beta_0 \pm \beta_p \sqrt{\frac{j\omega\epsilon\sigma}{Y(\omega, \beta)}} \quad (2.14)$$

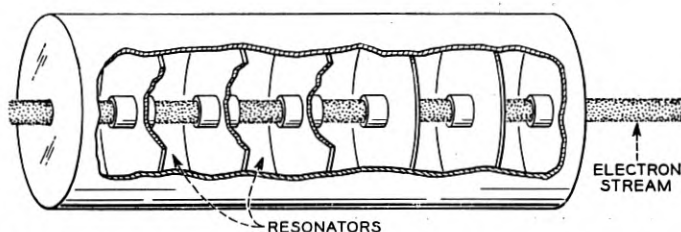


Fig. 2.2—An electron stream passing through a series of resonators, as in a multiresonator klystron.

Suppose, for instance, that the circuit admittance is capacitive and is equal to that for a longitudinal electric field in vacuum of area equal to the beam area  $\sigma$ . Then

$$Y(\omega, \beta) = j\omega\epsilon\sigma \text{ mho meter}$$

and we have two unattenuated waves

$$\beta = \beta_0 \pm \beta_p$$

We see that whenever (1) the circuit admittance is inductive or (2) the circuit admittance has a dissipative component,  $\beta$  will be complex, and there will be increasing and decreasing waves. Either of these conditions can be achieved, for instance, by surrounding the electron stream by a succession of essentially uncoupled resonators, tuned to be inductive, or with dissipation, as shown in Fig. 2.2. This is merely a continuous multi-resonator klystron.

In a transmission-line type of circuit such as we have considered and such as is used in the traveling-wave tube, for instance, the circuit admittance depends strongly on the phase constant  $\beta$ , and in solving (2.14) for  $\beta$  we must take cognizance of this fact.

We can, for instance, derive the circuit admittance from (1.12). We can use

$$E = j\beta V$$

and rewrite (1.12) as

$$\sigma i = \frac{-j}{\omega L \beta^2} (\omega^2 LC - \beta^2) E$$

Now, if the impressed current  $\sigma i$  is zero,  $\beta$  must have a value  $\beta_1$  such that

$$\beta_1 = +\sqrt{\omega^2 LC}$$

Also, the characteristic impedance of the line,  $K$ , is

$$K = +\sqrt{L/C}$$

In terms of these quantities

$$\sigma i = \frac{-j}{K \beta_1 \beta^2} (\beta_1^2 - \beta^2) E$$

and the circuit admittance  $Y(\omega, \beta)$  is

$$Y(\omega, \beta) = -\frac{\sigma i}{E} \tag{2.15}$$

$$Y(\omega, \beta) = \frac{j}{K \beta_1 \beta^2} (\beta_1^2 - \beta^2) \text{ mho meter}$$

Here  $K$  and  $\beta_1$  are positive quantities. We note that this admittance is capacitive for  $\beta < \beta_1$ , that is, for waves with a phase velocity greater than the natural phase velocity of the circuit, and inductive for  $\beta > \beta_1$ , that is for waves with a phase velocity less than the natural phase velocity of the circuit. This is easily explained. For small values of  $\beta$  the wavelength of the impressed current is long, so that current flows into and out of the circuit at widely separated points. Between such points the long section of series inductance has a higher impedance than the shunt capacitance to ground; the capacitive effect predominates and the circuit impedance is capacitive. However, for large values of  $\beta$  current flows into and out of the circuit at points close together. The short section of series inductance between such points provides a path of lower impedance than that through the capacitances and ground; the inductive impedance predominates and the circuit is inductive. Thus, for fast waves ( $\beta$  small) the circuit is capacitive and for slow waves ( $\beta$  large) the circuit is inductive.

We can, then, immediately make one observation. For a lossless circuit, any increasing or decreasing wave must have a phase velocity less than the natural phase velocity of the circuit.

We can make another observation as well; if the circuit has loss,  $Y(\omega, \beta)$  will have a real component, and from (2.14) all the waves must have an imaginary component of  $\beta$ , that is, they must be increasing or decreasing.

If we like, we can combine (2.15) with (2.14). Doing this directly, we obtain

$$(\beta - \beta_0)^2 = \omega \epsilon \sigma K \beta_1 \beta^2 \frac{\beta_p^2}{(\beta_1^2 - \beta^2)} \quad (2.16)$$

Unless the electron velocity is near the wave velocity ( $\beta_0$  near to  $\beta_1$ ) we will expect two sorts of solutions: one sort, for which  $\beta$  is near to  $\beta_0$  corresponding to "space-charge" waves; and the other, for which  $\beta$  is near to  $\pm\beta_1$ , corresponding to "circuit" waves. If  $\beta_0$  is not near to  $\beta_1$ , we can easily obtain approximate values of  $\beta$  for these two types of wave.

To obtain  $\beta$  for the space-charge waves we put  $\beta = \beta_0$  on the right-hand side of (2.16) and obtain

$$\beta = \beta_0 \pm \beta_p \sqrt{\frac{\omega \epsilon \sigma K \beta_1 \beta_0^2}{\beta_1^2 - \beta_0^2}} \quad (2.17)$$

If (2.17) gives a value of  $\beta$  differing by a small fraction from  $\beta_0$ , then (2.17) is to be trusted.

To obtain  $\beta$  for the forward circuit wave we put  $\beta = \beta_1$  on the left of (2.16) and in the numerator on the right. This gives for the forward wave

$$\beta = \beta_1 \left( 1 - \frac{\omega \epsilon \sigma K \beta_p^2 \beta_1}{(\beta_1 - \beta_0)^2} \right)^{1/2} \quad (2.18)$$

To obtain the backward wave, we put  $\beta = \beta_1$  on the left of (2.16) and in the numerator on the right, and obtain

$$\beta = -\beta_1 \left( 1 + \frac{\omega \epsilon \sigma K \beta_p^2 \beta}{(\beta_1 - \beta_0)^2} \right)^{1/2} \quad (2.19)$$

Again, (2.18) and (2.19) are to be trusted as long as  $\beta$  as given by (2.18) differs by a small fraction only from  $\beta_1$ .

We see that according to (2.19) the space-charge waves are unattenuated (real  $\beta$ ) for  $\beta_0 < \beta_1$ , that is, for electrons traveling faster than the circuit phase velocity, while there are increasing and decreasing waves for  $\beta_0 > \beta_1$ , that is, for electrons traveling more slowly than the circuit phase velocity. We see from (2.18) and (2.19) that the circuit waves are unattenuated (for lossless circuits), and travel a little more slowly than in the absence of electrons.

Further, we see that (2.17) and (2.18) are not to be trusted when  $\beta_0$  is close to  $\beta_1$ , that is, when the electron velocity is near to the circuit phase velocity. As a simple example, let

$$\beta_0 = \beta_1 \quad (2.20)$$

It will turn out that  $\beta$  will be very nearly equal to  $\beta_0$ . Hence,

$$\beta = \beta_0 + j\delta - j\beta = -j\beta_0 + \delta \quad (2.21)$$

Then, from (2.16) we have

$$j\delta = \pm\beta_p \sqrt{\frac{\omega\epsilon\sigma K\beta_0(\beta_0 + j\delta)}{\delta(-2j\beta_0 + \delta)}}$$

If we neglect  $\delta$  with respect to  $\beta_0$  in the sums inside the radical we obtain the equation

$$\delta^3 = -j\beta_p^2\beta_1\omega\epsilon\sigma K \quad (2.22)$$

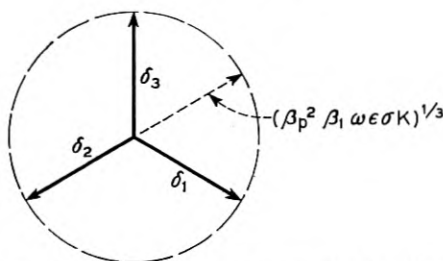


Fig. 2.3—Values of  $\delta$  for the three forward waves of a traveling-wave tube when the electron velocity is equal to the velocity of the undisturbed wave.

This yields the usual three forward waves of the traveling-wave tube.

$$\delta = (\beta^2\beta_1^2\omega\epsilon\sigma K)^{1/3}e^{j(-\pi/2+2n\pi)/3}$$

$$\delta_1 = (\beta^2\beta_1^2\omega\epsilon\sigma K)^{1/3}(\sqrt{3}/2 - j/2)$$

$$\delta_2 = (\beta^2\beta_1^2\omega\epsilon\sigma K)^{1/3}(-\sqrt{3}/2 - j/2)$$

$$\delta_3 = (\beta^2\beta_1^2\omega\epsilon\sigma K)^{1/3}(j)$$

We see that  $\delta_1$  represents an increasing wave slower than the natural phase velocity of the circuit,  $\delta_2$  represents a decreasing wave slower than the natural phase velocity of the circuit, and  $\delta_3$  represents an unattenuated wave faster than the natural phase velocity of the circuit. The 3  $\delta$ 's are illustrated in Fig. 2.3.

If  $\beta_0 \neq \beta_1$ , and if  $\beta_1$  is complex (a lossy circuit) the equation for  $\delta$  is more complicated, but  $\delta$  can be obtained numerically.

In addition to the three forward waves, that is, waves in the direction of electron motion, there is a backward wave. This is very much out of synchronism with the electron stream, and the backward wave is essentially the same as the wave in the absence of electron flow.

## III. FITTING BOUNDARY CONDITIONS; GAIN

So far the discussion has been concerned with a differential equation and wave-type solutions of it. Let us now consider an overall problem. Suppose that we inject an unmodulated electron stream into a circuit of some finite length and apply a signal to the end of the circuit nearest the source of electrons. Suppose that we adjust the output termination so that there is no backward wave.\* How will the field strength vary along the circuit? To answer this question, we must find out what combination in phase and amplitude of the three forward waves corresponds to these conditions. In terms of solving differential equations, we must fit the boundary conditions.

From Section I we have

$$\frac{\partial v}{\partial t} + u_0 \frac{\partial v}{\partial z} = -\frac{e}{m} E \quad (1.2)$$

or

$$v = \frac{j \frac{e}{m}}{u_0(\beta_0 - \beta)} E \quad (3.1)$$

with which we couple

$$i = \frac{-\beta_p^2}{(\beta_0 - \beta_p)^2} j \omega \epsilon E \quad (2.5)$$

In terms of

$$\beta = \beta_0 + j\delta$$

these relations become

$$\left( \frac{1}{u_0} \frac{e}{m} \right) v = \frac{1}{\delta} E \quad (3.2)$$

$$\left( \frac{-j}{\omega \epsilon \beta_p^2} \right) i = \frac{1}{\delta^2} E \quad (3.3)$$

These relations hold for each of the waves separately. Now, let us denote by  $E_1, E_2, E_3$  the fields of the three waves, and by  $E$  the actual field on the circuit. Then at the beginning of the circuit, where  $E$  is  $E_0$ , the applied field, the amplitudes  $E_{10}, E_{20}, E_{30}$  of  $E_1, E_2$  and  $E_3$  must satisfy

$$E_{10} + E_{20} + E_{30} = E_0 \quad (3.4)$$

\* This is a very special case, requiring a unique impedance terminating the  $+z$  end of the output circuit. See Section V.

Also, at the beginning of the helix, a-c. velocity and the a-c. convection current must be zero. This means that

$$\frac{E_{10}}{\delta_1} + \frac{E_{20}}{\delta_2} + \frac{E_{30}}{\delta_3} = 0 \quad (3.5)$$

$$\frac{E_{10}}{\delta_1^2} + \frac{E_{20}}{\delta_2^2} + \frac{E_{30}}{\delta_3^2} = 0 \quad (3.6)$$

For the case we have considered,  $\beta_1 = \beta_0$ ,  $\beta_1$  real,

$$\delta_2 = \delta_1 e^{-j(2\pi/3)}$$

$$\delta_3 = \delta_1 e^{+j(2\pi/3)}$$

and our equations become

$$E_{10} + E_{20} + E_{30} = E_0$$

$$E_{10} + E_{20} e^{j\frac{1}{3}(2\pi)} + E_{30} e^{-j\frac{1}{3}(2\pi)} = 0$$

$$E_{10} + E_{20} e^{j\frac{2}{3}(2\pi)} + E_{30} e^{-j\frac{2}{3}(2\pi)} = 0$$

We easily see that the solution is

$$E_{10} = E_{20} = E_{30} = \frac{1}{3} E_0 \quad (3.7)$$

If  $E$  is the field at a distance  $z$  along the helix

$$E = \frac{1}{3} E_0 e^{-j\beta_0 z} (e^{\delta_1 z} + e^{\delta_2 z} + e^{\delta_3 z}) \quad (3.8)$$

In Fig. 3.1,

$$20 \log_{10} \left| \frac{E}{E_0} \right|$$

is plotted vs  $CN$ , a factor proportional to distance.

We see that initially the amplitude does not change. This is necessarily so. The strength of the field can grow only through the electron stream giving energy to the circuit. The electron stream can give energy to the circuit only if it has an a-c. convection current. Initially the electron stream is unmodulated and hence it can give energy to the circuit only after it has traveled far enough to become modulated.

In the case we have considered, the amplitudes of the three wave components of the field are initially equal. Now,  $E_1$  increases with distance, while  $E_2$  decreases with distance and  $E_3$  is unattenuated. Hence, if the tube is long enough,  $E_2$  and  $E_3$  will be negligible near the output of the tube; and the field at the output, a distance  $\ell$  from the input, will be very nearly

$$E = \frac{1}{3} E_0 e^{-j\beta_0 \ell} e^{\delta_1 \ell}$$



Under these circumstances the gain  $G$  in db will be very nearly

$$G = 20 \log_{10} \left| \frac{E}{E_0} \right| = 20 \log_{10} \frac{1}{3} e^{\text{Re}(\delta \ell)} \text{ db}$$

$$G = -9.54 + \frac{20}{2.3} \frac{\sqrt{3}}{2} (\beta^2 \beta_1^2 \omega \epsilon \sigma K)^{1/3} \ell$$

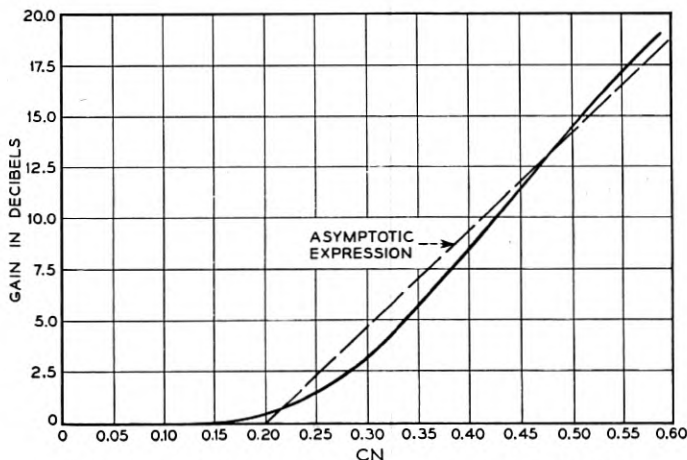


Fig. 3.1—Signal level along the helix of a traveling-wave tube.

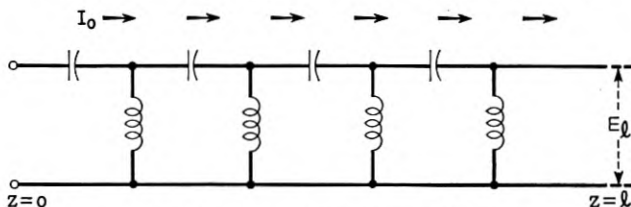


Fig. 4.1—A high-pass structure in which the phase velocity is in a direction opposite to that of power flow.

#### IV. BACKWARD WAVES AND OTHER PECULIAR WAVES

It is important to notice that, for the usual traveling-wave tube, it is possible to express the overall gain in terms of the increasing wave alone only because of the relative amplitudes of the three waves which make up the solution of the particular problem considered. That this is by no means a trivial point can be demonstrated by considering a case in which the circuit is a high-pass filter, as shown in Fig. 4.1. For such a circuit, the phase constant  $\beta_1$  is negative for a wave excited at the left end of the line which carries energy to the right. Such a wave will not interact with elec-

trons moving to the right. A wave excited at the right with power flow to the left has a positive value of  $\beta$  and will interact with electrons traveling to the right. Let us consider such an interaction.

First, as to the  $\delta$ 's. We see that, for a wave which varies with distance as

$$e^{-i\beta z}$$

where  $\beta$  is a positive number and has power flow to the left, the sign of  $V/I$  must be the opposite of what it would be if the power flowed to the right. This can be taken into account by reversing the sign of  $K$  in (2.19), and making

$$\delta^3 = +j\beta_p^2\beta_1\omega\epsilon\sigma K \quad (4.1)$$

where  $K$  is now taken as a positive number. We can then take

$$\delta_1 = (\beta_p^2\beta_1^2\omega\epsilon\sigma K)^{1/3} \left( -\frac{\sqrt{3}}{2} + j/2 \right)$$

$$\delta_2 = (\beta_p^2\beta_1^2\omega\epsilon\sigma K)^{1/3} \left( \frac{\sqrt{3}}{2} + j/2 \right)$$

$$\delta_3 = (\beta_p^2\beta_1^2\omega\epsilon\sigma K)^{1/3}(-j)$$

Here  $\delta_1$  represents a wave whose amplitude increases to the left, that is, a wave which grows in the direction of energy flow. We might think that this would immediately imply a gain similar to that obtained for energy flow in the direction of electron motion, but this would be jumping at conclusions.

Suppose we taken  $z = 0$  at the left-hand or output end of the circuit. There the electron stream enters unmodulated. There also we will assume the circuit to be terminated so as to prevent reflection of power. At the right-hand or input end of the circuit power will be fed in, giving an impressed field  $E_t$ .

Suppose  $\delta_1$ ,  $\delta_2$  and  $\delta_3$  are the appropriate  $\delta$ 's for this case. We see that our boundary conditions are

$$e^{-j\beta_0\ell}(E_{10}e^{j\delta_1\ell} + E_{20}e^{j\delta_2\ell} + E_{30}e^{-j\delta_3\ell}) = E_t$$

$$\frac{E_{10}}{\delta_1} + \frac{E_{20}}{\delta_2} + \frac{E_{30}}{\delta_3} = 0$$

$$\frac{E_{10}}{\delta_1^2} + \frac{E_{20}}{\delta_2^2} + \frac{E_{30}}{\delta_3^2} = 0$$

We have a relation between the  $\delta$ 's

$$\begin{aligned}\delta_1 &= \frac{\sqrt{3}}{2} + j\frac{1}{2} \\ \delta_2 &= \delta_1 e^{j(2\pi/3)} \\ \delta_3 &= \delta_1 e^{-j(2\pi/3)}\end{aligned}$$

From this we easily see that a solution of the last two equations is

$$E_{10} = E_{20} = E_{30}$$

Accordingly, the first equation becomes

$$\begin{aligned}E_{10} e^{-j\beta_0 \ell} (e^{j\delta_1 \ell} + e^{j\delta_2 \ell} + e^{j\delta_3 \ell}) &= E\ell \\ E_{10} &= \frac{E\ell e^{j\beta_0 \ell}}{(e^{j\delta_1 \ell} + e^{j\delta_2 \ell} + e^{j\delta_3 \ell})}\end{aligned}\quad (4.2)$$

Let us now assume that the tube is very long. We easily see that in this case

$$\begin{aligned}|e^{j\delta_1 \ell}| &\gg |e^{j\delta_2 \ell}| \\ |e^{j\delta_1 \ell}| &\gg |e^{j\delta_3 \ell}|\end{aligned}$$

So very nearly

$$E_{10} = E_{20} = E_{30} = \frac{E\ell e^{j\beta_0 \ell}}{e^{j\delta_1 \ell}}\quad (4.2)$$

and the total field at the output end of the tube is

$$E = E_{10} + E_{20} + E_{30} = 3E\ell e^{j\beta_0 \ell} e^{-j\delta_1 \ell}\quad (4.4)$$

This, however, is much smaller than the field  $E\ell$  at the input end of the tube.

What is the physical picture? The electrons are injected into the circuit as an unmodulated stream. In order to fit the boundary conditions at this point, the three waves must have comparable magnitudes at the point of injection. If this is the output, then any wave which "grows" from input toward output must be relatively very small at the input.

If boundary conditions are fitted for other cases, as, for an electron speed not equal to the circuit phase velocity ( $\beta_0 \neq \beta_1$ ), it may be found that the output may be a little greater than the input under some circumstances; this represents a small gain achieved through a spatial interference of the three wave components.

A sure way of distinguishing conditions which will allow amplification from conditions which will not is through a solution of the differential equations together with a fitting of the boundary conditions. In the case of backward waves there are, however, considerations concerning the source of energy and its transfer to the circuit (or field) which are useful.

Suppose that an unmodulated electron stream enters a microwave amplifier, travels for some distance through it, and emerges. If the electromagnetic output power of the amplifier is greater than the input power, the additional power must have come from the kinetic energy of the electron stream. The average electron must leave the amplifier with less energy of motion than it had on entering it.

We can say a little more. Let us call the total velocity of electrons, a-c. and d-c.,  $u$ . Then we have, corresponding to (1.1)

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial z} = -\frac{e}{m} E \quad (4.5)$$

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial z} (u^2) = -\frac{e}{m} E$$

We will consider an amplifier in which the  $u$  and  $E$  at any  $z$ -position are truly periodic. Let us integrate over the period of a cycle,  $\tau$ , and divide by  $\tau$

$$\frac{1}{\tau} \Big|_t^{t+\tau} u + \frac{1}{\tau} \frac{\partial}{\partial z} \int_t^{t+\tau} u^2 dt = \frac{1}{\tau} \int_t^{t+\tau} E dt \quad (4.6)$$

As  $u$  will be the same at  $t$  and  $t + \tau$ , the first term on the left is zero, and we have

$$\frac{\partial}{\partial z} \overline{u^2} = \overline{E} \quad (4.7)$$

Here  $\overline{u^2}$  and  $\overline{E}$  are time averages.

The field  $E$  is produced in a linear circuit by (1) the application of an a-c. signal, (2) by the presence of the electron stream. Certainly, the applied signal can produce no average field in a linear circuit. Further, unless electrons are turned back, the average electron convection current is independent of r-f level. In a linear circuit the average field must be proportional to the average impressed current, so the average field  $\overline{E}$  must be zero or independent of r-f level. Thus, the *time* average of  $\overline{u^2}$  at a given point must be independent of r-f level.\*

This means that the electron stream cannot be slowed down bodily by

\* L. A. MacColl pointed this out to the writer.

the r-f field. Energy is extracted from the stream only by a bunching process in which in the emerging beam the charge density is higher when the velocity is below average than it is when the velocity is above average. In other words, the kinetic energy averaged over electrons is reduced, even though the *time* average of  $u^2$  is not changed. This means that the emerging beam must be strongly bunched if much power is to be abstracted.

In the conventional traveling-wave tube all is well. At the input the r-f field is small and the beam is unbunched. At the output the r-f field is high, and the beam is strongly bunched, having lost energy to the circuit.

Imagine a tube using a backward wave, however. The electrons are injected unbunched at the output, where the signal level is high. They emerge at the input where the signal level is low. If the tube is to give high power, the stream must emerge strongly bunched. The disturbance in the electron stream cannot gradually increase as the field amplitude increases.

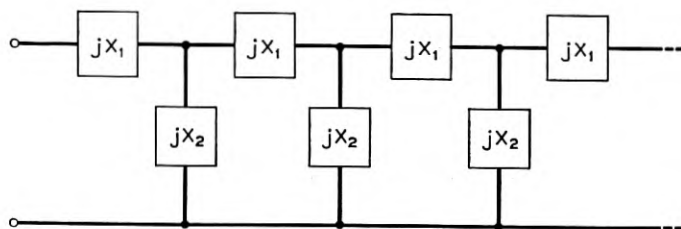


Fig. 4.2—A ladder network.

We have seen that one cannot draw conclusions about gain just by looking at the propagation constants of the waves. Waves are merely solutions of a differential equation connected with a physical system. To find the properties of the system one must examine, not various solutions of the differential equation, but the particular solution (which may be a combination of simple solutions) which applies to the system in question.

As a further example, we will examine another system whose differential equations yield "growing" solutions which turn out to be backward waves. Consider the ladder network of Fig. 4.2. This propagates an unattenuated wave if  $X_1$  and  $X_2$  have opposite signs, ( $X_1$  inductive and  $X_2$  capacitive, for instance). If, however,  $X_1$  and  $X_2$  are both capacitive or both inductive, then a wave excited in the circuit decays exponentially with distance. If we speak in terms of  $\beta_1$ , then

$$\begin{aligned} -j\beta_1 &= -\alpha_1 \\ \beta_1 &= -j\alpha_1 \end{aligned}$$

where  $\alpha_1$  is a real number.

The characteristic impedance  $K$  is reactive, inductive if both impedances are inductive and capacitive if both impedances are capacitive. Let

$$K = jX_0$$

where  $X_0$  is a real number. Then a positive value of  $X_0$  means inductive elements, and a negative value capacitive elements.

The space-charge waves are given by (2.17)

$$\beta = \beta_0 \pm \beta_p \sqrt{\frac{-\omega\epsilon\sigma X_0 \alpha_1 \beta_0^2}{\alpha_1^2 + \beta_0^2}} \quad (4.8)$$

We see that the waves are unattenuated for negative, capacitive values of  $X_0$ , and are increasing and decreasing for positive, inductive values of  $X_0$ . It can be shown that the increasing space-charge waves can be used to obtain gain.

The forward circuit waves are given by using  $K = jX_0$ ,  $\beta_1 = -j\alpha_1$  in (2.16),  $\beta = -j\alpha_1$  on the left and in the numerator on the right and  $\beta = -j\alpha_1$  in the denominator on the right.

$$\alpha = \alpha_1 \left( 1 + \frac{\omega\epsilon\sigma X_0 \alpha_1 \beta_p^2}{(j\alpha_1 + \beta_0)^2} \right)^{1/2} \quad (4.9)$$

As  $\alpha = j\beta$ , the variation with distance is as

$$e^{-\alpha z}$$

The backward wave is given by using  $\beta = +j\alpha_1$  on the left of (2.16) and in the numerator on the right

$$\alpha = -\alpha_1 \left( 1 - \frac{\omega\epsilon\sigma X_0 \alpha_1 \beta_p^2}{(j\alpha_1 - \beta_0)^2} \right)^{1/2} \quad (4.9)$$

If  $\alpha$  differs little from  $\pm\alpha_1$ , we can expand the square root in (4.6) and (4.7), separate real and imaginary parts, and write:

Forward wave:

$$\alpha = \alpha_1 \left( 1 + \frac{\omega\epsilon\sigma\beta_p^2 X_0 \alpha_1 (\beta_0^2 - \alpha_1^2)}{2(\beta_0^2 + \alpha_1^2)} - \frac{j\omega\epsilon\sigma\beta_p^2 \beta_0 X_0 \alpha_1^2}{(\beta_0^2 + \alpha_1^2)^2} \right) \quad (4.10)$$

Backward wave:

$$\alpha = -\alpha_1 \left( 1 + \frac{\omega\epsilon\sigma\beta_p^2 X_0 \alpha_1 (\beta_0^2 - \alpha_1^2)}{2(\beta_0^2 + \alpha_1^2)} + \frac{j\omega\epsilon\sigma\beta_p^2 X_0 \alpha_1^2}{(\beta_0^2 + \alpha_1^2)^2} \right) \quad (4.11)$$

The circuit "waves" which were rapidly attenuated in the absence of electrons ( $\beta_p = 0$ ) are a little more or less rapidly attenuated in the presence of electrons (more or less depending on whether  $X_0$  is positive or negative, and on the relative magnitudes of  $\beta_0$  and  $\alpha_1$ ), and they now have a phase constant, that is, an imaginary component of the propagation constant.

The phase velocity may be either positive or negative, depending on the sign of  $X_0$ . This added feature gives the solution a more "wavelike" quality, but physically we have merely a slight perturbation of the disturbance natural to the non-propagating ladder network.

In the absence of electrons, there is no real power flow in the modes of propagation of a purely reactive ladder network in which the shunt and series reactances have the same sign. Such a network can of course transmit power to a resistive load, but it transmits no power when terminated in its (reactive) characteristic impedance.

In the presence of electrons, there is a small power flow in the circuit. We can easily evaluate this. If, in (1.11), we assume a variation of the quantities with time and distance as

$$e^{-\alpha z} e^{j\omega t}$$

we obtain

$$I = \frac{\alpha}{j\omega L} V$$

Here  $\omega L$  stands for the series reactance, which we may call  $X_1$

$$\omega L = X_1$$

A positive value of  $X_1$  means series inductance. For non-propagating ladders,  $X_1$  and the characteristic reactance  $X_0$  have the same sign.

We then have

$$I = -j \frac{\alpha}{X_1} V$$

The quantity  $-j\alpha/X_1$  as evaluated in the presence of electrons will be the "hot" characteristic admittance.

The complex power flow  $P$  is

$$P = VI^*$$

So, in this case

$$P = \frac{j\alpha^*}{X_1} VV^*$$

Now, the "backward" wave, for which  $\alpha$  is given by (4.12), "increases" in the direction of electron flow. For it, the real part of the power  $\text{Re } P$  is given by

$$\text{Re } P/VV^* = - \frac{\omega \epsilon \sigma \beta_p^2 \alpha_1^3 X_0}{(\beta_0^2 + \alpha_1^2) X_1} \quad (4.12)$$

Note that  $X_0$  and  $X_1$  must have the same sign. Thus, the power flow for the wave which "increases" in the direction of electron flow is always in the direction opposite to the electron flow. The circuit power does not flow in the direction of increasing amplitude for the wave which "grows" in the

direction of electron flow. We might have deduced this from the fact that the phase velocity for the wave is greater than the electron velocity (see (2.9)).

While the wave which increases in the direction contrary to electron flow has its power flow in the direction of increasing amplitude, it is a backward wave and hence not suitable for producing gain.

The disturbance on the non-propagating ladder is closely related to a passive or cut-off mode of a waveguide excited at a frequency less than the cutoff frequency for the mode in question. In this case, the analogue of the circuit power  $VI^*$  is the integral of the Poynting vector over the guide cross section. When electrons flow through a waveguide these cut-off modes are perturbed much as indicated by (4.19) and (4.12). Because the perturbed modes have a "wavelike" character in that the propagation constant is no longer purely real, and because the amplitude may increase in the direction of electromagnetic power flow, some workers have proposed to obtain gain from these "growing waves."<sup>2</sup>

#### V. FURTHER CONSIDERATIONS CONCERNING BOUNDARY CONDITIONS

How necessary is it to fit boundary conditions in order to deduce what will happen? The suspect waves we have examined so far might be rejected as increasing in a direction contrary to the direction of electron flow,\* or as having electromagnetic power flow in a direction opposite to the direction of growth. Can we find some method for separating waves useful in producing gain from waves which are not, without consideration, explicit or implicit, of boundary conditions?

Let us consider the problem of fitting boundary conditions for a circuit plus an electron stream. Imagine that the end of the circuit near to the electron source ("near" end) is connected to a load impedance  $Z$  and that the end away from the electron source ("far" end) is driven by a voltage  $V$ . Let the wave which increases most rapidly in the direction of electron flow vary in amplitude as  $\exp(\alpha z)$ . Suppose that the length of the circuit is great, so that  $\alpha L$  is a large number and  $\exp(\alpha L)$  is a very large number.

At the near end the various wave components must be so related that  $i$  and  $v$  are zero and that the circuit voltage is  $ZI$ . At the near end, all four waves must be used to fit the boundary conditions at a given voltage level. Disallowing very special values of  $Z$ , we would expect that at the near end the four waves will have comparable amplitudes (the amplitudes are related by linear simultaneous equations). Thus, at the far end of the circuit, the wave which increases most rapidly with distance should strongly predominate. It seems that the most rapidly increasing wave is naturally connected with excitation of the circuit at the far end.

\* Though rejected only through considering boundary conditions.



On the other hand, assume that the far end of the circuit is terminated in some impedance  $Z$ . Consider the case in which the electron stream is velocity modulated at the source end and no exciting voltage is applied at the far end. We would expect that the required boundary conditions at the source end could be satisfied by using the waves excepting the one which increases most rapidly with distance. At the far end, to make  $V = IZ$  it is necessary to add a component of the wave which increases most rapidly with distance, a component of magnitude comparable to the sum of other components present *at the far end*. However, this added component is so small at the near end that there it can be disregarded. Thus, the manifestation of large forward gain comes not from the mere presence of a wave which increases in the forward direction, but from special properties of the waves and/or the terminating impedances which can be determined with certainty only by fitting boundary conditions.

Are not these arguments at variance with the usual analyses of operation of the traveling-wave tube? Suppose, for instance, that the helix is terminated in an arbitrary impedance at the input (near) end and that a voltage  $V$  is applied at the output (far) end. What wave will predominate? For a lossless helix, the true answer is that the increasing (forward) wave, not the unattenuated backward wave, will predominate. This can be avoided only by (1) choosing a particular (matched) value of source impedance or (2) making the helix lossy enough so that the backward wave "increases" more rapidly in the  $+z$  direction than any forward wave does. In tubes with a uniform loss along the helix, expedient (2) is adopted; when a center lossy section is used, both (1) and (2) are invoked, (1) in the output section and (2) in the center lossy section.

It is dangerous to consider the solutions of the linear differential equations of a physical system singly rather than in the combination which satisfies the boundary conditions. This sort of reasoning might lead one to believe that the problem of obtaining high voltages can be solved by finding a solution of Laplace's equation (say  $V = 1/r$ ) for which the potential goes to infinity at some point.

Cautions against neglecting the problem of boundary conditions apply equally well to problems of instability (increase of disturbances with time) as to problems of amplification. Thus, electron flow may be unstable when none of the waves grows with time for real values of  $\beta^6$ . On the other hand, in criticizing the work of Bohm and Gross,<sup>1</sup> R. Q. Twiss has shown<sup>9</sup> that electron flow is not necessarily unstable merely because some of the waves grow exponentially with time for real values of  $\beta$ .

<sup>9</sup> R. Q. Twiss, "On the Theory of Plasma Oscillations" Services Electronics Research Laboratory, Extracts from Quarterly Report No. 20, Oct. 1950, pp. 14-28.

## Interaxial Spacing and Dielectric Constant of Pairs in Multipaired Cables

By J. T. MAUPIN

(Manuscript Received April 24, 1951)

A major handicap in the evaluation of different designs and manufacturing processes, in respect to efficiency of space utilization inside the sheath of multipair telephone cable, has been the lack of a simple and accurate method of measuring the dielectric constant of such cable pairs. This paper describes a simple non-destructive method of determining both the interaxial spacing between conductors of a cable pair and the dielectric constant. An important by-product of the work is the demonstration of the fact that  $\epsilon = L \times C$  is not a valid means of determining the dielectric constant of cable pairs.

### INTRODUCTION

A cable pair consists of two individually-insulated conductors, of nominally equal circular cross-section, which have been twisted together in a long helix and stranded into a cable core with similar pairs. It has not been possible to analyze rigorously the electrical characteristics of such a circuit in terms of its rather complex physical configuration. For this reason, methods largely of an empirical nature have been used in the past to correlate physical and electrical characteristics of multipaired cables.

The capacitance of any system of conductors immersed in a homogeneous medium is directly proportional to the dielectric constant of the medium. The dielectric of a cable pair is not homogeneous, but it can be described in terms of a homogeneous dielectric which would produce the same capacitance. In addition to the dielectric properties of the insulating medium, the capacitance of a cable pair is determined by the disposition of the paired conductors with respect to each other and with respect to the surrounding pairs or sheath.

In particular, the interaxial separation between the wires of a pair has a critical effect on capacitance. The interaxial separation is determined by the ability of the insulation to resist deformation due to compressive forces encountered in cabling operations. Thus, the capacitance of a cable pair is largely dependent on the mechanical and dielectric properties of the conductor insulation, and a criterion of the relative efficiency of an insulation is the capacitance level, for a particular conductor gauge, resulting from a given cable space allowance or space-per-pair.

Experience with paper ribbon and paper pulp insulated cables has shown that occasional wide deviations in capacitance can occur even though the space-per-pair allowance is substantially constant. The aforementioned em-

pirical methods of capacitance-space analysis do not provide any insight as to whether these deviations are due to anomalies in the mechanical properties or in the dielectric properties of the insulation, or both.

With respect to some electrical and physical characteristics, it is evident that there is a close analogy between the cable pair and an "ideal" balanced shielded pair consisting of two straight and parallel solid cylindrical conductors enclosed in a cylindrical conducting shield, with the center line of the pair coinciding with the axis of the shield. A cross-section of such a circuit is shown in Fig. 1. The conductors are insulated from one another and from the shield by a homogeneous dielectric.

Rigorous mathematical expressions for the capacitance and inductance of the ideal pair in terms of its dimensions and dielectric constant have been

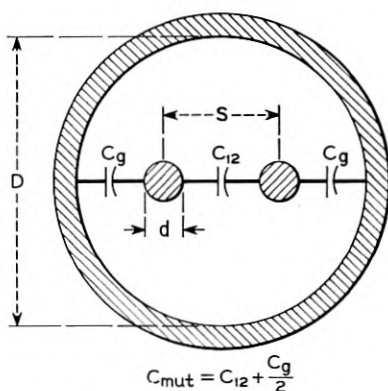


Fig. 1—A balanced, shielded pair.

derived by the Mathematical Research Group of the Laboratories and others. Measured values of the capacitance and inductance of a given cable pair, when substituted into these expressions, determine a set of dimensions and a dielectric constant value which describe an ideal pair having the same capacitance and inductance as the cable pair. The extent to which these idealized values represent actual cable conditions depends on the accuracy of the assumed equivalence of the two structures.

The purpose of this paper is to describe a few simple but direct experiments, the results of which demonstrate that these idealized parameters are closely representative of actual cable conditions in so far as interaxial spacing and dielectric constant are concerned. Thus emerges a simple technique, based on easy-to-make low-frequency measurements, for quantitative evaluation of these two important cable pair parameters. Applications of the

method to commercial designs of multipaired cable are included in what follows.

#### LOW FREQUENCY INDUCTANCE AND INTERAXIAL SPACING

The self-inductance of a pair of straight parallel wires in free space or within any non-magnetic shield is a function only of the ratio of conductor diameter to the interaxial spacing if the frequency is so low that proximity, shielding, and skin effects are negligible. The formula is as follows:

$$L = 1.482 \log \frac{2S}{d} + 0.1609 \text{ (mh/mi)} \quad (1)$$

$S$  = interaxial spacing

$d$  = conductor diameter

The first term gives the self inductance due to net external flux-linkage and the 0.1609 constant represents flux-linkage within the non-magnetic conductors.

Some textbooks give the formula with  $(2S-d)/d$  as the argument of the logarithm. This form is not valid when  $d$  is not small compared to  $S$ , as is the case with cable pairs. The derivation of formula (1) in Russell's "Alternating Currents" shows that it is valid for any  $S$ , provided only that the current is uniformly distributed across the conductor cross-section.

While it was believed that formula (1) was valid for cable pairs at, say, 1000 cps, it was desirable to establish experimentally whether or not the twist in the pair, and magnetic coupling between pairs, affected the measured inductance at a frequency in this range. The effect of coupling between pairs is greatest when all of the cable pairs except the pair under test are shorted together at both ends of the cable, thus providing a large number of closed loops for any induced currents. In making laboratory inductance tests at 1000 cps on lengths of 19-gauge cable with 0.084  $\mu\text{f}/\text{mi}$  capacitance (Type CNB) it was found that opening or shorting the surrounding pairs did not affect the measured inductance. Also, grounding or floating the far end of the test pair made no difference. The tests were repeated with the same results on lengths ranging from 300' to 5000'.

It should be noted that a correction term must be added to the measured 1000 cps inductance ( $L'$ ) to obtain the true distributed inductance ( $L$ ) when the cable length is such that propagation effects become appreciable. When the correction term is included, the equation for  $L$  at 1000 cps becomes:

$$L = L' + 1/3(R')^2C'$$

Where  $L'$ ,  $R'$ , and  $C'$  are measured inductance, resistance, and mutual capacitance, respectively.

Expressed as a percentage of  $L$  per length, the correction term varies as the square of length. It is only 0.5% for 750' of CNB but is 20% for 5000' of the same type of cable. The single correction term given is not sufficiently accurate for lengths of CNB greater than about 5000'. The comparable limiting length for smaller gauges would be less; it is about 1000' for 26-gauge pairs.

It is also necessary to allow for stranding takeup effects when converting from per length to per mile values. Stranding takeup is defined as the increment in length of a given pair compared to the cable length, due to the helix introduced in the pair during the cabling operation. It depends upon the size of cable and stranding lay and is negligible for cables of 50 small gauge

TABLE I  
COMPARISON OF MEASURED 1000 CPS INDUCTANCE WITH THAT CALCULATED FROM THE THEORETICAL FORMULA

Spacing ( $S$ ) in mils	$d/2S$	Calculated $L$	Measured $L$	% Difference
		mh/mi	mh/mi	
13 Gauge Wire				
75.6	0.474	0.641	0.641	0.00
106.2	0.337	0.860	0.856	-0.46
145.1	0.247	1.060	1.063	+0.28
19 Gauge Wire				
38.8	0.464	0.656	0.656	0.00
69.0	0.261	1.027	1.021	-0.60
Theoretical Formula				
$L = 1.482 \log 2S/d + 0.1609$ (mh/mi)				

pairs or less but causes an increase of 2.7% in pair length over cable length for pairs in the outside layer of a full size CNB cable ( $2\frac{3}{8}$ " diameter over the paper wrapped core).

Some short lengths of pair with wires parallel and approximately straight and with accurately known dimensions were made up in the laboratory. Inductance measurements at 1000 cps on these pairs agree very closely with inductances calculated from formula (1), for  $d/2S$  ratios from about 0.25, which is close to the nominal value for cable pairs, up to nearly 0.50, which is the highest possible value. These results are shown in Table I. The 19-gauge pair with a  $d/2S$  ratio of 0.261 was twisted fixed carriage style under constant tension. There was no measurable change in inductance for twist lengths as short as 2", compared with the straight, parallel condition. Measurement accuracy was about  $\pm 0.25\%$ .

The foregoing results lead to the conclusion that the effective  $d/2S$  ratio along the length of a cable pair can be determined from formula (1) using 1000 cps inductance, with an accuracy equal to that of the inductance tests. Since  $d$  is generally known or can be found with greater precision than the inductance,  $S$  can likewise be found with an accuracy dependent on the precision of the inductance tests. This conclusion applies regardless of the type of conductor insulation. However, it should be noted that formula (1) would not be accurate if there happened to be magnetic materials in close proximity with the pair under test, such as, for instance, pairs surrounding a core of steel tape bound coaxials.

### THE DIELECTRIC CONSTANT

The dielectric of a cable pair consists of a non-homogeneous mixture of solid insulating material and air. The dielectric constant of a cable pair can be defined as the ratio of the actual mutual capacitance of the pair to the mutual capacitance which would result if the solid insulating material were removed leaving a 100% air dielectric in the otherwise undisturbed cable. The problem is, of course, to find out what the mutual capacitance would be with an air dielectric, or with any other homogeneous dielectric of known dielectric constant. Because of the complex configuration of the cable structure it is not calculable.

The shielded balanced pair (referred to herein as the "ideal" pair) structure and its components of mutual capacitance are illustrated in Fig. 1. Although the cable pair structure is different from that of the ideal pair, it has essentially the same components of mutual capacitance. The static shield for the cable pair is not solid and perfectly cylindrical but consists of the other cable pairs or sheath which are immediately adjacent to the pair under consideration. The direct capacitances of the two wires to this ground or shield are very nearly equal.

Rigorous mathematical formulas have been derived by Mrs. S. P. Mead of the Bell Telephone Laboratories for the mutual capacitance ( $C_{mut}$ ) and the capacitance to ground ( $C_g$ ) of the ideal pair. These formulas are given in Appendix I. For discussion purposes they can be written as follows:

$$C_{mut} = \frac{\epsilon}{f_m \left( \frac{d}{2S}, \frac{S}{D} \right)} \quad (2)$$

$$C_g = \frac{\epsilon}{f_g \left( \frac{d}{2S}, \frac{S}{D} \right)} \quad (3)$$

where  $\epsilon$  = dielectric constant and  $f_m$  and  $f_g$  are the functions defined in appendix I.

Dividing equation (3) by equation (2):

$$C_g/C_{mut} = \frac{f_m\left(\frac{d}{2S}, \frac{S}{D}\right)}{f_g\left(\frac{d}{2S}, \frac{S}{D}\right)} \quad (4)$$

Thus, the  $C_g/C_{mut}$  ratio is independent of the dielectric constant, being a function only of the dimensional ratios describing the ideal pair configuration. The ratio  $d/2S$  also appears in formula (1). Therefore, knowledge of  $L$ ,  $C_g$ , and  $C_{mut}$  for an ideal pair is sufficient to find its dimensional ratios and dielectric constant, using the following procedure:

1. Equation (1) is solved for  $d/2S$ .
2. Knowing  $C_g/C_{mut}$  and  $d/2S$ , equation (4) is solved implicitly for  $S/D$ .
3. Knowing  $C_{mut}$ ,  $d/2S$ , and  $S/D$ , equation (2) is solved for  $\epsilon$ .

Curve sheets to facilitate the solution of equations (2) and (4) for the applicable ranges of  $d/2S$  and  $S/D$  are shown in Figs. 2 and 3.

When this same procedure is applied to a cable pair using its measured 1000 cps  $L$ ,  $C_g$ , and  $C_{mut}$ , the dielectric constant value obtained is truly representative of the pair in accordance with our definition of dielectric constant, only if the ideal and actual structures are equivalent. An absence of rigor in this method would be expected due to differences in configuration, and non-homogeneity in the cable dielectric.

The magnitude of the error in the dielectric constant of a cable pair, when determined by this method, was evaluated by comparison of tests on a cable having a known, homogeneous dielectric with tests on an identical cable structure having a non-homogeneous dielectric but a known dielectric constant.

These tests were made on a short length of cable containing twenty-two 19-gauge pairs. The conductors were insulated with solid polyethylene. The core wrap was polyethylene tape under an alpeph\* sheath. A low molecular weight polyethylene compound known as DXL-1 has the same dielectric constant as solid polyethylene (2.26). The compound is in a semi-solid state at room temperature and flows easily at 150° Fahrenheit. By filling the interstitial air space in the cable with DXL-1, a cable having a homogeneous dielectric with a dielectric constant of 2.26 is obtained. The dielectric constant of the cable structure before filling with DXL-1 is found from the ratio of mutual capacitance before and after filling, and a check on the accuracy of the ideal pair formulas as applied to cable pairs is available.

\* *Bell Laboratories Record*, November 1948.



These objectives were not realized with the hoped-for finesse, because of difficulty encountered in filling the interstitial air space with DXL-1. Both vacuum and pressure injection methods were tried, and the best that was

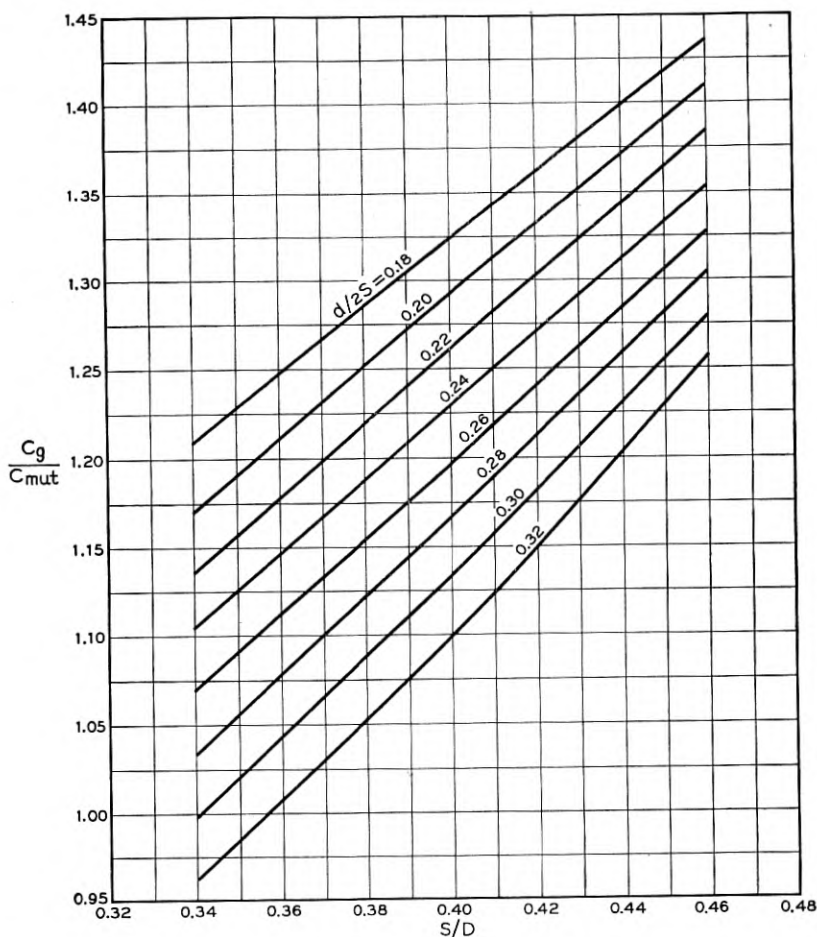


Fig. 2— $C_g/C_{mut}$  as a function of  $d/2S$ ,  $S/D$ .

obtained was a cable with from 85 to 90% of the interstitial space filled with DXL-1, or about 93 to 95% of the total dielectric space filled with polyethylene. These percentages are based on the measured increase in cable weight as a result of filling with DXL-1, and the air volume before filling calculated from nominal insulated conductor and core diameters. Evidently



the cable did not fill completely, because small bubbles of either air or ethylene gas formed during the injection process and would not "wash out" by continued flow. While these results are short of the objective, they are close enough to be useful.

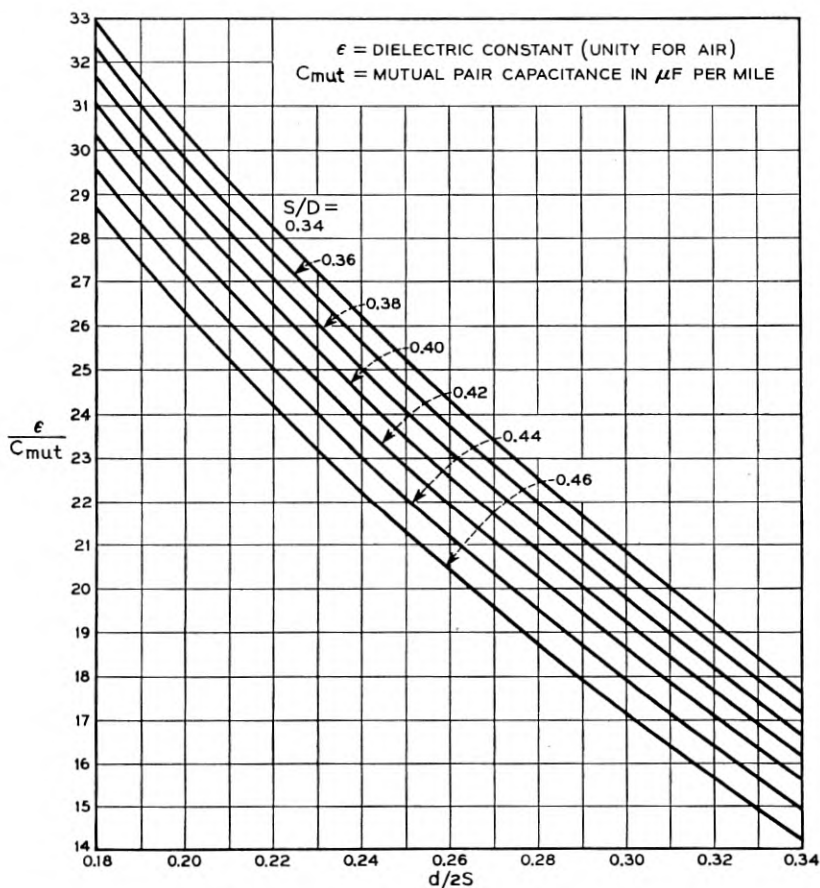


Fig. 3— $\epsilon/C_{mut}$  as a function of  $d/2S$ ,  $SD$ .

The remaining air in the cable, occupying interstitial space, should not reduce the dielectric constant below that of polyethylene (2.26) in proportion to the percentage of air volume. This is because the electric field is weaker in the interstitial space than in the space immediately surrounding the conductors. The dielectric constant of the experimental cable should, therefore, be somewhere between 2.26 and  $1 + (93\% \text{ of } 1.26) = 2.17$ .

Table II shows the results of application of the ideal pair formulas to 1000 cps inductance and capacitance measurements on the experimental cable both before and after injection of DXL-1 compound. The dielectric constant so obtained for the cable pairs after injection of DXL-1 is 2.20. This figure is within the limits set forth previously, and it seems probable that it is accurate to  $\pm 1\%$ , although there is evidently no way to make an independent check. If we accept 2.20 for the dielectric constant of the cable after filling with DXL-1, the dielectric constant for the unfilled or normal condition can be evaluated. Since only the dielectric was altered in filling the cable, the dielectric constant before filling is given by:

$$\epsilon_{mut} = 2.20 \times \frac{0.0881}{0.1070} = 1.81,$$

Where:  $0.0881 \mu\text{f}/\text{mi.} = C_{mut}$  before filling.

$0.1070 \mu\text{f}/\text{mi.} = C_{mut}$  after filling.

Table II shows, however, that application of the ideal pair formulas gives 1.88 for the normal dielectric constant. This figure is 4% higher than the correct 1.81 value. Examining the changes in mutual capacitance and capacitance to ground which occurred when the cable was filled with DXL-1, it is seen that  $C_g$  increased by 27.9% whereas  $C_{mut}$  increased by only 21.4%. The direct capacitance between wires of a pair,  $C_{12}$ , increased by only 12.5%. Since  $C_{12}$  is a component of  $C_{mut}$  but not of  $C_g$ , this accounts for the lower increase in  $C_{mut}$  as compared with  $C_g$ . The amount of air in the cable after filling is so small that it can be assumed that the dielectric is substantially homogeneous. Had the dielectric been homogeneous before filling, the percentage changes in  $C_g$ ,  $C_{mut}$ , and  $C_{12}$  would all have been equal. Thus it is evident that in the normal condition there is a dielectric constant ( $\epsilon_g$ ) applicable for  $C_g$  which is different from  $\epsilon_{mut}$ . We find that  $\epsilon_g$  is:

$$\epsilon_g = 2.20 \times \frac{0.0983}{0.1256} = 1.72,$$

Where  $0.0983 \mu\text{f}/\text{mi.} = C_g$  before filling.

$0.1256 \mu\text{f}/\text{mi.} = C_g$  after filling.

For a given set of capacitance and inductance data, there is, of course, only one value of dielectric constant which will satisfy the ideal pair formulas. This value will not be truly representative of actual cable conditions if non-homogeneity exists. Non-homogeneity, as evidenced by the experimentally determined inequality in  $\epsilon_g$  and  $\epsilon_{mut}$ , accounts for the 4%



error in the 1.88 figure. This follows from the fact that if the  $C_0/C_{mut}$  ratio obtained in the filled or homogeneous condition is used in the determination of the normal dielectric constant, then the correct 1.81 figure is obtained. Of course, these statements regarding the accuracy of the normal dielectric constant are predicated on the absolute accuracy of the 2.20 figure for the dielectric constant of the filled cable. However, as pointed out previously, there is but little room for uncertainty in this connection.

To summarize, it appears that the principal error involved in the use of ideal pair formulas to determine the dielectric constant of cable pairs is due to non-homogeneity of the dielectric. For polyethylene insulated 19-gauge pairs the error is about  $\pm 4\%$ . The magnitude of error expected for paper and pulp insulated pairs is less, as is explained below.

#### a. *Effective Diameter of Shield*

It has been shown that solution of equation (1) provides a value for  $S$  which does in fact equal the average interaxial spacing between wires of a cable pair. In the process of finding  $\epsilon$  using the ideal pair formulas, a value for  $S/D$  is found from which a value for  $D$  can be determined. The relation between the  $D$  value and the physical configuration of the cable pair structure is not precise. It is necessary to consider that  $D$  represents an effective diameter of shield; it is the diameter of the cylindrical shield of an ideal pair structure having the same  $d/2S$  and  $C_0/C_{mut}$  ratios and the same  $d$  as the cable pair.

The value of effective  $D$  is larger if the dielectric is non-homogeneous, as shown by comparing the figures for  $D$  in Table II obtained before and after filling with DXL-1. The  $D$  value obtained under homogeneous conditions is more representative of the average physical placement of conductors comprising the shield.

#### b. *Paper Ribbon and Paper Pulp Insulated Cable*

The distribution of solid insulating material in paper ribbon or pulp insulated cable is different from that in polyethylene insulated cable. Whereas in the latter there is a solid sheath of insulation around the conductor with air occupying interstitial space only, the paper ribbon or pulp insulated cable has some air dispersed among the insulating material in all portions of the dielectric space. "Interstitial space" is not well defined since the paper insulation tends to deform during cabling operations. This sort of dielectric would seem to be more homogeneous than that of the polyethylene insulated cable.

A method for approximating the  $\epsilon_0/\epsilon_{mut}$  ratio in terms of power factor measurements on the balanced and grounded circuits is discussed in Ap-

pendix II.  $\epsilon_0/\epsilon_{mut}$  ratios obtained for three pulp insulated cables do not differ from unity by more than  $\pm 2\%$ . These results, although not conclusive, tend to substantiate the above qualitative comparisons. It is believed, therefore, that  $\epsilon$  values obtained for paper ribbon or pulp insulated cables are probably accurate to about 1 or 2%.

Table III shows results of the use of the methods herein described to obtain the average  $S$ ,  $\epsilon$  and effective  $D$  for examples of some cable designs now in use or under experimental investigation. These data are included primarily to illustrate the method for a variety of cases, and are not comprehensive enough to serve as the basis for a study of the various cable designs and insulations.

### c. Dielectric Constant from $L \times C$ at High Frequencies

For the ideal pair structure the mutual capacitance and inductance are inversely related when the frequency is so high that current does not penetrate either conductors or shield. The relation between inductance and capacitance, in practical units, is then:

$$\epsilon = 34.70 CL_{\infty} \quad (5)$$

$L_{\infty}$  = limiting value approached by inductance as the frequency increases indefinitely—mh/mi.  
 $C$  = capacitance— $\mu$ f/mi.

It is known that  $L_{\infty}$  of individually shielded pairs, that is, pairs having metallic tape shields applied with an overlapped longitudinal or helical seam, can be accurately approximated from a series of inductance tests using the relation:

$$L_F = L_{\infty} + \sqrt{\frac{M}{F}} \quad (6)$$

Where  $L_F$  = distributed inductance at frequency  $F$ .  
 $M$  = a constant  
 $F$  = frequency

The tests are usually made in the range from 2 to 5 mc. Application of formulas (5) and (6) to individually shielded pairs is known to be an accurate technique for evaluating the dielectric constant of non-homogeneous combinations of air and solid materials and has been in use for many years.

This same method has sometimes been used with cable pairs. However, for the filled cable having a dielectric constant of about 2.20, these techniques gave a dielectric constant value of 2.58, which is about 17% too high. The reason for this large discrepancy is apparent when it is considered that the inverse relationship of  $L$  and  $C$  obtains only when the static and magnetic

TABLE III  
DIELECTRIC CONSTANT, INTERAXIAL SPACING, AND EFFECTIVE DIAMETER OF SHIELD FOR SOME TYPICAL MULTI-PAIRED CABLES

AWG	No. pairs	Description	Cable Average at 1000 cps			Space per pr. KCM*	d/2S	S mils	$C_E/C_{mut}$	S/D	Effective D mils	$\epsilon$
			$C_{mut}$ $\mu$ f/mi.	$C_g$ $\mu$ f/mi.	L mh./mi.							
19	61	Paper ribbon	0.0610	0.0706	1.071	18.3	0.243	73.8	1.157	0.369	200	1.53
19	51	Paper ribbon	0.0826	0.0869	0.910	12.2	0.313	57.5	1.052	0.373	154	1.57
19	51	Paper pulp	0.0811	0.0810	0.892	11.75	0.320	56.0	0.999	0.357	157	1.53
22	51	Paper pulp	0.0773	0.0779	0.907	6.5	0.314	40.3	1.008	0.354	114	1.50
19	26	Polyethylene	0.0831	0.0915	1.003	14.3	0.275	64.7	1.101	0.366	177	1.85

\* KCM = 1000 circular mils.

Note: The dimensional ratios and dielectric constant values listed above are based on arithmetic averages of capacitance and inductance for all pairs of a given cable. If these parameters were determined for each pair individually and averaged, the results would not precisely match the above figures, due to the non-linear functions involved. However, the distinction is slight since the RMS deviation of capacitance or inductance for the pairs of a cable is usually about 3%.

fields are completely terminated on the same shielding surface. The shield around a cable pair resembles a Faraday "cage" of wires which is inherently transparent to the magnetic field at low frequencies. Furthermore, at the highest frequency investigated (about 5 mc), the magnetic shielding of a cable pair by the adjacent surrounding conductors is still much less effective than the essentially perfect static shielding. Thus, it is concluded that the  $L \times C$  method should not be used for cable pairs when good accuracy is desired.

## ACKNOWLEDGMENTS

The writer wishes to acknowledge the valued assistance of his associates in this work, which was directed by Mr. O. S. Markuson.

## APPENDIX I

The following formulas were developed by Mrs. S. P. Mead:

I. Formula for the mutual capacitance of a balanced shielded pair:

$$C_{mut} = \frac{0.01944 \epsilon}{\log\left(\frac{1 - v^2}{u(1 + v^2)}\right) - 0.1086 \delta_{12}}$$

$C_{mut}$  is in  $\mu\text{f}/\text{mi}$ .

$u = d/2S$

$v = S/D$

$d$  = conductor diameter

$S$  = interaxial separation

$D$  = inside diameter of shield

$\epsilon$  = dielectric constant (unity for air)

$\delta_{12}$  is a complicated function of  $u$  and  $v$ . It increases for larger values of  $u$  and/or smaller values of  $v$ . The term  $0.1086 \delta_{12}$  amounts to from  $\frac{1}{2}\%$  to about  $7\%$  over the ranges of  $u$  and  $v$  values plotted in Fig. 3.

II. Formula for the capacitance to ground of a balanced shielded pair:

$$C_g = \frac{0.03889 \epsilon}{\log\left(\frac{[3 - \sqrt{1 + 4u^2}] [1 - v^4(1 + 4u^2)]}{8uv^2}\right) + 0.4343\Delta_1}$$

$C_g$  is in  $\mu\text{f}/\text{mi}$ .

$\epsilon$ ,  $u$  and  $v$  are defined in section I.

$\Delta_1$  is also a complicated function of  $u$  and  $v$ . The term  $0.4343 \Delta_1$  amounts to from less than  $\frac{1}{2}\%$  to about  $3\%$  over the ranges of  $u$  and  $v$  plotted in Fig. 2. It increases for larger values of  $u$  and/or larger values of  $v$ .

## APPENDIX II

It has been shown that if there is a dielectric constant for capacitance to ground ( $\epsilon_g$ ) which is different from  $\epsilon_{mut}$ , then values of  $\epsilon_{mut}$  as found from the ideal pair formulas will be in error. This appendix describes results of an attempt to evaluate the  $\epsilon_{mut}/\epsilon_g$  ratio for paper ribbon and pulp insulated cable using a method suggested by Mr. M. C. Biskeborn. It was reasoned that whatever non-homogeneity exists in the cable dielectric would be caused by variations in the insulation density in various portions of the dielectric space. Compressive forces encountered in twisting and stranding the pairs tend to make the density for the space between wires of a pair high as compared with an average density for the total dielectric space. Since ( $\epsilon - 1$ ) and the power factor, or simply  $G/C$ , go to zero approximately in direct

TABLE A  
EVALUATION OF  $\epsilon_g/\epsilon_{mut}$  FOR PULP INSULATED CABLE

AWG	Description	Cable Averages 1000 cps		A = G <sub>mut</sub> C <sub>mut</sub>	Cable Averages 1000 cps		B = G <sub>g</sub> C <sub>g</sub>	F = B/A Power Factor Ratio	$\frac{\epsilon_g}{\epsilon_{mut}}$
		C <sub>mut</sub> μi/mi	G <sub>mut</sub> μmhos/mi.		C <sub>g</sub> μi/mi.	G <sub>g</sub> μmhos/mi.			
19	51 Prs. Pulp	0.0858	2.00	23.3	0.0850	1.86	21.9	0.94	0.980
19	51 Prs. Pulp	0.0811	1.12	13.8	0.0810	1.19	14.7	1.065	1.022
22	51 Prs. Pulp	0.0773	1.09	14.1	0.0779	1.08	13.9	0.985	0.995

proportion to the amount of solid dielectric material, the following equation may be used to find an approximate value for  $\epsilon_g/\epsilon_{mut}$ :

$$\frac{\epsilon_g - 1}{\epsilon_{mut} - 1} = \frac{\text{Power Factor of Grounded Circuit}}{\text{Power Factor of Balanced Circuit}}$$

$$\frac{\epsilon_g - 1}{\epsilon_{mut} - 1} = \frac{G_g/C_g}{G_{mut}/C_{mut}} = F \text{ (power factor ratio)}$$

Dividing by  $\epsilon_{mut}$  and rearranging:

$$\frac{\epsilon_g}{\epsilon_{mut}} = F + \left( \frac{1 - F}{\epsilon_{mut}} \right)$$

Substituting 1.50 for  $\epsilon_{mut}$  on the right hand side introduces but small error in the eventual result since  $F$  is known to be closely equal to unity and permits solution for  $\epsilon_g/\epsilon_{mut}$  as follows:

$$\frac{\epsilon_g}{\epsilon_{mut}} = 1 - \left( \frac{1 - F}{3} \right)$$



This relation was used to evaluate  $\epsilon_g/\epsilon_{mut}$  for 3 pulp insulated cables for which the necessary data were available. The results are shown in Table A. None of the 3 cables has a  $\epsilon_g/\epsilon_{mut}$  ratio different from unity by more than 2%. The average for the 3 cables is 1.0 and it is felt that the  $\pm 2\%$  deviation could be due to inaccuracies inherent to conductance measurements. These data represent the only attempt to quantitatively evaluate non-homogeneity in pulp or paper insulated cable dielectrics and, while not by any means conclusive, they tend to substantiate the belief that the effects of any such non-homogeneity on the  $\epsilon_g/\epsilon_{mut}$  ratio is small.

# ***N*-Terminal Switching Circuits**

By **E. N. GILBERT**

(Manuscript Received Feb. 14, 1951)

The circuits considered have  $N$  accessible terminals and are operated by gangs of selector switches. Synthesis of any  $N$ -terminal switching function is accomplished. The synthesis method is proved to be economical in the sense that the switching functions which can be synthesized by any other method using much fewer contacts comprise a vanishingly small fraction of the total of all possible switching functions.

## INTRODUCTION

In a recent issue of The Bell System Technical Journal<sup>1</sup>, C. E. Shannon discussed the synthesis of two-terminal relay contact networks. Some of his results will be generalized in this paper to  $N$ -terminal networks which use selector switches with any number of positions instead of the two of a relay. The kind of circuit which will be considered may be visualized as a black box with  $N$  accessible terminals and with  $M$  shafts extending from it. Each shaft operates a selector switch (which will usually consist of several simple selector switches ganged together) inside the box. The rotors and contacts inside the box are connected electrically to one another and to the  $N$  terminals so that each way of setting the  $M$  shafts determines a pattern of interconnection of the  $N$  terminals.

We do not permit the black box to contain relay magnets or other devices which would allow the circuit to operate sequentially. Because of this restriction our results apply only to the simplest kind of switching circuit in which the state of the  $N$  terminals depends only on the present state of the  $M$  shafts, and not on the past history of the shafts. We may then use the term *N-terminal switching function* to mean a rule which assigns to each way of setting the  $M$  shafts a state of the terminals. We are concerned with the problem of synthesis: given an  $N$ -terminal switching function  $f$ , to find a switching circuit for which the states of the shafts and terminals correspond in the way indicated by  $f$ .

Let  $p_1, \dots, p_M$  be the numbers of positions which the  $M$  shafts can assume. Then there are  $p_1 \dots p_M$  different states of the shafts and the shafts have a memory<sup>2</sup>

<sup>1</sup> C. E. Shannon, *B.S.T.J.*, 28, pp. 59-98 (1949).

<sup>2</sup> C. E. Shannon, *B.S.T.J.*, 29, pp. 343-349 (1950).

$$H = \sum_{i=1}^M \log p_i$$

bits (in this paper "log" stands for "logarithm to base 2"). The results to follow include an estimate of the minimum number of contacts needed for almost all  $N$ -terminal switching functions and a network synthesis method which uses a number of contacts of the same order of magnitude as the minimum number. The number of contacts needed for almost all  $N$ -terminal switching functions is about  $\frac{N \log N 2^H}{H + \log N}$  when  $H$  and  $N$  are large. The words "almost all" are used here in the sense that the fraction of switching functions which can be synthesized using fewer contacts than the given number tends to zero as  $H$  and/or  $N$  increases. The number of contacts used by the synthesis method is about  $\frac{N^2 P 2^H}{H}$  where  $P$  is the number of positions on the largest switch. The factor  $P$  can be reduced in most cases. The analogous expressions found in Shannon's paper are  $\frac{2^n}{n}$  and  $\frac{2^{n+3}}{n}$  where  $n$  is the number of switching variables.

One of the most surprising facts about switching functions is that, if  $H$  is moderately large, almost none of them can be synthesized without using fantastically many contacts. This is already true of Shannon's two-terminal networks, and for  $N$ -terminal networks the situation is even worse. The reader may first turn to page 685 where a numerical example illustrating this phenomenon is given.

These paradoxical results are explained by noting that switching functions in general are much different from the usual kinds of switching functions which have practical applications. One concludes that the invention of better methods for synthesizing any imaginable function whatsoever will be of little help in practice. Almost all these functions are impossible to build (because of contact cost) and would be of no use if built. Instead one must try to isolate classes of useful switching functions which are easy to build.

## PART I: TWO-TERMINAL NETWORKS

### *Selector Switches*

A typical selector switch is shown in Fig. 1. It consists of a number of rotors turned by a shaft which can be set in any one of  $p$  positions. In each position of the shaft, certain of the rotors touch contacts, thereby closing those branches in the network containing the touched contacts. However, the only kinds of switches to be considered here are those with the property

that, if a contact is touched by a rotor when the shaft is in position number  $j$ , then this contact remains untouched for all other positions of the shaft.

Networks built from two-position switches are analyzed with the aid of Boolean Algebra. It is possible to construct an algebra which is appropriate for selector switch circuits. A detailed account of this algebra has been given by H. Piesch.<sup>3</sup>

The state of a switch with  $p$  positions can be associated with a switching variable  $x$  which ranges over the values  $1, 2, \dots, p$ . Then " $x = k$ " means the same as "the switch is in its  $k^{\text{th}}$  position." The state of a two-terminal network, using  $M$  switches with  $p_1, \dots$ , and  $p_M$  positions, is a hindrance function  $f(x_1, \dots, x_M)$  of the  $M$  switching variables  $x_1, \dots, x_M$  with  $x_i$

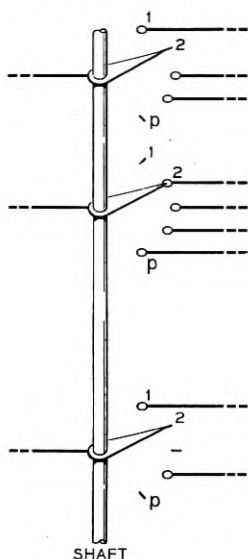


Fig. 1—Selector switch.

ranging from 1 to  $p_i$ . As usual  $f = 1$  means the circuit is open and  $f = 0$  means the circuit is closed. Then  $f(x_1, \dots, x_M) + g(x_1, \dots, x_M)$  is the function representing the series connection of two networks whose functions are  $f$  and  $g$  while  $f(x_1, \dots, x_M) g(x_1, \dots, x_M)$  represents the networks  $f$  and  $g$  in parallel.

The circuit which consists of just a rotor which touches a contact in its  $i^{\text{th}}$  position has hindrance function

$$e_i(x) = \begin{cases} 1 & \text{if } x \neq i \\ 0 & \text{if } x = i \end{cases}$$

<sup>3</sup> H. Piesch, *Archiv für Electrotechnik*, 33, pp. 674, 686 and pp. 733-746 (1939).

There is no simple identity which corresponds to the Boolean Algebra expansion by sums. An identity analogous to the Boolean Algebra expansion about  $x_1$  by products is

$$(1) \quad f(x_1, \dots, x_M) = \prod_{i=1}^p [e_i(x_1) + f(i, x_2, \dots, x_M)]$$

where the range of  $x_1$  is from 1 to  $p$ . To prove (1) we need only observe that in the product all the terms for which  $i \neq x_1$  have the value 1. The remaining term, for which  $i = x_1$ , has the value  $f(x_1, \dots, x_M)$ . The switching interpretation of (1) is illustrated in Fig. 2. By repeated use of (1) it follows that any function  $f(x_1, \dots, x_M)$  can be written as an expression involving

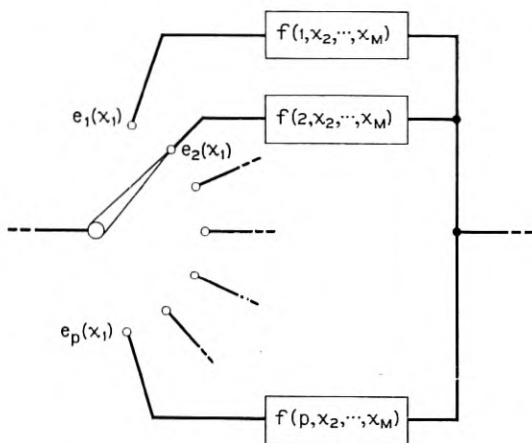


Fig. 2—Expansion of  $f(x_1, \dots, x_M)$  about  $x_1$ .

parentheses, addition signs, multiplication signs, the  $e_i(x_j)$ , and nothing else. Such expressions may be regarded as Boolean functions with the  $e_i(x_j)$  as variables; they may be rearranged and factored according to the usual rules of Boolean Algebra. However, one should keep in mind that the  $e_i(x_j)$  are subject to the constraints that a selector switch can be in only one position at any given time. The effect of these constraints is to add a cancellation law

$$e_h(x_j) + e_i(x_j) = 1 \quad \text{if } h \neq i.$$

The inverse  $e'_i(x_j)$  of the Boolean variable  $e_i(x_j)$  is the Boolean function

$$e'_i(x_j) = \begin{cases} 1 & \text{when } e_i(x_j) = 0 \\ 0 & \text{when } e_i(x_j) = 1. \end{cases}$$

Regarded as a hindrance function of the switching variable  $x_j$ ,

$$e'_i(x_j) = \begin{cases} 0 & \text{when } x_j \neq i \\ 1 & \text{when } x_j = i \end{cases}$$

Then by (1),

$$e'_i(x_j) = \prod_{h \neq i} e_h(x_j).$$

If the switch  $x_j$  has  $p_j$  positions, it takes  $p_j - 1$  contacts to build a circuit with hindrance function  $e'_i(x_j)$ .

### Synthesis

Suppose that  $M$  shafts, governed by switching variables  $x_1, \dots, x_M$ , are given, together with a two-terminal hindrance function  $f(x_1, \dots, x_M)$ . The synthesis problem is to design a network with hindrance function  $f(x_1, \dots, x_M)$ , adding suitable rotors and contacts to form selector switches from the given shafts.

One solution can be found immediately:

(i) As described above, express the hindrance function  $f(x_1, \dots, x_M)$  as a Boolean function  $B(\xi_1, \dots, \xi_R)$  of Boolean variables  $\xi_1, \dots, \xi_R$  (which are the  $e_i(x_j)$  with new labels). Here  $R = p_1 + p_2 + \dots + p_M$ .

(ii) Any of the well known methods of synthesizing relay networks can be used to design a network operated by the Boolean variables  $\xi_1, \dots, \xi_R$  and with hindrance function  $B(\xi_1, \dots, \xi_R)$ .

(iii) In the network found in (ii) replace each contact  $\xi_a$  by the appropriate  $e_i(x_j)$  and each back contact  $\xi'_a$  by the appropriate circuit  $e'_i(x_j)$ .

The solution found in (iii) will ordinarily use up a number of contacts which is unnecessarily large by many orders of magnitude. From Shannon's theorems on relay networks we know that the probability is high, that as many as

$$\frac{2^R}{R}$$

contacts will be needed in step (ii). The final circuit (iii) will have even more contacts if some circuits  $e'_i(x_j)$  are used.

The synthesis process which follows replaces the exponent  $R = \sum_{i=1}^M p_i$  in the estimate of the number of contacts by the smaller number  $\sum_{i=1}^M \log p_i$ . The reader may recognize the process as essentially the same as the one given by Shannon for two-terminal relay networks. The network will again take the form of a tree connected to a circuit which produces all functions of the switching variables which govern it.

If one expands  $f(x_1, \dots, x_M)$  about  $x_1, \dots, x_k$  one finds

$$(2) \quad f = \prod_{r,s,\dots,z} [e_r(x_1) + e_s(x_2) + \dots + e_z(x_k) + f(r, s, \dots, z, x_{k+1}, \dots, x_M)]$$

where the multiple product is taken over the entire range of the variables  $x_1, \dots, x_k$ . The different functions  $e_r(x_1) + e_s(x_2) + \dots + e_z(x_k)$  can all be realized with one tree circuit as shown in Fig. 3. If the expansion had been performed about all the variables  $x_1, \dots, x_M$ , identity (2) would

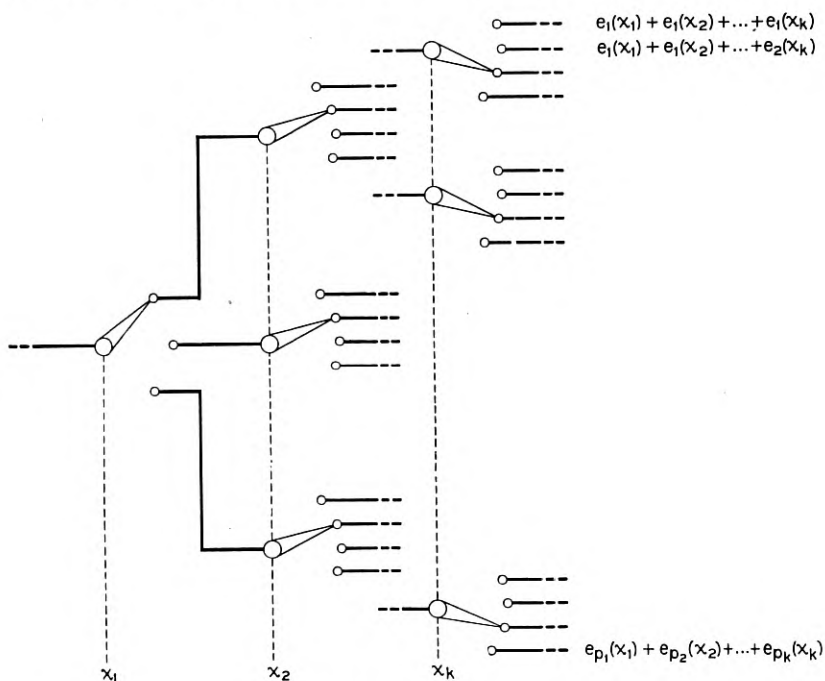


Fig. 3—Tree.

show that the desired function  $f$  could be synthesized by connecting one terminal to the input lead of a tree and the other terminal to certain of the tree's output leads. The number of contacts used would have been about  $2^{H+1}$ . The method which follows uses still fewer contacts.

The network which we use to synthesize the function  $f$  is shown in Fig. 4. It consists of the tree of Fig. 3 with its output leads connected to the input leads of a network on the right which is designed so that the hindrances from its input leads to its output lead are the functions  $f(r, s, \dots, z, x_{k+1},$

$\dots, x_M$ ). For given values  $r_0, s_0, \dots, z_0$  of the switching variables  $x_1, \dots, x_k$  of the tree, there is only one closed path through the tree; this path ends at the output lead labelled  $e_{r_0}(x_1) + \dots + e_{z_0}(x_k)$ .

The hindrance from this point to the output lead of the right-hand network is the hindrance of the network of Fig. 4. This hindrance is just  $f(r_0, s_0, \dots, z_0, x_{k+1}, \dots, x_M)$  (note that the connections to the disjunctive tree do not cause any interconnections among the other leads of the right hand network), which proves that the network has the required hindrance. By proper choice of the number  $k$  of switches in the tree we will obtain an economical design.

#### Network to Produce All Functions

To produce all of  $f(r, s, \dots, z, x_{k+1}, \dots, x_M)$  it suffices to build a circuit which produces every function of  $(x_{k+1}, \dots, x_M)$ . Let these variables be

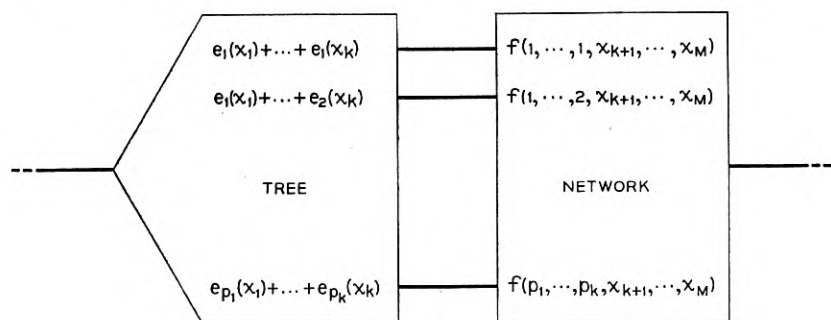


Fig. 4—Network for  $f(x_1, \dots, x_M)$ .

relabelled  $y_1, \dots, y_L$  and have ranges  $p_1, \dots, p_L$ . Let the largest of  $p_1, \dots, p_L$  be called  $P$ .

*Theorem I.* A network which produces every function of  $(y_1, \dots, y_L)$  can be built with a number  $\psi_L$  of contacts satisfying

$$(3) \quad \psi_L \leq P2^{p_1 \cdots p_L}.$$

The proof is by induction on  $L$ . Suppose that a network to produce every function of  $(y_1, \dots, y_{j-1})$  has been built with  $\psi_{j-1}$  contacts and try to build one for every function of  $(y_1, \dots, y_j)$ . The number of functions which the network must produce is  $2^{p_1 \cdots p_j}$ , for there are  $p_1 \cdots p_j$  different ways of setting the switches and two choices (0 or 1) for the value of the function for each state of the switches. Of these functions, the  $\psi_{j-1}$  network itself provides  $2^{p_1 \cdots p_{j-1}}$  functions with no additional contacts (these are the functions independent of  $y_j$ ). Any one of the remaining  $(2^{p_1 \cdots p_j} - 2^{p_1 \cdots p_{j-1}})$  func-



tions  $f$  can be obtained by connecting to the functions  $f(y_1, \dots, y_{j-1}, 1), \dots, f(y_1, \dots, y_{j-1}, p_j)$  through  $e_1(y_j), \dots, e_{p_j}(y_j)$  as shown in Fig. 5. In this way a new network is found which produces all functions of the  $j$  variables and uses  $\psi_j$  contacts where

$$(4) \quad \psi_j - \psi_{j-1} \leq P(2^{p_1 \cdots p_j} - 2^{p_1 \cdots p_{j-1}}).$$

If we now assume that formula (3) holds for  $\psi_{j-1}$  we obtain

$$\psi_j \leq P2^{p_1 \cdots p_j}.$$

Thus the theorem will follow by induction when we prove (3) for the case  $L = 1$ . Since  $\psi_0 = 0$  (no contacts are needed to synthesize the two functions 0 and 1) the inequality (4) reduces, when  $L = 1$ , to

$$\psi_1 \leq P(2^{p_1} - 2)$$

and the theorem is proved.

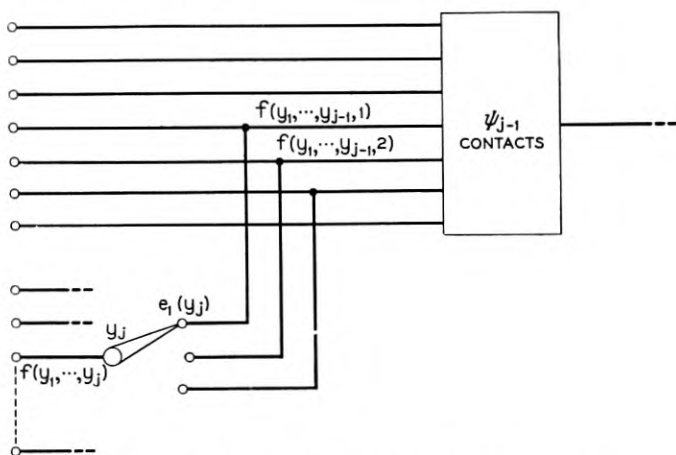


Fig. 5—Network to produce all functions of  $(y_1, \dots, y_j)$ .

The induction process we have just described will use up the smallest number of contacts when the large switches are used up first and the small switches last. If, in the process,  $p_j > p_{j+1}$ , then the number of contacts which would have been saved by making switch  $y_{j+1}$  precede switch  $y_j$  is found to be

$$(p_j - p_{j+1})2^{p_1 \cdots p_{j-1}}(2^{p_j} - 1)(2^{p_{j+1}} - 1).$$

By adding switches in order of decreasing size in the induction process, the factor  $P$  in (3) can be reduced to nearly  $p_L$ , the smallest of the  $L$  ranges. This refinement is unnecessary for the theory which follows.

### The Tree

The number of contacts used in the tree of Fig. 3 is

$$p_1 \cdots p_k \left( 1 + \frac{1}{p_k} + \frac{1}{p_{k-1} p_k} + \cdots + \frac{1}{p_k \cdots p_2} \right).$$

It can be shown that the most economical way to build the tree is to put the small switches at the narrow end of the tree. If the smallest number of positions of any of the switches in the tree is  $p_1$  then the number of contacts in the tree is less than

$$(5) \quad \frac{p_1 \cdots p_k}{1 - 1/p_1} \leq 2p_1 \cdots p_k.$$

### Upper Bound

Having counted the number of contacts which are used in the tree and in the network which produces all functions in Fig. 4, it only remains to decide how many of the given switches  $x_1, \cdots, x_M$  are to be put in each of these two parts.

*Theorem II.* Let  $P$  be the largest of the numbers  $p_1, \cdots, p_M$  of values which the variables  $x_1, \cdots, x_M$  can assume. Then any switching function of  $(x_1, \cdots, x_M)$  can be synthesized using no more than

$$(6) \quad \left( P + \frac{1}{2} \right) \frac{2^{H+1}}{H - 2 \log H}$$

contacts when  $H > 4$  bits.

To prove the theorem we consider two cases according as  $P$  is greater or less than  $H - 2 \log H$ .

Case 1: ( $P \geq H - 2 \log H$ )

In this case we use the synthesis process described above, putting all the switches into the tree and none in the network which produces all functions. The number of contacts used is less than  $2 \cdot 2^H$  and the theorem follows because

$$P \geq H - 2 \log H.$$

Case 2: ( $P < H - 2 \log H$ )

In this case we use the synthesis process described above, putting into the right-hand network a collection  $S$  of switches so chosen that  $\prod_s p_i$  comes as close as possible to  $H - 2 \log H$  without actually exceeding it. Then if

$$\prod_s p_i = (H - 2 \log H)F,$$

we have  $F \leq 1$ . Also,  $F \neq 0$  since any  $p_i$  satisfies

$$p_i \leq P < H - 2 \log H.$$

Since  $F \neq 0$  it follows that  $F > 1/P$ . For if  $0 < F \leq 1/P$ , adding another switch to the collection  $S$  will increase  $\prod_s p_i$  without making it exceed  $H - 2 \log H$ .

Using (3) and (5), the number of contacts in the network is less than

$$P2^{(H-2\log H)F} + \frac{2^{H+1}}{(H-2\log H)F} \leq \left( \frac{P}{H^2} + \frac{2P}{H-2\log H} \right) 2^H.$$

Since  $P < H - 2 \log H$ ,

$$\frac{P}{H^2} < \frac{1}{H} < \frac{1}{H - 2 \log H}$$

and the theorem is proved.

Only a small fraction of the functions will use up this many contacts. In any particular case, the number of contacts used will be about

$$\left( \frac{1}{F} + \frac{1}{2} \right) \frac{2^{H+1}}{H - 2 \log H}$$

and, if many different sizes of switches are used in the network, one should be able to make  $1/F$  much closer to 1 than  $P$ . Even when all the switches are the same size, one expects

$$\frac{1}{F} < \sqrt{P}$$

in about half the cases.

## PART II: N-TERMINAL NETWORKS

### Synthesis

Let the accessible terminals be labelled  $1, 2, \dots, N$ . To each pair  $i, j$  of terminals of an  $N$ -terminal network there corresponds a hindrance function  $B_{ij}(x_1, \dots, x_M)$  which tells whether or not there is a closed path between  $i$  and  $j$ . The  $B_{ij}$  satisfies a consistency requirement

$$(7) \quad "B_{ia} + B_{ab} + \dots + B_{de} + B_{ej} = 0 \text{ implies } B_{ij} = 0".$$

The number of consistency requirements (7) is

$$\sum_{r=3}^N \frac{N!}{2(N-r)!} \approx \frac{eN!}{2}.$$

However, one can show that all the requirements (7) hold if and only if the  $\frac{N(N-1)(N-2)}{2}$  requirements

$$"B_{ia} + B_{aj} = 0 \text{ implies } B_{ij} = 0"$$

hold.

Conversely, any set of  $\frac{N(N-1)}{2}$  hindrance functions which satisfy (7) determine a realizable  $N$ -terminal network. One way of synthesizing the network is just to connect, between each pair  $i, j$  of terminals, a two-terminal network with hindrance function  $B_{ij}$ . It follows from theorem II that

*Theorem III.* Any  $N$ -terminal switching function of switches with  $P$  or fewer positions can be synthesized with no more than

$$N(N-1) \left( P + \frac{1}{2} \right) \frac{2^H}{H - 2 \log H}$$

contacts when  $H > 4$  bits.

The network can also be synthesized using  $N$  trees, each of which produces all of the possible functions  $e_r(x_1) + e_s(x_2) + \dots + e_s(x_M)$ . Each terminal is connected to the input lead of one of the trees; and the output leads, to which the terminals are connected in any given state  $(x_1, \dots, x_M)$ , are interconnected in the way one wants the terminals to be interconnected in that state. The number of contacts used in this type of synthesis is less than

$$N2^{H+1}.$$

The synthesis using two-terminal networks ordinarily requires fewer contacts than the one using trees as long as

$$H - 2 \log H > \frac{1}{2}(N-1)(P + \frac{1}{2})$$

An example illustrating the design of a typical three-terminal network is given in the appendix.

### *Number of Functions*

Every  $N$ -terminal switching function determines a realizable matrix of hindrance functions  $B_{ij}(x_1, \dots, x_M)$ . It is important to know the number of different switching functions of  $(x_1, \dots, x_M)$ .

A state of the  $N$  terminals is determined by specifying the groups of terminals which are connected together. The number  $\phi(N)$  of such states is the number of ways that  $N$  different objects can be distributed into 1, 2,  $\dots$ , or  $N$  parcels when the parcels are indistinguishable from one another and no parcel is left empty.

A switching function represents one of these  $\phi(N)$  different states for each of the  $2^H$  different switch settings. Hence the number of switching functions is

$$(\phi(N))^{2^H}.$$

Although there is no simple formula for  $\phi(N)$ , a generating function for  $\phi(N)$  is well known:<sup>4</sup>

$$(8) \quad e^{e^z-1} = \sum_{n=0}^{\infty} \frac{\phi(n)}{n!} z^n.$$

A recursion formula which can be used to calculate  $\phi(N)$  is

$$(9) \quad \phi(N + 1) = \sum_{k=0}^N C_{N,k} \phi(k).$$

When  $N$  is large  $\phi(N)$  can be estimated with the help of the upper and lower bounds to be derived. These bounds will be of use to us later mainly because they show that, for large  $N$ ,  $\log \phi(N)$  is approximately  $N \log N$ .

*Theorem IV.*

$$(10) \quad \phi(N) \leq \frac{N!}{e} \left[ \frac{e^{N/\log_e N}}{\log_e \frac{N}{\log_e N}} \right]^N.$$

*Proof.* The maximum value of  $|e^{e^z-1}|$  on the circle  $|z| = r$  is  $e^{e^r-1}$ . Using (8) and Cauchy's inequality for the  $N^{\text{th}}$  coefficient in a power series,

$$(11) \quad \phi(N) \leq \frac{N! e^{e^r-1}}{r^N}$$

for all  $r > 0$ . The best estimate of  $\phi(N)$  will be obtained by minimizing (11) on  $r$ . To do this one sets  $r = r_0$  where

$$r_0 e^{r_0} = N.$$

The simpler result (10) is obtained from (11) by setting

$$r = \log_e \frac{N}{\log_e N}$$

*Theorem V.*

$$(12) \quad \frac{A^N}{A!} \leq \phi(N) \quad \text{for all integers } A.$$

*Proof.* Let  $Q(N, A)$  be the number of ways that  $N$  different objects can be distributed into 1, 2,  $\dots$ , or  $A$  indistinguishable parcels. Then  $Q(N, A)A!$

<sup>4</sup> W. A. Whitworth, Choice and Chance, p. 88, Cambridge, Bell, 1901.

must be greater than the number of ways  $N$  different objects can be placed in  $A$  different boxes (labelled  $1, \dots, A$ ); i.e.

$$A^N \leq Q(N, A)A! \leq \phi(N)A!.$$

To obtain the best lower bound from (12) one may maximize on  $A$ . The best value of  $A$  to use is one which comes close to satisfying

$$\left(1 + \frac{1}{A}\right)^{N-1} = A.$$

For large  $N$  the solution is approximately

$$A = \frac{N}{\log_e N}.$$

To perform the minimization in theorem IV more carefully one would solve

$$A_0 \log_e A_0 = N$$

for  $A_0$ . This is the minimizing equation given in theorem IV with  $r_0 = \log_e A_0$ . It is also very nearly the minimizing equation given in theorem V.

Then our proofs of theorems IV and V show that

$$\frac{A_0^N}{A_0!} \leq \phi(N) \leq N! \frac{e^{A_0-1}}{(\log_e A_0)^N}.$$

For large  $N$  these bounds differ by a factor of about

$$\frac{2\pi}{e} \frac{N}{\sqrt{\log_e N}}$$

More accurate information about the behavior of  $\phi(N)$  for large  $N$  is provided by an asymptotic series found by L. F. Epstein.<sup>5</sup> The first term in his series is

$$\phi(N) \sim \frac{a^{N-a} e^{a-1}}{\sqrt{\log_e a}}$$

where  $a$  is found by solving

$$a \log_e (a + 1) = N.$$

Figure 6 is a graph of  $\phi(N)$  vs.  $N$  using a log log scale for  $\phi(N)$ . The points are exact values and the curves show the upper and lower bounds.

<sup>5</sup> L. F. Epstein, *J.M.P.*, 18, 3, pp. 153-173 (1939).

*Number of Graphs*

Let  $G(N, K)$  be the number of topologically distinct linear graphs which can be drawn interconnecting the  $N$ -terminals and using  $K$  branches.  $G(N, K)$  counts *all* graphs including graphs with dangling branches and disconnected pieces. It also counts graphs in which any or all of the  $N$ -terminals are connected to no branches. Figures 7a, b, c, d, e show some topologically distinct graphs which would be counted in finding  $G(3, 10)$ .

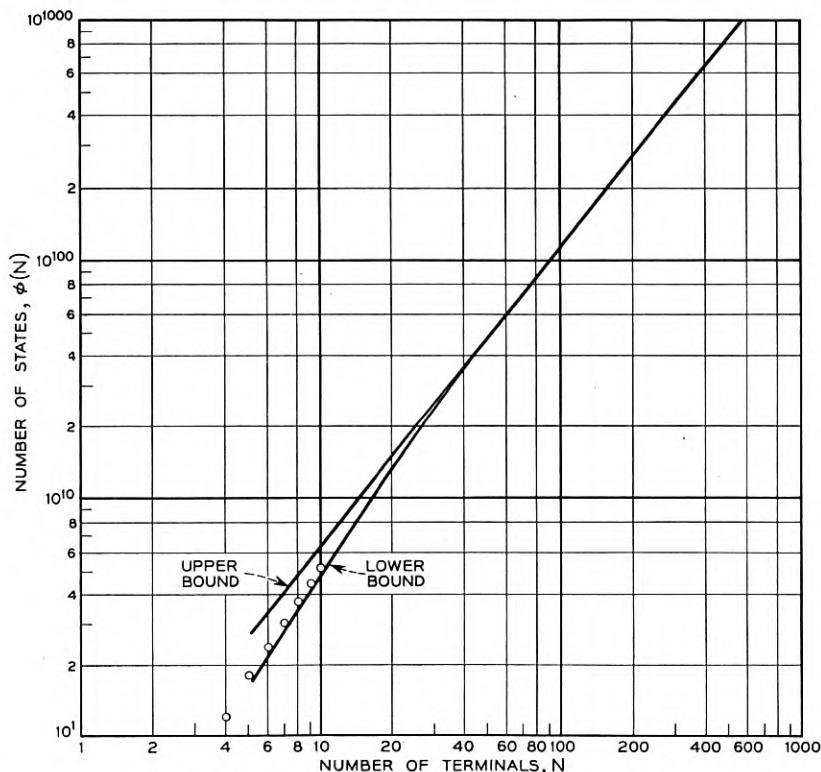


Fig. 6—Number of possible states of  $N$  terminals.

Graph 7f is topologically identical with graph 7b and so is not to be counted again. The first step toward finding a lower bound on the number of contacts which almost all switching functions require is to find an upper bound on  $G(N, K)$ .

*Theorem VI.*

$$G(N, K) < 2^{N+K} (N + 2K)^K .$$

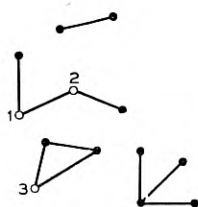
*Proof.* Every linear graph can be constructed by the following process. Let the branches be numbered 1, 2,  $\dots$ ,  $K$  and let the end points of the  $k^{\text{th}}$  branch be called  $A_k$  and  $B_k$ . There are  $K - 1$  places where partition marks can be inserted in the sequence  $A_1, \dots, A_k$  and hence there are  $2^{K-1}$  ways of partitioning the  $A_k$ 's into groups of the form

$$G_1 = (A_1, A_2, \dots, A_a)$$

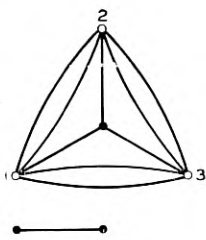
$$G_2 = (A_{a+1}, \dots, A_b)$$

$$G_3 = (A_{b+1}, \dots, A_c)$$

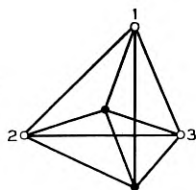
$$\vdots$$



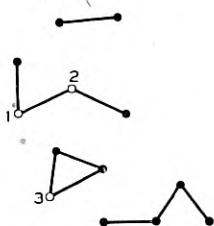
(a)



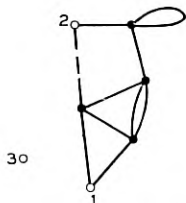
(b)



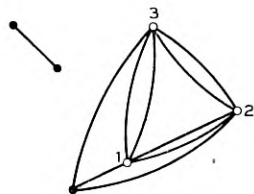
(c)



(d)



(e)



(f)

Fig. 7—Examples of graphs.

There are  $2^N$  ways of selecting some of the terminals 1, 2  $\dots$ ,  $N$ . Suppose that  $m$  of the terminals have been selected; then pick one of the partitions of the  $A_k$ 's which has  $m$  or more groups  $G_1, \dots, G_{m+s}$ . Connect all the end points in  $G_1$  to the first selected terminal, all the end points in  $G_2$  to the second selected terminal, etc. Next connect the terminals in  $G_{m+1}, \dots, G_{m+s}$  together to form  $s$  nodes. The number of ways of performing all these operations is less than

$$2^{N+K-1}.$$



Connect  $B_1$  to one of the  $N$ -terminals or to one of the nodes just made or else use  $B_1$  to make a new node. Connect  $B_2$  to one of the terminals or nodes or else use  $B_2$  to make a new node, etc. The number of ways of connecting  $B_1, \dots, B_K$  is less than

$$(N + K + 1)(N + K + 2) \cdots (N + 2K) < (N + 2K)^K,$$

which proves the theorem.

Since most graphs can be constructed in many different ways by this process, theorem VI gives a very poor estimate of  $G(N, K)$ . In the application which we will make of  $G(N, K)$  it is enough to know that  $\log G(N, K)$  behaves something like  $K \log K$ . To prove that  $K \log K$  cannot be replaced by anything much smaller we now give a lower bound for  $G(N, K)$ .

*Theorem VII.*

$$\frac{(\phi(K))^2}{2^K K!} \leq G(N, K)$$

*Proof.*  $G(N, K)$  is larger than the number of graphs which can be drawn without specifying certain nodes as terminals  $1, 2, \dots, N$ . Of these graphs let us count only those which have the property that no cycle in the graph has an odd number of branches. Another characterization of these graphs is that their nodes can be divided into two classes A and B such that no branch joins two nodes of the same class.

To construct such graphs we first number the branches  $1, 2, \dots, K$  and give them an orientation (say by putting an arrow head at one end of each branch). The front ends of the branches can be grouped together into nodes in  $\phi(K)$  ways. Then the tail ends of the branches are grouped together in one of  $\phi(K)$  ways. In this way a total of  $(\phi(K))^2$  different graphs can be drawn, in which the branches are numbered and oriented. If we now ignore the numbers on the branches we still have at least  $\frac{(\phi(K))^2}{K!}$  distinct graphs with oriented branches. If the orientation is ignored, the number of topologically different graphs which remain is greater than

$$\frac{(\phi(K))^2}{2^K K!}.$$

*Lower Bound*

We have seen that any switching function can be realized with no more than about

$$\frac{N^2 P 2^H}{H - 2 \log H}$$

contacts. To show that this number cannot be improved very much we will now show that almost all switching functions require a number of contacts of this order of magnitude.

*Theorem VIII.* Let any  $\epsilon > 0$  be given. The fraction of switching functions which can be synthesized using less than

$$(13) \quad (1 - \epsilon) \frac{2^H \log \phi(N)}{H + \log \log \phi(N)}$$

contacts approaches zero uniformly as the number  $M$  of switches becomes large.<sup>6</sup>

*Proof.* The number of switching functions which can be constructed with  $K$  contacts or less is certainly smaller than the number of ways the  $K$  branches of the  $G(N, K)$  graphs can be replaced by contacts  $e_r(x_i)$  or open circuits; i.e. smaller than

$$G(N, K) \left( \sum_{i=1}^M p_i + 1 \right)^K.$$

By theorem VI the fraction  $F(K)$  of the  $(\phi(N))^{2^H}$  switching functions which can be built using  $K$  contacts or less satisfies

$$\begin{aligned} F(K) &\leq 2^{N+K} (N + 2K)^K \left( \sum_{i=1}^M p_i + 1 \right)^K (\phi(N))^{-2^H} \\ &\leq 2^{2N+K} \left( \log K + 2 + \log \left[ \sum_{i=1}^M p_i + 1 \right] \right)^{-2^H \log \phi} \end{aligned}$$

where we have used

$$\begin{aligned} \log_2 (N + 2K) &\leq \log_2 2K + \frac{N}{2K} \log_2 e \\ &\leq \log_2 2K + \frac{N}{K}. \end{aligned}$$

When  $K$  is the expression (13), one finds

$$F(K) \leq 2^{2N} (\phi(N))^{2^H} \left( -\epsilon + \frac{\log(\sum p_i + 1) + 2}{H + \log \log \phi} \right).$$

Since  $\frac{\log \sum p_i}{H} = \frac{\log \sum p_i}{\log \prod p_i}$  approaches zero as the number of switches  $M$  gets large, it follows that for sufficiently large  $M$  and any  $N$

<sup>6</sup> The word *uniformly* is used to indicate that the fraction in question can be made smaller than any given number  $\delta > 0$  by making  $M$  larger than a certain number  $M(\delta)$  which depends on  $\delta$  but *not on*  $N$ .

$$(14) \quad \frac{\log(\sum p_i + 1) + 2}{H + \log \log \phi(N)} < \frac{\epsilon}{2}.$$

Then

$$(15) \quad F(K) \leq 2^{2N} (\phi(N))^{-\epsilon 2^{H-1}}$$

which approaches zero uniformly as  $M$  increases.

For most of the switching functions of practical interest  $H$  is much bigger than  $\log \log \phi(N)$ . In these cases the number

$$\frac{N^2 P 2^H}{H - 2 \log H}$$

is larger than (13) by a factor of about  $NP/\log N$ . In the case of two-terminal relay circuits the corresponding factor found by Shannon was only 8. It is not clear whether this difference indicates that there is a wider range of complexity for  $N$ -terminal networks than for two-terminal networks or that our methods for obtaining upper and lower bounds lose some of their effectiveness as  $N$  increases. Nevertheless, (13) is surprisingly large, as we shall see in the example which follows.

*Example.* Consider a telephone central office with 10,000 lines. If the office must be able to connect the lines together in pairs in any arrangement and to remember which line of a pair originated the call, a count of the number of different states which must be produced reveals that the office needs a memory of at least  $H = 64,000$  bits, which can be supplied using 19,200 switches with 10 positions each. The number of other switching functions that one might ask these 19,200 switches to perform is

$$\phi(10,000)^{2^{64,000}} = (10^{30,000})^{10^{19,200}} = 10^{10^{19,200}} \text{ approx.}$$

To apply theorem VIII to these other functions we first note that (14) will be satisfied as long as we pick  $\epsilon$  greater than .006. Then, substituting in (13) and (15) we discover that the chance that one of these switching functions chosen at random can be synthesized with less than about

$$10^{19,000} \text{ contacts}$$

is less than some number of the order of

$$10^{-10^{19,200}}.$$

If the same calculation is repeated for a 10,000-line office which is capable

of handling only 1,250 calls at one time we find

$$H = 22,000 \text{ bits,}$$

$$\# \text{ switches} = 6,666$$

$$\# \text{ functions} = 10^{106.670}$$

$$\epsilon = .015 \text{ or larger}$$

$$\# \text{ contacts} = 10^{6.671} \text{ or more for all but a fraction } 10^{-106.668} \text{ of the functions.}$$

Although these numbers of contacts appear incredible at first sight, there is no reason to expect the number of contacts for almost all switching functions to be a good indication of the number of contacts needed for the switching functions of practical use. This phenomenon has been discussed in detail by Shannon for the case of two-terminal networks. For  $N$ -terminal networks there are at least two other factors which may be mentioned.

Almost all switching functions can assume states which are not typical of the functions encountered in practice. For example, it can be shown that of the  $\phi(N)$  possible ways of distributing  $N$  different things into parcels, almost all of them use a number of parcels which is near  $N/\log_e N$ . Thus, in a typical state of a typical switching function the terminals are connected together in groups which average about  $\log_e N$  terminals per group in size. Telephone switching equipment ordinarily connects terminals together in pairs or in small groups.

A big difference between the design of two-terminal networks and of  $N$ -terminal networks is that, in the former case, one wants to obtain one specific switching function while, in the latter case, one is usually satisfied with a network which can produce certain desired states. There are many switching functions which all produce the same states but for different settings of the switches. For example, if the  $2^H$  desired states are all different, the designer will be content with any one of  $(2^H)!$  different switching functions. We believe that it actually would require something like  $10^{6.671}$  contacts to build a central office if the designer first listed all the desired states at random  $S_1, S_2, \dots, S_{2^H}$  and then required the office to be in state  $S_1$  for switch setting  $(1, 1, \dots, 1)$ , in  $S_2$  for the switch setting  $(1, 1, \dots, 2)$ , etc.

#### *Number of Selector Switch Rotors*

Since our estimate (13) of the number of contacts is the same whether the memory  $H$  is stored in two-position switches or in larger selector switches, one might hope that the selector switch circuits could be built using fewer rotors than the corresponding two-position switch circuits. We believe that this is not true. A typical node in one of the graphs constructed by the process of theorem VI has only about three or four branches connected to

it. This is so regardless of how large  $K$  is. If the rotor of a selector switch is connected to such a node, the chance is great that none of the other branches at the node are operated by this same switching variable. Hence we suspect that a typical switching network requires almost as many rotors as contacts.

## ACKNOWLEDGMENT

The author has discussed this switching problem with B. D. Holbrook, A. W. Horton, J. Riordan, and C. E. Shannon and has received valuable comments and ideas from them. He is also indebted to A. E. Joel and W. Keister for their suggestions for improving the readability of the manuscript.

TABLE I

$x$	$y$	$z$	$f(x, y, z)$	$f_{12}$	$f_{13}$	$f_{23}$
0	0	1	(12) (3)	0	1	1
0	0	2	(123)	0	0	0
0	0	3	(1) (2) (3)	1	1	1
0	0	4	(13) (2)	1	0	1
0	1	1	(13) (2)	1	0	1
0	1	2	(1) (23)	1	1	0
0	1	3	(123)	0	0	0
0	1	4	(1) (23)	1	1	0
1	0	1	(1) (23)	1	1	0
1	0	2	(13) (2)	1	0	1
1	0	3	(1) (2) (3)	1	1	1
1	0	4	(12) (3)	0	1	1
1	1	1	(123)	0	0	0
1	1	2	(123)	0	0	0
1	1	3	(1) (23)	1	1	0
1	1	4	(1) (2) (3)	1	1	1

## APPENDIX

To illustrate how the network synthesis method operates in a typical case consider a three-terminal network using three switches  $x$ ,  $y$ ,  $z$ . Switches  $x$  and  $y$  have two positions 0 and 1, and  $z$  has four positions 1, 2, 3, 4. A three-terminal switching function  $f(x, y, z)$  is defined by means of the first four columns of Table I. The sixteen entries in column four represent the states of the terminals which the network must produce for the corresponding switch settings given in the columns labelled  $x$ ,  $y$ ,  $z$ . In column four, parentheses are used to group terminals which are connected together; for example  $f(1, 0, 4)$  is the state in which terminals 1 and 2 are connected together and 3 is left free.

A network with switching function  $f(x, y, z)$  will be designed by connecting two-terminal networks between the pairs of terminals 1, 2; 2, 3; and 1, 3. The hindrance functions of these three two-terminal networks will be called

$f_{12}$ ,  $f_{23}$ , and  $f_{13}$ . We determine them, as shown in the last three columns of Table I, by setting  $f_{ij} = 0$  whenever terminals  $i$  and  $j$  are to be connected and  $f_{ij} = 1$  otherwise. Our methods of two-terminal network synthesis will produce the  $f_{ij}$  networks.

Our criterion ( $P > H - \log H$ ) for deciding which of the two synthesis methods to use (case 1 or case 2 of theorem II) is not the best rule when  $H$  is as small as it is in this example. By actually trying the different ways of apportioning switches with 2, 2, and 4 positions between a tree and a network to produce all functions, one finds that the most economical way is to

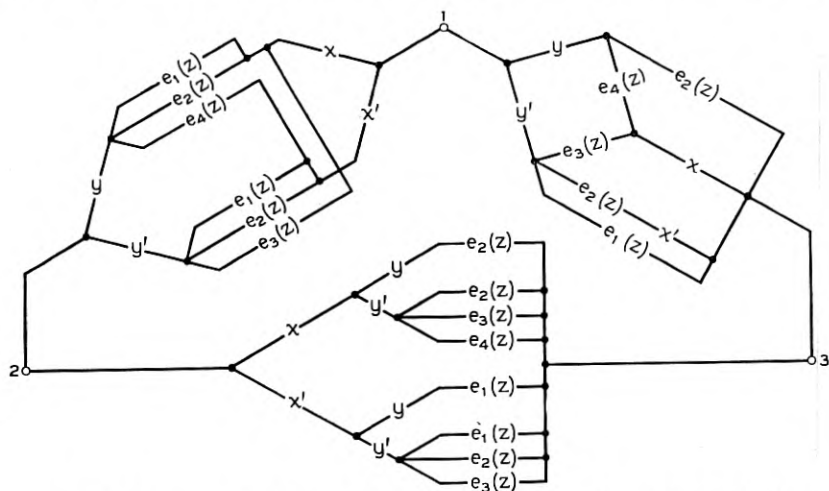


Fig. 8—Network with the 3-terminal switching function  $f(x, y, z)$  of Table I.

put a two-position switch, say  $x$ , into a network which provides all functions of  $x$  (0, 1,  $x$ , and  $x'$ ), and the other switches into a tree. When this procedure is adopted we next express  $f_{ij}$  in the form of identity (2). For example  $f_{13} = [y + e_2(z)][y + e_4(z) + x][y' + e_3(z) + x][y' + e_2(z) + x'][y' + e_1(z)]$ .

The synthesis method described in the text then leads directly to a network for  $f_{13}$  which is shown joining terminals 1 and 3 in Fig. 8. The network for  $f_{12}$  which is shown in Fig. 8 was obtained by the same process. For the sake of illustration the  $f_{23}$  network was found using a tree only.

# Coaxial Impedance Standards

By R. A. KEMPF

(Manuscript Received Mar. 7, 1951)

The calibrations of bridge networks used in developmental tests on coaxial cable are obtained by comparison of the networks with calculable standards of impedance consisting of a group of short-length precision copper coaxial lines. The standards are calculable by reason of the availability of precise formulae relating the distributed primary constants to the measurable physical constants and dimensions of the coaxials. This paper outlines the constructional problems and design features of a group of such standards of impedance which provide a range of values over a broad band of frequencies.

## INTRODUCTION

The "mile of standard cable" was for a long time the basis for rating the transmission qualities of telephonic apparatus and networks.<sup>1, 2</sup> The title of this paper suggests that a return to the old standard has been accomplished. This is true in a restricted sense, but with important differences. The standards here described consist of varying lengths of a rigid coaxial transmission line structure. Their sole function is to supply primary references of resistance, inductance, capacitance and conductance which are numerically comparable to typical unknowns encountered in laboratory cable measurements. Unlike the mile of standard cable, the rigid coaxial is simple structurally, its physical constants and dimensions may be determined accurately, and precise formulae are available for translating these properties into electrical constants at any frequency. It is thus an excellent means for the objective—calculable radio-frequency laboratory standards of R, L, G, and C of the restricted numerical range needed to calibrate the bridge networks used in measurements on the short lengths of cable available to the cable development engineer.

Because developmental cables are not usually available in the longer lengths on which the secondary constants of attenuation, phase, and characteristic impedance may be measured directly, laboratory measurements on a cable sample are usually confined to determination of the four distributed primary parameters or constants. From these the secondary constants may then be calculated.

Measurement of the distributed primary constants of a given line structure is an indirect process, except under limited or restricted circumstances.

<sup>1</sup> R. V. L. Hartley, "The Transmission Unit," *Electrical Communication*, Vol. I, No. 1, July, 1924.

<sup>2</sup> W. L. Everitt, *Communication Engineering*, pp. 101-2.

This is because (1) an impedance bridge can do no more than measure the unknown impedance which may be placed across its terminals; and (2) the line structure can be measured only by making bridge readings at its input or output terminals, from which points the true distributed series properties of the line appear to be altered by the shunt properties, and vice-versa. Statement (2) applies to all observations at the end of a section of transmission line, except when the line is very short electrically.

Assuming that accurate bridge measurements of the impedance at the terminals of a line are available, standard transmission formulae may be used to calculate rigorously the distributed primary constants. Uncertainty as to the accuracy of impedance bridge measurements led to the develop-

TABLE I  
DIMENSIONS AND PHYSICAL PROPERTIES WHICH DETERMINE THE PRIMARY ELECTRICAL CONSTANTS OF ANY COAXIAL TRANSMISSION LINE

Distributed Primary Electrical Constants	Determining Physically Measurable Quantities
Series Constants, $R$ and $L$	
Center Conductor Outer Conductor	$\rho, d, F$ $\rho, ID, t, F$
Shunt Constants, $G$ and $C$	
Capacitance Conductance	$d, ID, \epsilon$ $d, ID, F_p, \epsilon, F$

$R, L, G,$  and  $C$  are distributed resistance, inductance, conductance and capacitance, respectively, at any frequency,  $F$ .

$\rho$  is the dc volume resistivity of the copper conductors which have diameters  $ID$  and  $d$ , and wall thickness  $t$ .

$\epsilon$  and  $F_p$  are the composite dielectric constant and power factor respectively, assumed independent of frequency.

ment of coaxial impedance standards as a means of checking the accuracy of test apparatus. As this work progressed, and the merit of the standards was more fully appreciated, it came about that the bridges were not merely checked against the coaxial standards, but instead the bridge calibrations were derived from the coaxial standards.

#### *Input Impedance of a Coaxial*

The distributed primary constants of any coaxial with a uniform structure may be precisely computed in terms of dimensions and physical constants using formulae which have been developed by Schelkunoff<sup>3</sup> and others. Table I indicates the physically measurable quantities used to compute the respective distributed electrical constants,  $R, L, G,$  and  $C$ .

<sup>3</sup> S. A. Schelkunoff, "Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields," *B.S.T.J.*, Oct. 1934.



It is the open ( $Z_{op}$ ) and short circuit ( $Z_{sh}$ ) input impedances which are of utility for bridge calibration work, however, and except for lines much less than quarter wave in length,  $Z_{op}$  and  $Z_{sh}$  must be computed from the distributed constants using the transmission line equations.

The propagation constant and characteristic impedance may be computed from the distributed constants by means of the equations:

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)}, \text{ and} \quad (1)$$

$$Z_o = \sqrt{\frac{R + j\omega L}{G + j\omega C}}, \quad (2)$$

where the numerical values of the distributed constants are, of course, dependent on the appropriate quantities as in Table I and the length of the line. Further, for any coaxial line terminated in an open circuit or a short circuit, respectively:

$$\frac{1}{Z_{op}} = G' + j\omega C' = \frac{\tanh \gamma}{Z_o}, \text{ and} \quad (3)$$

$$Z_{sh} = R' + j\omega L' = Z_o \tanh \gamma \quad (4)$$

Equations (3) and (4) rigorously relate the input impedances for open and short circuited far-end conditions to  $\gamma$  and  $Z_o$  in (1) and (2), and thus back to the physically measurable quantities of the coaxial structure. Precise values of the apparent distributed primary constants  $R'$ ,  $L'$ ,  $G'$  and  $C'$  are the final objective, as these quantities comprise the standards. As is shown, they are computed from basic data on the dielectric of the coaxials and on a single metal comprising the conductors of the coaxials.

It is of interest to note that, because of the mutual effects of the distributed constants on each other, the conductance component of the input admittance of a coaxial line becomes increasingly a function of the dimensions and resistance of the conductors of the line, as frequency is increased. Thus calculable standards of conductance are obtained which are essentially independent of losses in the insulating material used to support the center conductor.

## COAXIAL STANDARDS FOR LABORATORY USE

### *General Description*

Although short-length coaxials have been used by the Laboratories for some years as impedance standards for cable measurements, refinements in measurement techniques have made it desirable to construct a new set of standards with very uniform components and improved structural qualities.

It is with the new set of impedance standards that this paper is chiefly concerned.

The new standards have been constructed in lengths varying from six inches to thirteen feet in increments of length so that, at a given frequency and far-end condition, eighteen standard impedances are available.

The completed standards have been provided with a permanent storage cabinet, Fig. 1, located in an air-conditioned cable development laboratory adjacent to measurement facilities. The special tools developed for construction, assembly, and use of the standards are also stored in the cabinet together with spare materials for maintenance.

Each standard consists of a seamless hard drawn copper tube  $\frac{3}{8}$ " I. D. as outer conductor and a straight hard drawn copper wire, nominally No. 10 A.W.G., as center conductor. The insulation is expanded polystyrene in the form of spaced cylinders. An aluminum tube is used over the copper tube for mechanical protection but is insulated from it. Stainless steel fittings are provided at each end to exclude dust, to facilitate connection for circulation of dry air, and to provide the short-circuit necessary for  $Z_{sh}$  measurements. The properties, selection, and preparation of the three components—wire, tubing, and insulation and the provision of a repeatable method for short circuiting the coaxials are the basic problems in construction, and are discussed in the following paragraphs.

### *Physical Constants*

The measured physical constants of the copper wire and tubing which are of interest in this application are given in Table II, and those of the expanded polystyrene insulation in Table III. Wherever practicable the absolute accuracies of the measuring instruments were checked against secondary standards of weights and measures, periodically referred to the U. S. Bureau of Standards laboratory for calibration.

### *Dimensions*

The I. D. dimension quoted in Table II is the average of a number of tests on end samples of tubing and was obtained from dimensional and weight relationships as expressed by the equation:

$$D = \sqrt{D_o^2 - \frac{4V}{\pi\ell}}, \quad (5)$$

where  $V$  is the volume of copper in the sample as measured by the displacement technique,  $D_o$  is the measured O. D. of the sample and  $\ell$  is its length.

The I. D. was also determined by direct measurements of O. D. and wall thickness, and agreement obtained with the figure quoted for the above

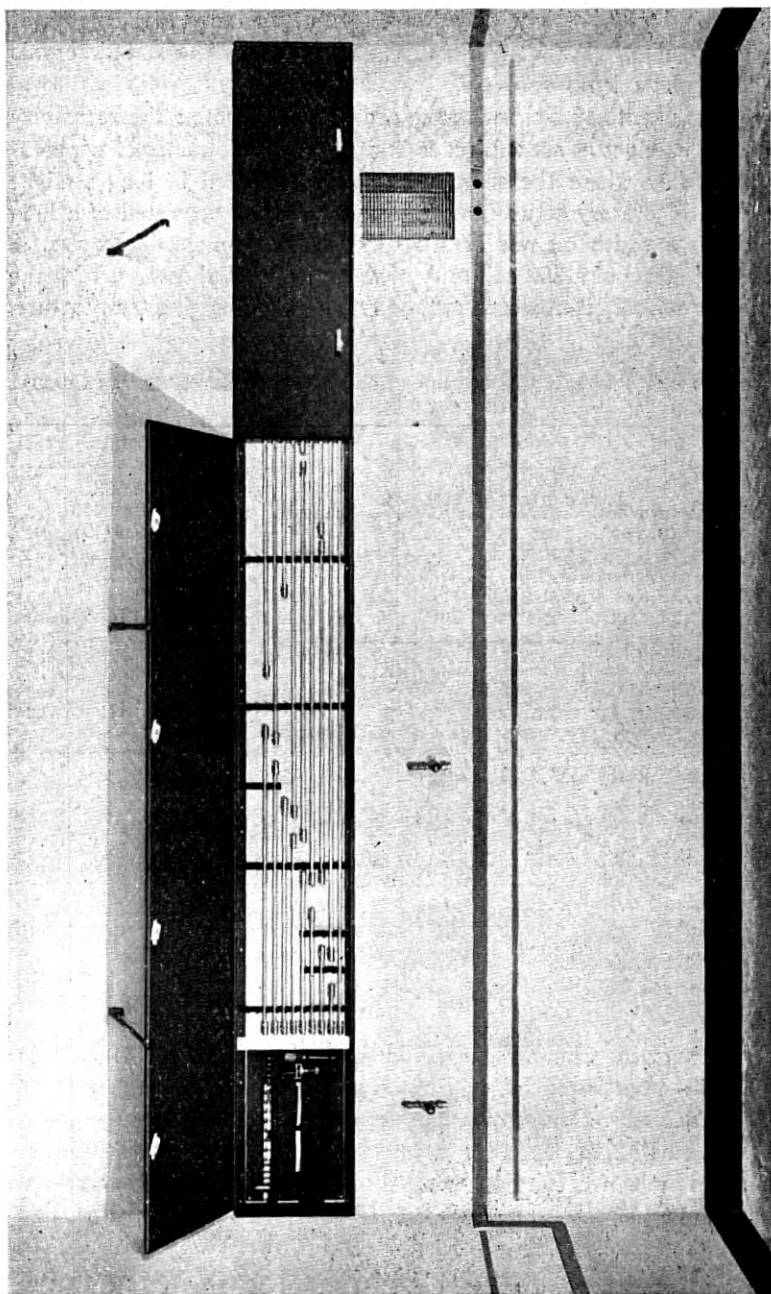


Fig. 1—Coaxial impedance standards and storage cabinet.

method. Variations of weight and dimensions along individual lengths of either wire or tubing are not significant as determined by checks at intervals on a number of the lengths.

The measurements of d-c. resistance needed to determine the resistivity of the copper conductors are subject to limitations of knowledge of temperature, particularly where the samples must be measured in long straight lengths. The laboratory setup used to obtain the data consisted of a long trough through which oil was circulated over the immersed samples. A feature was the use of a 200 gal. tank of water in thermal contact with the oil circuit to maintain it at a constant, or very slowly changing temperature.

TABLE II

MEASURED AVERAGE PHYSICAL CONSTANTS OF COPPER TUBING AND WIRE FOR COAXIAL IMPEDANCE STANDARDS

	Wire	Tubing
Density (gm/cc).....	8.89	8.938
Outside Diameter (inches).....	0.10042	0.50016
Inside Diameter (inches).....	—	0.37137
Wall Thickness (inches).....	—	0.06439
Volume Resistivity (microhm cm).....	1.7480	1.7209
Mass Conductivity (per cent of IACS).....	98.76	99.76
Temperature Coefficient of Resistance, 20°C..	0.003886	0.003938

TABLE III

MEASURED AVERAGE PHYSICAL CONSTANTS OF EXPANDED POLYSTYRENE INSULATING CYLINDERS FOR COAXIAL STANDARDS

Volume Expansion Ratio.....	41
Weight (gm/disc).....	0.1789
Length (inches).....	0.396
Density (gm/cc).....	0.0257
Dielectric Constant.....	1.033

A Kelvin double bridge used for the tests was operated in accordance with minimum-error principles.<sup>4</sup>

### Components-Wire

The center conductor of each coaxial must be drawn straight so that only light spaced support need be used to keep it in axial alignment with the tube. Since available commercial wire drawing machines normally depend on a driven small-diameter capstan to pull the wire through the final reducing die, it was necessary to draw the wire in the laboratory so that straight-out drawing could be achieved. Commercial machines were, however, used to reduce the supply from  $\frac{3}{8}$ " rod to 0.110" dia. wire without

<sup>4</sup> Electrical Measurements, Laws, McGraw-Hill, 1917.

intermediate annealing. The laboratory operation consisted of drawing the 0.110" stock straight out through a 0.1004" I. D. diamond die pre-selected as to circularity and finish of the bore.

The copper which was used is known in the trade as "electrolytic tough pitch" chemically composed of 99.95 per cent copper, 0.02 per cent oxygen and 0.03 per cent divided between six other minor contaminants.

There has always been some concern as to the possibility that a wire drawn from annealed stock might well have a thin full hard shell over a relatively annealed core. The d-c. resistivity measurements would then determine a weighted average resistivity, instead of the surface resistivity needed for calculations. To settle this point, tests were made on full-hard wire, semi-hard wire as used in commercial coaxials, fully annealed wire, and on annealed wire plated with a 0.2 mil layer of silver. The tests consisted of (a) comparing the a-c. resistance of the wires by precise methods at 1 mc where transmission is at a skin depth of 2 mils, and (b) microscopic study of grain structure of polished-etched cross sections of the samples at magnifications to 2000 diameters. The conclusion from both studies was that no thin skin exists. The wire is therefore treated as homogeneous throughout its cross section in computation of d-c. resistivity and in computation of a-c. resistance.

### *Copper Tubing*

The effects on the resistance of a coaxial traceable to the physical constants of the outer conductor are scaled down about 5:1 so that the requirements on the tubing are not so severe as on the wire. However, stock copper tubing from a distributor's warehouse cannot normally be used. Such tubing may have unacceptable inside surface roughness, ellipticity of bore, and a high and variable resistivity. Roughness and ellipticity are the result of worn plug dies frequently used in the drawing of commercially acceptable tubing, or omission of the plug die in drawing the tubing to final diameter. Most stock tubing contains phosphorous and, even though the percentage of phosphorous may be very small, the effect in increasing the resistivity is marked. The full-hard tubing used in the coaxial standards here described was procured directly from a mill, and was largely drawn in consecutive lengths from a single casting of oxygen-free electrolytic copper using selected dies.

### *Insulation*

Expanded polystyrene is the dielectric material used in the standards and its applicable properties are shown in Table III. In solid form it has a dielectric constant,  $\epsilon'$ , about 3% greater than that of air and, when used

to insulate the coaxial standards in the form of spaced cylinders, the composite dielectric constant is increased only about 0.4% over theoretical for air. Errors in determining the constant of the expanded material are thus scaled down by a factor of nearly 10:1 in their effect on the composite constant. The figure quoted for dielectric constant in Table III was obtained by adding incremental amounts of dielectric to a 12" length of standard coaxial and plotting the measured capacitance as increments above the computed capacitance for air dielectric, as in Fig. 2.

The distributed conductance,  $G$ , is derived from the power factor of the dielectric which, in the case of any reasonably good material, is so small

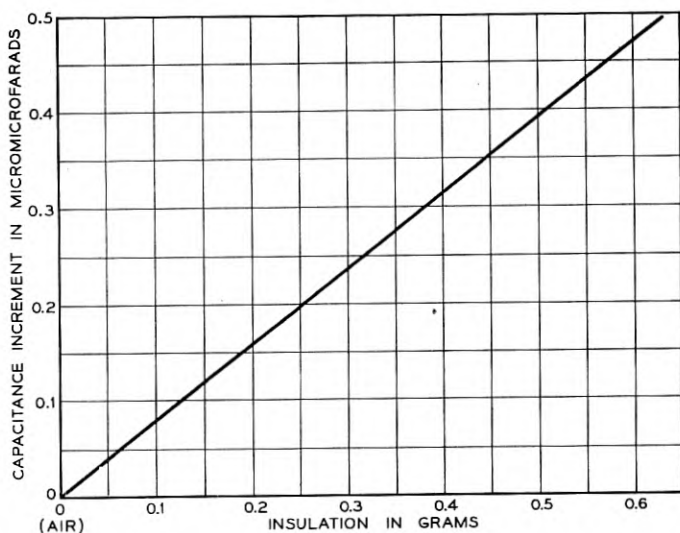


Fig. 2—Data for determination of dielectric constant of expanded polystyrene.

that experimental determination is subject to large error. The apparent conductance  $G'$  from (4) is the value at the input terminals of the coaxial. An important feature of the standards is that  $G'$  becomes independent of  $G$  at high frequencies and, therefore, it is desirable to reduce  $G$  (representative of the loss in the dielectric) to as small a value as possible in order that  $G'$  may become independent at the lowest possible frequency. This is accomplished by the use of expanded polystyrene as the dielectric of the standards.

Polystyrene is of such molecular structure that it is not hygroscopic. However, under certain conditions, water vapor may condense in cells of the expanded material in sufficient amount to increase the dielectric con-

stant and the conductance losses. For this reason all insulating cylinders have been pierced longitudinally so that low pressure dry air may be circulated through the assembled coaxials when they are in use.

The cylinders were cut with a high-speed fly cutter, and the center hole drilled in the same operation. It was determined by sensitive electrical tests that centering precision in the assembled coaxials was equivalent to that obtained in coaxials with lathe turned discs of solid materials.

Cylinder spacing of 3" on centers was determined as about the maximum permissible to prevent detectable sag in the center wire between points of support. There is no specific strength requirement on the cylinders except that they must support the weight of the straight drawn wire. However, a single polyethylene disc is used at the test end to resist radial thrusts which may occur in making test connections.

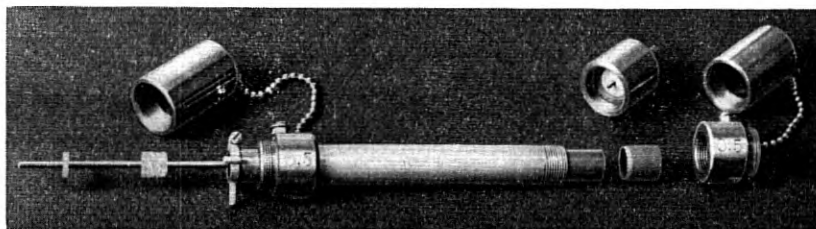


Fig. 3—Partially assembled coaxial standard.

### *Assembly Features*

It is necessary to have mechanical protection for the copper tubes of the standards, and this is provided for each by an aluminum tube slipped over the copper tube, with insulation consisting of a helical wrap of  $0.0015 \times \frac{1}{2}$ " paper ribbon first applied to the copper tube. The aluminum tube is locked on the copper tube by short lengths of fibre tubing wedged between at each end. The wedging action occurs when the end fittings are screwed to the aluminum tube. Figure 3 is a photograph of a partially assembled standard to illustrate further the construction. The presence of the aluminum tube may be disregarded in so far as its electrical effects are concerned because of the self-shielding qualities of the copper tube.

### *End Effects*

Experimental data indicate that if a coaxial tube is longer than its center conductor and the wire is then lengthened incrementally until it is as long as the tube, the capacitance of the coaxial remains directly proportional to



the length of the wire to better than  $0.01 \mu\text{mf}$ . The "fringing effect" at an open-end of a coaxial is considered negligible.

The individual center wires of the standards have been cut  $0.500''$  longer than the tubes and the extra length utilized at the test end for connection to the test equipment. In computing the impedances, the length of the tube has been used. The  $0.500''$  center wire projection is considered as part of the test-equipment leads and is separately accountable.

#### *Disc Short Circuiting of Coaxial Standards*

A shortcoming of coaxial standards previously developed has been in the use of solder to attach a short circuiting disc at the far end when using the length to provide calculable series impedance ( $Z_{sh}$ ). The use of a disc is the best means to short-circuit the inner and outer conductors with a minimum and calculable terminal impedance, and the accomplishment of this objective by repeatable mechanical means is an important feature of the new standards.

A very thin disc pressed against the end of a coaxial is effectively contacted along two concentric rings representing the peripheral edges of wire and tube of diameters  $d$  and  $D$  respectively. The d-c. resistance of the metallic area between the rings in terms of its resistance,  $r$ , in ohms per square is:

$$R = \frac{r}{2\pi} \int_{d/2}^{D/2} \frac{dx}{x} \quad (6)$$

(6) reduces to:

$$R = \frac{r}{2\pi} \ln D/d. \quad (7)$$

Figure 4 shows a cross-section of the short-circuiting device developed for the standards. Figure 3 includes a view of the assembled device. It consists of a stainless steel housing carrying a trapped silver disc and other parts to effect a repeatable, minimum-impedance short circuit when the housing is screwed to the end fitting of the coaxial. The operation is as follows:

- (a) When the housing is screwed on the end fitting, the disc is forced against the end of the copper tube, with pressure equalized by and derived from the rubber grommet.
- (b) Assuming the housing is always screwed on the end fitting as far as it will go, the total pressure on the disc is regulated on initial assembly by control of the amount of projection of the tube beyond the end fitting.



- (c) The center wire is drilled and tapped to a depth of  $\frac{5}{16}$ " at one end to accept the 0-80 screw carried by the stud. After the disc is tightly pressed against the tube as in (a), the center wire screw stud is turned until the center of the disc is drawn tightly against the end of the center conductor. The stud projects from the end of the housing and is milled at its end to accept a torque wrench. The wrench provides the means for development of repeatable pressure of the end of the wire against the center of the disc.

Experimental data on the relative pressure versus d-c. resistance characteristics are illustrated by the curves of Fig. 5. Disc pressure on the center wire is controlling as would be expected. The total d-c. resistance in-

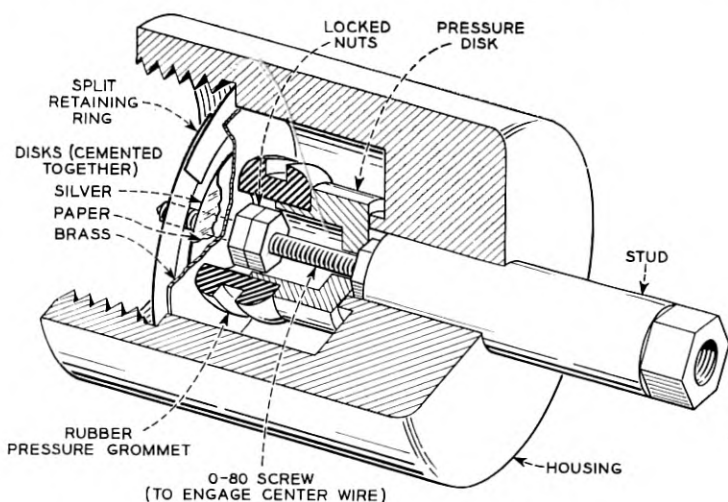


Fig. 4—Assembly detail, end-shortening device.

roduced by the short circuit is of the order of 0.1 milliohm, which includes the contact resistance. On the basis of experimental observation the resistance at radio frequencies appears to be more stable because the area of contact becomes of less importance as frequency is increased. The inductance introduced by the disc is negligible for all lengths and frequencies. The a-c. resistance of the disc is equal to its d-c. resistance up to about 1 mc where complete wave penetration ceases and skin effect becomes apparent. Above 2 mc the disc resistance increases as the square root of frequency.

#### Connection—Test End

A connector currently in use at the test end of the outer conductor is shown in position in the photograph, Fig. 3. The projection normal to the

axis of the assembly is used for clamp-type connection. A guillotine clamp for short-circuiting the input of the coaxial, as required in substitution type measurements, is placed across the center wire and the projection parallel to the axis of the assembly. The center conductor of each coaxial extends 0.500" beyond the tube, and is thus available for clamp type connections. Soldered connections are thus completely eliminated in the use of the standards.

### Eccentricity

Concentricity of the axes of wire and tube is assumed in most published formulae which relate the dimensions of a coaxial structure to its electrical

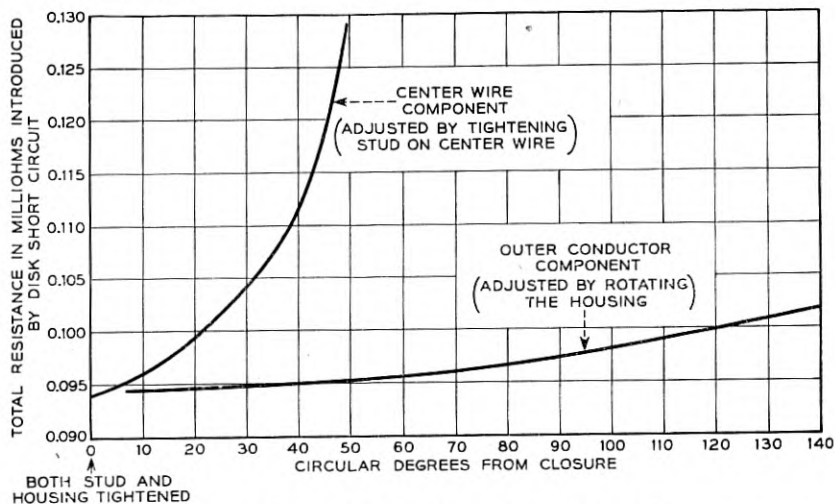


Fig. 5—Measured d-c resistance introduced by end shorting device as a function of adjustment.

constants. Consistent departures from concentricity, i.e. eccentricity, are subject to analysis as regards effects on the electrical constants. In the case of the standard coaxials, the use of straight, hard-drawn wire and tubing combined with close-spaced support of the wire are all factors which reduce eccentricity. However, it is obviously desirable to check the degree of residual eccentricity of the assembled coaxials point by point along each length. A method developed to do this makes use of Biot and Savart's law.<sup>5</sup>

The external field, at any point P, of a long circular wire or tube carrying a current I is given by

$$H = \frac{2I}{r} \quad (8)$$

<sup>5</sup> W. R. Smythe, *Static and Dynamic Electricity*, p. 272.

where  $r$  is the distance from the axis of the wire or tube to point  $P$ . This assumes that the current density on any concentric circle is uniform.

If the wire and tube are carrying the same current  $I$  but in opposite directions and if the axes of the wire and tube coincide, there will be no magnetic field at any external point. Therefore an alternating current in the conductors will not induce a voltage in a pick-up coil placed external to the coaxial. If, however, the two axes do not coincide a voltage will be induced in the pick-up coil and will be a maximum when the pick-up coil is in the

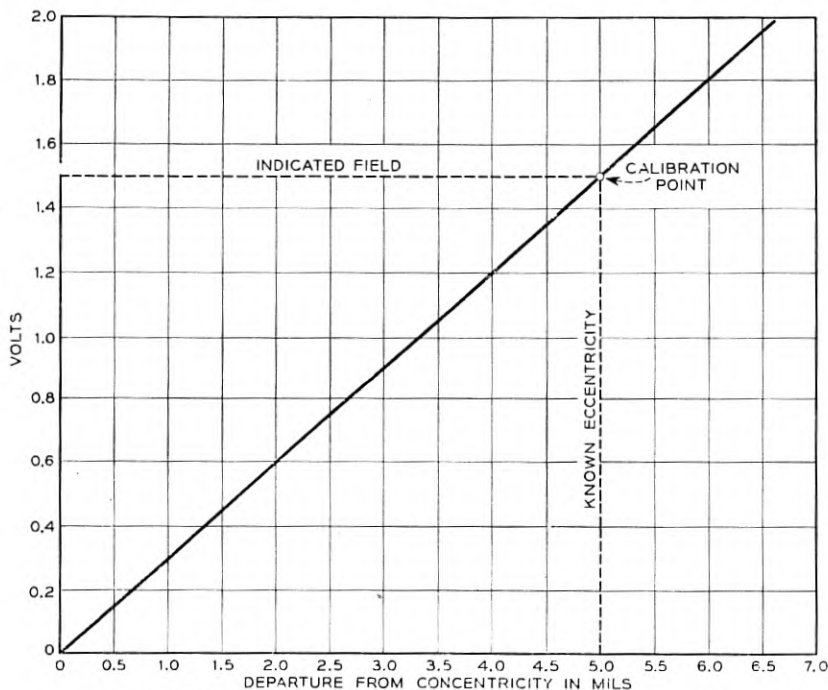


Fig. 6—Typical calibration curve in measurements of eccentricity of coaxial standards.

plane of the two axes. The magnitude of this maximum voltage will be proportional to the distance between the two axes if the measuring distance is large compared to this separation distance.

In practical use, the maximum field at  $P$  was measured in terms of volts relative to a known eccentricity obtained by insulating a coaxial with discs of the known eccentricity. Advantage was taken of the linear relationship of eccentricity and detectable field so that, with the distance to the tube maintained constant, the calibration curve of Fig. 6 was used to evaluate all standard coaxials after final assembly. The sensitivity was such that the

effect of an 0.08 mil known deviation in diameter of the wire was detectable in coaxials with otherwise perfect symmetry. Such deviations in symmetry as were observed developed primarily from variations in wire diameter as already stated, and from deviations in wall thickness of the copper tubing.

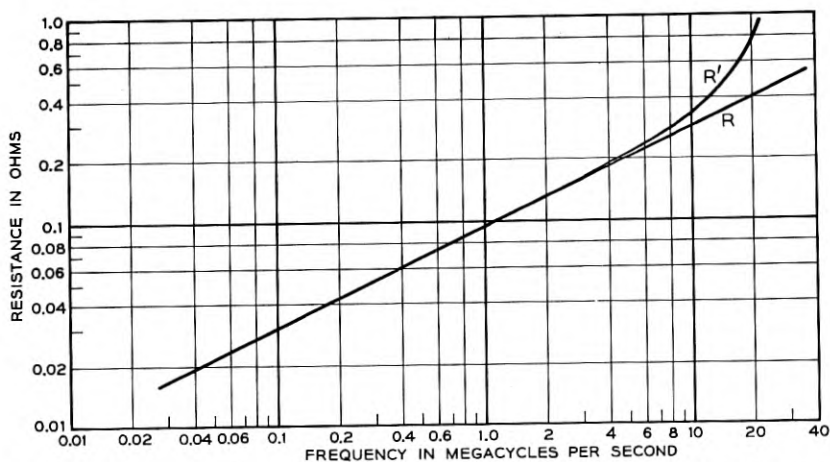


Fig. 7—Distributed resistance,  $R$ , and  $R'$  component of  $Z_{sh}$  of 7.0 ft. length, coaxial impedance standard.

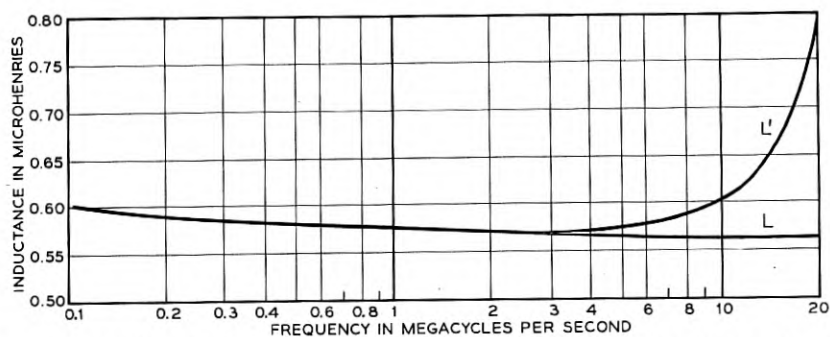


Fig. 8—Distributed inductance,  $L$ , and  $L'$  component of  $Z_{sh}$  of 7.0 ft. length, coaxial impedance standard.

### Numerical Values of Standard Impedances

#### Graphical Example

As an example, values for the apparent series and shunt primary constants  $R'$ ,  $L'$ ,  $G'$  and  $C'$  for a length of 7.0 feet are presented graphically in Figs. 7-10. For comparison, the distributed primary constants have also

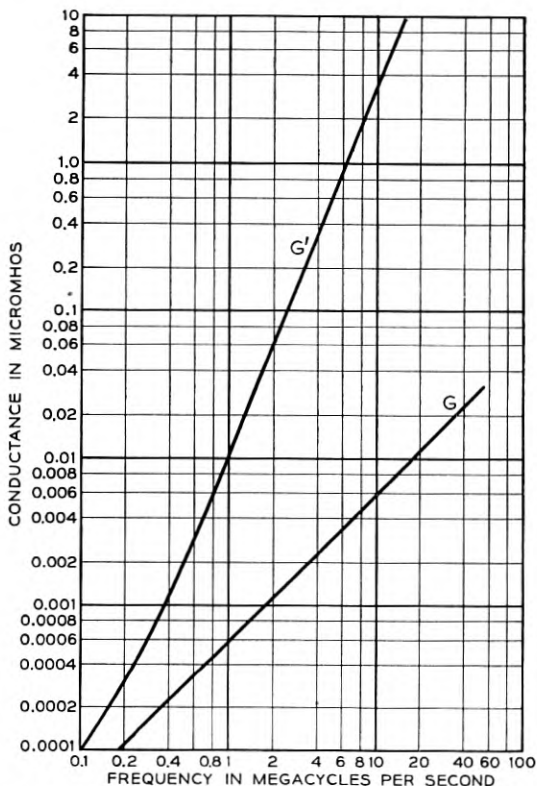


Fig. 9—Distributed conductance,  $G$ , and  $G'$  component of  $Z_{op}$  of 7.0 ft. length, coaxial impedance standard.

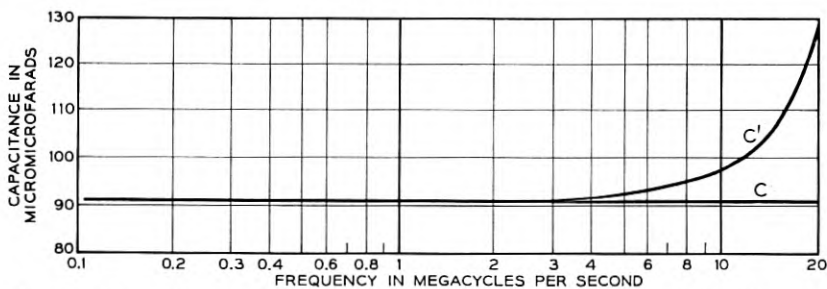


Fig. 10—Distributed capacitance,  $C$ , and  $C'$  component of  $Z_{op}$  of 7.0 ft. length coaxial impedance standard.

been plotted. It should again be emphasized that it is the "primes" which are available at the terminals of each standard for calibration purposes.

The plotted values of the primes were computed for a frequency range

of 50 kc to 20 mc using equations (3) and (4). From an error standpoint it is not practicable to use lengths much longer than 7 ft. as standards at 20 mc. However, shorter lengths of which there are twelve may be used to provide standards at higher frequencies. There are six longer lengths, and the entire group of eighteen may be used to provide eighteen numerical values of each "prime" in the frequency range of 50 kc to 10 mc. There is the restriction, from the standpoint of application, that the impedance components are available only in "pairs"; each  $R'$  has a definite  $L'$  associated with it, or each  $G'$  has a value of  $C'$  in association.

### Error Considerations

Examination of Figs. 7-10 shows that the "primes" as computed from (3) and (4) exceed the distributed constants by a gradually increasing

TABLE IV  
ESTIMATED ERRORS IN COMPUTED INPUT IMPEDANCE COMPONENTS OF 7.0 FT. LENGTH COAXIAL STANDARD

Input Impedance Component	Per Cent Error ( $\pm$ )	
	1 mc	20 mc
$R'$	0.05	0.1
$L'$	0.04	0.1
$G'^*$	2.	0.3
$C'$	0.06	0.1

\* Assumes possible error of  $\pm 25\%$  in knowledge of power factor of the dielectric material.

amount as frequency is increased. Except for  $G'$  the excess is largely proportional to the product  $\omega^2 LC \ell^2$ , where  $\ell$  is the length of the line, and  $L$  and  $C$  are in terms of unit length. The total error in a "prime" at a given frequency is then approximately the error in the distributed value combined with the proportioned error in  $\omega^2 LC \ell^2$ . Table IV shows the computed errors for the 7.0 ft. standard at 1 mc where the contribution of  $\omega^2 LC \ell^2$  is small and at 20 mc where it is relatively large. The errors quoted were computed from estimated errors involved in determination of the various physical quantities associated with the constants of Tables II and III.<sup>6</sup>

$G'$  was mentioned above as an exception. This results from the fact that  $G'$  is largely proportional to  $\omega^2 C^2 R \ell^3$  above 1 mc. That is, the excess of  $G'$  over  $G$  rapidly increases with frequency so that the value of  $G$  may be neglected above 1 mc. The precision of  $G'$  is that of the determination of  $\omega^2 C^2 R \ell^3$ , a quantity which can be determined with very good precision as

<sup>6</sup> "Electrical Measurements, and the Calculation of the Errors Involved," D. Karo, Macdonald & Co., London.

compared to determination of the power factor of the dielectric of the standard which is controlling in determination of  $G$ . Table IV shows the effect of these circumstances in the case of the 7.0 ft. length. The error in  $G'$  is computed as less than 2% at 1 mc. for 25% error in power factor determination. At 20 mc,  $G$  is less than one thousandth of  $G'$  and, as indicated in the table,  $G'$  has an error of 0.3% due entirely to errors in  $\omega^2 C^2 R l^3$ .

#### ACKNOWLEDGEMENT

The writer is greatly indebted to Mr. M. C. Biskeborn who contributed much valuable advice during the development of the standards.

# Instantaneous Compandors

By C. O. MALLINCKRODT

(*Manuscript Received Apr. 9, 1951*)

Instantaneous compandors have found useful application in time-division systems. This paper discusses the theory of the instantaneous compandor and evaluates the noise advantage when instantaneous companding is applied to telephone channels. The noise advantage depends upon the noise standard of the system. If the standard corresponds to that of a toll telephone system, a noise advantage of about 20 db is possible.

## INTRODUCTION

A compandor is characterized by compression followed by expansion. To achieve noise reduction by compandor action,\* compression is applied before and expansion after the noise exposure. By compression one means that the effective gain which is applied to a signal varies as a function of the magnitude of the signal, the effective gain being greater for small than for large signals. In the process of expansion the effective gain also varies as a function of the signal but is greater for large than for small signals.

There are two general classes<sup>1</sup> of compandors, "syllabic" and "instantaneous." For many years, because of theoretical and practical reasons, only the syllabic type was used to any appreciable extent. Although utilized primarily in special situations<sup>2-5</sup>, syllabic compandors have in these instances served to improve telephone operation by providing a substantial noise advantage. More recently<sup>6-11</sup> the instantaneous type also has begun to find important applications to time-division systems. Since an instantaneous compandor produces effective gain variations in response to instantaneous values of the signal wave, the instantaneous type is well adapted to pulse systems. Moreover, in time-division pulse-modulation systems, one instantaneous compandor usually serves a plurality of channels thereby affording additional economies.

## THEORY

Noise advantage due to compandor action arises primarily because it is the weak signals that are most susceptible to degradation by noise or other unwanted interference. Accordingly, weak signals are highly amplified by the compressor and are carried at a relatively high level through the noise

\* For all numbered references, see list at end of paper.



exposure. Stronger signals are amplified less highly. Loss, therefore, is removed from the expander as the signal increases and the noise increases correspondingly.

When the signals are conventional speech signals, loss is removed from the expander as the speech volume increases and consequently the noise volume increases correspondingly. An instantaneous compandor has the important advantage that level adjustments are frequent, for example, in a pulse-modulation system at the rate of about 8000 times per second for a message channel whose bandwidth approaches 4000 cycles. Consequently, the increased noise will be continuously masked by increased speech sound. During all silent periods, unwanted noise and interference receive maximum noise suppression in the expander. For an ordinary message channel these advantages are substantial.

Viewed broadly, an instantaneous compandor provides a ready means for making the noise susceptibility a function of the magnitude of the signal. If the noise susceptibility is made less than that of a linear system in one portion of the range, then it must be greater than that of a linear system in some other portion of the range. Whether an over-all improvement results depends entirely upon the nature of the signal. For example, in certain types of picture transmission systems a given value of noise produces about as much harm whether the signal be weak or strong. In this instance no benefit would accrue from making the noise susceptibility a function of signal strength.

An important consideration, therefore, is the evaluation of the noise advantage due to instantaneous companding. The theoretical treatment will give relationships for signal-to-noise ratio and noise susceptibility. Application of the theory to a particular example including a numerical evaluation of the noise advantage will be deferred to the last section.

### *Method of Analysis*

The analysis is based upon deductions<sup>12</sup> related to the sampling principle and is illustrated by Fig. 1 which shows a schematic of one channel of a multi-channel time-division system.

The incoming signal (Fig. 1) is filtered by a low-pass filter designated  $F_1$ . At the output of  $F_1$  the signal should be regarded as an arbitrary signal occupying the band of all frequencies slightly less than  $B$ . Brief samples of the signal are taken uniformly at the rate of  $2B$  samples per second. In this manner the signal is converted into a series of PAM (pulse amplitude modulated) pulses as indicated in Fig. 2. There is a unique relationship<sup>12</sup> between signals and samples (PAM pulses); if we are given the signal wave we can

determine the samples; if we are given the samples we can determine the signal wave.

In Fig. 1, if we ignore the compressor, the samples are immediately filtered by another low-pass filter designated  $F_2$  which temporarily will be assumed to be similar to  $F_1$ . If  $F_2$  attenuates all frequencies higher than  $B$  and if each filter includes accurate in-band equalization including correction for phase distortion, then the wave at the output of  $F_2$  except for delay will be an attenuated replica<sup>12</sup> of the wave at the output of  $F_1$ .

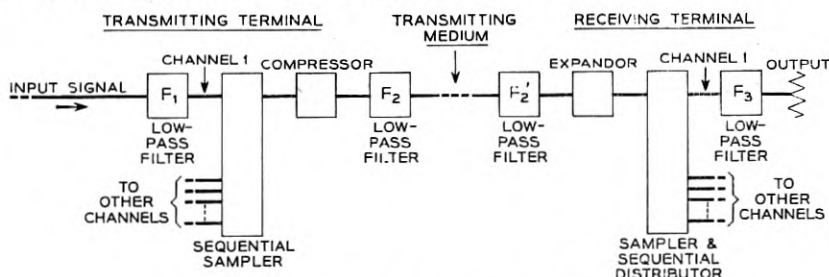


Fig. 1—Block schematic of multi-channel PAM system.

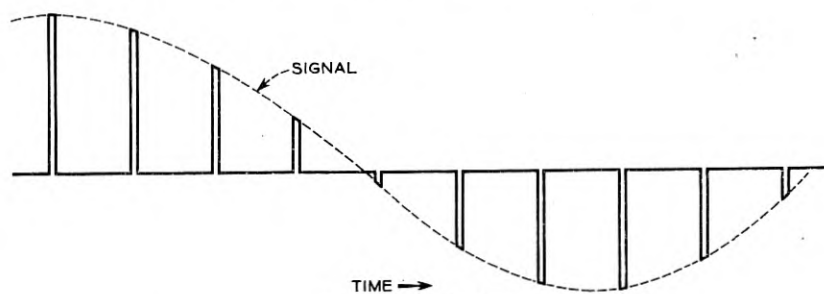


Fig. 2—PAM pulses.

If the cut-off frequency of  $F_2$  is raised sufficiently, then in the output of this filter one finds PAM pulses clearly separated in time. These pulses are samples, on an enlarged scale, of the signal that would have existed had the cut-off frequency been in the neighborhood of  $B$ .

If now the compressor (Fig. 1) is taken into account, then the PAM pulses at the output of the sequential sampler will be impressed upon the input of the compressor. The general form of a compressor characteristic is indicated in Fig. 3. The compressor is essentially instantaneous if its bandwidth is wide enough so that it can effect the required change in the magnitude of each pulse without increasing its duration. It is also significant to note that, theoretically, no more bandwidth<sup>12</sup> is needed to transmit the samples

after they have been compressed than before. Clearly, if the samples are compressed in accordance with an arbitrary but known law, and if the receiver expands them by an exactly inverse operation, the wanted information can be recovered.

Pulses from the sending end are fed to the transmitting medium (Fig. 1) and conveyed to the receiving terminal. This might be done by any of a number of different ways and the details of this portion of the system are not important to this discussion. What is important is that the analysis will assume that the signal at the input to the receiving end, except for noise accumulated along the way, is an exact but delayed copy of the signal leaving the transmitting end. The low-pass filter  $F_2'$  (Fig. 1) is similar to  $F_2$  and has been inserted to reject unwanted high-frequency noise.

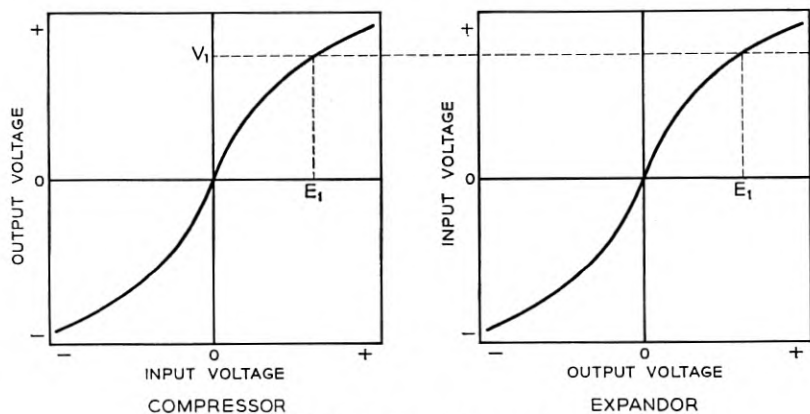


Fig. 3—Instantaneous compandor.

The filtered PAM pulses go to the input of an instantaneous wide-band expander whose characteristic is the inverse of that of the compressor. By interchanging the designations "input" and "output" on the compressor characteristic (Fig. 3) the characteristic becomes that of an expander. The combination of compressor and expander makes the over-all system linear as illustrated by the dotted lines in Fig. 3. The pulses from the expander go to the sampler which is accurately synchronized<sup>6, 7, 10, 11</sup>. Channel 1 pulses from the sampler go to  $F_3$ , a low-pass filter similar to  $F_1$ , and produce in the output of  $F_3$  a copy of the original signal<sup>12</sup> together with noise accumulated for the most part in the transmission medium.

#### Signal-to-Noise Ratio

To understand how the compandor affects the signal-to-noise ratio of the system, consider a single operation at the receiver. The magnitude of a

particular pulse may be represented as  $V_1 + v_1$ , where  $V_1$  corresponds to the signal voltage and  $v_1$  to the noise voltage. This pulse is impressed upon the input of a wide-band expander (Fig. 4) and produces a new pulse at the output of the expander. It will be assumed throughout that the maximum values of expander input and output voltages are equal, and for convenience will be taken as unity. The solid curve is the positive portion of an assumed expander characteristic. Since  $V_1$  represents the magnitude that the input pulse would have if the noise voltage were zero,  $E_1$  represents the corresponding magnitude of the output pulse. When the effect of noise is taken into account, the magnitude of the output pulse is  $E_1 + \Delta E_1$ . This goes by way of the sampler and distributor to the input of  $F_3$ .

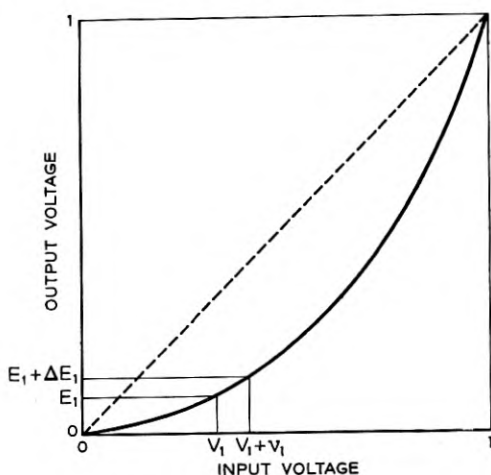


Fig. 4—Instantaneous expander.

From the sampling principle one deduces<sup>12</sup> that each pulse which appears at the input to  $F_3$  is directly proportional to the signal which occurs  $\Delta t$  seconds later at the output, where  $\Delta t$  is the delay in  $F_3$ . This delay will be neglected. Thus, in response to  $E_1 + \Delta E_1$  at the input, the value of the voltage at the output of  $F_3$  is  $k(E_1 + \Delta E_1)$  where  $k$  depends upon the design details of the system. Represent instantaneous signal and noise voltages at the output of  $F_3$  by  $S_1$  and  $N_1$ . Then,

$$S_1 = kE_1 \quad (1)$$

and

$$N_1 = kv_1 \frac{\Delta E_1}{v_1} \quad (2)$$

It is apparent (Fig. 4) that  $\frac{\Delta E_1}{v_1}$  is a function of the slope of the expander characteristic. Represented by  $s_1$ , this ratio is an important quantity and will be referred to as the "noise susceptibility of the system." Dividing (1) by (2) and dropping subscripts,

$$\frac{S}{N} = \frac{E}{vs} \quad (3)$$

where  $S/N$  is the ratio of instantaneous signal to instantaneous noise at the output of  $F_3$  and  $\frac{E}{v}$  the corresponding ratio without a compander.

### Noise Susceptibility

If it is assumed that the maximum signal is large compared to the noise, then

$$s = \frac{dE}{dV} \quad (4)$$

Because the characteristic of the expander is nonlinear, the noise susceptibility,  $s$ , varies as a function of signal input. When  $s$  is unity the noise susceptibility equals that of a linear system. The object of instantaneous companding is to make  $s$  a predetermined function of the magnitude of the signal. However, the predetermined choice is not entirely arbitrary. To avoid ambiguous signals at the receiver, as the input to the compressor varies continuously from zero to unity, the input-output characteristic must be single valued.

One notes that if  $s$  is averaged with respect to the expander input voltage, the average value is unity regardless of the shape of the characteristic. Similarly, if  $\frac{1}{s}$  is averaged with respect to expander output voltage, the value obtained is always unity.

### Important Difference Between Syllabic and Instantaneous Types of Companders

At this point it seems advisable to emphasize an important difference between syllabic and instantaneous types of companders. Signals compressed on a syllabic basis can be transmitted in a band not significantly different from that occupied by the original signal. Moreover, the requirements on the phase and attenuation-frequency characteristics of the path between compressor and expander are about the same as if the signal were not compressed. Accordingly, syllabic companders can and have been applied to a wide variety of existing types of systems.

This is not true of the instantaneous compandor. While instantaneous companding theoretically does not require an increase in bandwidth<sup>12</sup> between compressor and expander, additional transmission requirements<sup>13</sup> which this path must satisfy usually would be regarded as very severe when this band is no more than the bandwidth of the signal entering the compressor.\* This means, for example, that for practical reasons instantaneous compandors cannot be applied to existing types of single-sideband carrier telephone systems. On the other hand, if a pulse modulation or other type of multi-channel time-division system is capable of operating without compandors, the addition of instantaneous compandors will not alter matters. Of course, the net over-all transmission will change more than one db for each db change in the propagation from compressor output to expander input, but this is true for either type of compandor and depends essentially upon the properties of the expander.

#### APPLICATION OF THEORY

The theory will be used to evaluate the noise advantage of an instantaneous compandor in the PAM system shown in Fig. 1 when the signal is speech. The result is applicable<sup>10</sup> to other types of multi-channel pulse-modulation systems.

##### *Choice of Expander Characteristic*

The first step is to select a suitable characteristic. If the characteristic of the compandor is logarithmic, the signal-to-noise ratio at the output of the system will be independent of speech volume. It appears reasonable for talkers of different volume to be treated alike and so a logarithmic characteristic will be chosen. Results of experimental observations on this type of characteristic will be discussed in a later paragraph.

A logarithmic compandor is one in which the output voltage of the compressor is a logarithmic function of its input voltage. Conversely, the output voltage of the expander is an exponential function of its input voltage. This relationship may be written:

$$E = ae^{bv}$$

where  $a$  and  $b$  are arbitrary constants,  $E$  is the expander output voltage, and  $V$  the expander input voltage.

\* This implies suitable instrumentation which, for convenience, may utilize sampling. If a signal be compressed by an instantaneous compressor, the bandwidth occupied by the compressed signal obviously will be considerably increased. However, the information content of the compressed signal is no more than before and accordingly may be represented by another appropriate information signal whose frequency range is restricted to the bandwidth of the signal entering the compressor.

It is apparent that the characteristic cannot follow the exponential law at low values of input voltage, because, if the relationship is exponential,  $E$  is not zero when  $V$  is zero. This difficulty is avoided by using a characteristic which is linear for input voltages below a given value and exponential for input voltages above this value. A characteristic of this type is illustrated in Fig. 5. The point at which the characteristic changes from a linear to an

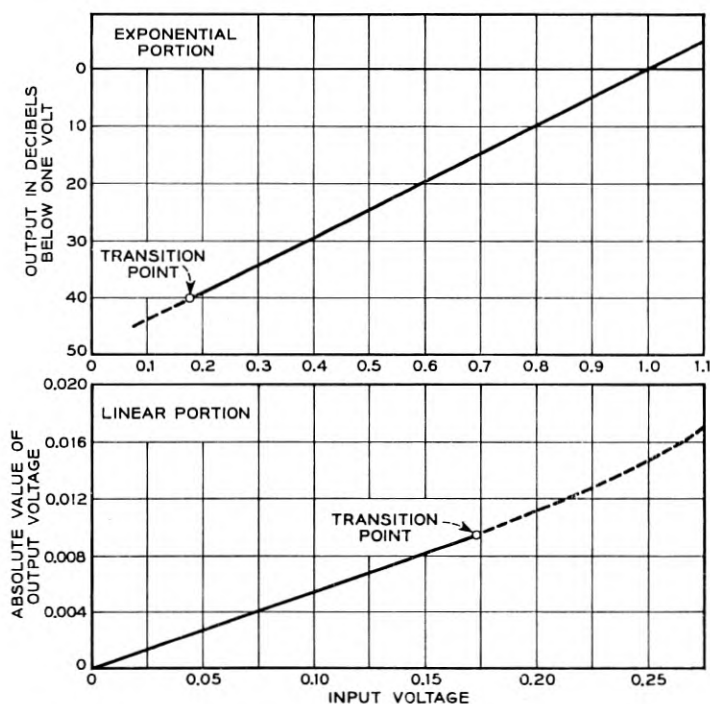


Fig. 5—Expander characteristic.

exponential relationship is referred to as the “transition point.” The transition from one function to the other occurs smoothly and the first derivative of the output with respect to the input is continuous at the transition point.

#### *Logarithmic Compandor*

Since the characteristic of the compandor is an odd function, all formulas will be limited to the positive portion. The exponential portion of the expander characteristic is given by

$$E = e^{(V-1)/V_1} \quad (5)$$

provided  $E = 1$  when  $V = 1$  and provided  $\frac{dE}{dV} = \frac{E_t}{V_t}$  at the transition point where the expander voltages are designated  $V_t$  and  $E_t$ . From (5) we write

$$E_t = e^{(V_t-1)/V_t}. \quad (6)$$

Let "expansion ratio" be defined by the ratio of  $E_m/E_t$  to  $V_m/V_t$  where  $E_m$  and  $V_m$  are the maximum values of the expander output and input voltages. Recall that  $E_m = V_m = 1$ , set  $K$  equal to the expansion ratio and write

$$K = \frac{V_t}{E_t}. \quad (7)$$

The expansion ratio may be represented as a function of  $V_t$  by replacing  $E_t$  in (7) by its value in (6), viz.,

$$K = V_t e^{(1-V_t)/V_t}. \quad (8)$$

Expressed in decibels,

$$K(\text{in db}) = 20 \log_{10} (V_t e^{(1-V_t)/V_t}). \quad (9)$$

When the value of  $K$  given by (9) refers to the compressor, it is called "compression ratio." This follows from the identical (Fig. 3) compressor and expander characteristics after input and output designations are interchanged.

The manner in which  $K$  and  $E_t$  are related to  $V_t$  is given by (8) and (6) respectively. These relationships are plotted in Fig. 6. Clearly, if any one of the three parameters is fixed, the entire expander characteristic is known.

#### Signal-to-Noise Ratio

Let  $s_2$  represent the noise susceptibility of the system during intervals when the magnitude of the signal voltage is within the exponential range of the expander. By differentiating (5) with respect to  $V$  and using (4) we get

$$s_2 = \frac{1}{V_t} e^{(V-1)/V_t} \quad (10)$$

which relates noise susceptibility to expander input voltage,  $V$ , which in turn equals the compressed signal voltage. To express  $s_2$  as a function of the normal signal voltage, apply (5) to (10), viz.,

$$s_2 = \frac{E}{V_t}. \quad (11)$$



Noise susceptibility, therefore, is directly proportional to the magnitude of the signal voltage. When  $s$  in (3) is replaced by the right-hand side of (11) we get

$$\frac{S}{N} = \frac{V_t}{v}. \quad (12)$$

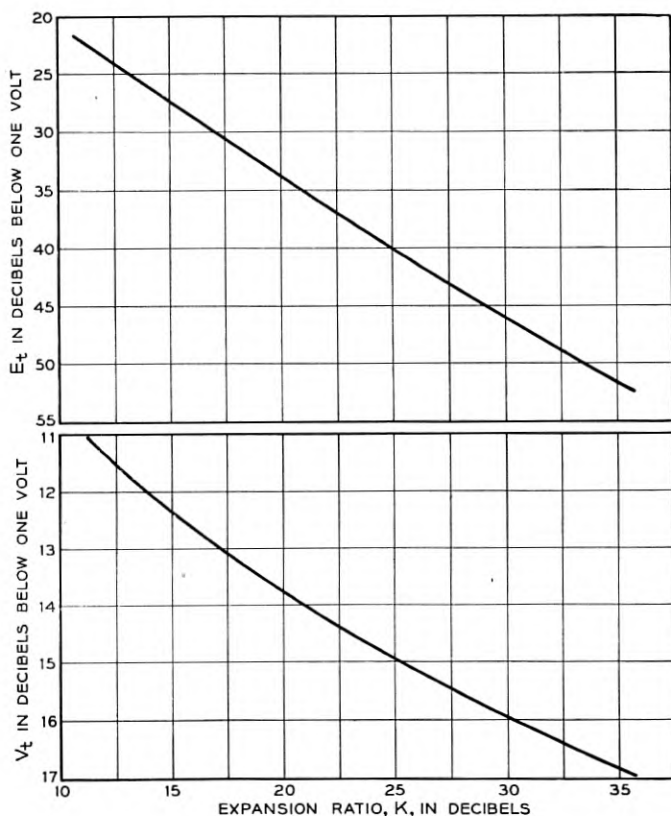


Fig. 6—Expander parameters.

This shows that the ratio of instantaneous signal to instantaneous noise voltages at the output of the system is independent of the magnitude of the signal for voltages within the exponential range of the expander.

Noise in the transmission medium is assumed to be fluctuation noise of uniform power density. It is convenient to replace  $v$  in (12) with the rms value of the noise voltage at the input to the expander.<sup>12</sup> Designate this

voltage  $v_r'$ , replace  $N$  by  $N_r$ , and (12) becomes

$$\frac{S}{N_r} = \frac{V_t}{v_r'} \quad (13)$$

The physical significance of this ratio may be explained as follows. The voltage which appears at the output of the system after the occurrence of a PAM pulse may be represented as the sum of a noise voltage and a signal voltage. Suppose that a very large number of measurements were made of the noise voltage which occurs along with a preassigned value of signal voltage. If the magnitude of the signal voltage is within the exponential range of the expander, then the ratio of signal voltage to the rms value

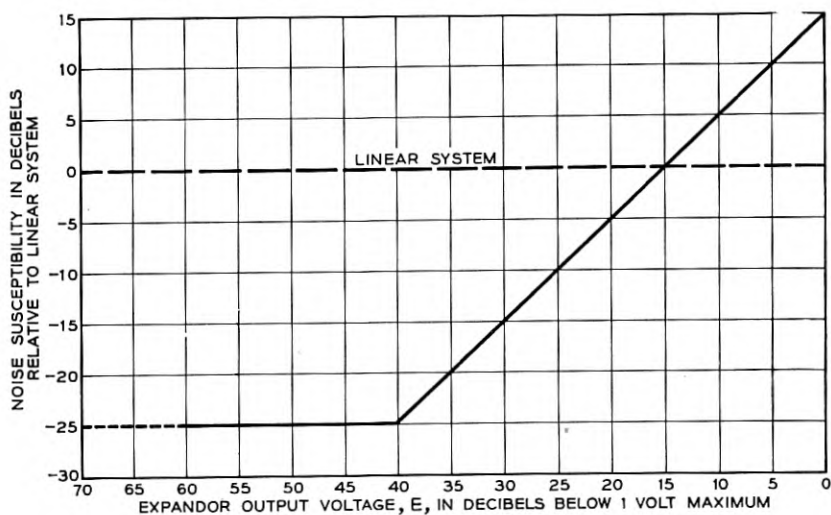


Fig. 7—Noise susceptibility.

of observed noise voltages would equal  $S/N_r$  in (13). The same value of  $S/N_r$  would be obtained if the measurements just described were repeated for any other preassigned magnitude of signal voltage within the exponential range of the expander.

When the voltages impressed upon the input of the expander are within the linear range of the characteristic, the noise susceptibility, designated  $s_3$ , is equal to  $\frac{E_t}{V_t}$ . From (7) we write

$$s_3 = \frac{1}{K} \quad (14)$$

This shows that, within the linear range of the compander, noise susceptibility is inversely proportional to expansion ratio.

Figure 7 illustrates the relationship between noise susceptibility and signal voltage. Noise susceptibility (Fig. 7) is expressed in db relative to that of a linear system.

#### *Value of $S/N_r$*

To evaluate the noise advantage which results from the use of an instantaneous compandor, it is necessary to know what requirement to place on  $S/N_r$ . This ratio refers to noise at the output of the system during intervals when the signal magnitude is within the exponential range of the expander. During these intervals the noise susceptibility of the system is proportional to the signal magnitude, so that the character of the noise is entirely different from that encountered in a linear system. People listening to speech transmitted through a system equipped with an instantaneous compandor have mistaken this type of noise for the distortion produced by an overloaded amplifier. Accordingly, experiments were made to determine how small  $S/N_r$  could be in a telephone channel whose frequency range was 200 to 3500 cycles.

A test circuit was devised which simulated the noise performance of a system equipped with a logarithmic compandor, and arrangements were provided so that the signal-to-noise ratio  $S/N_r$  could be adjusted over a wide range of values. The test procedure was to allow an observer to listen, during two consecutive intervals of time, to speech from the output of a linear system and from the compandor system. Conditions were arranged so that the noise at the outputs of the two systems was the same when the signal voltages were within the linear range of the compandor. The sequence in which the two conditions were presented to the observers was changed in a random manner, so that there was no way of identifying the compandor system except for the effect of the enhanced noise susceptibility during intervals when the signal magnitude was within the exponential range of the expander characteristic. Twenty-two observers participated in these tests and different speech volumes were used covering a range of 26 db.

Experimental results showed that the compandor system could be readily identified when  $S/N_r$  was 16 db or smaller, whereas the difference between the two systems was difficult to detect when  $S/N_r$  was 24 db or greater. An acceptable value of  $S/N_r$  for a typical telephone system is therefore somewhere between these two limits. A value of 22 db\* was selected as a conservative estimate. To confirm this, several people experienced in rating the quality of telephone systems were asked to listen to the output of the test circuit with  $S/N_r$  adjusted to 22 db. The consensus was that the quality was satisfactory.

\* This value is in agreement with the one used by C. B. Feldman and W. R. Bennett in studies of bandwidth and transmission performance, reported in reference 10.

Another point brought out by these tests was that the difficulty or ease with which the difference between the two systems could be detected was substantially independent of speech volume.

#### Noise Advantage

The value of  $S/N_r$  given above may be used to evaluate the noise advantage. Basically the problem is to find the permissible db increase in noise at the output of the transmission medium when the PAM system of Fig. 1 is equipped with an instantaneous compandor instead of linear networks having characteristics indicated by the dotted line in Fig. 4. For the comparison to be valid, the noise at the output of the system during intervals when the signal voltage is zero must be the same for the two conditions, and when the compandor is used  $S/N_r$  must equal 22 db.

Recall that  $v_r'$  represents the rms value of the noise voltage at the output of the transmitting medium when the instantaneous compandor is used, and let  $v_r$  represent the corresponding value when the linear networks are used. The noise susceptibility of the linear system is unity. Therefore, the noise at the output of the system during intervals when the signal voltage is zero will be the same for the two conditions provided  $v_r = v_r' s_3$ . When  $s_3$  is replaced by its value in (14) we require that

$$v_r = \frac{v_r'}{K}. \quad (15)$$

The equation which specifies that  $S/N_r$  equals 22 db is

$$12.59 = \frac{V_t}{v_r} \quad (16)$$

obtained by replacing  $S/N_r$  in (13) with the voltage ratio corresponding to 22 db. As shown by the lower curve of Fig. 6,  $V_t$  is a function of the expansion ratio,  $K$ . The quality of the two systems will be the same provided (15) and (16) are satisfied simultaneously.

Values of  $v_r$ ,  $v_r'$ , and  $K$  which simultaneously satisfy (15) and (16) are plotted in Fig. 8. Larger values of  $K$  would yield values of  $S/N_r$  smaller than the specified value of 12.59 db. Smaller values of  $K$  correspond to less noise improvement and make  $S/N_r$  larger than assumed necessary. The use of these curves will be illustrated by the following example.

It will be assumed that the rms value of the noise voltage at the output of a typical telephone channel is approximately 56 db below the highest signal voltage which the system is called upon to transmit. In the PAM system of Fig. 1 one volt was arbitrarily taken as the peak signal voltage at the output of the transmitting medium so that  $v_r$  is 56 db below one volt. From the upper

curve of Fig. 8, it is apparent that the noise at the output of the transmission medium,  $v_r$ , can be 35.8 db below one volt, and the corresponding value of  $K$  is 20.2 db. The noise advantage of the compandor equals  $K$ , and is about 20 db. Figure 6 shows that, in a 20 db expander,  $V_i$  is 13.8 db below one volt. The noise voltages at the expander input are well within the linear range of its characteristic. Otherwise (15) would not be valid.

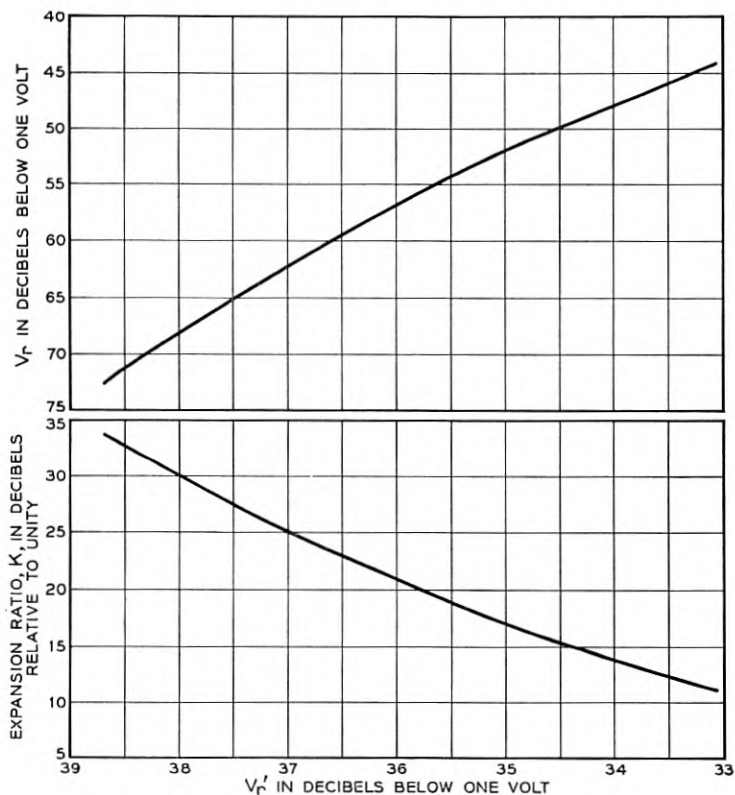


Fig. 8—System parameters used to evaluate noise advantage.

To illustrate that in the preceding example the noise advantage depends upon the noise standard, suppose the noise standard,  $v_r$ , had been 64 instead of 56 db below one volt. Proceeding as before, the noise advantage is about 26.7 instead of 20 db.

The general principles used in determining the noise advantage of instantaneous companding as applied to telephone channels presumably may be applied with equal convenience to other systems also using instantaneous companding and operating under different conditions.

## ACKNOWLEDGMENTS

The writer gratefully acknowledges the valuable assistance rendered to him in the preparation of this paper by Mr. H. S. Black, and the help of many others who participated in the experiments and offered valuable comments on the technical aspects of the analysis. These include: Messrs. C. B. Feldman, W. R. Bennett, J. O. Edson, and E. K. Van Tassel.

## REFERENCES

1. R. C. Mathes and S. B. Wright, "The Compondor—An Aid Against Static in Radio Telephony," *Bell Sys. Tech. Jour.*, Vol. XIII, page 315, July 1934.
2. N. C. Norman, "The Voice Operated Compondor," *Bell Laboratories Record*, Vol. 12, page 98, December 1934.
3. C. W. Carter, Jr., A. C. Dickieson, D. Mitchell, "Application of Compondors to Telephone Circuits," *Trans. A. I. E. E.*, Vol. 65, page 1079, December 1946 Supplement.
4. J. Lawton, "The Reduction of Cross-talk on Trunk Circuits by the Use of the Volume Range Compressor and Expander," *Post Office Electrical Engineers Journal*, Vol. 32, page 32, April 1939.
5. R. S. Caruthers, "The Type N-1 Carrier Telephone System: Objectives and Transmission Features," *Bell Sys. Tech. Jour.*, Vol. XXX, page 1, January 1951.
6. H. S. Black, "Pulse Code Modulation," *Bell Laboratories Record*, Vol. 25, page 265, July 1947.
7. L. A. Meacham and E. Peterson, "An Experimental Multi-Channel Pulse Code Modulation System of Toll Quality," *Bell Sys. Tech. Jour.*, Vol. XXVII, page 1, January 1948.
8. P. A. Reiling, "Companding in PCM," *Bell Laboratories Record*, Vol. 26, No. 12, page 487, December 1948.
9. W. R. Bennett, "Noise in PCM Systems," *Bell Laboratories Record*, Vol. 26, page 495, December 1948.
10. C. B. Feldman and W. R. Bennett, "Band Width and Transmission Performance," *Bell Sys. Tech. Jour.*, Vol. XXVIII, page 490, July 1949.
11. C. B. Feldman, "A 96-Channel Pulse Code Modulation System," *Bell Laboratories Record*, Vol. 26, page 364, September 1948.
12. W. R. Bennett, "Spectra of Quantized Signals," *Bell Sys. Tech. Jour.*, Vol. XXVII, page 446, July 1948.
13. W. R. Bennett, "Time Division Multiplex Systems," *Bell Sys. Tech. Jour.*, Vol. XX, page 199, April 1941.

# The Evolution of Inductive Loading for Bell System Telephone Facilities

By THOMAS SHAW

(Continued from April 1951 issue)

## PART IV: CABLE LOADING COIL CASES

### *General*

Up to this point this review of coil loading has been primarily in terms of transmission features of the loading systems and the coils, and development economics, and for this reason the references to potting developments have been brief, so as to minimize diversions from the main theme. A complete chronological review of all important aspects of the potting development work would require much more space than is available for this present article. On the other hand, because of the substantial importance of the potting developments in the economics of loading, more should be said than was included in the brief references in Parts II and III of the story.

This particular part of the review accordingly describes the more important high spots of the potting developments. It is limited to cable loading coil cases because of the early obsolescence of open-wire loading. The discussion is in terms of the changes from time to time in the various important design features, as indicated by the side headings and paragraph headings, and is thus a departure from the individual project-description procedure followed in other parts of the review.

At this point it should be emphasized that the work on the cases, which has been more nearly continuous than that on the coils, has kept pace in design ingenuity with the work on the coils, and has been very much more than the mere accommodation of the case designs to the changing sizes of the loading coils and of the loading complements.<sup>(1)</sup>

### *Casing Materials*

Until the late 1920's, cast iron casings were used for housing the coils. The moisture-proof seal between the case top and the main casing was obtained by "tongue and gutter" details, supplemented by metal gaskets. For the first decade or so, short lengths of wrought iron pipe with "pipe

<sup>(1)</sup> Additional information on potting developments is given in Bibliography items (6), (8), (26) and (30).

cap" ends were used in potting the smallest complements of coils. For all cases, lead-sheathed stub cables contained the coil terminal leads.

During the late 1920's, new case designs using thick, copper-bearing, steel plate ( $\frac{3}{8}$  inch thick) with welded seams were introduced for reasons of economy and to simplify manufacture. Because of the very great current demand for loading, and the extensive use of the large-size loading coil cases required by the larger loading complements, the foundry problems had become

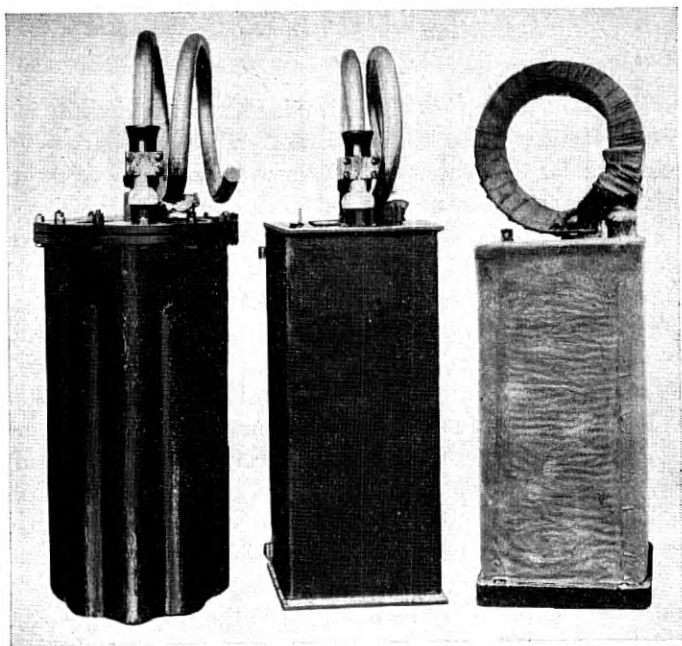


Fig. 19—Cast iron vs. welded steel loading coil cases. Complements of 84 P-type phantom loading units. At left: Cast iron case; In center: welded steel aerial case; At right: welded steel underground case.

very formidable; also, the production schedules had become somewhat irregular because of the great difficulties encountered in getting enough satisfactory castings of the largest sizes. The use of heavy machinery for shearing the steel plates and forming them to obtain cases of rectangular cross-section gave adequate dimensional design-flexibility. The welded steel designs, however, are not generally so satisfactory as the cast iron designs with respect to resistance to corrosion of accidentally exposed steel surfaces. Accordingly different types of protective coatings were provided for cases intended for underground cable and buried cable installations, and those intended for



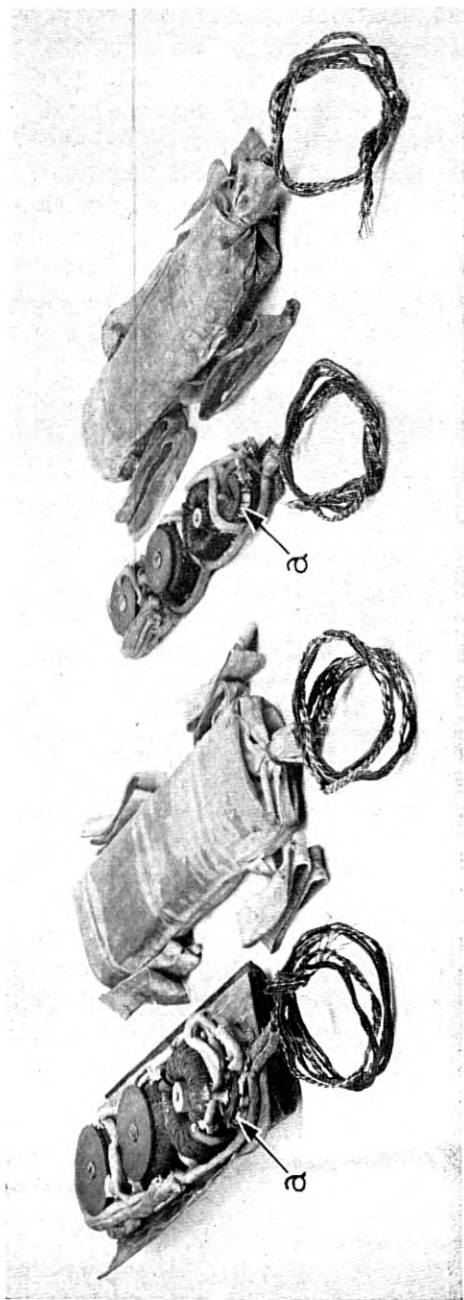


Fig. 20—Splice loading assemblies of individual phantom loading units. Assemblies before and after placement in canvas protective bags.

At left: M-type unit; At right: MF type loading unit.

“a” designates midget inductance coils used in crosstalk adjustments.

aerial cable jobs. Certain differences also were necessary in the mounting details. These various differences were recognized in the case code designations.

Beginning in the early 1930's, sections of lead sleeving with soldered lead (and later on, brass) tops and bottoms came into general use for small loading complements. To add mechanical support to the inherently weak, cylindrical lead sleeve, the larger-size lead cases were equipped with inner-lining steel tubes when used on cables maintained under gas pressure. The lead cases<sup>(u)</sup> are less expensive than rectangular-shaped welded-steel designs of equivalent potting capacity, and are suitable for use on underground and on aerial cables. On the lead cases used in buried cable projects, and on their

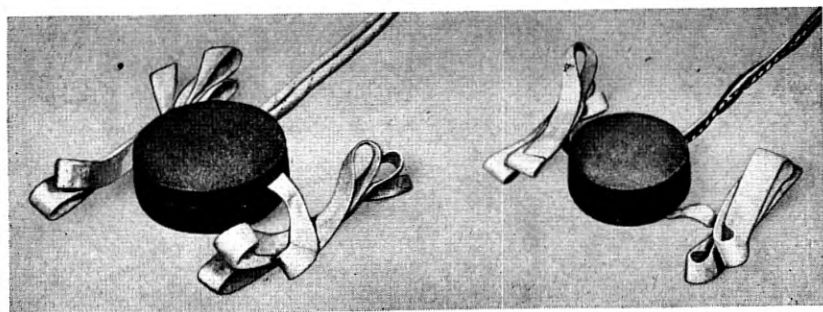


Fig. 21—Splice loading cases for exchange area loading. 88 mh coils potted in cardboard containers, and equipped with fabric tapes for fastening to cable core at splice. At left: 622 coil. At right: 632 coil.

stub cable sheaths, a special finish reinforced with armor provided protection against injury by rodents. Corresponding protection was also provided on the sheaths of stub cables of cast iron and welded steel cases intended for use on buried cables.

Another general change in the design of loading coil cases started about 1940 with the introduction of cylindrical,  $\frac{1}{8}$ -inch steel-tubing in place of thick steel-plate rectangular designs.<sup>(v)</sup> While initially this was a steel conservation measure, it was found to be very advantageous with respect to manufacturing techniques and economy. This development will eventually reduce the use of the previously mentioned lead-sleeve designs.

The basic problem of securing the most economical designs for different potting complements and different installation conditions has included the provision of special case designs for placement in cable splices, which do not

<sup>(u)</sup> Some lead sleeve cases for exchange area loading are shown in Fig. 17 (page 467).

<sup>(v)</sup> Some of these cylindrical thin steel cases are shown in Fig. 18 (page 468) and in the installation photographs Figs 26, 27, 28A and 28B (pages 728, 729, 730 and 731, respectively).

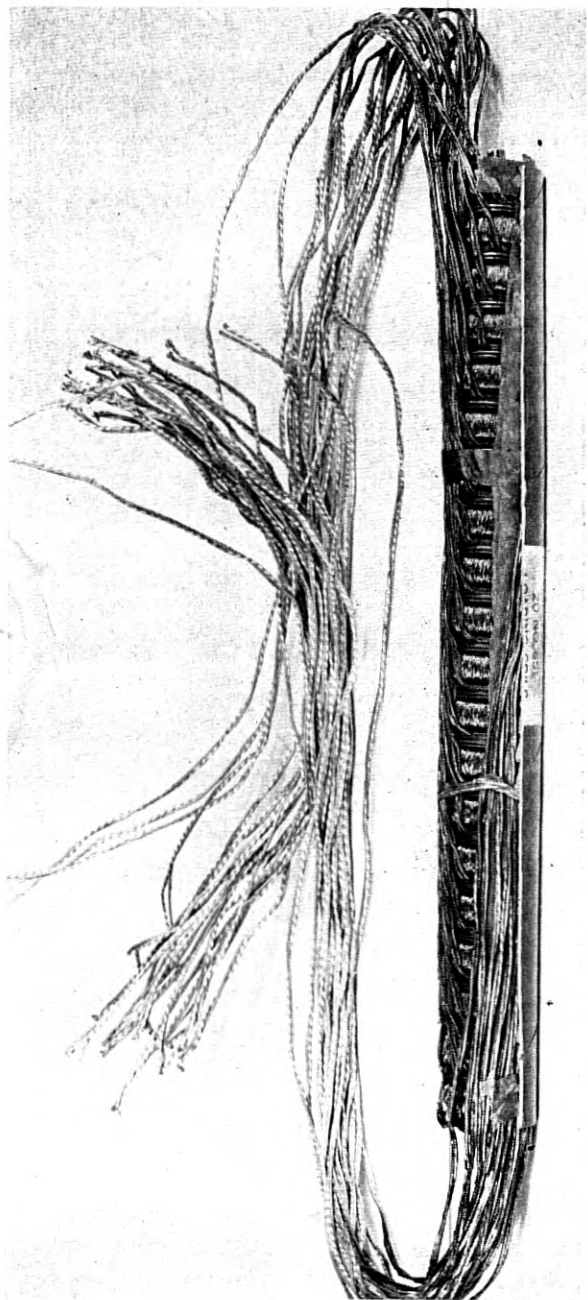


Fig. 22—20-Coil spindle assembly of No. 651 coils. Cable splice installations—long subscriber loops. View of assembly after impregnation, with part of plastic case removed to show the coils. The IN terminal pairs are bunched at one end, and the OUT pairs at the opposite end.

require stub cables, for office rack-installations, for submarine cables, and for buried, insulated wire. The splice-loading designs include individual cardboard containers for small exchange area (Fig. 21) and toll cable coils (Fig. 20), and baked varnish impregnated spindle-assemblies of coils for coaxial cable order-wire circuits, and for small exchange area cables (Fig. 22).

### *Case Sizes and Shapes*

Where involved, the case size and shape limitations have generally been imposed by underground cable installation conditions. The circular tops of

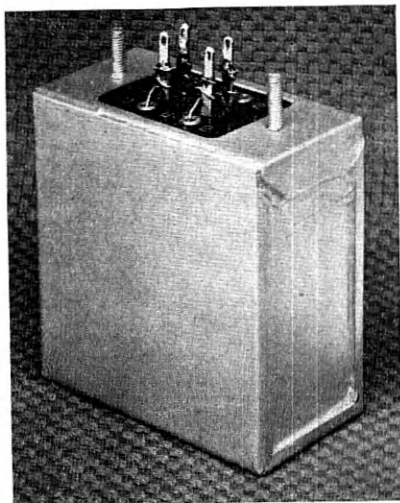


Fig. 23—Office type loading coil case for installation on office mounting plates. Designed for potting molybdenum-permalloy core program circuit loading coils.

the cast iron cases and the rectangular tops of the welded steel cases had to be small enough to permit lowering the cases through the circular manhole openings in the loading manholes and loading vaults. In the early applications of loading, 26 and 27-inch manhole openings were very common; later on, 30-inch openings became standard for loading manholes. In recent years, in redesigning the large, thick-steel cases that required 30-inch manhole openings for their installation, the superseding thin-steel designs were proportioned to permit installation in line manholes having 27-inch openings.

The case bodies of the cast iron cases were approximately circular in cross-section with scallop-shaped contours corresponding to the compartments in which the coil spindle-assemblies were mounted. These ranged from 3 to 7 in number. In the rectangular cross-section, welded steel designs, there

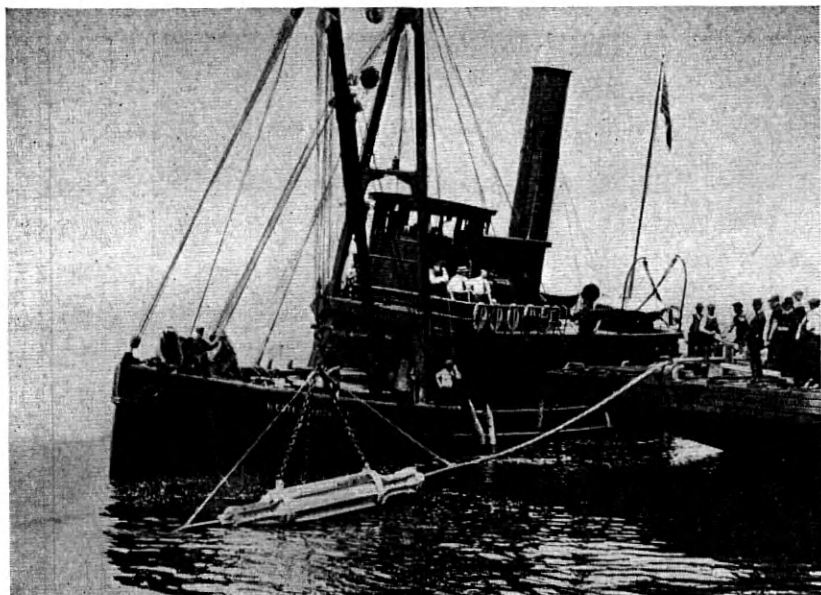


Fig. 24—Installation of submarine cable loading. An early type of cast iron case ready for lowering into water.

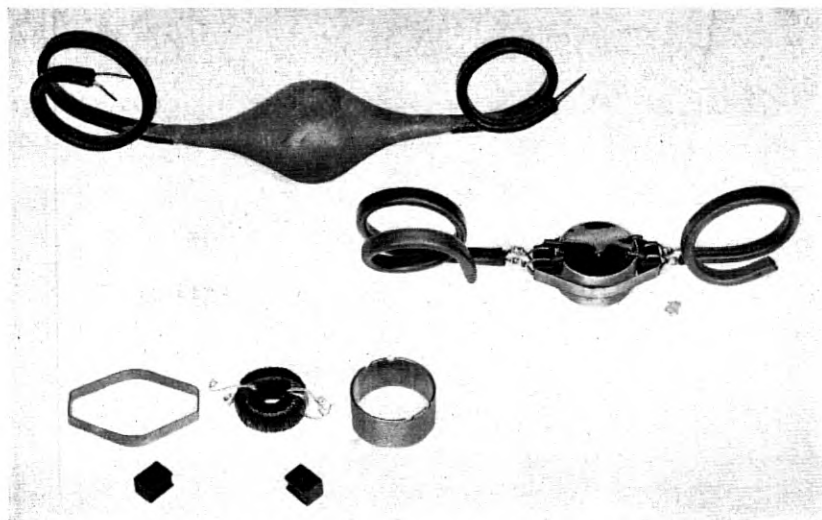


Fig. 25—Moulded rubber loading coil case for buried wire loading installations. Views of piece parts, partial assembly, and potted coil ready for installation.

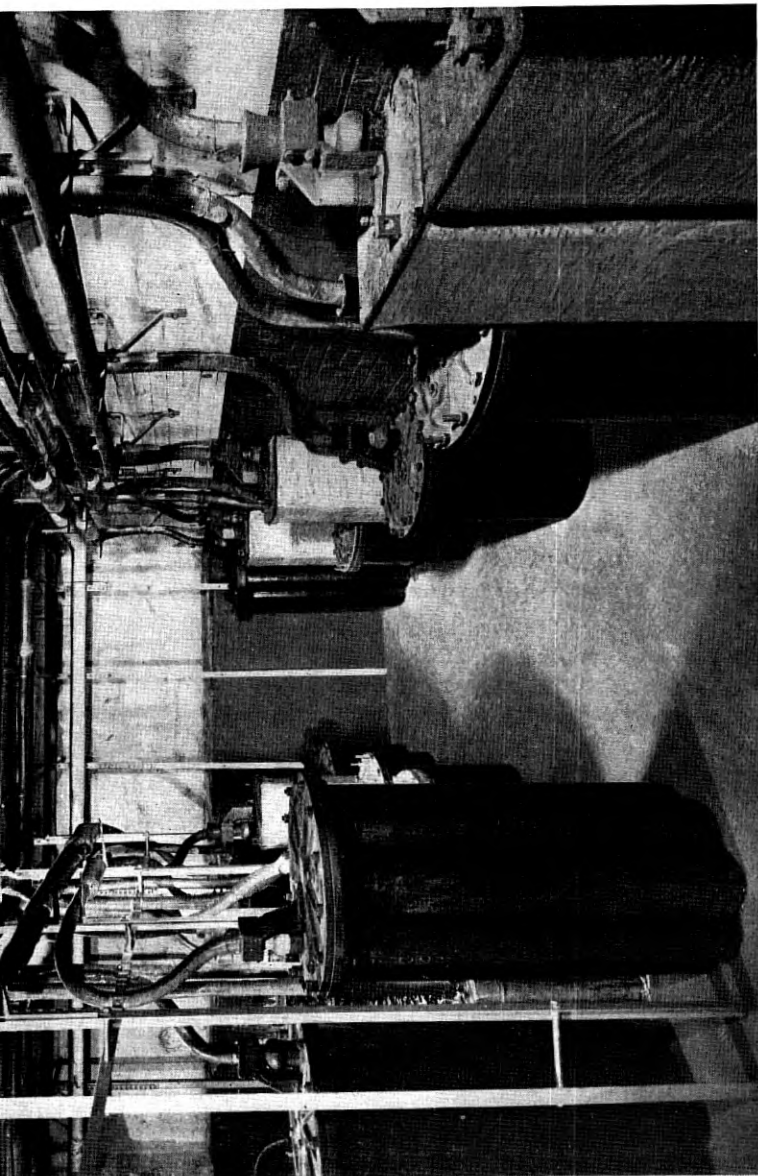


Fig. 26—Office vault installation of exchange area trunk cable loading. This vault is in the basement of a Manhattan office, close to the East River. The coils are used for loading trunks between Brooklyn and Manhattan offices routed through cables which cross the river in a rapid transit tunnel where space restrictions prevent the installation of loading coil cases. At time of photograph, a total of 36 cases containing a total of 11,404 coils were installed in this vault. The 36 cases comprise 20 cast iron, 9 rectangular welded steel, and 7 tubular thin steel designs. The view shows 17 cases; the other cases are hidden in background or are out of camera range. One case contains 456 coils; in the others the complement ranges from 303 to 306 coils. The individual coil codes are 601, 602, 603, 612, 613, 614 and 632. About 13% have iron dust cores, 63%

usually was design flexibility in proportioning the case heights and the cross-section dimensions to be approximately optimum for most efficient

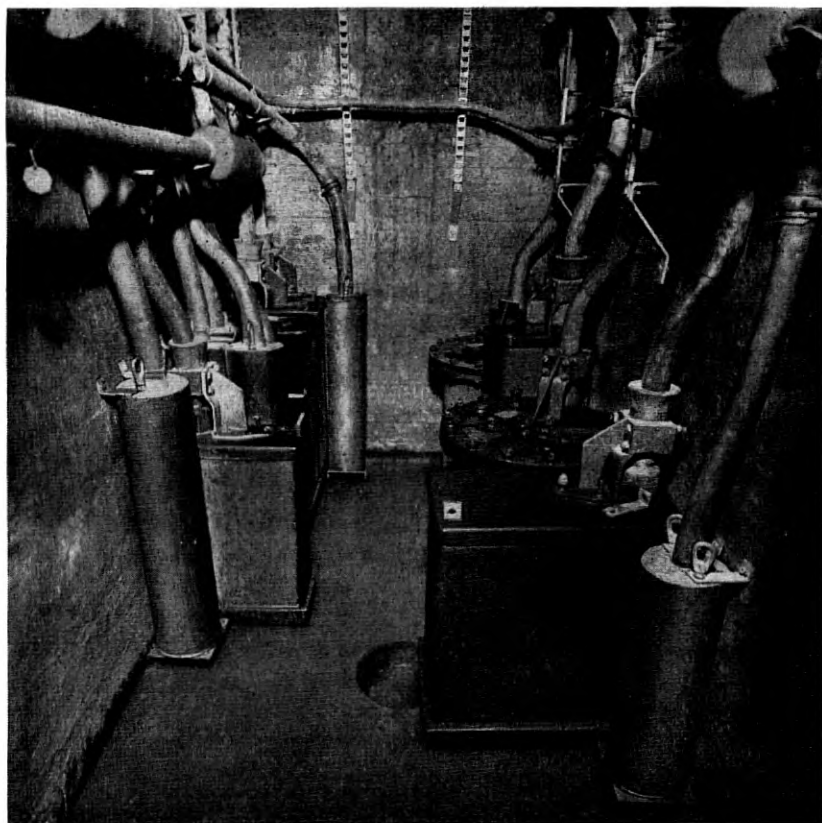


Fig. 27—Exchange area loading installation in side street auxiliary loading vault. In congested areas, space limitations frequently require the installation of underground cable loading coils in an auxiliary loading vault located in a side street near its intersection with the street under which the main cable conduit system is laid. Extensions of the case stub cables carry the coil terminal leads to the main cable splices. At time of photograph, a total of 18 cases containing a total of 7283 coils were installed in this auxiliary vault. Six cases had 303 or 304 coils, and 12 cases had 455 or 456 coils. The 18 cases include 2 cast iron cases, 5 rectangular welded steel cases, and 11 tubular, thin steel cases. The individual coil codes are Nos. 612, 632 and 643; 92% have 88 mh inductance, the remainder being 135 mh coils. 27% have permalloy cores; the remaining 73% have molybdenum-permalloy cores and formex insulated windings. This figure shows the far end of the loading vault where 14 cases are placed. Several tubular steel cases are hidden by larger cases in foreground.

use of the available mounting and splicing-space in the loading manholes, subject of course to the manhole-opening limitations previously mentioned.





Fig. 28A—Another auxiliary vault installation of exchange area loading. One end of vault. At time of photograph, 33 cases containing a total of 11,241 coils were installed in this vault. A majority of the cases appear both in this view and in the view of Fig. 28B. 25 cases have 303, 304 or 305-coil complements, each of the other 8 cases have 455 or 456 coils. The total complement comprises 2 cast iron cases, 17 rectangular welded steel cases, and 14 tubular thin steel cases. The individual coil codes are 612, 613, 614, 622, 623, and 643. The coil inductances are: 88 mh, 70%; 135, 27% and 175 mh, 2.7%. 35% of coils have permalloy cores; the others have molybdenum-permalloy cores. 57% of coils have formex insulated windings.

Many of the individual designs were relatively tall, with cross sections requiring small amounts of floor space. The recent trend in case design (1948,



1949) revolves about two requirements: (1) the cases shall be capable of being installed in manholes having 27-inch openings, and (2) the vertical



Fig. 28B—Other end of vault in Fig. 28A.

dimensions shall be as short as possible while meeting requirement (1). The primary purpose of these changes is to reduce the cost of manhole construction in new conduits.

The submarine cable loading cases previously mentioned are cylindrical

in shape, designed and installed so that the axis of the cylinder is in line with the axis of the submarine cable. A stub cable extends from each end, one having the IN leads to the coils, and the other the OUT leads. (Refer to Fig. 24, page 727.)

### *Potting Complement Sizes*

Usually, the initial traffic requirements and estimated rate of traffic growth are the most important factors in determining the potting complement sizes for particular projects. Frequently the initial loading complement is greater than subsequent complements. In some situations, two or more cases may be installed at the same time, because of the case-design size limitations.

*Non-Phantom Coils:* In the early loading applications of non-phantom coils, the most common potting complements were of the order of about 50 coils. The maximum complements prior to the development of inexpensive loading for 22 ga. exchange cables ranged up to 98 coils. Complements of 200 and 300 coils became quite common with the Nos. 601 and 602 coils. With the introduction of the No. 612 (permalloy-core) coils, complements of 450 and 600 coils became common, and occasionally a 900-coil complement was used. The maximum complements of the much smaller molybdenum-permalloy core coils have been held to about 450 coils. The foregoing is an interesting manifestation of the stubborn limits upon case-design cost-reduction that come into play as the coils become smaller and smaller. The labor-cost component is dominant and little saving in materials can be achieved. In consequence, it is frequently preferable to use two medium-size complements, rather than one over-size complement having the same total number of coils, especially if the second complement can be deferred for some time.

*Side and Phantom Coils:* In the early applications of loading to quadded 19 and 16 ga. toll cables, the side circuit coils and the phantom coils were some times potted in separate cases. In such applications, the side circuit coil complements ranged up to 98 coils, and the phantom coil complements ranged up to 48 coils, but the average-size complements were substantially smaller. Soon it became the common practice to pot associated side circuit and phantom circuit coils together in the same case, and in such instances loading complements for 24 cable quads were common. To help meet the increasing demand for toll cable loading in the early 1920's, maximum complements for loading 36 cable quads became available. When the phantom coils were reduced to side circuit coil-size (1923), the maximum potting complement was increased to 45 loading units. The large coil-size reduction that resulted from the use of compressed permalloy-powder cores made practicable during the late 1920's and early 1930's a further, very substantial, increase in the range of standard potting complements covering up to

84 phantom loading units of the P-type, and up to 108 loading units of the PB-type. These very large loading complements had an interesting economic significance. At the time they became available the demand for additional toll cable facilities was increasing by leaps and bounds, and it was not uncommon for standard-size and over-size cables to be fully loaded at the time of the installation of the cables. In this connection, full loading for a 50% over-size, quadded, composite 19 and 16 ga. cable could be provided with two of the maximum-size potting complements of phantom loading units. This contrasts with fairly common experience in early installations of full-size cables, where four, five or six cases were used to provide complete loading.

The further size-reductions in toll cable loading coils that were achieved with the standardization of the M-type phantom loading units and later with the MF-type units occurred during a period of greatly reduced demand for toll cable loading, influenced by the exploitation of carrier systems on conductors from which loading was removed and on new non-loaded cables. The maximum potting complements were accordingly held to 80-unit and 48-unit sizes for the M-type and MF-type loading units, respectively.

#### *Assembly Methods and Stub Cables*

*General:* From the beginning of commercial manufacture, multi-coil complements of cable loading were coaxially assembled on spindles which were held in fixed positions in the cases. Preceding this operation, the accumulated moisture was expelled from the coil windings and the coils were dipped in a moisture-resisting compound. In the multi-spindle cases, the different spindle-assemblies were mounted in separate compartments of the cases, with the compartment partitions providing shielding to control crosstalk among the spindle-assemblies. On the individual spindle-assemblies, crosstalk was controlled by using steel washers between adjacent coils and mounting the coils so that their small, external, magnetic fields would be substantially non-interfering. The winding ends of the coils were connected to textile-insulated twisted-pair leads in spindle unit-cables, treated with wax for moisture protection.

After the spindle-assemblies had been fixed in position in the case compartments, the spindle unit-cables were formed by hand into a stub cable core over which a somewhat loose-fitting lead sheath was drawn. A color code on the conductor insulation provided identification for "wire" and "mate" conductors, and for IN and OUT coil terminals. The final assembly operations included filling the case compartments with a viscous rosin-oil compound, and a top layer of asphalt compound; and the stub cable sheath was soldered to a nipple in the case cover. At various stages in the assembly

operations, suitable inspection tests were made to assure satisfactory conformation to specification requirements. After the final inspection tests, the outer end of the stub cable sheath was sealed to prevent entry of moisture.

When phantom loading started, the phantom coils were much larger than the side circuit coils. To conserve potting space in cases containing both types of coils, the individual spindle-assemblies consisted of only one type of coil. The phantom unit cross-connections between the phantom coils and their associated pairs of side circuit coils were made at the top of the case between quadded spindle unit-cables containing the OUT terminal leads of the phantom coils and the IN terminal leads of the side circuit coils. The quadded IN terminal leads of the phantom coils and the OUT terminal leads of the side circuit coils constituted the main line terminals of the complete loading units.

In general, all of the stub cable leads to the IN and OUT terminals of non-phantom type coils, and to the main line terminals of phantom unit combinations of side circuit and phantom circuit loading coils, were contained in a single stub cable sheath. Necessary exceptions to this practice occurred, however, in the submarine cable loading coil cases (which had two stub cables extending from opposite ends) and in the underground and aerial cable cases containing very large complements of P-B type loading units.

The first important change from the original potting practices followed soon after the introduction of phantom group loading. The original wax-dipped textile-insulated stub cables were found to be seriously objectionable sources of phantom-to-side and side-to-side crosstalk. To reduce crosstalk, and also to improve transmission, the practice started of using strip-paper insulated quadded conductors in a machine-stranded stub cable. This required a splice to be made inside the case, at the top, between the paper-insulated stub and the textile-insulated spindle unit-cables.

*Apparatus Group-Segregation, Four-Wire Circuits:* Several years later, when the development work on long-distance four-wire repeatered circuits got well under way, the assembly arrangements in the loading coil pots and the stub cable designs were changed to provide crosstalk segregation between the groups of coils used on the opposite-direction branches of the four-wire circuits. The segregation arrangements in the stub cables included shielding between the groups of terminal quads associated with the opposite-direction circuit groups. These loading coil case and stub segregation-arrangements were details of a fundamental plan for complete group-segregation between the opposite-direction branches of four-wire transmission systems (including the main cables themselves and the repeater office circuits) in order to control the intergroup near-end crosstalk coupling and prevent it from being a serious factor in the over-all crosstalk. (The relatively high amplification

in the four-wire repeaters made very desirable the rigorous control of this type of crosstalk coupling.) Initially, the loading apparatus and associated stub cable quads intended for use on two-wire repeated circuits were segregated in the coil cases from the two opposite-direction groups of four-wire circuit apparatus. Later on, this apparatus segregation between two-wire circuit coils and four-wire circuit coils was discontinued.

In mixtures of the two types of coils, the two-wire circuit apparatus was divided into approximately equal groups, each of which was combined with one of the four-wire opposite-direction groups. In order to simplify potting practices, the two-group segregation plan became the standard plan for all toll cable loading cases and was used in loading complements containing mixtures of two-wire circuit and four-wire circuit loading apparatus, and for complements consisting wholly of two-wire circuit coils, or of four-wire circuit coils.

In the most recent (1948-1949) general redesign of toll cable cases and stub cables, the segregation arrangements just described have been simplified, largely because of the present very small demand for additional H44-25 four-wire circuits. Group segregation of the apparatus within the cases is used only when four-wire H44-25 loading is involved. The use of shielding and of group-segregation arrangements in the stub cables has been discontinued. In complements of units for H44-25 loading, one of the opposite-direction groups is connected to a group of terminal quads having contiguous "quad counts" at the low end of the quad counting scheme described on page 737, and the other group is connected to the quads having "quad counts" at the high end of the quad counting scheme. When two-wire circuit coils are included in a loading complement containing four-wire circuit coils, the terminal quads of the two-wire coils use terminal quads having "quad counts" intermediate between those of the two opposite-direction groups of four-wire coils.

*Phantom Coil Size-Reduction:* The next important potting-practice change resulted from the size-reduction of the phantom coils to side circuit coil-size. From then on, the phantom coils were mounted on the same spindles as the side circuit coils, with each phantom coil immediately adjacent to its associated pair of side circuit coils. This permitted a more efficient use of potting space and made unnecessary the use of spindle-unit cabling for the cross connections between the coil components of the phantom loading units. The miniature inductance coils used in the loading-unit crosstalk adjustments were mounted on a frame located at the top of the coil spindle-assemblies and were connected in the circuit at the splices made between the stub cable and the spindle unit-cables leading to the coil line terminals.

*Assembly Redesign, Exchange Area Coils:* The first major change in the

assembly and case-wiring arrangements for non-phantom exchange area coils was made during the middle 1920's in solving difficult problems that arose in the design of cases for 200-coil, 300-coil, and larger complements of the new small-size loading coils. The spindle-assemblies of coils were fastened to a skeleton frame to which the case pot-cover and the machine-stranded textile-insulated stub cable were attached. The part of the stub cable that extended within the case was subdivided into unit cables which were associated with the individual spindle-assemblies, to which the coil leads were connected, thus eliminating the intermediate spindle cables. After these connections had been made, the complete coil-assembly, with stub cable, was lowered into the coil casing. The case cover was then fastened to the coil casing, and the case filling-compound was poured through a small temporary opening in the cover.

*Assembly and Cabling Changes; Beginning of Use of Loading Units in Individual Shielding Containers:* Some of the improved assembly-arrangements, above described, were made available for phantom loading units soon after the permalloy-core toll cable coils became available, along with the standardization of the P-B type loading units. Certain differences were necessary, however, in order to permit the assembly of the 3-coil loading units in individual, cylindrical, shielding containers, having one end open. The associated three coils of a unit were mounted on a short hollow dowel, at one end of which was mounted a frame supporting terminal posts for all of the line terminals of the individual coils and the crosstalk adjustment-elements (small inductances and resistances, preselected to meet crosstalk requirements). The cross-connections between the phantom and side circuit coils were made between appropriate terminal posts, and the loading unit main-line terminals were connected to the stub cable conductors at this point, after the crosstalk adjustment-elements were connected in the circuit. The individual loading units in their shielding containers were fastened to a vertical frame by bolts extending through the hollow spindles. This frame was attached to the case cover. The stub cable was a machine-stranded, single piece of paper-insulated cable having the part below the case cover separated into unit cables for the connections to horizontal rows of loading units.

The loading unit and potting assembly methods are illustrated in Figs. 10 and 11 (pages 186 and 188, respectively). Generally similar arrangements were used with the M-type and SM-type loading units. With the standardization of the MF-type loading units, the adoption of cylindrical steel case bodies resulted in some potting assembly changes which are illustrated in Figs. 12, 13, and 14 (pages 192, 193 and 194, respectively). In the medium and large size potting complements, the individual loading units are mounted



near the outer periphery of circular mounting plates, and the inner end of the stub cable extends through a circular opening at the center of these plates, (Fig. 13). In the very small complements, the MF units are stacked one above the other as shown in Fig. 14.

The concentric layer-type stub cables, used with new cases potting P-B type loading units and subsequently with the M-type, SM-type, and MF-type loading units, had an improved color-code for the conductors of the terminal quads which provided a counting-scheme type of identification for each of the individual loading units potted in a given case. To facilitate this full-scale identification of the individual units, it was necessary to have a precise wiring coordination between the positions of the individual units on the case assembly frames and the positions of their terminal quads in the stub cable (which were identifiable in terms of the quad-count color-code). This permitted the necessary coordination of the coil-grouping arrangements which were desirable for crosstalk reasons in complements of four-wire circuit loading (as previously discussed) with the group segregation and shielding arrangements of the associated stub cable terminal quads. A simplifying factor was the use of adjacent quads for the IN and OUT terminals of same loading unit.

The improved stub designs greatly simplified the manufacturing problems involved in providing, (a) full flexibility as regards loading complement sizes, and (b) full flexibility for desirable combinations of different types of loading units.

(a) Using a relatively small number of case sizes, provision was made for obtaining any total-complement size, ranging from one up to the maximum-complement size, in steps of one loading unit. A different size of stub cable was used for each different size of case. When less than a full complement was desired in a particular size of case, the unused terminal quads were left open at the inner end of the stub cable and were tagged at the outer end. In terms of "quad counts" these non-used quads had contiguous numbers at the "high" end of the quad counting-scheme, and were readily identified by means of the quad-count color-code and the tags previously mentioned. (A name-plate on each case recorded the number and the code types of loading units potted in the case.) In the prior art, different stub cables had been provided for fitting the different partial and full potting-complements in particular sizes of cases.

(b) For several years prior to the standardization of the improved assembly and stub design it had been a common practice to use mixed potting complements of different types of loading units, in order to realize the maximum potting and installation economies inherent in the use of larger-size loading complements made practicable by size-reduction of the loading

coils. These mixtures usually comprised combinations of coils for four-wire long-haul circuits with one or two types of coils for short-haul or medium-haul two-wire circuits. When this practice of mixed potting complements started, it was customary to use different stub cables for each different potting-mixture, even when the same total number of loading units were involved. In these stub cables, each of the different types of loading units had its own individual color-code identification. In the new set-up, the relatively simple color-code counting-scheme provided full flexibility for all desirable combinations of different types of loading units.

Loading units of a given type made use of terminal quads having *contiguous numbers* in the quad counting-scheme. In mixtures involving two types of loading units, for example P1B and P11B units, the units having the lower-number component in their code-designation used the low-numbered terminal quads, and the units having the higher code-number used the high-numbered terminal quads. In mixtures involving three different types of loading units, the units having the intermediate code-number in their code designation used a contiguous group of terminals which were intermediate in position between the low-numbered quads and the high-numbered quads which were respectively associated with the loading units having the lowest and the highest code numbers. These procedures were followed in each of the two segregated groups of opposite-direction loading units previously mentioned.

*Quadded Stub Cables for Non-Phantom Coils:* During the 1930's, the practice of using quadded stub cables for cases potting non-phantom type exchange area and program circuit loading coils was started. One pair of each terminal quad was connected to the IN terminals of a coil, and the associated pair was connected to the OUT terminals of the same coil. Previously, the IN and OUT terminals had been grouped in different unit cables. The close association of IN and OUT terminals in the new quadded stub cables reduced the factory testing-time, and simplified the preparatory phases of the field splicing of the stub cables to the main cables. Other subsequent improvements included the use of paper-pulp insulation on the stub cable conductors, in place of textile insulation. By this time it had become a common practice to terminate the coil windings on terminal clips mounted in close proximity to the coils. The inner ends of the stub cable conductors were soldered directly to these clips.

*Stub Cable Conductor Sizes:* Since the case stub cables are extensions of the main cables, transmission considerations have generally led to the use of about the same sizes of conductors. However, notable exceptions to this rule have been accepted in situations where conformation to the rule would have made the stub cable unduly expensive, or unduly large and difficult to



handle at the factory or during installation. The stub cable conductor sizes have ranged from 13-gauge in the cases containing coils designed for composite coarse-gauge toll cables to 24-gauge in the standard cases for the coils used principally on 22 and 24-gauge non-quadded exchange cables. For

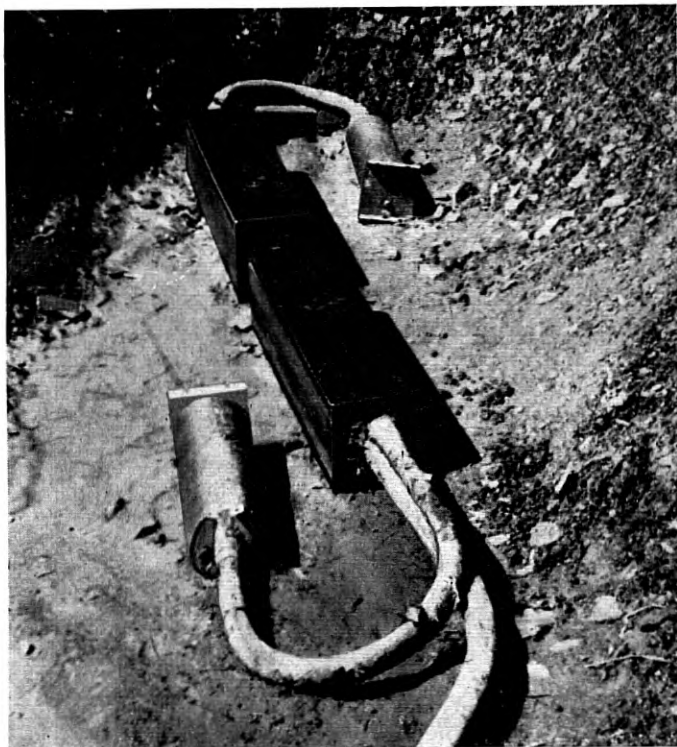


Fig. 29—Buried coaxial cable installation of voice-frequency loading on outer layer quads. View of installation after completion of the splicing work, and prior to filling in the excavation. The loading coils are potted in two tubular, thin steel, cases. Each of black boxes near center covers a cable splice, and furnishes protection against mechanical injury. At each splice, connections are made to the stub cable conductors of a single loading coil case. Splicing difficulties prevent all of the connections from being concentrated at a single cable splice.

several decades, 19-gauge stub cables were used for the toll cable loading cases. The most recent case designs are using 22-gauge conductors.

### *Dielectric Strength*

From the beginning of the use of cable loading, a fundamental design requirement has been that the insulation of the loading coils and of the

associated stub cable and pot wiring should have a dielectric breakdown-strength as high as that of the cables for which the loading was intended, and preferably somewhat higher, to assure that the loading apparatus would not be dielectrically weak points in the loaded cable systems.

When dielectric-strength improvements in toll cable design started during the late 1930's in order to reduce damage by lightning, especially on buried and aerial cables, equivalent improvements were also made in the loading apparatus. These cable and apparatus improvements were primarily concerned with raising the dielectric strength of the insulation between core and sheath.

During recent years, the extensive installation of buried toll cables in areas where the ground resistance is high has led to the use of cables having very much higher dielectric strength (wire to ground) than those used during the 1930's. The development of the copper-jacketed toll cable having a thermoplastic protective covering between the lead sheath and the jacket, which was capable of withstanding a dielectric-strength test of 10,000 volts d-c, between the sheath and the jacket, made it necessary to apply an equivalent insulation to the exterior of the buried loading coil cases. The more recent development of the "Lepeth" sheathed toll cable has made it possible to approach a dielectric breakdown-strength of the order of 25,000 volts d-c between cable core and sheath. This is achieved by extruding a sheath of polyethylene of suitable thickness over the cable core, and over this a thin lead sheath. Loading coil cases were redesigned to match this construction, using an *inner lining* of thermoplastic insulation to provide equivalent insulation between the coils and the case. The stub cables have dielectric design-features corresponding with those of the Lepeth toll cable.

### *Potting Costs*

The potting cost per coil, or loading unit, varies considerably with the potting complement-size, and is a maximum in small complements. These general relations apply for all types of coils.

In the early designs, the average potting cost per coil was much smaller than the coil costs. Over the years, the case cost-reduction that has resulted from coil size-reductions, increased complement-size, and other design changes, has been smaller on a percentage basis so that in the present designs the average per coil potting costs are somewhat greater than those of the coils. Notwithstanding the changes in cost relations just mentioned, the direct and indirect savings that have resulted from the potting development work constitute a substantial fraction of the aggregate plant cost-reduction which has been achieved by the use of coil loading.

## PART V: LOADING FOR INCIDENTAL CABLES IN OPEN-WIRE LINES

### INTRODUCTION

From the earliest days of telephony, when it became necessary to use pieces of cable in long-distance lines to provide toll entrance facilities at toll centers or for other purposes at intermediate points, such cables have had more or less objectionable effects on the over-all transmission-system performance. These impairments resulted from the much greater transmission loss per unit length in the inserted cable, and from reflection effects occurring at the cable junctions with the open-wire—these being due to the large differences between the cable and open-wire impedances.

Prior to the advent of loading, the losses in incidental cables could be reduced to low unit-length values only by using expensive coarse-gauge cables. Cable loading became available just in time to head off the installation of some very expensive coarse-gauge cables that had been proposed for unusually long entrance facilities in the New York and Boston areas. The use of loading on the open wires greatly increased the economic importance of attenuation reduction in the incidental cables occurring in such lines. By substantially raising the line impedance, loading also increased the magnitude of the reflection losses at junctions with non-loaded cables.

Standard "heavy" loading (Table II, page 156) came into general use on long entrance cables in the loaded lines. While this loading did not have a sufficiently high impedance to match that of the loaded line, it was close enough to reduce the junction reflection losses to acceptable values. A special light-weight loading found some use on incidental cables in non-loaded lines.

When satisfactory types of telephone repeaters became available for extensive use on loaded open wires, the cable junction impedance-irregularities, and other irregularities, had to be reduced to very small values so as to avoid repeater circuit unbalances that would objectionably restrict the repeater gains. These severe requirements put a high premium upon the use of an improved type of cable loading having impedance characteristics that matched closely those of the associated open-wire circuits. This "extra-high" impedance loading also had very satisfactory attenuation properties.

Subsequently, when the exploitation of the vacuum-tube repeater started on non-loaded open-wire lines it became necessary to use a new, low-impedance type of impedance-matching loading on their associated incidental cables. Because of the low impedance, the attenuation reduction was considerably less than that provided by the extra-high impedance loading just

mentioned. This, however, was not a serious limitation in the non-loaded repeater circuits. After open-wire loading became obsolete, further improvements in telephone repeaters and in transmission standards led to progressive improvements and refinements in the loading used on cables in non-loaded lines.

Beginning around 1920, the rapidly increasing use of open-wire telephone and telegraph carrier systems made it necessary to use in the associated incidental cables improved loading systems that provided good impedance-matching and attenuation-reduction properties over the complete voice and carrier-frequency bands used by the carrier transmission systems. The use of several different carrier telephone systems employing materially different frequency-band widths made it economically desirable in due course to use several different types of loading to provide the necessary transmission bandwidth through the incidental cables.

The various types of cable loading mentioned above are separately considered under suitable headings in the following pages. The two principal subdivisions of Part V are devoted to voice-frequency loading and to carrier loading, respectively. A third subdivision briefly describes a special type of voice-frequency phantom loading which is used in coordinated phantom-group combinations with side-circuit carrier loading systems that provide 10-kc and 30-kc transmission bands.

The importance of the incidental cable loading described in Part V of this article is due to its substantial, beneficial contributions to the transmission service-performance of the relatively expensive open-wire facilities, rather than from the amount of loading so employed. This is quite small relative to that used in voice-frequency toll cables and exchange cables.

#### (V-A) VOICE-FREQUENCY IMPEDANCE-MATCHING LOADING

Since the most important early uses of the vacuum-tube repeaters on open-wire facilities were on loaded lines, the first new impedance-matching loading system was developed for this particular use. As noted later, this had an important effect on the loading system subsequently developed for use on cables in non-loaded lines.

##### *Loading for Cables in Loaded Open-Wire Lines*

The new phantom-group loading for this use was designed to have closely the same values of nominal impedance and theoretical cut-off frequency as those of the loaded lines. The cable coil inductances had to be a little higher than the open-wire coil inductances, in consequence of the smaller amount of distributed inductance in the cable. A standard cable coil-spacing of about

5575 ft. (0.062 mf/mi cable) was adopted so as to have loading section capacitances close to those of the open-wire loading sections. This cable loading system originally known as "extra-heavy" loading, and later designated E248-154, was used on coarse-gauge cable conductors and had a slightly lower attenuation loss than that of the then standard "heavy" loading for coarse-gauge toll cables. (In the loading designation, *E* is the symbol for 5575-ft. spacing.)

The loading coils used 65-permeability iron-wire cores with two short, series air-gaps, to secure good magnetic stability.

The long obsolescence of open-wire loading makes further description of the E248-154 cable loading unimportant.

#### *Loading for Cables in Non-Loaded Open-Wire Lines*

When open-wire repeaters first came into general use, it was a common situation for entrance and intermediate cables to have one group of circuits associated with loaded open-wire pairs, and another group connected to non-loaded pairs. In such situations, it was obviously very desirable that the different types of cable loading associated with the loaded and the non-loaded lines should be installed at the same cable loading points.

#### *Early Standard Loading Systems*

*E28-16 Loading:* It was found that a satisfactory, low-impedance type of impedance-matching loading could be obtained by using 28 mh side circuit coils and 16 mh phantom coils at the spacing used for E248-154 loading. Some quantitative data regarding this low-impedance loading, designated E28-16, are included in Table XV (page 746) along with corresponding data on other voice-frequency loading systems subsequently standardized for cables in non-loaded lines.

*M44-25 Loading:* In many situations where the impedance-matching requirements were not so severe, and where loaded open-wire lines were not involved in the incidental cables along with the non-loaded lines, a somewhat cheaper type of low-impedance loading using a longer coil-spacing was utilized. Data regarding this loading, designated M44-25, are included in Table XV. (It is of interest that this type of loading had been used on a small scale prior to the extensive utilization of telephone repeaters.)

From Table XV it will be noted that the two loading systems had the same nominal impedance and that the better system, E28-16, had much higher cut-off frequencies. A brief digression regarding the important part which the cut-off frequency plays in the impedance-matching problem in the upper speech-frequency band is appropriate at this point.

### *Importance of Cut-Off Frequency*

Basically, the general impedance-matching problem under discussion is complicated by the fact that the non-loaded line is a "smooth," i.e., a uniform, line, whereas the loaded cable is a "lumpy" line. On the one hand, the sending-end impedance of the non-loaded line is substantially a constant resistance with negligible reactance over the frequency range above about 1 kc. On the other hand, the high-frequency impedance of the loaded cable may vary substantially in its resistance and reactance components with rising frequency, depending upon the type of loading termination employed. "Half-coil" and "mid-section" terminations have the important advantage of substantially negligible reactance, for which reason one or the other of them was used in the early applications of impedance-matching loading.<sup>(w)</sup> With these particular loading terminations, the resistance component of the loaded cable impedance changes with rising frequency, at a rapidly accelerating rate as the cut-off frequency is approached. The reference impedance in these changes is the nominal impedance of the loaded cable, which for optimum impedance-matching should be equal to that of the non-loaded line. (Numerically, the nominal impedance in ohms is equal to the square root of the ratio of the total circuit inductance, in henrys, to the total mutual capacitance, in farads, per unit length.) The resistance changes with rising frequency go up when mid-section termination is used, and drop down when half-coil termination is used.

The important practical significance of the foregoing is that the high-frequency impedance irregularities at the open-wire cable junction become progressively smaller as the loading cut-off frequency is raised (provided that the nominal impedances of the line and cable are closely alike). With the simple types of loading terminations above described, the requirements for good impedance-matching make it desirable to have much higher cut-off frequencies than those which are necessary from the standpoint of attenuation-frequency distortion in entrance and intermediate cables.

### *H28-16 Loading*

The discontinuance of the manufacture of open-wire loading coils about 1924, and the decreasing importance of open-wire loading, made it desirable to discontinue the use of the E-spaced loading solely for entrance and intermediate cables. Plant simplicity and flexibility requirements made it desirable to use H-spaced loading to permit coordination with the loading

<sup>(w)</sup> Half-coil termination involves the use of coils having one-half of the regular "full-coil" inductance at the end of the cable, followed in regular periodic sequence by "full" loading sections and "full" loading coils. In "mid-section" termination, the first full-coil is located one-half of a full loading section away from the end of the cable.



used on toll cable circuits along the same routes. Studies of these coordination possibilities resulted in the standardization of H28-16 entrance cable loading during 1927. Referring to Table XV, it will be seen that the change from 5575-ft. spacing to 6000-ft. spacing, using the same loading inductance values, resulted in a small drop (about 4%) in nominal impedance and theoretical cut-off frequency. A contemporary allied development made available new types of balancing networks which simulated the iterative impedances of the H28-16 loaded cables. The use of these new networks with repeaters at the office ends of long H28-16 loaded cables gave considerably better repeater balances than those obtained with open-wire balancing networks at the office ends of long E28-16 loaded cables. Up to this time, balancing networks which simulated the impedances of the associated open-wire lines had been used with the open-wire repeaters. This early practice was continued on open-wire lines having short entrance cables with H28-16 loading.

### *H31-18 Loading*

*General:* It was known prior to the standardization of H28-16 loading that the 28-16 mh loading inductances were not optimum from the impedance-matching standpoint for use at 6000-ft. spacing. However, it was appreciated that the concurrent development work on the compressed permalloy-powder core-material previously described (Section 9.1) was approaching completion and that a general size-reduction redesign of all toll cable and toll entrance loading coils would soon be undertaken. These considerations made it undesirable to develop for the H-spaced 28-18 loading new iron-dust core loading coils which would in all probability be superseded in a short time by permalloy-core coils. Thus it happened that the development work for the improved H31-18 loading system was coordinated with that on smaller-size, permalloy-core, loading coils having the necessary new inductance values for use in that system.

The H31-18 loading was designed to have a slightly higher nominal impedance than E28-16 loading, to make it more suitable for use on incidental cables in 104-mil open-wire lines, which were expected to be its principal field of use, since the more expensive, larger conductors (128 and 165-mil) were destined for use principally on a carrier basis and would require carrier loading on their incidental cables.

*Improved Loading Terminations:* It was also considered desirable to provide better impedance-matching characteristics at high voice-frequencies to assist in obtaining more satisfactory repeater operation on long-haul, multi-repeater, voice-frequency circuits which were becoming more common and more important in the rapid expansion of the open-wire plant. This require-

ment was met by providing improved loading terminations, which kept the resistance component of the cable impedance fairly close to the nominal impedance over the upper part of the frequency-band transmitted by the voice-frequency repeaters, and which also had satisfactory low reactance. Two different but equally satisfactory terminations<sup>12</sup> were developed, to provide flexibility and economy in the loading layouts. One of these was theoretically based on the mid-section termination previously described. This half-section termination was extended to about an 0.83-fractional section, followed at the open-wire junction by a terminal loading unit having inductance values about .36 of the full-weight loading inductances used in the loading designations. The other new loading termination was theo-

TABLE XV  
VOICE-FREQUENCY LOADING FOR INCIDENTAL CABLES IN NON-LOADED OPEN-WIRE LINES

Loading Designations	Coil Spacing (ft.)	Type of Circuit	Nominal Impedance (ohms)	Theoretical Cut-off Frequency (cycles)
E28-16	5575	Side	650	7250
		Phantom	400	7650
M44-25	8750	Side	650	4600
		Phantom	400	4900
H28-16	6000	Side	630	7000
		Phantom	380	7400
H31-18	6000	Side	666	6700
		Phantom	403	7000

Note: The full-coil inductances in millihenrys are given in the loading designations. The first number applies to the side circuits and the second number to the phantom circuit.

retically based on the mid-coil termination previously described. It used 0.86-fractional coils instead of half coils, and had a 0.36-fractional loading section adjacent to the open-wire side. These new loading terminations were known as "Fractional coil" or "Fractional section" terminations, depending on whether the fractional coil or the fractional section was the terminal element closest to the open-wire line. At the office ends of loaded entrance cables a mid-section loading termination was frequently used, and the repeater balancing network-circuits were adjusted to correspond with this situation in the line.

The H31-18 loading system was standardized in 1928 and is still the standard voice-frequency loading system for incidental cables in open-wire circuits which are not arranged or used for carrier operation.

*Loading Systems Data:* General transmission data regarding the loading systems briefly described above are given in Table XV.



*Attenuation Data:* The relatively low cable impedances which are necessary for good impedance-matching limit the attenuation-loss reduction to smaller values than those obtained with the higher impedance loading systems used on toll cable facilities. The theoretical 1000-cycle attenuation values for H31-18 loading (on a db/mi basis) are 0.56, 0.30, and 0.16, respectively, for 19 ga., 16 ga., and 13 ga. cables. The phantom circuit attenuation is nearly 20% lower, being about 0.47, 0.24, and 0.13 db/mi.

The attenuation losses in the other loading systems of Table XV are close to those for H31-18 loading. This follows from the fact that their impedances are nearly the same in magnitude.

#### *Low-Frequency Impedance Matching*

Before ending the discussion of transmission system characteristics, it is important to note that the attainment of optimum impedance-matches at low voice-frequencies, with the types of loading under discussion, involves the use of the so-called "optimum" cable conductor sizes. This follows from the fact that at these low frequencies the circuit resistances are important factors in determining the open wire and cable impedances. The optimum conductor combinations are 13 ga. cable for use in association with 165-mil open-wire lines, 16 ga. cable with 128-mil lines, and 19 ga. cable with 104-mil lines. Allowing for the loading coil resistances, these combinations of cable and open-wire conductor-sizes closely conform to the fundamental theoretical requirement that the unit-length ratio of series resistance (ohms) to shunt capacitance (farads) to total linear inductance (henrys) in the loaded cables should be close to the corresponding linear ratio in the associated non-loaded lines.

#### *Loading Coils and Cases for Incidental Cables*

In their general design features, excepting inductance and effective resistance, the voice-frequency loading coils for incidental cables corresponded with those currently used in toll cable circuits. When the toll cable coils were redesigned to take advantage of new core-materials, or in other important features, the entrance cable coils were included in the general redesign work.

The first loading units developed for H31-18 loading were coded in the "P" series. The code designation P4 applied to the "full-weight" loading unit. The fractional-weight loading units developed for use in the "fractional section" and the "fractional coil" loading termination were coded P5 and P6, respectively. These numerical code components have been retained in the code designations of all standard replacement designs, namely the PB, M, SM, and MF series of loading units.

The potting practices used with the entrance cable coils were generally

similar to those used with similar-sized toll cable coils. The potting complements for incidental cables, however, were small relative to the complements most generally used in the toll cables. Occasionally, in situations where toll entrance facilities and long-distance cable facilities shared the same cable for a short distance, the potting complements would include both types of loading. The color code used on the coil terminal quads in the stub cables of the loading coil cases facilitated identification of the different types of loading in the cable splicing-operations.

#### (V-B) CARRIER LOADING FOR INCIDENTAL CABLES IN OPEN-WIRE CARRIER SYSTEMS

##### *Historical*

The first open-wire carrier telephone system was installed late in 1918, and in the early 1920's general commercial use began to expand rapidly. A comprehensive account of the pioneering development work is given in a 1921 *A.I.E.E.* paper<sup>36</sup> by E. H. Colpitts and O. B. Blackwell.

Experimental types of carrier loading were made available for use on incidental cables in the open-wire lines on which the first carrier systems were installed. In general, these early carrier loading installations were engineered to specific job requirements.

*C4.1 and C4.8 Loading:* From this experience there evolved a quasi-standard loading treatment which served the current service needs, pending completion of the development of the first standard carrier loading systems, C4.1 and C4.8, late in 1923. These were designed to provide good impedance-matching up to a top frequency of about 30 kc. During the intervening years this loading has remained standard for incidental cables in carrier systems using this frequency-band, even though important changes have been made in the carrier systems themselves, notably the first Type C carrier systems<sup>37</sup> during the middle 1920's, and the improved Type C systems<sup>38</sup> during the late 1930's.

*B15 Loading:* During the late 1920's a lower cut-off carrier loading system designated B15 was designed especially for use with carrier facilities operating below a top frequency of about 10 kc. This loading served a double purpose. It was suitable for use with the old standard Type B carrier telegraph system<sup>36</sup> and with the new standard, single-channel, Type D carrier telephone system.<sup>39</sup> (In many of its early applications the Type B telegraph system used the frequency space between the voice circuit and the carrier telephone channels.) The B15 loading system is still in good standing. When an improved single-channel telephone system, Type H<sup>40</sup>, was developed during the late 1930's, its frequency allocation was chosen so that it could use "spare" B15 circuits which had become available on a substantial mileage of incidental cables.

*A2.7 and 3.0 Loading:* During the late 1920's the rapid expansion of carrier working led to extensive studies of the practicability of obtaining a larger number of telephone channels in long-haul carrier systems. These studies indicated a good prospect of using a wider frequency-band extending up to a top frequency of about 50 kc. In order to secure a much better control of intersystem crosstalk over the wider frequency-band, plans were made for spacing the wires of individual pairs much closer together, and for spacing adjacent pairs at greater distances apart. Also, improved transposition systems were designed for these new open-wire arrangements. In the period of interest, the open-wire plant was expanding very rapidly, and as a part of this expansion several entirely new pole lines were required for important long-haul service. These lines incorporated the improved construction features above mentioned. Even though the proposed new broader-band carrier systems were still in the "discussion stage" of development, it seemed desirable that a new type of broader-band loading should be installed on the incidental cables in the new pole lines, in order to avoid the larger expense of eventually replacing the 30-kc Type C loading, if it should be used initially. These considerations resulted in the rush development of the Types A2.7 and A3 carrier loading systems specifically to meet the impedance-matching requirements over the proposed 50-kc band. This loading was duly installed according to plan, but fate decreed that it should never be used for its originally intended purpose. Type C carrier telephone systems were immediately installed on the new lines, in the expectation of removal when broader-band systems became available, and the Type A loading was actually used only for 30-kc transmission.

The explanation for this turn of events was that before the final development requirements could be established for the proposed new 4 or 5-channel systems, some entirely new factors<sup>(x)</sup> entered the continuing studies and eventually resulted in a decision to develop a 12-channel system.<sup>41</sup> This was designed for placement above a Type C system on the same open-wire pair, making a total of 15 carrier channels above the voice-frequency circuit. The new broad-band carrier telephone system was coded in the "J" series. Its top working-frequency was about 143 kc.

*Type J Loading:* In due course, the development of new carrier loading was coordinated with the work on the new carrier telephone system. Three loading systems, designated J-0.72, J-0.85, and J-0.94, became available during 1937-1938 and are still in good standing, although they are not extensively used.

In the following pages, the general transmission characteristics of the Type C, B, and J loading systems are described, and some general informa-

<sup>(x)</sup> Including high-gain, high-stability, negative-feedback repeaters, and crystal filters.

tion is given regarding the loading apparatus and the building-out apparatus. The Type A systems are not included.

### Loading Systems Characteristics

*General:* A summary of loading systems characteristics is given in Table XVI, below. Attenuation data are given in Table XVII.

*Compensated Loading Terminations:* These loading systems are commonly known as compensated loading by virtue of their use of compensated loading terminations<sup>12</sup> to provide the desired impedance-matching characteristics at about the minimum cost. Over the working carrier-band the impedances of these compensated circuits closely approximate the non-reactive, flat, frequency-resistance characteristic of their "corresponding smooth lines"; that

TABLE XVI  
CARRIER LOADING FOR INCIDENTAL CABLES IN OPEN-WIRE CARRIER TELEPHONE SYSTEMS

Loading Designation	Approx. Top Working Freq. (kc.)	Theoretical Cut-off Freq. (kc.)	Nominal Impedance (ohms)	Theoretical Total Loading Section Capacitance (mmf.)	Theoretical Coil Spacing (ft.)	Representative Coil Spacing (ft.)	Full-Coil Inductance (mh.)
C4.1	30	45	590	12100	929*	740	4.09
C4.8	30	41.5	640	12100	929*	740	4.78
B15	10	13.5	640	36850	3000*	2800	14.7
J-0.72	142	208	542	3027	633†	500	0.72
J-0.85	142	190	575	3105	648†	500	0.85
J-0.94	142	181	600	3105	648†	500	0.94

Notes: \* In ordinary quadded cable having 0.062 mf/mi side circuit capacitance.

† In special 16 ga. disc insulated cable having 0.025 mf/mi capacitance.

is to say, the "lumpiness-of-loading" effects on the loaded cable impedance are reduced to tolerable low values over a predetermined frequency-band. By also having the nominal impedance of the loaded cable close to that of the associated open-wire line, satisfactory impedance-matches are obtained up to a much higher fraction of the loading cut-off than is possible with the more simple loading terminations used with the voice-frequency loading. In some carrier loading designs, this impedance-matching band extends up to about 0.75 of the cut-off frequency, or a little higher. An extension to still higher frequencies, relative to the cut-off frequency, would tend to result in objectionable "lumpiness-of-loading" attenuation impairments. The compensated loading terminations achieve substantial economies in the loading costs by permitting the use of much lower cut-off frequencies than would otherwise be feasible, thus allowing the full-weight coils to be spaced at much longer intervals.

Additional information regarding the loading terminations is given in the description of the terminal loading units which are used in these terminations.

*Control of Impedance Irregularities; Loading Layouts:* The carrier loading systems are engineered and installed to meet unusually severe limits on impedance irregularity among the individual loading sections and at the terminals. In installations involving more than one carrier system, it is especially desirable to restrict the individual impedance irregularities in order to control intersystem reflection crosstalk. The significance of this is understandable when one appreciates that usually the dominating reason for using carrier loading on the incidental cable is to avoid the objectionable reflection crosstalk that would result from the impedance irregularities caused by non-loaded cables. An additional important reason for the control of impedance irregularities is to avoid large humps in the insertion loss-frequency charac-

TABLE XVII  
CARRIER LOADING ATTENUATION DATA

Loading Designation	Cable Conductor Gauge	Cable Capacitance (mf/mi)	Attenuation—db/mi				
			1 kc	10 kc	30 kc	80 kc	140 kc
C4.1	13	0.062	0.28	0.39	0.92	—	—
	16	0.062	0.42	0.52	1.04	—	—
C4.8	16	0.062	0.40	0.50	1.14	—	—
	19	0.062	0.67	0.78	1.37	—	—
B15	16	0.062	0.35	0.54	—	—	—
	19	0.062	0.62	0.80	—	—	—
J-0.85	16	0.025	0.41	0.51	0.64	0.93	1.36

teristics which might cause objectionable frequency-distortion within the individual channels.

The procedure for controlling impedance irregularities in the loaded incidental cables involves the adjustment of the total capacitances of the individual loading sections to values close to the theoretical design values by means of adjustable building-out condensers. Ordinarily, a precision limit of about  $\pm 1\%$  is involved. To make these limits economically practicable, precision types of capacitance measuring-instruments have been made available, along with low cost building-out devices capable of simple precision adjustments.

The theoretical total loading capacitances for the different carrier loading systems are given in Table XVI, along with theoretical values of coil spacing in terms of the "nominal capacitance" of the usual type of cable involved. The actual geographical coil-spacing is usually well below this theoretical spacing because of the unavoidable capacitance deviations that occur in commercial paper-insulated cables. The loading layout procedure is such

that the highest-capacitance cable pairs in the individual loading sections will not have too much capacitance. When the loading is installed, the capacitances of the various pairs in the individual loading sections are measured, as also are the mutual capacitances of the loading coils and their associated stub cables, and then enough shunt capacitance is added to obtain the desired theoretical total capacitance, per loading section.

In many installations, especially in underground cables, the theoretically best loading points for the carrier loading coils (after engineering allowances have been made for cable-capacitance deviations) frequently occur at points where it would be unduly expensive to install the loading. In such instances, shortened spacings are used, and the building-out adjustments are increased to correct for the geographical spacing-deficiency along with the cable-capacitance deviations.

For reasons above mentioned, the individual coil-spacings may vary considerably in the same project, and the average coil-spacing may be quite different on different projects involving the same type of loading. The "representative coil-spacings" given in Table XVI are representative job averages.

As with the voice-frequency loading, the choice of cable conductor-gauge is important in the impedance-matching performance of the C4.1, C4.8, and B15 loaded cables at low voice-frequencies. The optimum resistance relations between the cable conductors and the open-wire conductors are the same as in voice-frequency loading. In the use of the Type J loading as practiced on short cables, this resistance-ratio question is unimportant because such loaded cables are substantially "transparent" at voice frequencies.

The loading terminations are unimportant factors in voice-frequency impedance-matching. This follows from the fact that the voice frequencies are low relative to the loading cut-off, for which reason the carrier loaded circuits act as electrically smooth lines in this range.

*Type C Loading:* These loading systems were designed for use on cable pairs connected to 12-inch spaced open-wire pairs. The impedances of the open-wire pairs vary substantially with the conductor size and because of this a single cable-loading system would not be satisfactory as regards carrier-frequency impedance-matching for all types of open-wire. The C4.1 system is used on cable pairs connected to 165-mil open-wire pairs. The C4.8 is a compromise system for use on cables connected to the less important and less expensive 128-mil and 104-mil open-wire pairs.

It is of interest that the theoretical coil-spacing for Type C loading is one-sixth of that of the E-spacing described in the discussion of voice-frequency impedance-matching loading.



Considerable Type C loading has been used on cables associated with open-wire pairs which have their conductors spaced 8 inches apart. The closer wire-spacing reduced the open-wire impedances below the values for which the carrier loading was originally designed. To obtain better impedance-matches when used with these lower-impedance lines, the Type C carrier loading was "*modified*" to have lower impedances by systematically building-out each loading section to have a higher total loading-section capacitance. This procedure also reduced the loading cut-off by an amount proportional to the impedance reduction, which effect limited the allowable impedance reduction. The "standard" modification of C4.1 loading dropped the nominal impedance to 558 ohms, and the cut-off to 42.5 kc. The "modification" of C4.8 loading dropped the nominal impedance to 625 ohms, and the cut-off frequency to 40.5 kc. The standard Type C loading apparatus was used in these installations.

*B15 Loading:* The single-channel open-wire carrier system with which this type of incidental cable loading is associated is a short-haul transmission system, principally used on 104-mil open-wire pairs. Since the impedance-matching requirements are much more lenient than those for loaded cables in the multi-channel systems a single weight of loading is sufficient.

The cable-capacitance deviations tend to be considerably smaller than with Type C loading, because of "random" splicing at a considerably larger number of intermediate cable splices within the individual loading sections. In consequence, the average amount of capacitance building-out is much smaller (on a percentage basis).

*Type J Loading:* Because of the higher frequencies involved in the Type J carrier systems the impedance-matching requirements are even more severe than those for the Type C systems. For this reason, a series of three Type J loading systems were made available, as noted in Table XVI.

To make carrier loading economically feasible for 140-kc transmission it was necessary to develop an entirely new type of low-capacitance cable for use with the loading. The new cable makes use of shielded, "spiral-four" units of 16 ga. conductors. The conductors are supported by means of insulating discs at the corners of a square, and the diagonally opposite conductors are associated as working pairs. The spacing between these wires and between them and the quad shields is such as to obtain a mutual capacitance very close to 0.025 mf/mi in the individual carrier pairs. The structural relations between the associated pairs of the individual units are such as to minimize crosstalk coupling. The over-all dimensions of the shielded units are such that not more than 7 or 8 units can be provided in a single cable without using an over-size sheath. In consequence the cable cost per carrier pair is high.

The low-capacitance construction, above described, also results in a much higher ratio of distributed inductance to distributed capacitance than that of paper insulated cables. This makes it necessary to build out the series inductance in a proper ratio to shunt capacitance, when geographical spacing-deviations require the use of corrective building-out adjustments. These capacitance-inductance adjustment devices are considerably more expensive than the relatively simple condensers used in the adjustments on Type C and B loaded circuits. Closer coil-spacing also makes the Type J loading more expensive. All in all, the total cost of the Type J loaded cable pairs is very high relative to that of the Type C loading. Furthermore, the attenuation-reduction feature of the loading, although it is substantial in magnitude per unit length, does not have a large economic value in reducing the number and cost of repeaters required in a complete carrier system. These considerations have limited the use of Type J loading to short cables, seldom more than 0.5 mile long.

In entrance-cable installations of greater length it is common practice to use line filters at the outer end of the cable to separate the "J" frequencies from the lower frequencies. The "J" frequencies are then transmitted to the office over non-loaded pairs terminated at each end in impedance modifying transformers. Separate cable pairs having Type C loading transmit the "C" carrier channels and the voice frequencies.

In such installations a special type of adjustable loading is used on the short "lead-in" cables from the bare open wire to the line filters, when they are installed in "filter huts" at the outer end of the cable. At the filter hut, this loading uses a continuously variable air-core inductance coil of the solenoidal, inductometer type, with which adjustable condensers are associated, one on each side of the coil. This provides a variable impedance loading which is adjustable for a predetermined range of impedances and for a predetermined range of lengths of lead-in cable. Long lead-in cables also require a (non-adjustable) loading unit at their open-wire end. The adjustments for optimum impedance-matching are made in terms of return-losses measured at the open-wire end of the lead-in cable.

### *Carrier Loading Apparatus*

*General:* The initial, experimental designs used large-size, toroidal-shaped, non-magnetic cores, and finely stranded copper conductors. These coils were nearly as large as the biggest coil shown in the headpiece, (page 149). Their construction made it possible to secure lower effective resistances at the high carrier-frequencies than could be obtained for the same total cost using the best magnetic materials then available. An additional advantage was that their non-magnetic cores could not cause non-linear distortion. This particu-



lar advantage assumed critical importance in later years when it became necessary to work to stringent over-all limits of non-linear distortion in the long-haul carrier facilities. As an example of the importance of controlling non-linear distortion, it became necessary during the late 1920's to mount the carrier loading coils in individual, shielding containers in order to prevent the small leakage fields of the toroidal air-core coils from penetrating nearby magnetic parts of the loading coil cases, thereby causing objectionable inter-channel modulation interference.

The satisfactory control of non-linear distortion has made it necessary to continue the use of non-magnetic cores in the carrier loading coils, notwithstanding the large improvements that have been made in magnetic core-materials during the last three decades. These improvements would make it possible to use much smaller coils without objectionably degrading the steady-state transmission performance. However, the hysteresis characteristics of the best available magnetic materials are such that if these materials should be employed it would be necessary to use coils larger and more expensive than the non-magnetic core coils, in order to meet present-day severe limits on allowable intermodulation interference in the Type C telephone systems.

#### *Types C and B Loading*

*Full Coils:* These loading systems use the same general types of full-weight loading coils and terminal loading units, except as regards their electrical parameters. The over-all dimensions of the full-weight coils are about  $6\frac{3}{8}$  inches in diameter and  $2\frac{1}{2}$  inches axial height. The shielding container has an over-all diameter of about  $7\frac{3}{4}$  inches and an axial height of  $3\frac{1}{4}$  inches.

*Terminal Loading Units:* The terminal loading units which provide the compensated loading terminations, previously mentioned, include a 0.82 fractional-weight series loading coil. This is shunted on the open-wire side (or office side) by a two-element network consisting of a condenser in series with an inductance coil, and located between the two half-windings of the coil. The complete terminal network may be regarded as an extension of half-coil termination. The portion beyond the half-coil point in the series (fractional) coil functions as an impedance corrective-network to produce the approximate "corresponding smooth line" impedance, previously described. The correct electrical proportioning of the elements of this corrective network is very important. The series loading coil is much smaller than the regular full-weight loading coil. Its size and those of the other network-elements are such as to allow the assembly of the complete terminal loading unit in the same size of shielding container as that used for the full-weight

loading coils. The standard potting complements range up to 16 coils or terminal units. The loading units are installed at the ends of full-length



Fig. 30—Various stages in the assembly of toroidal type carrier loading coils. Several coils mounted in their shielding containers are piled on a bench at right center. On the bench at left center, an assembly of coils is being connected to the stub cable conductors. In diagonal center, a completely assembled and cabled complement of 16 coils is ready for placement in a tall, rectangular shaped, cast iron case.

terminal loading sections. When short cables have only one loading section, terminal loading units are used at each end.

*Building-Out Condensers:* As previously indicated, building-out capacitance adjustments are extensively required in the control of local impedance irregularities, especially in the installations of the Type C loading.

In the office adjustments of the capacitance of the terminal loading sections, multi-unit paper-insulated condensers are employed. These condensers consist of ten different unit-condensers having six different nominal capacitance values, the ratio between the highest and lowest being of the order of about 30 to 1. Parallel combinations of the individual units are

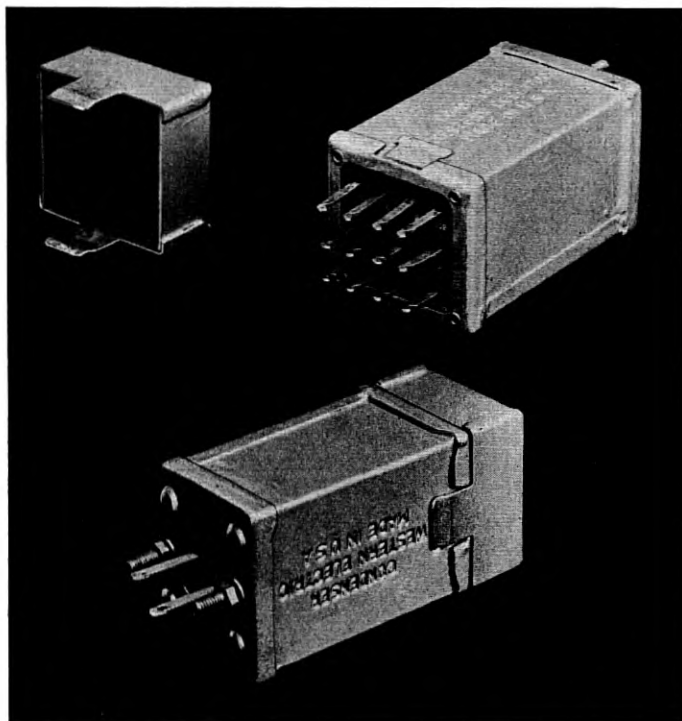


Fig. 31—Multi-unit, paper insulated building-out condenser for use in offices. Upper views show can cover for terminals of individual condensers removed to permit parallel cross connections, to obtain desired total capacitance. Lower view shows the complete assembly. The main terminals of the parallel connection of unit condensers appear at the left end in close proximity to the studs which are used in fastening the condenser case to the office mounting plates.

selected by measurement to provide the total required building-out capacitance, with the required precision.

The intermediate and open-wire terminal loading sections make use of small wire-wound and small mica condensers in the capacitance building-out adjustments. These are usually installed at a cable loading point, placed within the sleeve of the loading splice. The wire-wound condensers consist of parallel, insulated conductors wound in layer formation around small ceramic spools, and impregnated with moisture-resisting compound. Their

capacitances are continuously adjustable, by unwinding the outer end of the bifilar winding, and trimming off the excess length. The nominal capacitance

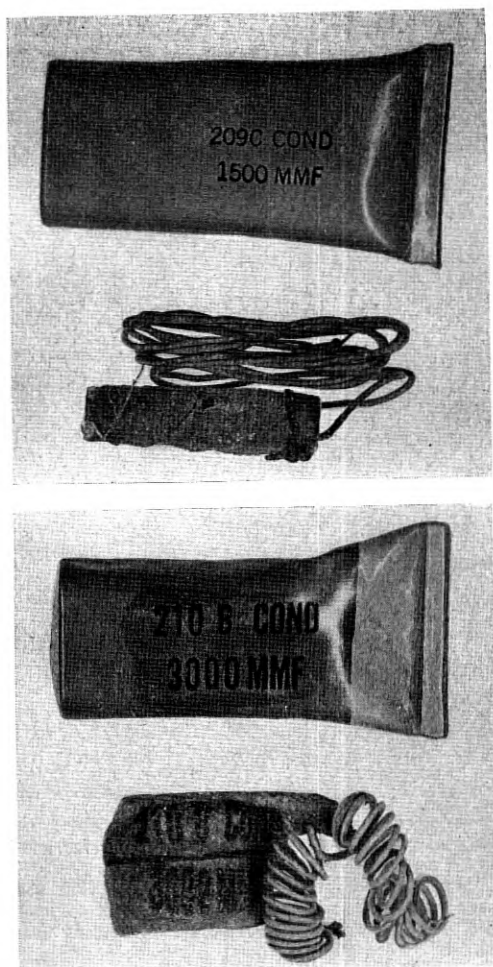


Fig. 32—Building-out condensers designed for installation within carrier loading splice sleeves. Upper view: Continuously adjustable, wire-wound condenser. Lower view: Non-adjustable, mica insulated condenser in small canvas bag. The upper part of each view shows the containers in which the condensers are placed, to protect them from moisture penetration and physical injury during the period intervening between manufacture and installation.

values, prior to adjustment, range from 500 mmf to 3000 mmf. In occasional instances where the total required capacitance cannot be provided by the highest-capacitance wire-wound condenser, a non-adjustable mica condenser

is used in parallel with a wire-wound condenser. In such combinations, the precision capacitance-adjustments are made with the wire-wound condenser. The nominal capacitances of the mica condensers range from 500 mmf to 4500 mmf.

Prior to the development of these small splice-installation types of building-out condensers (during the late 1930's), building-out stub cables were extensively used in the loading-section capacitance adjustments. Several pairs in these stubs were connected in parallel for use with an individual main cable pair. By varying the number of parallel pairs, and the length of

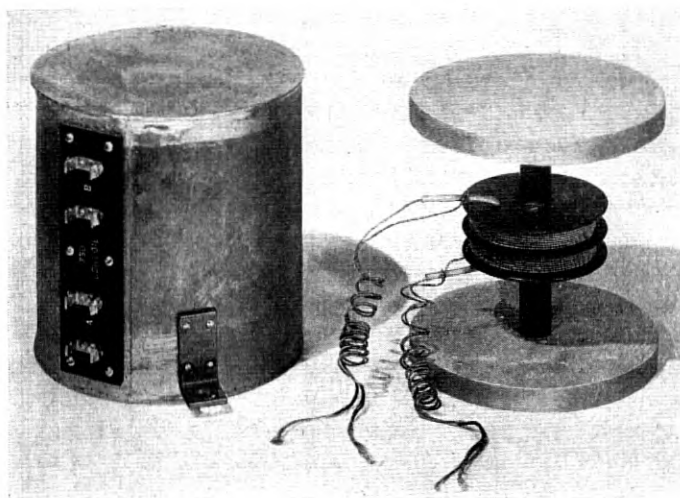


Fig. 33—Solenoidal type non-magnetic core loading coil used for type J carrier loading. At right: Internal coil structure and supports; At left: Copper shielding-container, with coil inside. This view shows the terminal strip on which the coil terminal clips are mounted, and also one of the brackets used in fastening the coil in position in the loading coil cases.

the stub cable, the necessary wide range of building-out capacitance was obtained with the required degree of precision.

### *Type J Loading*

*Full-Weight Loading Coils:* The full-weight loading coils are small solenoidal-type, air-core coils having a layer-type winding with a very finely stranded conductor, for control of coil resistance at the high "J" frequencies. The outside diameter and axial length are  $2\frac{3}{4}$  inches and 0.5 inch, respectively. To contain the external magnetic field, and control modulation effects and intercoil crosstalk that would otherwise result, a relatively large shielding container is required. Its over-all diameter is about  $5\frac{3}{8}$  inches and its axial

length about  $5\frac{3}{4}$  inches. The container dimensions are such that the small energy losses in the container-material have unobjectionable reactions upon the frequency-resistance and inductance characteristics of the coils.

A comparison of the effective resistance characteristics of representative toroidal and solenoidal types of carrier loading coils is of interest at this point.

Referring to table XVIII, the more favorable resistance values of the solenoidal coils at 30 kc. and above are due to greater refinements in the stranding of the copper conductors. The relatively small coil size, however, penalizes the low-frequency resistance; this is tolerable in the Type J systems because of the relative unimportance of the voice-frequency circuit.

*Terminal Loading Units:* For engineering flexibility in the loading layouts, and to minimize the cost of building out the terminal loading sections, two different, equally satisfactory, types of compensated loading terminations<sup>41</sup> are provided for use with Type J carrier loading systems. One of these is electrically analogous to that used with the 30-kc. and 10-kc. loading systems, and is theoretically based on the half-coil termination previously described. Lower inductance and capacitance elements are used because of the much wider carrier frequency-band. The alternative type of loading termination is theoretically based on half-section termination. It involves an extension of the terminal loading section from half-section to about 0.8 full-section and the use of a terminal loading unit which employs a fractional-weight loading coil (approx. 0.32 full-coil inductance) in series with the cable, and which has equal-capacitance condensers connected in parallel across each of the two line windings of the fractional coil.

TABLE XVIII  
EFFECTIVE RESISTANCE DATA—REPRESENTATIVE CARRIER LOADING COILS

Type of Coil	Nominal Inductance (mh)	Resistance in ohms per Millihenry at Specified Frequencies in Kilocycles				
		1	10	30	80	140
Toroidal.....	4.83	0.48	0.58	1.35	—	—
Toroidal.....	14.75	0.45	0.76	—	—	—
Solenoidal.....	0.85	0.95	0.95	1.1	1.3	2.0

At the junctions of cable and open wire, the cases which pot the shielded terminal units in pairs are mounted on crossarm fixtures in close proximity to the bare open wire. In office installations, the loading unit assemblies are mounted on individual panels for installation on an equipment bay in close proximity to the associated Type J system line filters.

*Building-Out Units:* The building-out apparatus used in conjunction with

Type J loading is radically different from that used with the Types C and B loading, primarily because it is usually desirable to include series inductance along with shunt capacitance, in proper proportions, because of the relatively high ratio of distributed inductance to distributed capacitance in the disc-

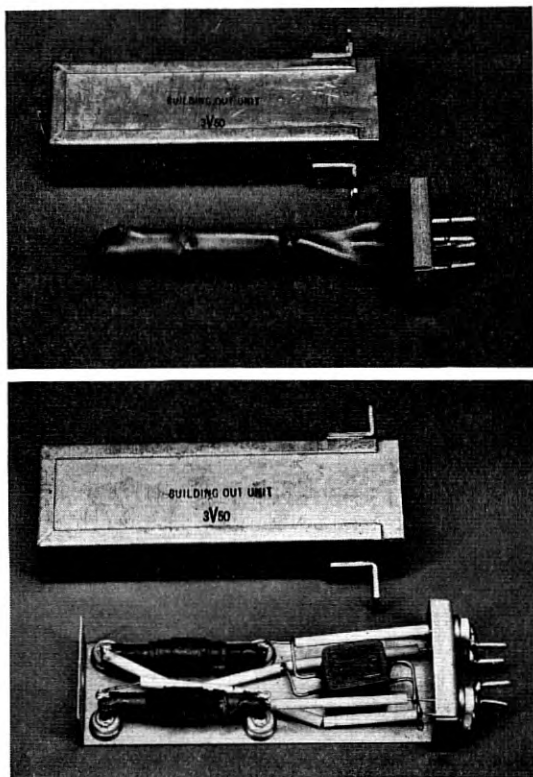


Fig. 34—Building-out units used in electrical adjustments of type J carrier loading sections. Upper View: Continuously-adjustable wire wound unit. Prior to adjustment it has a distributed shunt capacitance of about 275 mmf, and a series inductance of about 16.5 microhenrys; Lower View: Single section non-adjustable artificial line, providing a shunt capacitance of about 250 mmf (single condenser) and a total series inductance of about 15 microhenry (2-coils). This particular unit simulates a length of about 53 ft. of disc-insulated cable pair. Other (multi-section) artificial line units simulate longer lengths of cable pair.

insulated cable with which the loading is used (about 1.4 m.h. inductance and 0.025-mf capacitance, per mile).

Two types of building-out units are required, (1) a continuously adjustable, wire-wound unit which is used for making precision adjustments, and



(2) a graded series of non-adjustable artificial lines using lumped shunt condensers and lumped series inductances, these "lumps" being electrically small enough to avoid objectionable "lumpiness" effects. The required building-out units are connected in tandem with the loaded cable pair under adjustment.

The adjustable unit consists of a single-layer, bifilar winding on a non-magnetic spool about 4 inches long and  $\frac{3}{8}$  inch in diameter. The wiring of the unit to its two pairs of line terminals is such as to provide a series inductive-aiding connection of its line windings when the unit is serially inserted in the cable pair according to plan. The diameter of the spool is chosen to provide the desired ratio of inductance to capacitance per bifilar turn, taking into account the increase of capacitance of the winding which is caused by a wax-dipping process.

The inductance coils used in the non-adjustable units are very small air-core solenoids. Their inductances are adjusted (in manufacture) to have the correct electrical relations with the associated very small, mica-type, shunt condensers.

The complete building-out adjustment for an individual loading section usually involves the use of one or more artificial-line units in tandem with an adjustable unit. The adjustments are made in terms of capacitance measurements, since this procedure automatically provides the required series inductance. Capacitance measurements of the cable pair to be adjusted and of preselected non-adjustable units precede the precision adjustment of the wire-wound unit. This latter is accomplished by removing an integral number of bifilar turns from the winding, to meet the capacitance requirements, after which the shortened winding is reconnected to its main line terminals.

*Housing of Building-Out Units:* For flexibility in installation, the different electrical sizes of building-out units are "potted" in individual containers of the same size. These are much too large for installation in the loading splice-sleeves. Accordingly, in the cable-type cases for full-weight loading coils, and in the open-wire terminal pole cases for terminal loading units, space is provided in compartments with removable covers for the installation of the cable building-out units. The connections to the main cable circuit are made to terminal strips mounted in these compartments. Thus the installation of the building-out units can be made after the loading coils and loading units have been spliced to the disc-insulated incidental cables. The terminal loading units used at office ends of loaded entrance cables also include space and wiring provision for the installation of building-out units which may be required on the cable side of the terminal loading unit.



(V-C) VOICE FREQUENCY PHANTOM LOADING FOR COMBINATION WITH SIDE  
CIRCUIT CARRIER LOADING

Voice-frequency, impedance-matching, phantom circuit loading is available for use in coordinated combinations with C4.1, C4.8, and B15 carrier loading in situations where the need for improving the phantom circuit transmission in an incidental cable warrants the use of loading. This brings up a factor not previously mentioned, namely, that the early applications of carrier telephone systems made use of the side circuits of open-wire phantom groups. This is still the general situation, especially with the short-haul, single-channel systems. On the other hand, a substantial fraction of the Type C systems installed since the late 1920's, including those that now work in the frequency-band below a Type J system, uses open-wire pairs that are not arranged for phantom working.

The phantom loading under consideration is limited to voice-frequency operation because of the serious technical difficulties and high costs that would be involved in the satisfactory operation of carrier systems simultaneously on open-wire side circuits and their associated phantoms, and through the incidental cables.

The phantom group full-weight loading units and terminal loading units, which provide the phantom circuit loading, also include carrier loading apparatus for the associated side circuits; i.e., the phantom loading apparatus is not separately available. Thus, when phantom loading is required, it is necessary to engineer and install the loading on a carefully coordinated phantom-group basis.

The "full-coil" inductance of the phantom loading used in association with 30-kc. side circuit loading is 12.8 mh. and the full loading-section capacitance corresponds to that of "E" spacing. Thus its loading designation is E12.8, and the complete phantom-group loading designations become CE4.1-12.8 and CE4.8-12.8. The corresponding phantom circuit loading for use in association with B15 side circuit loading is designated H15, and the phantom group loading is designated BH15-15.

There must be an integral number of side circuit carrier loading sections in each voice-frequency phantom loading section. This ratio is 2 to 1 with *BH* loading. With *CE* loading, it may be 7 or 8 or 9, to 1, depending upon the average amount of building-out in the side circuits. This numerical variability with *CE* loading results from the fact that the condensers which are used primarily for building out the (carrier) side circuits add negligible capacitance to the phantom. An adjustable four-wire type of condenser is available for capacitance building-out adjustments of the phantom circuit. Depending upon the amount of capacitance building-out used in the carrier

side circuits, the average geographical spacing for the full-weight phantom loading units ranges from about one mile to nearly 6000 feet.

The voice-frequency attenuation in the loaded phantom circuits is appreciably lower than that in the associated side circuits. The small transmission impairments which the phantom loading apparatus introduces into the associated carrier side circuits are negligible. The voice-frequency impedance-matches between the loaded cable phantoms and the (non-loaded) open-wire phantoms are nearly as good as those obtained with voice-frequency phantom group loading.

An interesting feature of the phantom loading under discussion is that it uses a 2-coil scheme, with similar phantom coils in each side circuit at each phantom loading point. This scheme is one of several covered by the basic phantom loading patent (*U. S. No. 980,921*; Jan. 10, 1911) but was not used commercially in the Bell System until the late 1920's when very severe side-to-side crosstalk limits became necessary in phantom-group carrier installations for the control of high-frequency intersystem crosstalk. This control was strengthened by shielding the two associated phantom loading coils from one another.

#### BIBLIOGRAPHY (*Continued*)

6. B. Gherardi, "Commercial Loading of Telephone Circuits in the Bell System," *Trans. A.I.E.E.*, Vol. XXX, p. 1743, 1911.
8. Thomas Shaw and William Fondiller, "Development and Application of Loading for Telephone Circuits," *Trans. A.I.E.E.*, Vol. XLV; Published in *The Bell System Technical Journal*, Vol. V, pp. 221-281, April 1926.
12. R. S. Hoyt, "Impedance of Loaded Lines and Design of Simulating and Compensating Networks," *B.S.T.J.*, July 1924.
26. V. E. Legg and F. J. Given, "Compressed Powdered Molybdenum-Permalloy for High-Quality Inductance Coils," *B.S.T.J.*, Vol. XIX, p. 385, 1940.
30. S. G. Hale, A. L. Quinlan and J. E. Ranges, "Recent Improvements in Loading Apparatus for Telephone Cables," *Trans. A.I.E.E.*, Vol. 67, 1948.
36. E. H. Colpitts and O. B. Blackwell, "Carrier Current Telephony and Telegraphy," *Trans. A.I.E.E.*, Vol. XL, p. 205, 1921.
37. H. A. Affel, C. S. Demarest and C. W. Green, "Carrier Systems on Long Distance Lines," *Trans. A.I.E.E.*, Vol. 48, 1928; *B.S.T.J.*, Vol. VII, July 1928.
38. J. T. O'Leary, E. C. Blessing and J. W. Beyer, "A New Three-Channel Carrier Telephone System," *B.S.T.J.*, Vol. XVIII, Jan. 1939.
39. H. S. Black, M. L. Almquist and L. M. Ilgenfritz, "Carrier Telephone System for Short Toll Circuits," *Trans. A.I.E.E.*, Vol. 48, 1929.
40. H. J. Fisher, M. L. Almquist and R. H. Mills, "A New Single Channel Carrier Telephone System," *Trans. A.I.E.E.*, Jan. 1938; *B.S.T.J.*, Jan. 1938.
41. B. W. Kendall and H. A. Affel, "A Twelve Channel Carrier Telephone System for Open-Wire Lines," *B.S.T.J.*, Vol. XVIII, Jan. 1939.

(to be concluded)

## Contributors to This Issue

A. M. CLOGSTON, Massachusetts Institute of Technology, B.S. in Physics, 1938; Ph.D., 1941; from 1941-46 he worked on magnetrons in the Radiation Laboratory at M.I.T. Bell Telephone Laboratories, 1946-. Dr. Clogston is now doing research principally on electron tubes.

E. N. GILBERT, B.S. in Physics, Queens College, 1943; Massachusetts Institute of Technology Radiation Laboratory, 1944-46; Ph.D. in Mathematics, M.I.T., 1948. Bell Telephone Laboratories, 1948-. Dr. Gilbert has been concerned with mathematical problems of switching and communication theory.

F. K. HARVEY, B.E.E., New York University, 1939. Bell Telephone Laboratories, 1929-. In the Physical Research and Transmission Research Departments, Mr. Harvey has been chiefly concerned with the investigation and measurement of acoustical devices.

R. A. KEMPF, B.S. in Electrical Engineering, University of Illinois, 1937. Bell Telephone Laboratories, 1937-. Mr. Kempf is in the Outside Plant Development Department and has been with the Toll Cable group located at Point Breeze, Baltimore, Maryland since coming with the Laboratories, except for the period from 1941-45 when he was on active duty in the U. S. Navy.

WINSTON E. KOCK, B.E., University of Cincinnati, 1932; M.S., 1933; Ph.D., University of Berlin, 1934. Institute for Advanced Study, Princeton, New Jersey, 1935-36. Director of Electronic Research, Baldwin Piano Company, Cincinnati, Ohio, 1936-42. Bell Telephone Laboratories, Research Department, 1942-. Dr. Kock was engaged in radar antenna work in the Radio Research Department during the war. He is now engaged in microwave and acoustic research.

C. O. MALLINCKRODT, Washington University, B.S. in E.E., 1930. Bell Telephone Laboratories, 1930-April 1951. Mr. Mallinckrodt engaged in the development of carrier telephone repeaters and transmission regulators. During the war he worked on pulse modulation telephone systems which led to the development of the theory of instantaneous companders.

J. T. MAUPIN, B.S. in E.E., University of Kentucky, 1947. U. S. Air Force, 1943-46. Bell Telephone Laboratories, 1947-. Except for one year in the School for Communication Development Training, Mr. Maupin has been engaged in research and development work on telephone cables.

J. R. PIERCE, B.S. in Electrical Engineering, California Institute of Technology, 1933; Ph.D., 1936. Bell Telephone Laboratories, 1936-. Dr. Pierce has been engaged in the study of vacuum tubes.

WILLIAM J. PIETENPOL, University of Colorado, B.S. in Electrical Engineering, 1943; Radio Corporation of America, Lancaster, Pennsylvania, 1943-46. Ohio State University, 1946-49; Ph.D., 1949. Bell Telephone Laboratories, 1950-. Dr. Pietenpol is engaged in Transistor development.

H. H. SCHNECKLOTH, B.S. in Electrical Engineering, University of Minnesota, 1925. Western Electric Company, 1917-18; Northwestern Bell Telephone Company, 1918-29; American Telephone and Telegraph Company, Department of Development and Research, 1929-34; Bell Telephone Laboratories, 1934-. Mr. Schneckloth's work was in the manufacturing, maintenance, and equipment engineering phases of telephone switching prior to 1929. Since then he has been engaged in switching systems engineering and planning work.

THOMAS SHAW, S.B., Massachusetts Institute of Technology, 1905. American Telephone and Telegraph Company, Engineering Department, 1905-19; Department of Development and Research, 1919-33. Bell Telephone Laboratories, 1933-48. Mr. Shaw's active telephone career was mainly concerned with loading problems in telephone circuits, including the transmission and economic features of the loading apparatus. The article now being published was started shortly before his retirement in 1948.

ROBERT LEE WALLACE, JR., University of Texas, B.A. in Physics and Mathematics, 1936; M.A., 1939; Harvard University, 1939-45; Special Research Associate in the field of military communications, 1941-45. Bell Telephone Laboratories, 1946-. Mr. Wallace has been concerned with problems in magnetic recordings and with Transistors.