The Bell System Technical Journal

Vol. XX *April, 1941* *No. 2*

STEADY STATE SOLUTIONS OF TRANSMISSION LINE EQUATIONS

S. O. Rice

Methods of obtaining the steady state voltages and currents in a uniform transmission line consisting of several parallel wires are described in Part I. This line may or may not be acted upon by an externally impressed field distributed along its length. A square matrix Γ, which is a generalization of the propagation constant γ for a single circuit, is introduced. Matrix expressions obtained for the voltages and currents involve Γ in much the same way as the corresponding single circuit expressions involve γ. In Part II similar methods are described for obtaining the voltages and currents in a transmission line composed of a number of multi-terminal symmetrical sections connected in tandem. Expressions for the voltages and currents in a line composed of unsymmetrical sections are also given. These sections may or may not contain generators.

THE transmission lines considered here are of two kinds, namely the uniform transmission line, and the transmission line consisting of a number of identical sections connected in tandem. The problem discussed is that of determining the steady state electrical behavior of these lines when the terminal conditions are given. Often there arises the problem of determining the currents induced in a uniform transmission line by an arbitrary impressed field of some fixed frequency or of determining the currents produced by generators placed in the branches of the sections if the line is of the second kind. This is the type of problem with which we shall be particularly concerned.

In dealing with the uniform transmission line it is found convenient to introduce a matrix Γ, which is a generalization of the propagation constant γ for a single wire with ground return, or for a single circuit. This enables us to obtain matrix expressions for the currents and voltages which are similar in form to the single circuit expressions.

A similar situation exists for the transmission line composed of a number of symmetrical sections. However, when the sections are unsymmetrical the corresponding procedure does not appear to yield a corresponding simplification and the formulas are considerably more complicated than in the symmetrical case.

This paper is divided into two parts corresponding to the two kinds of transmission lines. The first part discusses the uniform line. After a statement of the transmission equations in matrix form, expressions for the voltages and currents are given. Two methods of evaluating these expressions are described. The first is based upon a property possessed by many transmission systems, namely that the various modes of propagation have nearly the same speed. The second method is based upon equations which may be obtained by the formal application of a theorem due to Sylvester. The first part concludes with the proof that these two methods lead to the correct results.

After a short introduction the second part discusses the difference equations which govern the transmission in a line composed of multi-terminal sections. The sections may contain generators. Expressions for the voltages and currents in a symmetrical section line, i.e. a line whose sections are symmetrical, are stated and proved in much the same order as the corresponding expressions for the uniform line. A discussion of the unsymmetrical section line concludes the second part.

A sketch of the solution of the uniform transmission line equations by the classical method is given in Appendix I. In Appendices II and III methods are described for solving the symmetrical section line difference equations. These methods are similar to the one of Appendix I. The method of Appendix III uses section constants which may be obtained from measurements made at one end of a typical section.

Part I

Uniform Transmission Lines

1.1 *Differential Equations*

For the sake of convenience in writing down equations we shall assume that the particular line under consideration consists of three parallel wires with ground return, or of three parallel circuits, denoted by the subscripts a, b, and c respectively. The differential equations for this line in an arbitrary impressed field are[1]

$$
\begin{aligned}
\frac{dv_a}{dx} &= -Z_{aa}i_a - Z_{ab}i_b - Z_{ac}i_c + l_a(x) \\
\frac{dv_b}{dx} &= -Z_{ba}i_a - Z_{bb}i_b - Z_{bc}i_c + l_b(x) \qquad (1.1) \\
\frac{dv_c}{dx} &= -Z_{ca}i_a - Z_{cb}i_b - Z_{cc}i_c + l_c(x)
\end{aligned}
$$

[1] These equations are given in substance by J. R. Carson and R. S. Hoyt, *B.S.T.J.*, Vol. 6, pp. 495–545 (1927). Equations (1.2) are equivalent to their equation (90) and equations (1.1) may be obtained by combining their equations (83), (84), and (94). We shall use the term "impressed field" to mean a field distributed along the line. According to our convention there is no impressed field when the line is energized only at the terminals.

and

$$\begin{aligned}\frac{di_a}{dx} &= -Y_{aa}v_a - Y_{ab}v_b - Y_{ac}v_c + t_a(x)\\ \frac{di_b}{dx} &= -Y_{ba}v_a - Y_{bb}v_b - Y_{bc}v_c + t_b(x)\\ \frac{di_c}{dx} &= -Y_{ca}v_a - Y_{cb}v_b - Y_{cc}v_c + t_c(x)\end{aligned} \tag{1.2}$$

where $Z_{ab} = Z_{ba}$, $Y_{ab} = Y_{ba}$, etc. If we are dealing with three parallel wires $l_a(x)$, $l_b(x)$, $l_c(x)$ are the longitudinal components of the electric force of the impressed field at the wire surfaces; $t_a(x)$, $t_b(x)$, $t_c(x)$ are specified by the admittance of the direct leakage paths and the values of the impressed potentials at the wires. If there are no direct leakage paths the t's are zero.

In order to put these equations in matrix form[2] we introduce the column matrices

$$v = \begin{bmatrix} v_a \\ v_b \\ v_c \end{bmatrix}, \quad i = \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix}, \quad l(x) = \begin{bmatrix} l_a(x) \\ l_b(x) \\ l_c(x) \end{bmatrix}, \quad t(x) = \begin{bmatrix} t_a(x) \\ t_b(x) \\ t_c(x) \end{bmatrix}, \tag{1.3}$$

and the symmetrical square matrices

$$Z = \begin{bmatrix} Z_{aa} & Z_{ab} & Z_{ac} \\ Z_{ba} & Z_{bb} & Z_{bc} \\ Z_{ca} & Z_{cb} & Z_{cc} \end{bmatrix} \qquad Y = \begin{bmatrix} Y_{aa} & Y_{ab} & Y_{ac} \\ Y_{ba} & Y_{bb} & Y_{bc} \\ Y_{ca} & Y_{cb} & Y_{cc} \end{bmatrix} \tag{1.4}$$

The equations (1.1) and (1.2) may now be written as

$$\begin{aligned}\frac{dv}{dx} &= -Zi + l(x)\\ \frac{di}{dx} &= -Yv + t(x)\end{aligned} \tag{1.5}$$

and these are the equations to be solved.

When there is no impressed field equations (1.5) give

$$\begin{aligned}\frac{d^2v}{dx^2} &= ZYv\\ \frac{d^2i}{dx^2} &= YZi\end{aligned} \tag{1.6}$$

[2] Cf. L. A. Pipes, *Phil. Mag.*, Vol. 24 (1937), p. 97.

and the analogy with the one circuit case leads us to put

$$\Gamma^2 = ZY, \qquad \Gamma = \sqrt{ZY} \tag{1.7}$$

where Γ is a square matrix representing a generalization of the propagation constant. Putting aside for the moment the question of interpreting the square root, we note that interchanging the rows and columns in $\Gamma^2 = ZY$ gives

$$\Gamma'^2 = Y'Z' = YZ, \qquad \Gamma' = \sqrt{YZ} \tag{1.8}$$

where the primes denote transposition. Y' and Z' are equal to Y and Z respectively because of their symmetry. We thus expect Γ' to be associated with the propagation of i in the same way that Γ is associated with the propagation of v.

1.2 Statement of Results for an Infinite Line—No Impressed Field

It is shown that when there is no impressed field the voltages and currents at any point x in a transmission line extending from $x = 0$ to $x = \infty$ are given by

$$\begin{aligned} v(x) &= e^{-x\Gamma}v(o) = e^{-x\Gamma}Z_oi(o) \\ i(x) &= e^{-x\Gamma'}i(o) \\ v(x) &= Z_oi(x) \end{aligned} \tag{1.9}$$

where $e^{-x\Gamma}$ is the square matrix defined by the convergent series of matrices[3]

$$e^{-x\Gamma} = I - \frac{x\Gamma}{1!} + \frac{x^2\Gamma^2}{2!} - \frac{x^3\Gamma^3}{3!} + \cdots \tag{1.10}$$

and $e^{-x\Gamma'}$ is the transposed of $e^{-x\Gamma}$. I denotes the unit matrix. Z_o is a square matrix and is called the characteristic impedance matrix:

$$Z_o = \Gamma^{-1}Z = \Gamma Y^{-1} \tag{1.11}$$

Additional expressions of the same type for Z_o are given by equations (1.45). The matrix $e^{-x\Gamma}Z_o$, being of the nature of a transfer impedance, is symmetrical.

The matrices $e^{-x\Gamma}$ and Z_o may be computed in several ways, the choice depending upon the circumstances. The first method to be described is useful when x is not too large and when the propagation constants of the various modes of propagation are nearly equal to each other. In the case of open-wire lines these propagation constants are grouped around the value $j\omega/v$ where v is of the order of 180,000 miles per second. The second method may be used for all cases, including those for which the series in

[3] Frazer, Duncan and Collar, "Elementary Matrices," Cambridge University Press, §2.5. In the work which follows, this text will be referred to as "F.D.C."

the first method converge too slowly to be of value. However, it requires the solution of an mth degree equation and the determination of the m modes of propagation where m is the number of circuits. For $m = 2$ this is no handicap and the method is quite convenient. In this case the method is closely related to one described by John Riordan in an unpublished memorandum.

First Method: Multiply the matrices Z and Y together to obtain ZY. Choose the number γ^2 in

$$ZY = I\gamma^2 + R, \tag{1.12}$$

where I is the unit matrix, so that the elements of R are small in comparison with γ^2. For many transmission lines it is possible to do this. Γ may be obtained by using the binomial theorem to expand the square root in the formula

$$\Gamma = \sqrt{ZY} = \gamma(I + \gamma^{-2}R)^{\frac{1}{2}}, \tag{1.13}$$

where γ is that square root of γ^2 whose real and imaginary parts are non-negative. In carrying out the work it is convenient to introduce the matrix S whose elements are small in comparison with unity.

$$\Gamma = \gamma(I + S) \tag{1.14}$$

To compute S, first compute the matrix $R/2\gamma^2$ and then use the power series

$$\begin{aligned} S = \left(\frac{R}{2\gamma^2}\right) - \frac{1}{2}\left(\frac{R}{2\gamma^2}\right)^2 + \frac{1}{2}\left(\frac{R}{2\gamma^2}\right)^3 - \frac{5}{8}\left(\frac{R}{2\gamma^2}\right)^4 \\ + \frac{7}{8}\left(\frac{R}{2\gamma^2}\right)^5 - \frac{21}{16}\left(\frac{R}{2\gamma^2}\right)^6 + \cdots \end{aligned} \tag{1.15}$$

This series will usually converge rapidly. The matrix $e^{-x\Gamma}$ is given by

$$e^{-x\Gamma} = e^{-z} \cdot e^{-zS} \tag{1.16}$$

where z is a number, $z = \gamma x$, and e^{-zS} is to be computed from

$$e^{-zS} = I - \frac{zS}{1!} + \frac{(zS)^2}{2!} - \frac{(zS)^3}{3!} + \cdots \tag{1.17}$$

$e^{-x\Gamma'}$ is obtained from $e^{-x\Gamma}$ by interchanging the rows and columns. The characteristic impedance matrix may be obtained from (1.11),

$$Z_o = \Gamma Y^{-1},$$

after computing Γ from S as in (1.14).

If only $e^{-x\Gamma}$ is required the following series may be used.

$$e^{-x\Gamma} = \sum_{p=0}^{\infty} \left(\frac{Rx}{2\gamma}\right)^p \frac{b_p(z)}{p!} \tag{1.18}$$

where R, γ, and z have the same meaning as above and the coefficients are computed from

$$b_o = e^{-z}, \qquad b_1(z) = -e^{-z}, \qquad b_2(z) = e^{-z}\left(1 + \frac{1}{z}\right)$$

$$b_{p+2}(z) = b_p(z) - \frac{2p+1}{z} b_{p+1}(z)$$

In the first term of the series $\left(\frac{Rx}{2\gamma}\right)^0$ denotes I.

Second Method: Γ, $e^{-x\Gamma}$ and Z_o may be regarded as functions of the square matrix ZY. In order to express these functions in a form suitable for calculation we apply Sylvester's theorem[4]. The characteristic matrix of ZY is

$$f(\gamma^2) = \gamma^2 I - ZY \tag{1.19}$$

where now γ^2 is regarded as a variable instead of a fixed number as in the first method. We shall suppose that ZY is a square matrix of order m and that the roots $\gamma_1^2, \gamma_2^2, \cdots \gamma_m^2$ of the characteristic function, i.e. of the determinantal equation

$$| f(\gamma^2) | = 0, \tag{1.20}$$

are distinct. Let the matrix $F(\gamma^2)$ be the adjoint of $f(\gamma^2)$ and denote the derivative of the characteristic function by

$$| f(\gamma^2) |^{(1)} = \frac{d}{d(\gamma^2)} | f(\gamma^2) | \tag{1.21}$$

Since $\gamma_1^2, \gamma_2^2, \cdots \gamma_m^2$ are all different $| f(\gamma_r^2) |^{(1)}$ is unequal to zero for $r = 1, 2, \cdots m$. Sylvester's theorem says that if $P(ZY)$ is any polynomial in ZY then

$$P(ZY) = \sum_{r=1}^{m} N(\gamma_r^2) P(\gamma_r^2) \tag{1.22}$$

where $P(\gamma_r^2)$ is a scalar (and thus deviates from our convention that capital letters denote square matrices). $N(\gamma_r^2)$ is a square matrix:

$$N(\gamma_r^2) = \frac{F(\gamma_r^2)}{| f(\gamma_r^2) |^{(1)}} \tag{1.23}$$

When $m = 2$, $N(\gamma_2^2)$ is equal to $I - N(\gamma_1^2)$.

[4] F.D.C. §3.9. The u and λ of the reference are the ZY and γ^2 of the present section.

Applying (1.22) to Γ, $e^{-x\Gamma}$ and Z_o even though they are not polynomials in ZY gives results which may be verified to be true.

$$\begin{aligned} \Gamma &= \sqrt{ZY} = \sum N(\gamma_r^2)\gamma_r \\ e^{-x\Gamma} &= e^{-x\sqrt{ZY}} = \sum N(\gamma_r^2)e^{-x\gamma_r} \\ Z_o &= (ZY)^{\frac{1}{2}}Y^{-1} = \sum N(\gamma_r^2)\gamma_r Y^{-1} \\ e^{-x\Gamma}Z_o &= \sum N(\gamma_r^2)\gamma_r e^{-x\gamma_r}Y^{-1} \end{aligned} \tag{1.24}$$

where the summations extend from $r = 1$ to $r = m$ and $\gamma_1, \gamma_2, \cdots \gamma_m$ are the square roots of $\gamma_1^2, \gamma_2^2, \cdots \gamma_m^2$ respectively whose real parts are non-negative. $\gamma_1, \gamma_2, \cdots \gamma_m$ are also the propagation constants of the "normal modes" of propagation. Some light is thrown on the physical significance of the matrix $N(\gamma_r^2)$ by supposing that only the rth normal mode is being propagated on the transmission line. $N(\gamma_r^2)$ is such that it can be expressed as a column matrix times a row matrix. The voltages in circuits 1, 2, $\cdots$ m are proportional to the first, second, $\cdots$ mth elements, respectively of the column matrix. The currents in circuits 1, 2, $\cdots$ m are proportional to the corresponding elements in the row matrix.

1.3 Results for Any Uniform Line—No Impressed Field

When the length of the line is finite the voltages and currents may be expressed as

$$\begin{aligned} v(x) &= \cosh x\Gamma\, v(o) - \sinh x\Gamma\, Z_o i(o) \\ i(x) &= -\sinh x\Gamma'\, Z_o^{-1}v(o) + \cosh x\Gamma'\, i(o) \end{aligned} \tag{1.25}$$

where Z_o and Γ have the same meaning as before. The matrices $\sinh x\Gamma Z_o$ and $\sinh x\Gamma' Z_o^{-1}$ are symmetrical. The square matrices $\cosh x\Gamma$ and $\sinh x\Gamma$ are defined by the series

$$\begin{aligned} \cosh x\Gamma &= I + \frac{x^2\Gamma^2}{2!} + \frac{x^4\Gamma^4}{4!} + \cdots \\ \sinh x\Gamma &= \frac{x\Gamma}{1!} + \frac{x^3\Gamma^3}{3!} + \cdots \end{aligned} \tag{1.26}$$

$\cosh x\Gamma'$ is obtained by interchanging the rows and columns of $\cosh x\Gamma$ and $\sinh x\Gamma'$ is obtained similarly from $\sinh x\Gamma$. Solving (1.25) for $v(o)$ and $i(o)$ gives

$$\begin{aligned} v(o) &= \cosh x\Gamma\, v(x) + \sinh x\Gamma\, Z_o i(x) \\ i(o) &= \sinh x\Gamma'\, Z_o^{-1}v(x) + \cosh x\Gamma'\, i(x) \end{aligned}$$

As in the case of the infinite line, we have two ways of computing the coefficients of $v(o)$ and $i(o)$ in the expressions (1.25) for $v(x)$ and $i(x)$.

First Method: Choose a number γ^2 and compute the matrices R, S, Γ, Z_o as described in the first method for the infinite line. The matrix $e^{x\Gamma}$ is given by

$$e^{x\Gamma} = e^{z} \cdot e^{zS}$$

where $z = \gamma x$ and e^{zS} is computed from the series

$$e^{zS} = I + \frac{zS}{1!} + \frac{z^2 S^2}{2!} + \cdots$$

If the elements of zS are so large that the series converges slowly it may be worthwhile to divide zS by 16, say, compute $\exp\left(\frac{zS}{16}\right)$ from the series, and then obtain e^{zS} by four matrix multiplications. When e^{zS} is known its inverse e^{-zS} can be computed and $e^{-x\Gamma}$ obtained from (1.16). The hyperbolic functions are given by

$$\begin{aligned} \cosh x\Gamma &= \tfrac{1}{2}\,(e^{x\Gamma} + e^{-x\Gamma}) \\ \sinh x\Gamma &= \tfrac{1}{2}\,(e^{x\Gamma} - e^{-x\Gamma}) \end{aligned} \tag{1.27}$$

which follow from the series definitions of the various matrices.

If only the coefficients in (1.25) are required we may choose γ^2 and compute R and powers of the matrix $Rx/2\gamma$. Then the coefficients in (1.25) are given by

$$\begin{aligned} \cosh x\Gamma &= \sum_{p=0}^{\infty} \left(\frac{Rx}{2\gamma}\right)^p \frac{a_p(z)}{p!} \\ \sinh x\Gamma\, Z_o &= \sum_{p=0}^{\infty} \left(\frac{Rx}{2\gamma}\right)^p \frac{a_{p+1}(z)}{p!\gamma}\, Z \\ \sinh x\Gamma' Z_o^{-1} &= \sum_{p=0}^{\infty} \left(\frac{R'x}{2\gamma}\right)^p \frac{a_{p+1}(z)}{p!\gamma}\, Y \end{aligned} \tag{1.28}$$

where R' is the transposed of R, and the scalar coefficient $a_p(z)$ is a function of $z = \gamma x$ given by

$$\begin{gathered} a_o(z) = \cosh z \qquad a_1(z) = \sinh z \\ a_2(z) = \cosh z - \frac{\sinh z}{z} \\ a_{p+2}(z) = a_p(z) - \frac{2p+1}{z}\, a_{p+1}(z), \end{gathered} \tag{1.29}$$

and it is understood that $(Rx/2\gamma)^0 = I$.

Second Method: Compute the propagation constants $\gamma_1, \gamma_2, \cdots \gamma_m$ and the square matrices $N(\gamma_r^2)$ given by (1.23) as in the second method for the infinite line. Then

$$\begin{aligned} \cosh x\Gamma &= \Sigma\, N(\gamma_r^2) \cosh x\gamma_r \\ \sinh x\Gamma\, Z_o &= \sinh x\Gamma\, \Gamma Y^{-1} \\ &= \Sigma\, N(\gamma_r^2) \sinh x\gamma_r\, \gamma_r Y^{-1} \\ \sinh x\Gamma'\, Z_o^{-1} &= \sinh x\Gamma'\, \Gamma'^{-1} Y \\ &= \sum N'(\gamma_r^2) \frac{\sinh x\gamma_r}{\gamma_r} Y \end{aligned} \tag{1.30}$$

where $N'(\gamma_r^2)$ is the transposed of $N(\gamma_r^2)$, $N(\gamma_r^2)$ being defined by (1.23), and the summations extend from $r = 1$ to $r = m$.

When the transmission line consists of perfectly conducting wires strung on perfect insulators over a perfectly conducting earth the magnetic and electrostatic fields are related so as to make Z equal to $\gamma_o^2 Y^{-1}$ where

$$\gamma_o = j\omega/c,$$

ω being 2π times the frequency and c the speed of light.

It is interesting to apply the first method of solution to this line. Even though the proof of the first method, which is given in §1.10, does not cover this case there seems to be little doubt that the correct answer is obtained. We have

$$ZY = \gamma_o^2 I$$

Choosing $\gamma = \gamma_o$ gives $R = 0$ and therefore $S = 0$. It follows that

$$\begin{aligned} &\Gamma = \gamma_o I,\ Z_o = \Gamma^{-1} Z = \gamma_o^{-1} Z \\ &\cosh x\Gamma = \cosh (x\gamma_o I) = \cosh x\gamma_o\, I \\ &\sinh x\Gamma\, Z_o = \sinh x\gamma_o\, \gamma_o^{-1} Z \\ &\sinh x\Gamma'\, Z_o^{-1} = \sinh x\gamma_o\, \gamma_o Z^{-1} \end{aligned}$$

When these are put into equations (1.25) the expressions for $v(x)$ and $i(x)$ in a perfect transmission line are obtained:

$$\begin{aligned} v(x) &= \cosh x\gamma_o v(o) - \frac{\sinh x\gamma_o}{\gamma_o} Zi(o) \\ i(x) &= -\gamma_o \sinh x\gamma_o Z^{-1} v(o) + \cosh x\gamma_o i(o) \end{aligned} \tag{1.31}$$

1.4 Results for Any Uniform Line—Impressed Field

The differential equations to be satisfied in this case are given by (1.5). A solution which reduces to $v(o)$ and $i(o)$ at $x = 0$ is

$$\begin{aligned} v(x) &= \cosh x\Gamma\, v(o) - \sinh x\Gamma\, Z_o i(o) \\ &\quad + \int_0^x \cosh (x - \xi)\Gamma\, l(\xi)\, d\xi - \int_0^x \sinh (x - \xi)\Gamma\, Z_o\, t(\xi)\, d\xi \\ i(x) &= -\sinh x\Gamma'\, Z_o^{-1} v(o) + \cosh x\Gamma'\, i(o) \\ &\quad - \int_0^x \sinh (x - \xi)\Gamma' Z_o^{-1} l(\xi)\, d\xi + \int_0^x \cosh (x - \xi)\Gamma' t(\xi)\, d\xi \end{aligned} \tag{1.32}$$

The matrices $\cosh x\Gamma$, $\sinh x\Gamma$ and Z_o are the same as the ones discussed in §1.2 and §1.3. The elements of the integral[5] of a matrix U (U is not necessarily a square matrix) are given by the integrals of the corresponding elements of U.

In many cases of practical interest the impressed field varies exponentially with respect to x. The column matrices $l(x)$ and $t(x)$ may then be expressed as

$$l(x) = e^{-x\theta} \begin{bmatrix} \lambda_a \\ \lambda_b \\ \lambda_c \end{bmatrix} \qquad t(x) = e^{-x\theta} \begin{bmatrix} \tau_a \\ \tau_b \\ \tau_c \end{bmatrix} \tag{1.33}$$

where the λ's and τ's are constants and θ is the propagation constant of the impressed field in the direction of the line. The integrations in the expressions (1.32) may be performed with the result

$$\begin{aligned} v(x) &= \cosh x\Gamma\, v(o) - \sinh x\Gamma\, Z_o i(o) \\ &\quad + \tfrac{1}{2}\,(e^{x\Gamma} - e^{-x\theta} I)\,(\Gamma + \theta I)^{-1}(\lambda - Z_o\tau) \\ &\quad - \tfrac{1}{2}\,(e^{-x\Gamma} - e^{-x\theta} I)\,(\Gamma - \theta I)^{-1}(\lambda + Z_o\tau) \\ i(x) &= -\sinh x\Gamma'\, Z_o^{-1} v(o) + \cosh x\Gamma'\, i(o) \\ &\quad + \tfrac{1}{2}\,(e^{x\Gamma'} - e^{-x\theta} I)\,(\Gamma' + \theta I)^{-1}(\tau - Z_o^{-1}\lambda) \\ &\quad - \tfrac{1}{2}\,(e^{-x\Gamma'} - e^{-n\theta} I)\,(\Gamma' - \theta I)^{-1}(\tau + Z_o^{-1}\lambda) \end{aligned} \tag{1.34}$$

provided that the inverse matrices exist. The matrix $(e^{x\Gamma'} - e^{-x\theta}I)(\Gamma' + \theta I)^{-1}$ is the transposed of $(e^{x\Gamma} - e^{-x\theta}I)(\Gamma + \theta I)^{-1}$, etc. If one of these matrices, say $\Gamma - \theta I$, has no inverse then it is necessary to evaluate the

[5] F.D.C. §2.10.

corresponding integral in some other way. Thus it may be advantageous to use the formula

$$
\begin{aligned}
-(e^{-x\Gamma} - e^{-x\theta}I)(\Gamma - \theta I)^{-1} &= e^{-x\Gamma} \int_0^x e^{\xi\Gamma - \xi\theta I}\, d\xi \\
&= e^{-x\Gamma}\left[xI + \frac{x^2}{2!}(\Gamma - \theta I) + \frac{x^3}{3!}(\Gamma - \theta I)^2 + \cdots\right]
\end{aligned}
\tag{1.35}
$$

Two special cases of (1.34) are of interest. When the line is shorted at both ends, $v(o) = v(x) = 0$, where x is the line length, and

$$
\begin{aligned}
i(o) = \tfrac{1}{2} Z_o^{-1} (\sinh x\Gamma)^{-1}[(e^{x\Gamma} - e^{-x\theta}I)(\Gamma + \theta I)^{-1}(\lambda - Z_o\tau) \\
- (e^{-x\Gamma} - e^{-x\theta}I)(\Gamma - \theta I)^{-1}(\lambda + Z_o\tau)]
\end{aligned}
$$

$$
\begin{aligned}
i(x) = \frac{e^{-x\theta}}{2} Z_o^{-1} (\sinh x\Gamma)^{-1}[(e^{x\Gamma} - e^{x\theta} I)(\Gamma - \theta I)^{-1}(\lambda + Z_o\tau) \\
- (e^{-x\Gamma} - e^{x\theta}I)(\Gamma + \theta I)^{-1}(\lambda - Z_o\tau)]
\end{aligned}
$$

When the line is terminated in its characteristic impedance at both ends, $v(o) = -Z_o i(o)$, $v(x) = Z_o i(x)$, and

$$
\begin{aligned}
i(o) &= \tfrac{1}{2} (I - e^{-x\theta}e^{-x\Gamma'})(\Gamma' + \theta I)^{-1}(Z_o^{-1}\lambda - \tau) \\
i(x) &= -\tfrac{1}{2} (e^{-x\Gamma'} - e^{-x\theta}I)(\Gamma' - \theta I)^{-1}(Z_o^{-1}\lambda + \tau)
\end{aligned}
$$

The matrices occurring in the expressions (1.34) for $v(x)$ and $i(x)$ may be computed by the first or second method described for the uniform line in the absence of an impressed field. The second method involves the use of expansions similar to

$$
\begin{aligned}
&(e^{x\Gamma} - e^{-x\theta}I)(\Gamma + \theta I)^{-1}(\lambda - Z_o\tau) \\
&\qquad = \sum N(\gamma_r^2)\left(\frac{e^{x\gamma_r} - e^{-x\theta}}{\gamma_r + \theta}\right)\left(\lambda - \frac{Z}{\gamma_r}\tau\right) \\
&(e^{x\Gamma'} - e^{-x\theta}I)(\Gamma' + \theta I)^{-1}(Z_o^{-1}\lambda - \tau) \\
&\qquad = \sum N'(\gamma_r^2)\left(\frac{e^{x\gamma_r} - e^{-x\theta}}{\gamma_r + \theta}\right)\left(\frac{Y}{\gamma_r}\lambda - \tau\right)
\end{aligned}
\tag{1.36}
$$

where the summations run from $r = 1$ to $r = m$ and $N'(\gamma_r^2)$ is the transposed of the square matrix $N(\gamma_r^2)$ given by (1.23). In obtaining these expansions by Sylvester's theorem, Z_o in the first is replaced by $\Gamma^{-1}Z$ and Z_o^{-1} in the second by $\Gamma'^{-1}Y$.

If we assume that an impressed field acts upon the perfect transmission line of equations (1.31), we see that $t(x) = 0$ because there are no direct leakage paths. We may also write

$$(e^{x\Gamma} - e^{-x\theta}I)(\Gamma + \theta I)^{-1} = (e^{x\gamma_o}I - e^{-x\theta}I)(\gamma_o I + \theta I)^{-1}$$
$$= \frac{e^{x\gamma_o} - e^{-x\theta}}{\gamma_o + \theta} I$$

From this and similar equations it follows that

$$v(x) = \cosh x\gamma_o v(o) - \frac{\sinh x\gamma_o}{\gamma_o} Zi(o) + \frac{1}{2}\left[\frac{e^{x\gamma_o} - e^{-x\theta}}{\gamma_o + \theta} - \frac{e^{-x\gamma_o} - e^{-x\theta}}{\gamma_o - \theta}\right]\lambda \tag{1.37}$$
$$i(x) = -\gamma_o \sinh x\gamma_o Z^{-1} v(o) + \cosh x\gamma_o i(o) - \frac{\gamma_o}{2}\left[\frac{e^{x\gamma_o} - e^{-x\theta}}{\gamma_o + \theta} + \frac{e^{-x\gamma_o} - e^{-x\theta}}{\gamma_o - \theta}\right] Z^{-1}\lambda$$

1.5 Results for Infinite Uniform Line—Impressed Field

When the line extends from $x = 0$ to $x = \infty$ and the impressed field is such that the voltages and currents remain finite at $x = \infty$, the appropriate solutions may be obtained from the results of §1.4 by a limiting process. The condition that $v(x)$ remain finite suggests that the coefficient of $e^{x\Gamma}$ be zero in the expression (1.32) for $v(x)$. This gives a relation between $v(o)$ and $i(o)$ which must be satisfied:

$$v(o) = Z_o i(o) - \int_0^\infty e^{-\xi\Gamma}[l(\xi) - Z_o t(\xi)]\, d\xi \tag{1.38}$$

If the impressed field varies exponentially with x expression (1.34) gives

$$v(o) = Z_o i(o) - (\Gamma + \theta I)^{-1}(\lambda - Z_o\tau) \tag{1.39}$$

Expressions for $v(x)$ and $i(x)$ may be obtained by using relations (1.38) and (1.39) in (1.32) and (1.34) respectively. As these are somewhat lengthy we shall state only two which follow from (1.39).

$$v(x) = e^{-x\Gamma} v(o) + \tfrac{1}{2}(e^{-x\Gamma} - e^{-x\theta}I)[(\Gamma + \theta I)^{-1}(\lambda - Z_o\tau) - (\Gamma - \theta I)^{-1}(\lambda + Z_o\tau)]$$
$$i(x) = Z_o^{-1} e^{-x\Gamma} v(o) + \tfrac{1}{2}Z_o^{-1}(e^{-x\Gamma} + e^{-x\theta}I)(\Gamma + \theta I)^{-1}(\lambda - Z_o\tau) - \tfrac{1}{2}Z_o^{-1}(e^{-x\Gamma} - e^{-x\theta}I)(\Gamma - \theta I)^{-1}(\lambda + Z_o\tau) \tag{1.40}$$

Two similar expressions may be obtained in which the initial current $i(o)$ instead of $v(o)$ appears on the right. If the line is terminated in its characteristic impedance at $x = 0$, $v(o) = -Z_o i(o)$, and the voltages and currents produced by the impressed field are

$$\begin{aligned} v(o) &= -\tfrac{1}{2}(\Gamma + \theta I)^{-1}(\lambda - Z_o\tau) \\ i(o) &= \tfrac{1}{2} Z_o^{-1}(\Gamma + \theta I)^{-1}(\lambda - Z_o\tau) \end{aligned} \tag{1.41}$$

As in §1.4 these expressions may be computed by the first and second methods described in §1.3. For example, the application of the second method to the relation (1.39) which must exist between $v(o)$ and $i(o)$ in an infinite line gives

$$v(o) = \sum_{r=1}^{m} N(\gamma_r^2)\left[\frac{Z}{\gamma_r} i(o) - \frac{1}{\gamma_r + \theta}\left(\lambda - \frac{Z}{\gamma_r}\tau\right)\right] \tag{1.42}$$

where $N(\gamma_r^2)$ is the square matrix (1.23).

1.6 Outline of Proofs

The proof of the results which have been stated is divided into three parts. In the first part it is shown that if Γ is a matrix such that (a) its square is ZY and (b) every element in the matrix $e^{-x\Gamma}$ approaches zero as $x \to \infty$, then the expressions for $v(x)$ and $i(x)$ involving Γ and Z_o satisfy the transmission line equations. In the second part of the proof it is shown that if certain requirements are met Γ as obtained by the first method satisfies the conditions (a) and (b) and hence the expressions for $v(x)$ and $i(x)$ given by the first method are correct. The third part of the proof discusses a general procedure which may be used to prove the equations which constitute the second method.

Both the second and the third parts of the proof are based upon the solution of the transmission line equations which is sketched in Appendix I. This solution assumes that the propagation constants of the normal modes of propagation are unequal, and our proofs are limited accordingly. However, considerations of continuity seem to show that the first method is valid even when two or more propagation constants are equal. Under the same circumstances the second method suggests the use of the confluent form of Sylvester's theorem.[6]

1.7 Relations Obtained by Considering An Infinite Line

We suppose that we are going to deal with transmission lines possessing the non-singular, symmetrical impedance and admittance matrices Z and Y. We further suppose that, by some means or other, we have determined a matrix Γ which satisfies the two conditions; (a) the square of Γ is

$$\Gamma^2 = ZY, \tag{1.43}$$

and (b) every element in the matrix $e^{-x\Gamma}$ approaches zero as $x \to \infty$.

[6] F.D.C. §3.10.

Consider a line extending from $x = 0$ to $x = \infty$, there being no impressed field. Viewing the line at $x = 0$ as an n terminal network shows that there is a symmetrical matrix Z_o such that $v(o) = Z_o i(o)$. Let this be taken as the definition of the characteristic impedance matrix Z_o. We shall show from the differential equations of the line that

1. The voltages and currents in the infinite line are given by

$$\begin{aligned} v(x) &= e^{-x\Gamma} v(o) \\ i(x) &= e^{-x\Gamma'} i(o) \end{aligned} \tag{1.44}$$

2. The matrix Z_o satisfies the relations

$$\begin{aligned} Z_o &= \Gamma^{-1} Z = Z\Gamma'^{-1} = \Gamma Y^{-1} = Y^{-1}\Gamma' \\ Z_o^{-1} &= Z^{-1}\Gamma = \Gamma' Z^{-1} = Y\Gamma^{-1} = \Gamma'^{-1} Y \end{aligned} \tag{1.45}$$

$$v(x) = Z_o i(x) \tag{1.46}$$

3. The matrices Z_o, Z, and Y obey the commutation rules

$$\begin{aligned} \Phi(\Gamma) Z_o &= Z_o \Phi(\Gamma') \\ \Phi(\Gamma) Z &= Z \Phi(\Gamma') \\ Y\Phi(\Gamma) &= \Phi(\Gamma') Y \end{aligned} \tag{1.47}$$

where $\Phi(\Gamma)$ is any square matrix, such as $e^{-x\Gamma}$, representable as a convergent power series in Γ with scalar coefficients. Furthermore, the matrices $\Phi(\Gamma)Z_o$, $\Phi(\Gamma)Z$, and $Y\Phi(\Gamma)$ are symmetrical.

The differential equations of the transmission line are

$$\frac{dv}{dx} = -Zi, \qquad \frac{di}{dx} = -Yv, \qquad \frac{d^2 v}{dx^2} = ZYv \tag{1.48}$$

the third following from the first two when i is eliminated. That $v(x) = e^{-x\Gamma} v(o)$ is a solution of the third equation may be verified by direct substitution and differentiation[7]. Since this expression for $v(x)$ approaches zero as $x \to \infty$ and reduces to $v(o)$ at $x = 0$, it represents the voltages in an infinite transmission line. Hence the first equation in (1.44) is true. Setting it in the first differential equation of (1.48), putting $x = 0$, replacing $v(o)$ by $Z_o i(o)$, and noting that $i(o)$ may be regarded as an arbitrary column gives

$$\Gamma Z_o = Z \tag{1.49}$$

Since Γ was assumed to be non-singular, Z_o is equal to $\Gamma^{-1}Z$. Z is symmetrical and the reciprocity theorem for electrical networks requires that Z_o

[7] The differentiation of the exponential function is discussed in F.D.C. §2.7.

be symmetrical, hence

$$Z_o = \Gamma^{-1} Z = Z \Gamma'^{-1}$$

The first group of equations in (1.45) follow from this together with the expression $\Gamma^2 Y^{-1}$ for Z obtained from (1.43). The second group in (1.45) is obtained from the first group.

The commutation rule for Z_o is obtained from (1.49) together with the equation obtained from (1.49) by transposition. Since Z is symmetrical

$$\Gamma Z_o = Z_o \Gamma', \qquad \Gamma^2 Z_o = \Gamma Z_o \Gamma' = Z_o \Gamma'^2,$$

$$\Gamma^n Z_o = Z_o \Gamma'^n$$

and the first of equations (1.47) follow from this. The second and third of equations (1.47) may be obtained similarly from the relations (1.45). The matrix $\Phi(\Gamma) Z_o$ is symmetrical since its transposed is $Z_o\ [\Phi(\Gamma)]'$ and this is equal to $Z_o \Phi(\Gamma') = \Phi(\Gamma) Z_o$. A similar argument applies to the other matrices in (1.47).

The expression for $i(x)$ in (1.44) may be obtained by Maclaurin's expansion. Setting $x = 0$ in the second differential equation of (1.48),

$$\left(\frac{di}{dx}\right)_o = -Yv(o) = -YZ_o i(o) = -\Gamma' i(o)$$

where we have used the equality between the first and last members of the first equation of (1.45) and where the subscript 0 denotes the value of the derivative at $x = 0$. Repeated differentiation gives

$$\frac{d^2 i}{dx^2} = -Y \frac{dv}{dx} = YZi = \Gamma'^2 i$$

$$\left(\frac{d^3 i}{dx^3}\right)_o = \Gamma'^2 \left(\frac{di}{dx}\right)_o = -\Gamma'^3 i(o)$$

and so on. Hence

$$i(x) = \left[I - \frac{x\Gamma'}{1!} + \frac{x^2 \Gamma'^2}{2!} - \cdots\right] i(o)$$

$$= e^{-x\Gamma'} i(o)$$

Equation (1.46) may now be obtained by using the commutation rule for Z_o:

$$v(x) = e^{-x\Gamma} v(o) = e^{-x\Gamma} Z_o i(o)$$

$$= Z_o e^{-x\Gamma'} i(o) = Z_o i(x)$$

This completes the proof of equations (1.44) to (1.47).

1.8 Proof of Relations for Any Uniform Line—Impressed Field

Here it is shown that if a matrix Γ satisfies the two conditions of §1.7 and if Z_o is the characteristic impedance matrix defined there, then the voltages and currents in any uniform line are given by the expressions (1.32). If suitable conditions are fulfilled the relation (1.38) between $v(o)$ and $i(o)$ for an infinite line may be obtained from (1.32).

First of all, $v(x)$ and $i(x)$ reduce to the required values of $v(o)$ and $i(o)$ at $x = 0$. All that remains to be shown is that $v(x)$ and $i(x)$ as given by (1.32) are solutions of the transmission line equations (1.5). By substituting (1.32) in (1.5) and using the formulas

$$\frac{d}{dx} \cosh x\Gamma = \Gamma \sinh x\Gamma = \sinh x\Gamma\, \Gamma$$

$$\frac{d}{dx} \sinh x\Gamma = \Gamma \cosh x\Gamma = \cosh x\Gamma\, \Gamma$$

which follow immediately from the series definitions (1.26) of the hyperbolic functions, we obtain two matrix equations corresponding to the two differential equations. The terms in these equations involving $v(o)$ may be canceled out provided

$$\begin{aligned} \Gamma \sinh x\Gamma &= Z \sinh x\Gamma' Z_o^{-1} \\ \Gamma' \cosh x\Gamma' Z_o^{-1} &= Y \cosh x\Gamma \end{aligned} \tag{1.50}$$

and these are seen to be true from (1.45) and (1.47). The terms involving $i(o)$ may be canceled by a similar argument. The terms involving $l(x)$ may be canceled provided

$$\int_0^x \sinh (x - \xi)\Gamma\, \Gamma l(\xi)\, d\xi = \int_0^x Z \sinh (x - \xi)\Gamma' Z_o^{-1} l(\xi)\, d\xi$$

$$\int_0^x \Gamma' \cosh (x - \xi)\Gamma' Z_o^{-1} l(\xi)\, d\xi = \int_0^x Y \cosh (x - \xi)\Gamma\, l(\xi)\, d\xi$$

and these are seen to be true when x in (1.50) is replaced by $(x - \xi)$. The terms involving $t(x)$ may be similarly canceled. Thus we have verified that $v(x)$ and $i(x)$ as given by (1.32) are solutions of the transmission line equation provided that the commutation rules (1.47) and the relations (1.45) involving Z_o of §1.7 are satisfied. This is the case when Γ is such that (a) Γ^2 is equal to ZY and also (b) every element in $e^{-\Gamma x}$ approaches zero as $x \to \infty$.

In order to establish equation (1.38) for the Γ of §1.7 several assumptions regarding the impressed field are required. Writing the hyperbolic functions in the first of equations (1.32) in exponential form and premultiplying both sides by $2e^{-\Gamma x}$ gives

$$2e^{-x\Gamma} v(x) = \left[v(o) - Z_o i(o) + \int_0^x e^{-\xi\Gamma} [l(\xi) - Z_o t(\xi)] \, d\xi \right] + e^{-2x\Gamma} [v(o) + Z_o i(o)] + e^{-x\Gamma} \int_0^x e^{-(x-\xi)\Gamma} [l(\xi) + Z_o t(\xi)] \, d\xi$$

When $x \to \infty$ equation (1.38) is obtained provided that the impressed field and the terminal conditions at the far end are such that (a) $v(x)$ remains finite, (b) the integral in (1.38) converges, and (c), the last expression on the right in the equation above approaches zero as $x \to \infty$.

1.9 Derivation of Equations (1.25)

Although equations (1.25) may be obtained by setting $l(x) = t(x) = 0$ in §1.8, it is of some interest to derive them directly. By repeated differentiation of the equations

$$\frac{dv}{dx} = -Zi, \qquad \frac{di}{dx} = -Yv \tag{1.48}$$

the second, third and higher order derivatives may be obtained. Using these in Maclaurin's expansion about $x = 0$ gives

$$\begin{aligned} v(x) &= \left[I + \frac{x^2}{2!} ZY + \frac{x^4}{4!} (ZY)^2 + \cdots \right] v(o) - \left[\frac{x}{1!} I + \frac{x^3}{3!} ZY + \frac{x^5}{5!} (ZY)^2 + \cdots \right] Zi(o) \\ i(x) &= -\left[\frac{x}{1!} I + \frac{x^3}{3!} YZ + \frac{x^5}{5!} (YZ)^2 + \cdots \right] Yv(o) + \left[I + \frac{x^2}{2!} YZ + \frac{x^4}{4!} (YZ)^2 + \cdots \right] i(o) \end{aligned} \tag{1.51}$$

These series converge for all values of x and could be used for computation were it not for the unfortunate fact that in most problems a great many terms would be required for a satisfactory answer. For the time being, let Γ be any matrix whose square is ZY. The definitions (1.26) of the hyperbolic functions enable us to write (1.51) as

$$\begin{aligned} v(x) &= \cosh x\Gamma \, v(o) - \sinh x\Gamma \, \Gamma^{-1} Zi(o) \\ i(x) &= -\sinh x\Gamma' \, \Gamma'^{-1} Yv(o) + \cosh x\Gamma' \, i(o) \end{aligned} \tag{1.52}$$

If in addition to being a matrix whose square is ZY, Γ is also such that every element in $e^{-x\Gamma}$ approaches zero as $x \to \infty$, then we may use the relations (1.45) for Z_o and obtain (1.25).

Incidentally, when we put $ZY = I\gamma^2 + R$ in (1.51) and rearrange the terms so as to get a power series in R we get the series (1.28).

1.10 Proof of the First Method

The first method consists essentially of determining Γ from the series expansion of (1.13):

$$\Gamma = \gamma \sum_{n=0}^{\infty} \frac{(-\frac{1}{2})_n}{n!} \frac{(-R)^n}{\gamma^{2n}}, \tag{1.53}$$

where $(-\frac{1}{2})_n = (-\frac{1}{2})(\frac{1}{2})(\frac{3}{2}) \cdots (n - \frac{3}{2})$ when $n > 0$ and $(-\frac{1}{2})_o = 1$, and then computing Z_o and the required exponential and hyperbolic functions of $x\Gamma$. From §1.7 and §1.8 it follows that the first method gives the correct result provided that Γ as determined by (1.53) satisfies the conditions: (a) its square is equal to ZY and (b) every element of $e^{-x\Gamma}$ approaches zero as $x \to \infty$.

These two conditions are satisfied by the matrix

$$\Gamma = PGP^{-1} \tag{1.54}$$

where P and G are matrices defined by equations (A1.1) and (A1.3) of Appendix I, G being a diagonal matrix whose rth element is γ_r. For from (A1.9) the square of Γ is

$$\Gamma^2 = PG^2P^{-1} = ZY$$

Furthermore,

$$\begin{aligned} e^{-x\Gamma} &= \sum_0^{\infty} \frac{(-x)^n}{n!} (PGP^{-1})^n \\ &= P \sum_0^{\infty} \frac{(-x)^n}{n!} G^n P^{-1} \\ &= PM(x)P^{-1} \end{aligned} \tag{1.55}$$

where $M(x)$ is diagonal matrix (A1.5) whose rth element is $e^{-\gamma_r x}$. Since the real part of γ_r is positive and the elements of P are independent of x it follows that the second condition is satisfied.

It will now be shown that PGP^{-1} may be expanded in the series (1.53) provided that γ may be chosen so as to make all of the points $\zeta_r = \dfrac{\gamma_r}{\gamma}$, $r = 1, 2, \cdots m$, in the complex ζ plane lie within that loop of the lemniscate $|\zeta^2 - 1| = 1$ which contains the point $\zeta = 1$. For then we may write the rth element in G as a convergent series:

$$\begin{aligned} \gamma_r &= \gamma\left(1 + \frac{\gamma_r^2 - \gamma^2}{\gamma^2}\right)^{\frac{1}{2}} \\ &= \gamma \sum_{n=0}^{\infty} \frac{(-\frac{1}{2})_n}{n!} \frac{(\gamma^2 - \gamma_r^2)^n}{\gamma^{2n}} \end{aligned} \tag{1.56}$$

and PGP^{-1} may be written as a convergent infinite series, the nth term of which contains the matrix (assuming only three circuits for the sake of simplicity)

$$P\begin{bmatrix} \gamma_1^2 - \gamma^2 & 0 & 0 \\ 0 & \gamma_2^2 - \gamma^2 & 0 \\ 0 & 0 & \gamma_3^2 - \gamma^2 \end{bmatrix}^n P^{-1} = R^n, \tag{1.57}$$

where the equality follows from the definition (1.12) of R and equation (A1.9) of Appendix I. This series for PGP^{-1} is exactly the same as the series (1.53), and this completes the proof of the first method.

The equations (1.18) and (1.28) which are incidental to the first method, will now be established for the case in which the matrix Γ occurring in them is equal to PGP^{-1}. For then we have equation (1.55) and the equations

$$\cosh x\Gamma = P\begin{bmatrix} \cosh x\gamma_1 & 0 & 0 \\ 0 & \cosh x\gamma_2 & 0 \\ 0 & 0 & \cosh x\gamma_3 \end{bmatrix} P^{-1} \tag{1.58}$$

$$\sinh x\Gamma\, Z_o = \sinh x\Gamma\, \Gamma^{-1} Z$$

where $\sinh x\Gamma\, \Gamma^{-1}$ may be expressed in the same fashion as $\cosh x\Gamma$, the rth element of the diagonal matrix being $\dfrac{\sinh x\gamma_r}{\gamma_r}$. The elements in the diagonal matrices occurring in these expressions may be expanded in series by replacing γ_r by its representation (1.56), assuming $\left|\dfrac{\gamma_r^2}{\gamma^2} - 1\right| < 1$, and using[8]

$$e^{-z\sqrt{1+r}} = \sum_0^\infty \left(\frac{rz}{2}\right)^p \frac{(-)^p}{p!} \sqrt{\frac{2z}{\pi}}\, K_{p-\frac{1}{2}}(z)$$

$$\cosh z\sqrt{1+r} = \sum_0^\infty \left(\frac{rz}{2}\right)^p \frac{1}{p!} \sqrt{\frac{\pi z}{2}}\, I_{p-\frac{1}{2}}(z)$$

$$\frac{\sinh z\sqrt{1+r}}{\sqrt{1+r}} = \sum_0^\infty \left(\frac{rz}{2}\right)^p \frac{1}{p!} \sqrt{\frac{\pi z}{2}}\, I_{p+\frac{1}{2}}(z)$$

where $I_{p-\frac{1}{2}}(z)$ and $K_{p-\frac{1}{2}}(z)$ are Bessel functions of the first and second kinds, respectively, for imaginary argument. Equations (1.18) and (1.28) are obtained when equation (1.57) and the Bessel function recurrence relations are used.

[8] These are special cases of formulas given in "Theory of Bessel Functions," by G. N. Watson, page 141.

1.11 Proof of the Second Method

To establish the second method we must prove the various formulas which are used. These formulas all involve the square matrix $N(\gamma_r^2)$ defined by (1.23).

Since $N(\gamma_r^2)$ is proportional to $F(\gamma_r^2)$ it follows that $N(\gamma_r^2)$ may be expressed as

$$N(\gamma_r^2) = p_r\rho_r \cdots \tag{1.59}$$

where p_r is the column matrix defined in Appendix I and ρ_r is a row matrix specified by p_r and $N(\gamma_r^2)$. Applying Sylvester's theorem to the unit matrix gives

$$I = \sum N(\gamma_r^2) = \sum p_r\rho_r = [p_1, p_2, p_3]\begin{bmatrix}\rho_1\\ \rho_2\\ \rho_3\end{bmatrix}$$

where the two matrices on the extreme right are partitioned square matrices. From the definition of P in Appendix I it follows that

$$[p_1, p_2, p_3] = P, \qquad \begin{bmatrix}\rho_1\\ \rho_2\\ \rho_3\end{bmatrix} = P^{-1} \tag{1.60}$$

These relations enable us to verify the equations (1.24) when Γ is equal to PGP^{-1}. Thus for the first of equations (1.24)

$$\Gamma = PGP^{-1} = [p_1, p_2, p_3]G\begin{bmatrix}\rho_1\\ \rho_2\\ \rho_3\end{bmatrix}$$

$$= [p_1, p_2, p_3]\begin{bmatrix}\gamma_1\rho_1\\ \gamma_2\rho_2\\ \gamma_3\rho_3\end{bmatrix} = \sum p_r\gamma_r\rho_r = \sum N(\gamma_r^2)\gamma_r$$

The second equation of (1.24) follows likewise from the expression (1.55) for $e^{-x\Gamma}$.

The third equation of (1.24) follows at once from the first when we use (1.45), $Z_o = \Gamma Y^{-1}$. The fourth equation is obtained by writing

$$e^{-x\Gamma}Z_o = PM(x)P^{-1}PGP^{-1}Y^{-1}$$
$$= PM(x)GP^{-1}Y^{-1}$$

and proceeding as in the case of the first equation.

All of the other equations connected with the second method may be proved in a similar way. Incidentally, the formulas obtained by the second method are closely related to the "special form of solution" described in §6.5 of F.D.C.

PART II

TRANSMISSION LINES COMPOSED OF MULTI-TERMINAL SECTIONS

2.1 Introductory

Some transmission systems may be regarded as consisting of a number of identical sections connected in tandem. The question of determining the steady state electrical behavior of such a system from a knowledge of the properties of a single section will be considered here.

Each section will have a certain number, say $m + 1$, terminals on its left end and an equal number on the right. The case in which there are only two terminals ($m = 1$) has been completely worked out, and some studies of more general cases have been made. The ones which most nearly approach the point of view of the present paper are those due to S. Koizumi[9].

In the present work difference equations are used to solve the general case in much the same manner as they have been used in studying the two-terminal case. This approach differs from that used by Koizumi and throws additional light on the problem.

In several lists of formulas, particularly in Appendix IV, I have included a number of results due to Koizumi for the sake of completeness.

2.2 Transmission Equations for a Typical Section

We consider the equations connecting the input and output currents and voltages for the nth section which is shown in Fig. 1. The directions which are assumed for positive current flow are indicated by arrows. The leads marked 0 play a special role in that all the voltages are

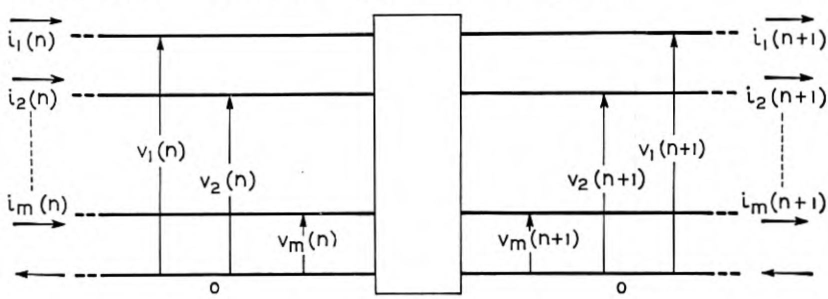

FIG. 1

measured with respect to them, and the currents which they carry are the sum of the currents flowing into or out of the remaining terminals at the end under consideration. In applications to transmission lines the terminals 0 would correspond to the ground or the cable sheath.

The currents and voltages shown in Fig. 1 are related by a number of

[9] Archiv für Electrotechnik, Vol. 33, pp. 171–188, 609–622 (1939). See also a paper by M. G. Malti and S. E. Warschawski, Trans. A.I.E.E., Vol. 56, pp. 153–158 (1937).

sets of $2m$ linear equations which may be conveniently written in matrix form. One such set is

$$
\begin{aligned}
v(n) &= Z_{11}i(n) - Z_{12}i(n+1) + v^\circ(n) \\
v(n+1) &= Z_{21}i(n) - Z_{22}i(n+1) + u^\circ(n)
\end{aligned}
\tag{2.1}
$$

Z_{11}, Z_{12}, Z_{21}, Z_{22} are square matrices of order m whose elements are impedances. $v(n)$ and $i(n)$ are the column matrices

$$
v(n) = \begin{bmatrix} v_1(n) \\ v_2(n) \\ \vdots \\ v_m(n) \end{bmatrix} \qquad i(n) = \begin{bmatrix} i_1(n) \\ i_2(n) \\ \vdots \\ i_m(n) \end{bmatrix}
$$

The column matrices $v^\circ(n)$ and $u^\circ(n)$ arise from generators which may be acting within the nth section. If both ends of the section are open circuited so that $i(n) = i(n+1) = 0$ the equations show that $v(n) = v^\circ(n)$, $v(n+1) = u^\circ(n)$. Consequently, $v^\circ(n)$ and $u^\circ(n)$ give the open circuit voltages produced on the left and right ends of the nth section by the internal generators. If the section is a passive network then $v^\circ(n) = u^\circ(n) = 0$ and they do not appear in the equations. The subscripts on the square matrices, the Z's, are chosen so as to preserve the analogy for the simple case $m = 1$, where the left and right ends of the section are denoted by the subscripts 1 and 2, respectively.

Solving the equation (1.1) for $i(n)$ and $i(n+1)$ gives

$$
\begin{aligned}
i(n) &= Y_{11}v(n) + Y_{12}v(n+1) + i^\circ(n) \\
-i(n+1) &= Y_{21}v(n) + Y_{22}v(n+1) - j^\circ(n)
\end{aligned}
\tag{2.2}
$$

where the elements of the Y's are admittances and $i^\circ(n)$, $j^\circ(n)$ are the currents produced by the internal sources when the terminals on the right and left are short-circuited so that $v(n) = v(n+1) = 0$.

A third set of equations is

$$
\begin{aligned}
v(n) &= Av(n+1) + Bi(n+1) - Bj^\circ(n) \\
i(n) &= Cv(n+1) + Di(n+1) - Cu^\circ(n)
\end{aligned}
\tag{2.3}
$$

Solving these equations for $v(n+1)$ and $i(n+1)$ gives

$$
\begin{aligned}
v(n+1) &= D'v(n) - B'i(n) + B'i^\circ(n) \\
i(n+1) &= -C'v(n) + A'i(n) + C'v^\circ(n)
\end{aligned}
\tag{2.4}
$$

There are a great many relations between the square matrices appearing in the equations (2.1) to (2.4). These are discussed in Appendix IV.

For symmetrical sections $Y_{21} = Y_{12}$, $Y_{22} = Y_{11}$, $Z_{21} = Z_{12}$, $Z_{22} = Z_{11}$ and equations (2.1) and (2.2) become

$$\begin{aligned} v(n) &= Z_{11}i(n) - Z_{12}i(n+1) + v^\circ(n) \\ v(n+1) &= Z_{12}i(n) - Z_{11}i(n+1) + u^\circ(n) \end{aligned} \tag{2.5}$$

$$\begin{aligned} i(n) &= Y_{11}v(n) + Y_{12}v(n+1) + i^\circ(n) \\ -i(n+1) &= Y_{12}v(n) + Y_{11}v(n+1) - j^\circ(n) \end{aligned} \tag{2.6}$$

Eliminating $i(n)$ from (2.5) and $v(n)$ from (2.6) and using, from (A4.4), $A = Z_{11}Z_{12}^{-1} = -Y_{12}^{-1}Y_{11}$ leads to the difference equations

$$v(n+1) + v(n-1) - 2Av(n) = B[i^\circ(n) - j^\circ(n-1)] \tag{2.7}$$

$$i(n+1) + i(n-1) - 2A'i(n) = C[v^\circ(n) - u^\circ(n-1)] \tag{2.8}$$

Since we also have $B' = B$, $C' = C$, $D' = A$ for symmetrical sections equations (2.4) become

$$\begin{aligned} v(n+1) &= Av(n) - Bi(n) + Bi^\circ(n) \\ i(n+1) &= -Cv(n) + A'i(n) + Cv^\circ(n) \end{aligned} \tag{2.9}$$

We assume that the distribution of the sources in the branches of a symmetrical network need not be symmetrical with respect to the two ends, even though the impedances of the branches are.

2.3 *Statement of Results for Infinite Symmetrical Section Line—Passive*

When the sections are passive the equations to be solved are, from (2.9),

$$\begin{aligned} v(n+1) &= Av(n) - Bi(n) \\ i(n+1) &= -Cv(n) + A'i(n) \end{aligned} \tag{2.10}$$

If the line extends from $n = 0$ to $n = \infty$ the solution is

$$\begin{aligned} v(n) &= e^{-n\Gamma}v(o) \\ i(n) &= e^{-n\Gamma'}i(o) \\ v(n) &= Z_o i(n) \end{aligned} \tag{2.11}$$

where the matrix $e^{-\Gamma}$ is such that (a) the equation

$$e^{-\Gamma} + e^{\Gamma} = 2A \tag{2.12}$$

is satisfied, e^{Γ} being the inverse of $e^{-\Gamma}$, and (b) all the elements of the matrix $e^{-n\Gamma}$ approach zero as $n \to \infty$. In dealing with sections we shall never have occasion to consider Γ itself but only its exponential and associated functions. The characteristic impedance matrix Z_o is defined by the relation between the initial currents and voltages in an infinite line

$$v(o) = Z_o i(o) \tag{2.13}$$

A formal solution of (2.12) may be obtained by writing it as

$$\cosh \Gamma = A \tag{2.14}$$

Then

$$\begin{aligned} e^{-\Gamma} &= \cosh \Gamma - \sinh \Gamma \\ &= A - (A^2 - I)^{\frac{1}{2}} = A - (BC)^{\frac{1}{2}} \end{aligned}$$

where the square root is to be chosen so that condition (b) for $e^{-\Gamma}$ is satisfied. The characteristic impedance matrix Z_o is given by equations (2.34) of which the following two are representative.

$$Z_o = (\sinh \Gamma)^{-1} B = \sinh \Gamma \, C^{-1} \tag{2.15}$$

where $\sinh \Gamma$ is given by $2 \sinh \Gamma = e^{\Gamma} - e^{-\Gamma}$.

The wide variety of sections makes it appear unlikely that there is a general method of determining $e^{-\Gamma}$ analogous to the first method discussed for the uniform line. However, in some cases rapidly convergent series for $e^{-\Gamma}$ and e^{Γ} may be obtained. For example, suppose that the elements of $(2A)^{-1}$ are small compared to those of 2A. Then, from (2.12),

$$\begin{aligned} e^{\Gamma} &= 2A - (2A)^{-1} - (2A)^{-3} - 2(2A)^{-5} - \cdots \\ e^{-\Gamma} &= (2A)^{-1} + (2A)^{-3} + 2(2A)^{-5} + \cdots \end{aligned}$$

Again, if $A^2 - I = BC$ is expressible as $I\gamma^2 + R$ where the elements of R are small in comparison with γ^2, we have (cf. equations (1.14), (1.15))

$$e^{\Gamma} = A + \gamma \left[I + \frac{R}{2\gamma^2} - \frac{1}{2}\left(\frac{R}{2\gamma^2}\right)^2 + \cdots \right]$$

$$e^{-\Gamma} = A - \gamma \left[I + \frac{R}{2\gamma^2} - \frac{1}{2}\left(\frac{R}{2\gamma^2}\right)^2 + \cdots \right]$$

Finally, it follows from a comparison of equations (2.11) and (A2.12) that a suitable $e^{-\Gamma}$ is given by

$$e^{-\Gamma} = P\Lambda P^{-1}, \qquad e^{-\Gamma'} = Q\Lambda Q^{-1} \tag{2.16}$$

where P, Q and Λ are the matrices designated by the same symbols in Appendices II and III.

The formal application of Sylvester's theorem leads to a method of solving the symmetrical section line which is analogous to the second method discussed for the uniform line. Thus, if $P(A)$ is any polynomial in A, then

$$P(A) = \sum_{r=1}^{m} N(\zeta_r) P(\zeta_r) \tag{2.17}$$

where $P(\zeta_r)$ is not a square matrix but a scalar and $N(\zeta_r)$ is the square matrix

$$N(\zeta_r) = \frac{F(\zeta_r)}{|f(\zeta_r)|^{(1)}}. \tag{2.18}$$

$F(\zeta)$ is the adjoint of the characteristic matrix

$$f(\zeta) = I\zeta - A \tag{2.19}$$

and $\zeta_1, \zeta_2, \cdots \zeta_m$ are the roots, assumed to be unequal, of the characteristic equation

$$|f(\zeta)| = 0.$$

The denominator in the expression for $N(\zeta_r)$ is the derivative of the characteristic function:

$$|f(\zeta_r)|^{(1)} = \left[\frac{d}{d\zeta}|f(\zeta)|\right]_{\zeta=\zeta_r}$$

The formal application of Sylvester's theorem then gives

$$\begin{aligned} \cosh \Gamma &= A = \Sigma N(\zeta_r)\zeta_r \\ e^{-\Gamma} &= A - (A^2 - I)^{\frac{1}{2}} = \Sigma N(\zeta_r)\lambda_r \\ e^{-n\Gamma} &= \Sigma N(\zeta_r)\lambda_r^n \\ Z_o &= (\sinh \Gamma)^{-1} B = \Sigma \frac{N(\zeta_r)\, 2}{(\lambda_r^{-1} - \lambda_r)} B \end{aligned} \tag{2.20}$$

where $N(\zeta_r)$ is given by (2.18), the summations run from $r = 1$ to $r = m$, and λ_r is related to ζ_r through

$$2\zeta_r = \lambda_r + \lambda_r^{-1}, \qquad \lambda_r = \zeta_r - \sqrt{\zeta_r^2 - 1} \tag{2.21}$$

where the sign of the square root is chosen so that $|\lambda_r| < 1$. λ_r is related to $e^{-\Gamma}$ in the same way that ζ_r is related to $\cosh \Gamma$.

2.4 Results for Any Symmetrical Section Line—Passive

The solutions of equations (2.10) which reduce to the given values $v(o)$, $i(o)$ at $n = 0$ are

$$\begin{aligned} v(n) &= \cosh n\Gamma\, v(o) - \sinh n\Gamma\, Z_o i(o) \\ i(n) &= -\sinh n\Gamma'\, Z_o^{-1}\, v(o) + \cosh n\Gamma'\, i(o) \end{aligned} \tag{2.22}$$

where $e^{-\Gamma}$ and Z_o are the matrices of §2.3. These may be put in slightly different form by using the relations

$$\sinh n\Gamma\, Z_o = Z_o \sinh n\Gamma'$$

$$\sinh n\Gamma'\, Z_o^{-1} = Z_o^{-1} \sinh n\Gamma$$

When (2.22) are interpreted by Sylvester's theorem we obtain

$$
\begin{aligned}
v(n) &= \Sigma N(\zeta_r)\left[\tfrac{1}{2}(\lambda_r^{-n} + \lambda_r^n)v(o) - \frac{\lambda_r^{-n} - \lambda_r^n}{\lambda_r^{-1} - \lambda_r}\, Bi(o)\right] \\
i(n) &= \Sigma N'(\zeta_r)\left[-\frac{\lambda_r^{-n} - \lambda_r^n}{\lambda_r^{-1} - \lambda_r}\, Cv(o) + \tfrac{1}{2}(\lambda_r^{-n} + \lambda_r^n)i(o)\right]
\end{aligned}
\tag{2.23}
$$

where $N'(\zeta_r)$ is the transposed of $N(\zeta_r)$ and $N(\zeta_r)$ is given by (2.18) and the summations run from $r = 1$ to $r = m$.

2.5 Results for Any Symmetrical Section Line—Active

When the sections contain generators the equations to be solved are those of (2.9). The solutions corresponding to the initial values $v(o)$ and $i(o)$ are, for $n \geqq 1$,

$$
\begin{aligned}
v(n) &= \cosh n\Gamma\, v(o) - \sinh n\Gamma\, Z_o i(o) \\
&\quad + \sum_{p=1}^{n} \{\cosh (n - p)\Gamma\, Bi^\circ(p - 1) - \sinh (n - p)\Gamma\, Z_o Cv^\circ(p - 1)\} \\
i(n) &= -\sinh n\Gamma'\, Z_o^{-1} v(o) + \cosh n\Gamma'\, i(o) \\
&\quad + \sum_{p=1}^{n} \{\cosh (n - p)\Gamma'\, Cv^\circ(p - 1) - \sinh (n - p)\Gamma'\, Z_o^{-1} Bi^\circ(p - 1)\}
\end{aligned}
\tag{2.24}
$$

These may be simplified somewhat by replacing Z_oC and $Z_o^{-1}B$ by $\sinh \Gamma$ and $\sinh \Gamma'$, respectively.

The series in the above expressions may be summed when the generators are such that

$$
v^\circ(n) = e^{-n\theta} i^\circ, \qquad i^\circ(n) = e^{-n\theta} v^\circ \tag{2.25}
$$

where θ is a scalar and i° and v° are column matrices whose elements are independent of n. Thus

$$
\begin{aligned}
v(n) &= \cosh n\Gamma\, v(o) - \sinh n\Gamma\, Z_o i(o) \\
&\quad + \tfrac{1}{2}(e^{n\Gamma} - e^{-n\theta}I)(e^{\Gamma} - e^{-\theta}I)^{-1}(Bi^\circ - Z_oCv^\circ) \\
&\quad + \tfrac{1}{2}(e^{-n\Gamma} - e^{-n\theta}I)(e^{-\Gamma} - e^{-\theta}I)^{-1}(Bi^\circ + Z_oCv^\circ) \\
i(n) &= -\sinh n\Gamma'\, Z_o^{-1} v(o) + \cosh n\Gamma'\, i(o) \\
&\quad + \tfrac{1}{2}(e^{n\Gamma'} - e^{-n\theta}I)(e^{\Gamma'} - e^{-\theta}I)^{-1}(Cv^\circ - Z_o^{-1}Bi^\circ) \\
&\quad + \tfrac{1}{2}(e^{-n\Gamma'} - e^{-n\theta}I)(e^{-\Gamma'} - e^{-\theta}I)^{-1}(Cv^\circ + Z_o^{-1}Bi^\circ)
\end{aligned}
\tag{2.26}
$$

provided that the inverse matrices exist.

We may interpret these expressions by Sylvester's theorem. For example,

$$v(n) = \sum_{r=1}^{m} N(\zeta_r)\left[\tfrac{1}{2}(\lambda_r^{-n} + \lambda_r^{n})v(o) - \frac{\lambda_r^{-n} - \lambda_r^{n}}{\lambda_r^{-1} - \lambda_r} Bi(o) + \frac{1}{2}\frac{\lambda_r^{-n} - e^{-n\theta}}{\lambda_r^{-1} - e^{-\theta}}\left(Bi^\circ - \frac{2BCv_o}{\lambda_r^{-1} - \lambda_r}\right) + \frac{1}{2}\frac{\lambda_r^{n} - e^{-n\theta}}{\lambda_r - e^{-\theta}}\left(Bi^\circ + \frac{2BCv_o}{\lambda_r^{-1} - \lambda_r}\right)\right] \quad (2.27)$$

where $N(\zeta_r)$ is given by (2.18).

When the line extends to $n = \infty$ and the sources and end conditions satisfy suitable conditions we have the relation

$$v(o) = Z_o i(o) - \sum_{p=1}^{\infty} e^{-p\Gamma}[Bi^\circ(p-1) - Z_o Cv^\circ(p-1)] \quad (2.28)$$

When the impressed field is of the form (2.25) this becomes

$$v(o) = Z_o i(o) - (e^{\Gamma} - e^{-\theta}I)^{-1}(Bi^\circ - Z_o Cv^\circ) \quad (2.29)$$

provided that the inverse matrix exists. Expressions for $v(n)$ and $i(n)$ in such an infinite line may be obtained by using (2.28) or (2.29) in (2.24) or (2.26).

Applying Sylvester's theorem to (2.29) gives

$$v(o) = \sum_{r=1}^{m} N(\zeta_r)\left(\frac{2Bi(o)}{\lambda_r^{-1} - \lambda_r} - \frac{Bi^\circ}{\lambda_r^{-1} - e^{-\theta}} + \frac{\lambda_r^{-1} - \lambda_r}{2(\lambda_r^{-1} - e^{-\theta})} v^\circ\right) \quad (2.30)$$

The last term within the braces may be replaced by

$$\frac{2BCv^\circ}{(\lambda_r^{-1} - \lambda_r)(\lambda_r^{-1} - e^{-\theta})}$$

2.6 Derivation of the Properties of an Infinite Line

We shall consider a symmetrical section line which is specified by the equations

$$\begin{aligned} v(n+1) &= Av(n) - Bi(n) \\ i(n+1) &= -Cv(n) + A'i(n) \end{aligned} \quad (2.10)$$

From these equations and the relations $A^2 - BC = I$, $AB = BA'$, $A'C = CA$ of (A4.6) it follows that

$$\begin{aligned} v(n+1) + v(n-1) &= 2Av(n) \\ i(n+1) + i(n-1) &= 2A'i(n) \end{aligned} \quad (2.31)$$

If $e^{-\Gamma}$ is a matrix satisfying the conditions of §2.3, namely, (a) $e^{-\Gamma}$ satisfies the equation

$$2 \cosh \Gamma = e^{\Gamma} + e^{-\Gamma} = 2A \tag{2.14}$$

and (b), every element in $e^{-n\Gamma}$ approaches zero as $n \to \infty$, and if Z_o is defined by $v(o)$ and $i(o)$ for an infinite line as in (2.13), then

1. In an infinite line

$$\begin{aligned} v(n) &= e^{-n\Gamma} v(o), \\ i(n) &= e^{-n\Gamma'} i(o), \end{aligned} \tag{2.32}$$

$$v(n) = Z_o i(n) \tag{2.33}$$

2. The characteristic impedance matrix Z_o is given by

$$\begin{aligned} Z_o &= (\sinh \Gamma)^{-1} B = B(\sinh \Gamma')^{-1} = C^{-1} \sinh \Gamma' = \sinh \Gamma \, C^{-1} \\ Z_o^{-1} &= B^{-1} \sinh \Gamma = \sinh \Gamma' \, B^{-1} = (\sinh \Gamma')^{-1} C = C(\sinh \Gamma)^{-1} \end{aligned} \tag{2.34}$$

3. The matrices Z_o, B and C obey the commutation rules

$$\begin{aligned} \Phi(e^{\Gamma}) Z_o &= Z_o \Phi(e^{\Gamma'}) \\ \Phi(e^{\Gamma}) B &= B \Phi(e^{\Gamma'}) \\ C \Phi(e^{\Gamma}) &= \Phi(e^{\Gamma'}) C \end{aligned} \tag{2.35}$$

where $\Phi(e^{\Gamma})$ is a square matrix representable as a sum of powers of $e^{\pm\Gamma}$. The matrices $\Phi(e^{\Gamma}) Z_0$, $\Phi(e^{\Gamma}) B$, and $C\Phi(e^{\Gamma})$ are symmetrical.

To prove these statements we proceed as follows: By direct substitution into (2.31) it is seen that $v(n) = e^{-n\Gamma} v(o)$ is a solution by virtue of condition (a) satisfied by $e^{-\Gamma}$. Since, by condition (b), $v(n) \to o$ as $n \to \infty$ it follows that $v(n)$ is the voltage in an infinite line. Similarly, $i(n) = e^{-n\Gamma'} i(o)$ is the current in such a line. Substituting the expressions (2.32) for $v(n)$ and $i(n)$ into the difference equations (2.10), setting $n = 0$, using the definition of Z_o, and regarding $v(o)$ and $i(o)$ as arbitrary columns gives

$$\begin{aligned} e^{-\Gamma} &= A - B Z_o^{-1} \\ e^{-\Gamma'} &= -C Z_o + A' \end{aligned} \tag{2.36}$$

Applying condition (a) in the form of (2.14) to these equations gives

$$B Z_o^{-1} = \sinh \Gamma, \qquad C Z_o = \sinh \Gamma' \tag{2.37}$$

Since the sections are symmetrical, B and C are symmetrical matrices, and from the reciprocal theorem for networks it follows that Z_o is also symmetrical. These remarks and (2.37) lead to (2.34). Setting the expressions (2.32) for $v(n)$ and $i(n)$ in the second of the difference equations (2.10) using the definition of Z_o, and regarding $i(o)$ as an arbitrary column gives

$$e^{-(n+1)\Gamma'} = -Ce^{-n\Gamma}Z_o + A'e^{-n\Gamma'}$$

$$(A' - e^{-\Gamma'})e^{-n\Gamma'} = Ce^{-n\Gamma}Z_o$$

Replacing $A' - e^{-\Gamma'}$ by CZ_o, as follows from the case $n = 0$, and pre-multiplying by C^{-1} gives

$$Z_o e^{-n\Gamma'} = e^{-n\Gamma}Z_o$$

and this leads to the first of equations (2.35). From (2.34) and the relations $AB = BA'$, $CA = A'C$ we have

$$\sinh\Gamma\, B = B\sinh\Gamma' \qquad \cosh\Gamma\, B = B\cosh\Gamma'$$

$$C\sinh\Gamma = \sinh\Gamma'\, C \qquad C\cosh\Gamma = \cosh\Gamma'\, C$$

Addition and subtraction leads to

$$e^{\pm\Gamma}B = Be^{\pm\Gamma'} \qquad Ce^{\pm\Gamma} = e^{\pm\Gamma'}C$$

from which the last two of equations (2.35) follow. Since each of equations (2.35) expresses the equality of a matrix and its transposed, it follows that the matrices are symmetrical.

Equation (2.33), which is almost self-evident on physical grounds, follows from

$$v(n) = e^{-n\Gamma}v(o) = e^{-n\Gamma}Z_o i(o)$$

$$= Z_o e^{-n\Gamma'}i(o) = Z_o i(n).$$

2.7 Proof of Relations for Any Symmetrical Section Line

The expressions (2.24) for $v(n)$ and $i(n)$ in a line whose sections contain generators may be verified to satisfy the difference equations (2.9). The expressions (2.34) for Z_o and the commutation rules (2.35) for B and C are used in the verification. Setting $n = 1$ in the expressions for $v(n)$ and $i(n)$ gives the difference equations (2.9) and hence $v(n)$ and $i(n)$ are the solutions which correspond to the initial values $v(o)$ and $i(o)$.

In order to derive the relation (2.28) between $v(o)$ and $i(o)$ for an infinite line we put the hyperbolic functions in the expression (2.24) for $v(n)$ in exponential form and multiply through by $2e^{-n\Gamma}$

$$2e^{-n\Gamma}v(n) = v(o) - Z_o i(o) + \sum_{p=1}^{n} e^{-p\Gamma}[Bi^\circ(p-1) - Z_o Cv^\circ(p-1)]$$

$$+ e^{-2n\Gamma}[v(o) + Z_o i(o)]$$

$$+ e^{-n\Gamma}\sum_{p=1}^{n} e^{-(n-p)\Gamma}[Bi^\circ(p-1) + Z_o Cv^\circ(p-1)]$$

Hence, letting $n \to \infty$ and using condition (b) satisfied by $e^{-\Gamma}$, equation (2.28) is obtained provided that (i) the terminal conditions at the far end are such that $v(n)$ remains finite, (ii) the sum in (2.28) converges, and (iii) the expression in the last line in the equation just above approaches zero as $n \to \infty$.

The results obtained by the formal application of Sylvester's theorem may be verified by using the results of Appendix II and writing $N(\zeta_r)$ as the product of a column matrix and a row matrix. They may also be verified more directly. For example, setting $n = 0$ in the expressions (2.23) for $v(n)$ and $i(n)$ in any passive symmetrical section line and using

$$\sum_{r=1}^{m} N(\zeta_r) = I, \tag{2.38}$$

which follows from Sylvester's theorem, we see that $v(n)$ and $i(n)$ reduce to the appropriate values $v(o)$ and $i(o)$ at $n = 0$. Substituting $v(n)$ and $i(n)$ into the difference equations (2.10) and using

$$\begin{aligned} BC &= A^2 - I \\ (I\zeta_r - A)N(\zeta_r) &= N(\zeta_r)(I\zeta_r - A) = 0 \\ BN'(\zeta) &= N(\zeta)B \\ CN(\zeta) &= N'(\zeta)C, \end{aligned} \tag{2.39}$$

shows that they are solutions. The second of the relations (2.39) follows from the fact that $N(\zeta_r)$ is proportional to the adjoint $F(\zeta_r)$ of $f(\zeta_r)$. In the third and fourth relations

$$N(\zeta) = \frac{F(\zeta)}{|f(\zeta)|^{(1)}}$$

which is in agreement with the definition (2.18) of $N(\zeta_r)$. To establish the third relation we start from,[10]

$$\begin{aligned} (\zeta I - A)\, F(\zeta) &= I\,|\,f(\zeta)\,| \\ (\zeta I - A)\, N(\zeta) &= I\,|\,f(\zeta)\,|/|\,f(\zeta)\,|^{(1)} \end{aligned}$$

Postmultiplication by B gives

$$(\zeta I - A)\, N(\zeta)\, B = B\,|\,f(\zeta)\,|/|\,f(\zeta)\,|^{(1)}$$

We also have

$$\begin{aligned} (\zeta I - A')F'(\zeta) &= I\,|\,f(\zeta)\,| \\ (\zeta I - A')N'(\zeta) &= I\,|\,f(\zeta)\,|/|\,f(\zeta)\,|^{(1)} \end{aligned}$$

[10] F.D.C. §3.5.

Premultiplication by B and use of $BA' = AB$ gives

$$(\zeta I - A)BN'(\zeta) = B \mid f(\zeta) \mid / \mid f(\zeta) \mid^{(1)}$$

Hence, the third equation in (2.39) holds except possibly for $\zeta = \zeta_r$, and from the concept of continuity it holds there also. The fourth equation in (2.39) may be proved in the same manner.

The expression (2.20) for Z_o may be obtained by letting n become very large in the expression (2.23) for $v(n)$. $v(o)$ and $i(o)$ must be related so that $v(n)$ remains finite. Since $\mid \lambda_r \mid < 1$ and the λ_r's are unequal the coefficients of λ_r^{-n} must vanish. This requires

$$N(\zeta_r)v(o) = \frac{2N(\zeta_r)Bi(o)}{\lambda_r^{-1} - \lambda r}$$

Summing r from 1 to m and using (2.38) gives the required expression for Z_o.

2.8 The Unsymmetrical Section Line

The method used here is analogous to those described in Appendices I and II for the uniform line and the symmetrical section line. The other methods apparently do not lead to the simplification which occurs in the symmetrical case.

Equations (2.2) and (2.1) lead to the difference equations

$$Y_{12}v(n + 2) + [Y_{11} + Y_{22}]v(n + 1) + Y_{21}v(n) = -i^\circ(n + 1) + j^\circ(n) \quad (2.40)$$

$$Z_{12}i(n + 2) - [Z_{11} + Z_{22}]i(n + 1) + Z_{21}i(n) = v^\circ(n + 1) - u^\circ(n) \quad (2.41)$$

Both of these equations are of the form

$$Gx(n + 2) + Hx(n + 1) + G'x(n) = g(n) \quad (2.42)$$

in which G and H are square matrices of order m, H being symmetrical and G' being the transposed of G. When the sections are passive equations (2.40) and (2.41) become

$$Y_{12}v(n + 2) + [Y_{11} + Y_{22}]v(n + 1) + Y_{21}v(n) = 0 \quad (2.43)$$

$$Z_{12}i(n + 2) - [Z_{11} + Z_{22}]i(n + 1) + Z_{21}i(n) = 0 \quad (2.44)$$

In the passive, unsymmetrical case the expressions for $v(n)$ and $i(n)$ are of the form

$$\begin{aligned} v(n) &= P\Lambda^n a + \bar{P}\Lambda^{-n}\bar{a} \\ i(n) &= Q\Lambda^n a - \bar{Q}\Lambda^{-n}\bar{a} \end{aligned} \quad (2.45)$$

Comparison with (A2.8) shows that in the symmetrical case $\bar{P} = P$ and $\bar{Q} = Q$. The minus signs over $\bar{P}$, $\bar{Q}$, and $\bar{a}$ indicate that they are associated with propagation in the negative direction. The propagation constants of the m modes of propagation are the same in the positive as in the negative direction, as indicated by the appearance of Λ^n and Λ^{-n} in (2.45). Corresponding to any given propagation constant say λ_r, there are two modes of propagation, one in a positive direction and the other in the negative direction. The distribution of the voltages corresponding to these two modes are given by the rth columns in P and $\bar{P}$, respectively. The fact that P and $\bar{P}$ differ shows that the distributions differ according to the direction of propagation even though the propagation constant is the same. Λ is still the diagonal matrix defined in (A2.3) but now the computation of the elements λ_r is more difficult than when the section is symmetrical. They are defined as the roots of the equation

$$| G\lambda^2 + H\lambda + G' | = 0 \tag{2.46}$$

which are less than unity in absolute value. The second of the equations (A4.5) shows that the roots of (2.46) are the same whether the Z's or the Y's are used in place of G and H. Of course, this is to be expected on physical grounds. The third of the equations (A4.5) may be used to show that the roots of (2.46) are also the roots of

$$\begin{vmatrix} \lambda A - I & \lambda B \\ \lambda C & \lambda D - I \end{vmatrix} = 0 \tag{2.47}$$

From the form of (2.46) it follows that if λ_r is a root so is λ_r^{-1}. This fact may be used to simplify the determination of λ_r. When the substitution

$$2\zeta = \lambda + \lambda^{-1}, \qquad \lambda = \zeta - \sqrt{\zeta^2 - 1} \tag{A2.4}$$

is made equation (2.46) may be written as

$$0 = | (G + G')\zeta + H + (G' - G)\sqrt{\zeta^2 - 1} |$$

$0 = | (G + G')\zeta + H |$

$+ (\zeta^2 - 1)$ times the sum of $\dfrac{m(m-1)}{2!}$ determinants each obtained by replacing two columns of $|(G + G')\zeta + H|$ by the corresponding columns of $(G - G')$

$+ (\zeta^2 - 1)^2$ times the sum of $\dfrac{m(m-1)(m-2)(m-3)}{4!}$ determinants each obtained by replacing four columns of $|(G + G')\zeta + H|$ by the corresponding columns of $(G - G')$

$+ \cdots$

The last equation is a polynomial of degree m in ζ which is to be solved for its roots $\zeta_1, \zeta_2, \cdots \zeta_m$. For simplicity we assume that these roots are distinct. λ_r is then determined from ζ_r by the relations (A2.4), the sign of the radical being chosen so that $|\lambda_r| < 1$ as in the symmetrical case. In his second paper Koizumi has given a procedure which amounts to an alternative method of determining Λ.

We shall first assume that the Y's are known and that our equations are

$$\begin{aligned} i(n) &= Y_{11}v(n) + Y_{12}v(n+1) \\ -i(n+1) &= Y_{21}v(n) + Y_{22}v(n+1) \end{aligned} \tag{2.48}$$

As described above Λ may be computed from the determinantal equation

$$|f(\lambda)| = 0$$

where $f(\lambda)$ represents the matrix

$$f(\lambda) = Y_{12}\lambda^2 + (Y_{11} + Y_{22})\lambda + Y_{21} \tag{2.49}$$

Let p_r be proportional to any non-zero column in $F(\lambda_r)$ where $F(\lambda)$ is the adjoint of $f(\lambda)$ and let $\bar{p}_r'$ be proportional to any non-zero row of $F(\lambda_r)$. Then the matrices P and $\bar{P}$ in the expressions (2.45) for $v(n)$ and $i(n)$ are given by

$$\begin{aligned} P &= [p_1, p_2, \cdots p_m] \\ \bar{P} &= [\bar{p}_1, \bar{p}_2, \cdots \bar{p}_m] \end{aligned} \tag{2.50}$$

where $\bar{p}_r$ is the column obtained by transposing the row $\bar{p}_r'$. The matrices Q and $\bar{Q}$ are obtained from P and $\bar{P}$ by means of the equations

$$\begin{aligned} Q &= Y_{11}P + Y_{12}P\Lambda = -Y_{22}P - Y_{21}P\Lambda^{-1} \\ \bar{Q} &= -Y_{11}\bar{P} - Y_{12}\bar{P}\Lambda^{-1} = Y_{22}\bar{P} + Y_{21}\bar{P}\Lambda \end{aligned} \tag{2.51}$$

which are derived from (2.45) and (2.48).

The properties of the individual columns of P and $\bar{P}$ lead to the relations

$$\begin{aligned} Y_{12}P\Lambda^2 + (Y_{11} + Y_{22})P\Lambda + Y_{21}P &= 0 \\ Y_{12}\bar{P}\Lambda^{-2} + (Y_{11} + Y_{22})\bar{P}\Lambda^{-1} + Y_{21}\bar{P} &= 0 \end{aligned} \tag{2.52}$$

and these guarantee that the difference equations (2.48) will be satisfied when the expressions (2.45) for $v(n)$ and $i(n)$ are used.

When the Z's are known instead of the Y's the procedure is much the same. The difference equations are

$$\begin{aligned} v(n) &= Z_{11}i(n) - Z_{12}i(n+1) \\ v(n+1) &= Z_{21}i(n) - Z_{22}i(n+1) \end{aligned} \tag{2.53}$$

and the equation to determine the λ_r's is

$$|f(\lambda)| = 0$$

where now $f(\lambda)$ represents the matrix

$$f(\lambda) = Z_{12}\lambda^2 - (Z_{11} + Z_{22})\lambda + Z_{21} \tag{2.54}$$

Let q_r be proportional to any non-zero column in $F(\lambda_r)$ where $F(\lambda)$ is the adjoint of $f(\lambda)$ and let $\bar{q}_r'$ be proportional to any non-zero row of $F(\lambda_r)$. The matrices Q and $\bar{Q}$ in the expressions (2.45) for $v(n)$ and $i(n)$ are given by

$$\begin{aligned} Q &= [q_1, q_2, \cdots q_m] \\ \bar{Q} &= [\bar{q}_1, \bar{q}_2, \cdots \bar{q}_m] \end{aligned} \tag{2.55}$$

where $\bar{q}_r$ is the column obtained by transposing the row $\bar{q}_r'$. From (2.45) and (2.53) equations for P and $\bar{P}$ in terms of Q and $\bar{Q}$ are obtained:

$$\begin{aligned} P &= Z_{11}Q - Z_{12}Q\Lambda = -Z_{22}Q + Z_{21}Q\Lambda^{-1} \\ \bar{P} &= -Z_{11}\bar{Q} + Z_{12}\bar{Q}\Lambda^{-1} = Z_{22}\bar{Q} - Z_{21}\bar{Q}\Lambda \end{aligned} \tag{2.56}$$

The difference equations (2.53) are satisfied by our expressions for $v(n)$ and $i(n)$ by virtue of the relations

$$\begin{aligned} Z_{12}Q\Lambda^2 - (Z_{11} + Z_{22})Q\Lambda + Z_{21}Q = 0 \\ Z_{12}\bar{Q}\Lambda^{-2} - (Z_{11} + Z_{22})\bar{Q}\Lambda^{-1} + Z_{21}\bar{Q} = 0 \end{aligned} \tag{2.57}$$

which are a consequence of the properties of the individual columns of Q and $\bar{Q}$.

If the system extends to $n = +\infty$ and if the voltages and currents are to remain finite at $n = \infty$ the elements of $\bar{a}$ must be zero and the expressions (2.45) for $v(n)$ and $i(n)$ become

$$\begin{aligned} v(n) &= P\Lambda^n a = P\Lambda^n P^{-1} v(o) \\ i(n) &= Q\Lambda^n a = Q\Lambda^n Q^{-1} i(o) \\ v(n) &= PQ^{-1} i(n), \qquad i(n) = QP^{-1} v(n) \end{aligned} \tag{2.58}$$

where we have assumed that P^{-1} and Q^{-1} exist. We accordingly introduce the characteristic impedance and admittance matrices Z_o and Y_o associated with propagation in the positive direction, i.e., in the direction of n increasing.

$$\begin{aligned} v(n) = Z_o i(n), \qquad i(n) &= Y_o v(n), \qquad Z_o = Y_o^{-1} \\ Z_o = PQ^{-1} = Z_{11} - Z_{12}Q\Lambda Q^{-1} &= -Z_{22} + Z_{21}Q\Lambda^{-1}Q^{-1} \\ Y_o = QP^{-1} = Y_{11} + Y_{12}P\Lambda P^{-1} &= -Y_{22} - Y_{21}P\Lambda^{-1}P^{-1} \end{aligned} \tag{2.59}$$

Incidentally, since Z_o must be a symmetrical matrix the above equations show that $Z_{12}Q\Lambda Q^{-1}$ and $Z_{21}Q\Lambda^{-1}Q^{-1}$ are symmetrical. Z_o and Y_o satisfy the relations

$$Z_oCZ_o + Z_oD - AZ_o - B = 0 \qquad Y_oBY_o + Y_oA - DY_o - C = 0$$
$$(Z_{22} + Z_o)Z_{12}^{-1}(Z_{11} - Z_o) = Z_{21}, \qquad (Y_{22} + Y_o)Y_{12}^{-1}(Y_{11} - Y_o) = Y_{21} \quad (2.60)$$
$$Z_oQ\Lambda Q^{-1} = P\Lambda P^{-1}Z_o \qquad Y_oP\Lambda P^{-1} = Q\Lambda Q^{-1}Y_o$$

The characteristic and admittance matrices $\bar{Z}_o$ and $\bar{Y}_o$ associated with propagation in the negative direction are introduced in a similar way. Suppose the system extends to $n = -\infty$. Then $a = o$ and

$$\begin{aligned} v(n) &= \bar{P}\Lambda^{-n}\bar{a} = \bar{P}\Lambda^{-n}\bar{P}^{-1}v(o) \\ i(n) &= \bar{Q}\Lambda^{-n}\bar{a} = -\bar{Q}\Lambda^{-n}\bar{Q}^{-1}i(o) \\ v(n) &= -\bar{P}\bar{Q}^{-1}i(n), \qquad i(n) = -\bar{Q}\bar{P}^{-1}v(n) \end{aligned} \quad (2.61)$$

Hence we write

$$\begin{aligned} v(n) &= -\bar{Z}_oi(n), \qquad i(n) = -\bar{Y}_ov(n) \\ \bar{Z}_o &= \bar{P}\bar{Q}^{-1} = -Z_{11} + Z_{12}\bar{Q}\Lambda^{-1}\bar{Q}^{-1} = Z_{22} - Z_{21}\bar{Q}\Lambda\bar{Q}^{-1} \\ \bar{Y} &= \bar{Q}\bar{P}^{-1} = -Y_{11} - Y_{12}\bar{P}\Lambda^{-1}\bar{P}^{-1} = Y_{22} + Y_{21}\bar{P}\Lambda\bar{P}^{-1} \end{aligned} \quad (2.62)$$

$\bar{Z}_o$ and $\bar{Y}_o$ satisfy the relations

$$\begin{aligned} &\bar{Z}_oC\bar{Z}_o - \bar{Z}_oD + A\bar{Z}_o - B = 0 \qquad \bar{Y}_oB\bar{Y}_o - \bar{Y}_oA + D\bar{Y}_o - C = 0 \\ &(Z_{11} + \bar{Z}_o)Z_{21}^{-1}(Z_{22} - \bar{Z}_o) = Z_{12} \qquad (Y_{11} + \bar{Y}_o)Y_{21}^{-1}(Y_{22} - \bar{Y}_o) = Y_{12} \end{aligned} \quad (2.63)$$

The fact that $Q'(Z_o + \bar{Z}_o)\bar{Q} = P'(Y_o + \bar{Y}_o)\bar{P}$ is a diagonal matrix may be used as a check on computations.

When the expressions (2.45) for $v(n)$ and $i(n)$ are placed in (2.3), $j^\circ(n)$ and $u^\circ(n)$ being zero, we obtain the relations

$$\begin{aligned} P\Lambda^{-1} &= AP + BQ \qquad P\bar{\Lambda} = A\bar{P} - B\bar{Q} \\ Q\Lambda^{-1} &= CP + DQ \qquad \bar{Q}\Lambda = -C\bar{P} + D\bar{Q} \end{aligned} \quad (2.64)$$

When the typical section contains generators the difference equation to be solved is of the form (2.42)

$$Gx(n + 2) + Hx(n + 1) + G'x(n) = g(n) \quad (2.42)$$

This is true for symmetrical as well as unsymmetrical sections, G being a symmetrical matrix in the former case so that $G' = G$. The expressions for $v(n)$ and $i(n)$ are those of (2.45) with the particular solutions added:

$$\begin{aligned} v(n) &= P\Lambda^na + \bar{P}\Lambda^{-n}\bar{a} + u(n) \\ i(n) &= Q\Lambda^na - \bar{Q}\Lambda^{-n}\bar{a} + j(n) \end{aligned} \quad (2.65)$$

where P, $\overline{P}$, Q, $\bar{Q}$ are determined as before and $u(n)$ and $j(n)$ depend upon the generators.

Here we shall consider only the physically important case in which the voltages of the generators in the nth section are proportional to $e^{-n\theta}$ where θ is a constant. In this case $g(n)$ may be expressed as

$$g(n) = ge^{-n\theta} \tag{2.66}$$

where g is a column matrix whose elements are independent of n. A particular solution is obtained by assuming

$$x(n) = ye^{-n\theta}$$

Setting this in (2.42) gives

$$(Ge^{-2\theta} + He^{-\theta} + G')y = g$$

Hence a particular solution is

$$x(n) = (Ge^{-2\theta} + He^{-\theta} + G')^{-1}ge^{-n\theta} \tag{2.67}$$

This method fails when θ is equal to one of the roots $\lambda_1, \cdots \lambda_m, \lambda_1^{-1}, \cdots \lambda_m^{-1}$. In this case, a particular integral may be obtained by a method similar to one described in §5.11 of F.D.C.

APPENDIX I

Classical Solution of Uniform Transmission Line Equations

For the sake of convenience we again assume that there are three circuits in the transmission line. The equations to be solved are:

$$\frac{dv}{dx} = -Zi, \qquad \frac{di}{dx} = -Yi \tag{1.48}$$

We adopt here the notation associated with equations (1.19) and (1.20), $f(\gamma^2)$ being the characteristic matrix of ZY, $F(\gamma^2)$ the adjoint of $f(\gamma^2)$, and $\gamma_1^2, \gamma_2^2, \gamma_3^2 (m = 3)$ being the roots, supposed distinct, of $|f(\gamma^2)| = 0$. The propagation constants $\gamma_1, \gamma_2, \gamma_3$ are those square roots of $\gamma_1^2, \gamma_2^2, \gamma_3^2$ which in physical systems have a positive real part.

The solution may be constructed[11] as follows: Let the column p_r be proportional (with any convenient constant of proportionality) to any non-zero column of $F(\gamma_r^2)$. The non-zero columns of $F(\gamma_r^2)$ are proportional to each other according to a theorem in matrix algebra.[12] Construct the square matrix P from the columns p_1, p_2, p_3:

$$P = [p_1, p_2, p_3] \tag{A1.1}$$

[11] The method is that described in F.D.C. §5.7(a) and §5.10

[12] F.D.C. §3.5 Theorem D.

and obtain the square matrix Q from P:

$$Q = Z^{-1}PG = YPG^{-1} \tag{A1.2}$$

where G is the diagonal matrix

$$G = \begin{bmatrix} \gamma_1 & 0 & 0 \\ 0 & \gamma_2 & 0 \\ 0 & 0 & \gamma_3 \end{bmatrix} \tag{A1.3}$$

The voltages and currents at any point x are

$$\begin{aligned} v(x) &= PM(x)a + PM(-x)\bar{a} \\ i(x) &= QM(x)a - QM(-x)\bar{a} \end{aligned} \tag{A1.4}$$

where a and $\bar{a}$ are arbitrary column matrices associated with propagation in the positive and negative directions of x and $M(x)$ is the diagonal matrix

$$M(x) = \begin{bmatrix} e^{-\gamma_1 x} & 0 & 0 \\ 0 & e^{-\gamma_2 x} & 0 \\ 0 & 0 & e^{-\gamma_3 x} \end{bmatrix} \tag{A1.5}$$

The values of a and $\bar{a}$ are to be determined from the boundary conditions. When the line extends to $x = \infty$

$$\begin{aligned} v(x) &= PM(x)P^{-1}v(o) = Z_0 i(x) \\ i(x) &= QM(x)Q^{-1}i(o) \end{aligned} \tag{A1.6}$$

where the characteristic impedance matrix Z_0 is given by

$$\begin{aligned} Z_0 = PQ^{-1} &= PG^{-1}P^{-1}Z = PGP^{-1}Y^{-1} \\ &= ZQG^{-1}Q^{-1} = Y^{-1}QGQ^{-1} \end{aligned} \tag{A1.7}$$

Since $v = p_r e^{\gamma_r x}$ and $i = q_r e^{\gamma_r x}$, where q_r is the rth column of Q, are solutions the differential equations give

$$(I\gamma_r^2 - ZY)p_r = 0, \qquad (I\gamma_r^2 - YZ)q_r = 0 \tag{A1.8}$$

and from these it follows that

$$P^{-1}ZYP = Q^{-1}YZQ = G^2 \tag{A1.9}$$

The relations (A1.8) may be used to prove the following:

1. The elements in the rth column of Q are proportional to those in the non-zero rows of $F(\gamma_r^2)$.
2. The matrix $P'Q$ is a diagonal matrix and from this it follows that if ψ is any diagonal matrix

$$(P\psi P^{-1})' = Q\psi Q^{-1} \tag{A1.10}$$

3. The characteristic impedance matrix Z_0 satisfies the relation

$$Z = Z_0 Y Z_0 \tag{A1.11}$$

4. The inverse matrices P^{-1} and Q^{-1} always exist if $\gamma_1, \gamma_2, \gamma_3$ are distinct.

APPENDIX II

Classical Solution of Symmetrical Section Line Equations—I

The method of this section is very similar to that of Appendix I. The equations to be solved are (2.10) or one of the sets

$$\begin{aligned} v(n) &= Z_{11}i(n) - Z_{12}i(n+1) \\ v(n+1) &= Z_{12}i(n) - Z_{11}i(n+1) \end{aligned} \tag{A2.1}$$

$$\begin{aligned} i(n) &= Y_{11}v(n) + Y_{12}v(n+1) \\ -i(n+1) &= Y_{12}v(n) + Y_{11}v(n+1) \end{aligned} \tag{A2.2}$$

which are obtained from (2.5) and (2.6). We shall use the notation associated with equation (2.19), $f(\zeta)$ being the characteristic matrix of A, $F(\zeta)$ the adjoint of $f(\zeta)$, and $\zeta_1, \zeta_2, \cdots \zeta_m$ the roots, assumed unequal, of the characteristic equation $|f(\zeta)| = 0$. The diagonal matrices Λ and Σ are defined by

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \\ 0 & & & \lambda_m \end{bmatrix},$$
$$\Sigma = \begin{bmatrix} \sqrt{\zeta_1^2 - 1} & 0 & \cdots\cdots & 0 \\ 0 & \sqrt{\zeta_2^2 - 1} & & \\ 0 & & & \sqrt{\zeta_m^2 - 1} \end{bmatrix} \tag{A2.3}$$

where

$$2\zeta_r = \lambda_r + \lambda_r^{-1}, \qquad \lambda_r = \zeta_r - \sqrt{\zeta_r^2 - 1} = \frac{1}{\zeta_r + \sqrt{\zeta_r^2 - 1}} \tag{A2.4}$$

In general, electrical energy will be dissipated in the typical section and from the physical significance of λ_r, as seen from equations (A2.8) below, it follows that the sign of the radical in (A2.4) may be chosen so that $|\lambda_r| < 1$. Since $\sqrt{\zeta_r^2 - 1} = \zeta_r - \lambda_r = \frac{1}{2}(\lambda_r^{-1} - \lambda_r)$ it follows that

$$\Sigma = \tfrac{1}{2}(\Lambda^{-1} - \Lambda) \tag{A2.5}$$

Let the column matrix s_r be proportional to any non-zero column in $F(\zeta_r)$ where $F(\zeta)$ is the adjoint of $f(\zeta)$. (It follows from the theory of matrices that the non-zero columns of $F(\zeta_r)$ differ from each other only by a multiplying factor.) The matrix S is then formed by taking s_1 to be the first column, s_2 the second and so on.

$$S = [s_1, s_2, \cdots s_m] \tag{A2.6}$$

Similarly let the row matrix t_r' be proportional to any nonvanishing row of $F(\zeta_r)$ and form the matrix T where

$$T = [t_1, t_2, \cdots t_m] \tag{A2.7}$$

in which t_r is the column matrix obtained by transposing t_r'.[13]

Solving our difference equations for the passive case by the customary method gives the expressions

$$\begin{aligned} v(n) &= P\Lambda^n a + P\Lambda^{-n}\bar{a} \\ i(n) &= Q\Lambda^n a - Q\Lambda^{-n}\bar{a} \end{aligned} \tag{A2.8}$$

for the voltages and the currents. P and Q are square matrices and a and $\bar{a}$ are column matrices whose elements are determined by the boundary conditions. a and $\bar{a}$ are of the same nature as constants of integration. The minus sign over $\bar{a}$ indicates that it is associated with propagation in the negative direction, i.e., in the direction of n decreasing.

P and Q may be chosen in a number of ways, each choice requiring different values of a and $\bar{a}$ to represent the same system. In all cases, however, the rth column of P may be expressed as $\alpha_r s_r$ where α_r is a scalar multiplier which may depend upon r. Similarly the rth column of Q may be expressed as $\beta_r t_r$. When either P or Q has been chosen the other one is fixed since equations (A2.2) and (A2.1) require

$$\begin{aligned} Q &= Y_{11}P + Y_{12}P\Lambda = -Y_{11}P - Y_{12}P\Lambda^{-1} \\ P &= Z_{11}Q - Z_{12}Q\Lambda = -Z_{11}Q + Z_{12}Q\Lambda^{-1} \end{aligned} \tag{A2.9}$$

Some useful choices are,

$$\begin{aligned} &1.\quad P = S, && Q = -Y_{12}S\Sigma = B^{-1}S\Sigma \\ &2.\quad P = S\Sigma && Q = Z_{12}^{-1}S = CS \\ &3.\quad Q = T, && P = Z_{12}T\Sigma = C^{-1}T\Sigma \\ &4.\quad Q = T\Sigma && P = -Y_{12}^{-1}T = BT \end{aligned} \tag{A2.10}$$

The particular choice to be made depends upon the system of difference equations which is being used. In choices 1 and 2, T is not required and in 3 and 4, S is not required. However, if both S and T are known some of

[13] Methods of determining s_r and t_r' are available. A description will be found in F.D.C. §4.12.

the matrix multiplication may be avoided. Taking choice 1 as an example, we may determine the rth row of Q from the expression $\beta_r t_r$. To determine β_r only one element in the rth column of $-Y_{12}S\Sigma$ need be known, for β_r is the quotient obtained by dividing this element by the corresponding element in t_r. The product $P'Q$ must be a diagonal matrix, and the same is true of $S'T$. This may serve to check computations.

That the expressions for $v(n)$ and $i(n)$ given by (A2.8) and (A2.10) satisfy the transmission equations (A2.1), (A2.6) and (2.10) may be verified by direct substitution and use of

$$S(\Lambda + \Lambda^{-1}) = 2AS \qquad T(\Lambda + \Lambda^{-1}) = 2A'T \tag{A2.11}$$

These relations follow from the properties of the individual columns of S and T.

When the system extends to $n = \infty$ $\bar{a}$ must be zero in order that the voltages and currents may remain finite. This is true because λ_r is chosen so that $|\lambda_r| < 1$. From equations (A2.8) it follows that

$$\begin{aligned} v(n) &= P\Lambda^n a = P\Lambda^n P^{-1} v(o) \\ i(n) &= Q\Lambda^n a = Q\Lambda^n Q^{-1} i(o) \\ v(n) = PQ^{-1} i(n) &\qquad i(n) = QP^{-1} v(n) \end{aligned} \tag{A2.12}$$

the reciprocal matrices Q^{-1} and P^{-1} always exist when the sections are symmetrical and the roots $\zeta_1, \zeta_2, \cdots \zeta_m$ distinct. The last equations in (A2.12) suggest the introduction of the characteristic impedance and admittance matrices Z_o and Y_o:

$$\begin{aligned} v(n) &= Z_o i(n), \qquad i(n) = Y_o v(n), \qquad Z_o = Y_o^{-1}. \\ Z_o &= PQ^{-1} = Z_{11} - Z_{12}Q\Lambda Q^{-1} = -Z_{11} + Z_{12}Q\Lambda^{-1}Q^{-1} \\ &= S\Sigma^{-1}S^{-1}B = S\Sigma S^{-1}Z_{12} \\ &= Z_{12}T\Sigma T^{-1} = BT\Sigma^{-1}T^{-1} \\ Y_o &= QP^{-1} = Y_{11} + Y_{12}P\Lambda P^{-1} = -Y_{11} - Y_{12}P\Lambda^{-1}P^{-1} \\ &= -Y_{12}S\Sigma S^{-1} = CS\Sigma^{-1}S^{-1} \\ &= T\Sigma^{-1}T^{-1}C = -T\Sigma T^{-1}Y_{12} \end{aligned} \tag{A2.13}$$

Not all of the expressions for Z_o and Y_o obtainable from (A2.10) have been included in (A2.13). Z_o and Y_o are symmetrical matrices. Although P and Q are arbitrary to some extent the same is not true of Z_o and Y_o. Computed values of Z_o and Y_o may be checked by use of the relations

$$\begin{aligned} A^2 - I &= (Z_o Z_{12}^{-1})^2 = (Y_{12}^{-1} Y_o)^2 \\ Z_o C Z_o &= B, \qquad Y_o B Y_o = C \\ Y_o Z_{12} &= -Y_{12} Z_o \end{aligned} \tag{A2.14}$$

Sometimes it is desirable to terminate a line consisting of a finite number of sections by a network which simulates an infinite line. As is known, the elements in one such network may be obtained from Y_o. Every terminal is joined to every other terminal, including the return terminal (denoted by the subscript o), by the branches of this network. The admittance of the branch connecting terminal i to terminal j, $i \neq 0, j \neq 0$, is $-y_{ij}$ where y_{ij} is the element in the ith row and jth column of Y_o. The admittance of the branch connecting terminal i to terminal o is $y_{i1} + y_{i2} + \cdots + y_{ii} + \cdots + y_{im}$, i.e., it is the sum of all the elements whose first subscript is i.

APPENDIX III

Classical Solution of Symmetrical Section Line Equations—II

When the electrical properties of a typical symmetrical section are to be determined by measurement, equations (A2.1) and (A2.2) show that Z_{11} and Y_{11} may be obtained by measurements at one end. In order to obtain Y_{12} and Z_{12} measurements have to be made at both ends. Expressions for $v(n)$ and $i(n)$ will now be given which depend only upon Z_{11} and Y_{11} and hence are useful in case the measurements are restricted to one end.

The method is based upon the equations

$$\begin{aligned} v(n+2) + v(n) &= Z_{11}[i(n) - i(n+2)] \\ i(n+2) + i(n) &= Y_{11}[v(n) - v(n+2)] \end{aligned} \tag{A3.1}$$

which may be derived from (A2.1) and (A2.2). Combining these equations leads to

$$[I - Z_{11}Y_{11}][v(n+2) + v(n-2)] + 2\,[I + Z_{11}Y_{11}]v(n) = 0$$

$$[I - Y_{11}Z_{11}][i(n+2) + i(n-2)] + 2\,[I + Y_{11}Z_{11}]i(n) = 0$$

The first step in the solution is to solve the equation

$$|\,\mu I - Z_{11}Y_{11}\,| = 0 \tag{A3.2}$$

for its roots $\mu_1, \mu_2, \cdots \mu_m$ which we shall suppose are distinct. The diagonal matrices M and $M^{\frac{1}{2}}$ are defined by

$$M = \begin{bmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & & \\ \vdots & & & \\ 0 & \cdots\cdots & & \mu_m \end{bmatrix}, \qquad M^{\frac{1}{2}} = \begin{bmatrix} \mu_1^{\frac{1}{2}} & 0 & \cdots & 0 \\ 0 & \mu_2^{\frac{1}{2}} & & \\ \vdots & & & \\ 0 & \cdots\cdots & & \mu_m^{\frac{1}{2}} \end{bmatrix} \tag{A3.3}$$

and Λ is defined as in (A2.3) where λ_r is given by

$$\lambda_r = \sqrt{\frac{\mu_r^{\frac{1}{2}} - 1}{\mu_r^{\frac{1}{2}} + 1}}, \qquad \mu_r = \left[\frac{1 + \lambda_r^2}{1 - \lambda_r^2}\right]^2 \tag{A3.4}$$

The sign of $\mu_r^{\frac{1}{2}}$ is chosen so that $|\lambda_r| < 1$, and this is the value to be used in $M^{\frac{1}{2}}$. However, there is an ambiguity in the sign of λ_r which is inherent in this method. A relation between Λ and $M^{\frac{1}{2}}$ is

$$M^{\frac{1}{2}} = (I + \Lambda^2)(I - \Lambda^2)^{-1} \tag{A3.5}$$

Let u_r be proportional to any non-zero column and w_r' be proportional to any non-zero row of the matrix adjoint to $[\mu_r I - Z_{11}Y_{11}]$ and form the matrices

$$U = [u_1, u_2, \cdots u_m]$$
$$W = [w_1, w_2, \cdots w_m]$$

(cf. equations (A2.6) and (A2.7) for S and T) where w_r is the column obtained from w_r'.

The voltages and currents are given, as before, by

$$\begin{aligned} v(n) &= P\Lambda^n a + P\Lambda^{-n}\bar{a} \\ i(n) &= Q\Lambda^n a - Q\Lambda^{-n}\bar{a} \end{aligned} \tag{A2.8}$$

and there is again a number of ways in which P and Q may be chosen. In all cases the rth column of P may be expressed as $\alpha_r u_r$ and the rth column of Q as $\beta_r w_r$. The equations fixing Q when P is chosen and vice versa are, from equations (A3.1)

$$\begin{aligned} Q &= Y_{11}PM^{-\frac{1}{2}} \\ P &= Z_{11}QM^{-\frac{1}{2}} \end{aligned} \tag{A3.6}$$

where $M^{-\frac{1}{2}}$ is the inverse of $M^{\frac{1}{2}}$. Equations (A3.6) may also be obtained from (A2.10).

Suitable choices for P and Q are

$$\begin{aligned} &1.\ P = U, && Q = Y_{11}UM^{-\frac{1}{2}} = Z_{11}^{-1}UM^{\frac{1}{2}} \\ &2.\ P = UM^{\frac{1}{2}}, && Q = Y_{11}U = Z_{11}^{-1}UM \\ &3.\ Q = W, && P = Z_{11}WM^{-\frac{1}{2}} = Y_{11}^{-1}WM^{\frac{1}{2}} \\ &4.\ Q = WM^{\frac{1}{2}}, && P = Z_{11}W = Y_{11}^{-1}WM \end{aligned} \tag{A3.7}$$

$P'Q$ and $U'W$ must be diagonal matrices. That the expressions for $v(n)$ and $i(n)$ just derived satisfy the difference equations (A3.1) may be verified by making use of

$$UM = Z_{11}Y_{11}U, \qquad WM = Y_{11}Z_{11}W \tag{A3.8}$$

Equations (A3.8) follow from the properties of the individual columns of U and W. The characteristic impedance and admittance matrices are given by

$$\begin{aligned} Z_o &= PQ^{-1} = Z_{11}QM^{-\frac{1}{2}}Q^{-1} = PM^{\frac{1}{2}}P^{-1}Y_{11}^{-1} \\ &= UM^{-\frac{1}{2}}U^{-1}Z_{11} = UM^{\frac{1}{2}}U^{-1}Y_{11}^{-1} \\ &= Z_{11}WM^{-\frac{1}{2}}W^{-1} = Y_{11}^{-1}WM^{\frac{1}{2}}W^{-1} \\ Y_o &= QP^{-1} = Y_{11}PM^{-\frac{1}{2}}P^{-1} = QM^{\frac{1}{2}}Q^{-1}Z_{11}^{-1} \\ &= Y_{11}UM^{-\frac{1}{2}}U^{-1} = Z_{11}^{-1}UM^{\frac{1}{2}}U^{-1} \\ &= WM^{-\frac{1}{2}}W^{-1}Y_{11} = WM^{\frac{1}{2}}W^{-1}Z_{11}^{-1} \end{aligned} \tag{A3.9}$$

The matrices Z_o and Y_o may be checked by means of the relations

$$Z_oY_{11} = Z_{11}Y_o, \qquad Z_oY_{11}Z_o = Z_{11}, \qquad Y_oZ_{11}Y_o = Y_{11} \tag{A3.10}$$

Another set of solutions may be based upon the equations

$$\begin{aligned} 2v(n) &= -Z_{12}[i(n+1) - i(n-1)] \\ 2i(n) &= Y_{12}[v(n+1) - v(n-1)] \end{aligned} \tag{A3.11}$$

which are derivable from (A2.1) and (A2.2). Combining these equations gives, upon using $Y_{12}^{-1}Z_{12}^{-1} = -BC$,

$$\begin{aligned} v(n+2) - 2v(n) + v(n-2) &= 4BCv(n) \\ i(n+2) - 2i(n) + i(n-2) &= 4CBi(n) \end{aligned} \tag{A3.12}$$

However, we shall not consider these equations here beyond pointing out that they lead to

$$\begin{aligned} P &= Z_{12}Q\Sigma, \qquad & Q &= -Y_{12}P\Sigma \\ P\Sigma^2 &= BCP, \qquad & Q\Sigma^2 &= CBQ \end{aligned}$$

which may also be derived from (A2.10).

APPENDIX IV

Relations Between the Square Matrices of a Multi-Terminal Section

When the reciprocal theorems of network theory are applied to equations (2.1) and (2.2) it is found that Z_{11}, Z_{22}, Y_{11}, Y_{22} are symmetrical and

$$Z_{21} = Z'_{12}, \qquad Y_{21} = Y'_{12} \tag{A4.1}$$

i.e., Z_{21} and Y_{21} are the transposed matrices of Z_{12} and Y_{12}, respectively.

Solving equations (2.1) for the currents and comparing the result with (2.2) shows that

$$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}^{-1} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}, \qquad \begin{bmatrix} i^\circ(n) \\ -j^\circ(n) \end{bmatrix} = -\begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}\begin{bmatrix} v^\circ(n) \\ u^\circ(n) \end{bmatrix}$$

These are partitioned matrices. The square matrices have $2m$ rows and columns and the column matrices have $2m$ elements. The first of these relations may be written as

$$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \tag{A4.2}$$

where I denotes the unit diagonal matrix of order m. Multiplying the two matrices on the left together and equating the elements of the product to the elements on the right gives

$$\begin{aligned} Z_{11}Y_{11} + Z_{12}Y_{21} &= I \\ Z_{11}Y_{12} + Z_{12}Y_{22} &= 0 \\ Z_{21}Y_{11} + Z_{22}Y_{21} &= 0 \\ Z_{21}Y_{12} + Z_{22}Y_{22} &= I \end{aligned} \tag{A4.3}$$

Transposing the matrices in these equations leads to other relations. Thus, from the first we obtain $Y_{11}Z_{11} + Y_{12}Z_{21} = I$. These equations also yield expressions for the Y's in terms of the Z's and vice versa.

A somewhat similar treatment involving equations (2.1) and (2.3) leads to expressions for the Z's in terms of A, B, C and D. The Y's may be likewise expressed. These relations are given in the following table.

$$\begin{aligned} Y_{11} &= DB^{-1} & Y_{11}^{-1} &= Z_{11} - Z_{12}Z_{22}^{-1}Z_{21} \\ Y_{12} &= C - DB^{-1}A = -B'^{-1} & Y_{12}^{-1} &= Z_{21} - Z_{22}Z_{12}^{-1}Z_{11} \\ Y_{21} &= -B^{-1} & Y_{21}^{-1} &= Z_{12} - Z_{11}Z_{21}^{-1}Z_{22} \\ Y_{22} &= B^{-1}A & Y_{22}^{-1} &= Z_{22} - Z_{21}Z_{11}^{-1}Z_{12} \end{aligned}$$

$$\begin{aligned} Z_{11} &= AC^{-1} & Z_{11}^{-1} &= Y_{11} - Y_{12}Y_{22}^{-1}Y_{21} \\ Z_{12} &= AC^{-1}D - B = C'^{-1} & Z_{12}^{-1} &= Y_{21} - Y_{22}Y_{12}^{-1}Y_{11} \\ Z_{21} &= C^{-1} & Z_{21}^{-1} &= Y_{12} - Y_{11}Y_{21}^{-1}Y_{22} \\ Z_{22} &= C^{-1}D & Z_{22}^{-1} &= Y_{22} - Y_{21}Y_{11}^{-1}Y_{12} \end{aligned} \tag{A4.4}$$

$$\begin{aligned} A &= Z_{11}Z_{21}^{-1} = -Y_{21}^{-1}Y_{22} \\ B &= Z_{11}Z_{21}^{-1}Z_{22} - Z_{12} = -Y_{21}^{-1} \\ C &= Z_{21}^{-1} = Y_{12} - Y_{11}Y_{21}^{-1}Y_{22} \\ D &= Z_{21}^{-1}Z_{22} = -Y_{11}Y_{21}^{-1} \end{aligned}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} D' & -B' \\ -C' & A' \end{bmatrix} \qquad \begin{aligned} AD' - BC' &= I \\ CD' &= DC' \end{aligned} \qquad \begin{aligned} AB' &= BA' \\ DA' - CB' &= I \end{aligned}$$

$$\begin{bmatrix} v^\circ(n) \\ -i^\circ(n) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} u^\circ(n) \\ -j^\circ(n) \end{bmatrix}$$

The following equations in which λ is an arbitrary scalar multiplier may be verified by equating coefficients of powers of λ and using the relations just given.

$$(Z_{21} - \lambda Z_{11})(Y_{11} + \lambda Y_{12}) = (\lambda Z_{12} - Z_{22})(\lambda Y_{22} + Y_{21})$$

$$(\lambda^2 Z_{12} - \lambda Z_{11} - \lambda Z_{22} + Z_{21})(Y_{11} + \lambda Y_{12})$$
$$= (\lambda Z_{12} - Z_{22})(\lambda^2 Y_{12} + \lambda Y_{11} + \lambda Y_{22} + Y_{21}) \qquad \text{(A4.5)}$$

$$\begin{bmatrix} -Y_{21} & 0 \\ 0 & Z_{21} \end{bmatrix} \begin{bmatrix} \lambda A - I & \lambda B \\ \lambda C & \lambda D - I \end{bmatrix} = \begin{bmatrix} \lambda Y_{22} + Y_{21} & \lambda I \\ \lambda I & \lambda Z_{22} - Z_{21} \end{bmatrix}$$

Sometimes it is of interest to obtain the elements of Y_{12}, say, when Z_{11}, Z_{22}, Y_{11}, Y_{22} are known. Relations helpful in studying this problem are

$$Y_{11}Z_{11}Y_{12} = Y_{12}Z_{22}Y_{22}, \qquad Y_{11}Z_{11}Y_{11} - Y_{11} = Y_{12}Z_{22}Y_{21}$$
$$Y_{12}Y_{22}^{-1}Y_{21} = Y_{11} - Z_{11}^{-1} \qquad Z_{12} = -Z_{11}Y_{12}Y_{22}^{-1}$$
$$Y_{21}Y_{11}^{-1}Y_{12} = Y_{22} - Z_{22}^{-1} \qquad Z_{21} = -Y_{12}^{-1}(Y_{11}Z_{11} - I)$$

When the typical section is symmetrical some simplification takes place and we have

$$Y_{11} = Y_{22} \qquad Z_{11} = Z_{22} \qquad A = D' \qquad AB = BA'$$
$$Y_{12} = Y_{21} \qquad Z_{12} = Z_{21} \qquad B = B' \qquad A'C = CA$$
$$C = C' \qquad A^2 - BC = I \qquad \text{(A4.6)}$$
$$Z_{11}Y_{11} + Z_{12}Y_{12} = I \qquad A'B^{-1}A - C = B^{-1}$$
$$Z_{11}Y_{12} + Z_{12}Y_{11} = 0$$

APPENDIX V

Properties of the Matrix $G\lambda^2 + H\lambda + G'$

The matrix

$$f(\lambda) = G\lambda^2 + H\lambda + G' \qquad \text{(A5.1)}$$

which entered the discussion of the case of unsymmetrical sections has a number of interesting properties which are given below. G and H are square matrices with m rows each, and H is required to be symmetrical. As before, we shall denote by $\lambda_1, \cdots \lambda_m, \lambda_1^{-1}, \cdots \lambda_m^{-1}$ the 2 m roots of the determinantal equation

$$|f(\lambda)| = 0$$

and we shall suppose these roots to be distinct. Let the column k_r and the row l_r be such that

$$k_r l_r = F(\lambda_r) \tag{A5.2}$$

where $F(\lambda)$ is the matrix adjoint to $f(\lambda)$, and let the square matrices K and L be defined by

$$K = [k_1, k_2, \cdots k_m], \qquad L = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_m \end{bmatrix} \tag{A5.3}$$

Comparison of (A5.3) and (2.50) suggests that when G and H are expressed in terms of the Y's we have the relations

$$K = P, \qquad L = \bar{P}' \tag{A5.4}$$

The method of choosing the column p_r and the row $\bar{p}_r'$ shows that they are related by

$$p_r \bar{p}_r' = \gamma_r F(\lambda_r)$$

instead of (A5.2) where γ_r may turn out to be any non-zero constant, and consequently equations (A5.4) are not satisfied in general. Nevertheless K and L may be regarded as particular choices for P and $\bar{P}'$. In the same way K and L may be regarded as particular choices for Q and $\bar{Q}'$ when G and H are expressed in terms of the Z's. There is still some arbitrariness connected with K and L since the product $k_r l_r$ is unchanged when the k_r is multiplied by some number and l_r is divided by the same number.

The relations which correspond to (2.52) and (2.57) are

$$\begin{aligned} GK\Lambda^2 + HK\Lambda + G'K &= 0 \\ GL'\Lambda^{-2} + HL'\Lambda^{-1} + G'L' &= 0 \end{aligned} \tag{A5.5}$$

where Λ is the diagonal matrix whose elements are $\lambda_1, \lambda_2, \cdots \lambda_m$. These relations are consequences of the properties of k_r and l_r. Two more relations may be obtained by transposition. From the first of (A5.5) and the transposed of the second it follows that

$$\begin{aligned} GK\Lambda K^{-1} + H + G'K\Lambda^{-1}K^{-1} &= 0 \\ L^{-1}\Lambda LG + H + L^{-1}\Lambda^{-1}LG' &= 0 \end{aligned} \tag{A5.6}$$

where it is assumed that the reciprocal matrices K^{-1} and L^{-1} exist. Combinations similar to $K\Lambda K^{-1}$, $K\Lambda^{-1}K^{-1}$, etc. enter the expressions (2.59) for Z_o and Y_o.

By differentiating the equation

$$f(\lambda)F(\lambda) = \Delta(\lambda)I,$$

where $\Delta(\lambda)$ is the determinant

$$\Delta(\lambda) = |f(\lambda)| = |G\lambda^2 + H\lambda + G'|,$$

it may be proved that

$$\begin{aligned} GK\Lambda K^{-1} + H + L^{-1}\Lambda LG &= L^{-1}EK^{-1} \\ G'K\Lambda^{-1}K^{-1} + H + L^{-1}\Lambda^{-1}LG' &= -L^{-1}EK^{-1} \end{aligned} \tag{A5.7}$$

in which E is the diagonal matrix

$$E = \begin{bmatrix} \overset{(1)}{\Delta}(\lambda_1) & 0 & \cdots & 0 \\ 0 & \overset{(1)}{\Delta}(\lambda_2) & & \\ \vdots & & & \\ 0 & & & \overset{(1)}{\Delta}(\lambda_m) \end{bmatrix}$$

and

$$\overset{(1)}{\Delta}(\lambda_r) = \left[\frac{d}{d\lambda}\Delta(\lambda)\right]_{\lambda=\lambda_r}$$

Since the roots λ_r are assumed to be distinct, $\overset{(1)}{\Delta}(\lambda_r) \neq 0$.

We also have the equations

$$\begin{aligned} KE^{-1}L &= L'E^{-1}K' \\ GK\Lambda E^{-1}L - GL'\Lambda^{-1}E^{-1}K' &= I \end{aligned} \tag{A5.8}$$

The first equation of (A5.8) shows that $KE^{-1}L$ is a symmetrical matrix. From this and the second equation it follows that

$$GK\Lambda K^{-1} - GL'\Lambda^{-1}L'^{-1} = L^{-1}EK^{-1} \tag{A5.9}$$

From the first of equations (A5.7) and the second of (A5.6)

$$GK\Lambda K^{-1} - L^{-1}\Lambda^{-1}LG' = L^{-1}EK^{-1} \tag{A5.10}$$

and the comparison with (A5.9) shows that the matrix $GL'\Lambda^{-1}L'^{-1}$ is symmetrical. The other matrices of this type are also symmetrical as may now be seen from equations (A5.6) and (A5.7). Results of this sort may be obtained from physical principles by noting that Z_o and Y_o must be symmetrical matrices.

As an example of the application of these formulas we assume that G and H are expressed in terms of the Y's. Then we may take K and L' to be particular choices of P and $\overline{P}$ and equation (A5.9) becomes

$$Y_{12}P\Lambda P^{-1} - Y_{12}\overline{P}\Lambda^{-1}\overline{P}^{-1} = (\overline{P}')^{-1}EP^{-1}.$$

From equations (2.59) and (2.62)

$$Y_o + \overline{Y}_o = Y_{12}P\Lambda P^{-1} - Y_{12}\overline{P}\Lambda^{-1}\overline{P}^{-1}$$

and hence

$$\overline{P}'(Y_o + \overline{Y}_o)P = E.$$

For the more general choice of P and $\overline{P}$ allowed in §2.8 the diagonal matrix E is replaced by a general diagonal matrix. Similarly it follows that

$$\overline{Q}'(Z_o + \overline{Z}_o)Q$$

is a diagonal matrix.

Engineering Problems in Dimensions and Tolerances

By W. W. WERRING

Dimensional Units

The basic unit in most considerations of dimensions in the United States is the inch. The value of the inch is so important that many companies including the Bell System maintain in their measurement laboratories a standard yard bar calibrated against the standard at the National Bureau of Standards. In spite of this it is an interesting and curious fact that though all have been much concerned over the legal value of the dollar there has been little interest even among engineers in the exact legal value of the inch. Actually there is no single answer to so simple a question as "What is an inch?" In fact, we have changed from a British inch and our own legal meter, to our inch and the International meter and now through action of the American Standards Association we are actually using an inch based on conversion from the International meter which is neither our own legal inch or the British legal inch—and the British are using it too. Table I shows this history of the legal inch in the United States.

It will be seen that under the present status there exists a difference of two parts in a million between the legal inch and the inch used in the dimensional work of industry. This difference is more theoretical than real in small dimensions and industrial use. The bill before Congress, sponsored by the Bureau of Standards is intended to eliminate this as well as any possible ambiguity in the U. S. inch.

Decimal Dimensioning

In subdividing the inch the modern trend in industry is toward the use of decimals instead of the older common fractions although fractions continue to be used, especially for dimensions of certain materials such as iron pipe, lumber, phenol fiber. In fact even a special decimal system based on using only the tenths and fiftieths of an inch is being considerably discussed by general industry. This system would use a scale on which the smallest division is $\frac{1}{50}''$ or $.020''$ instead of $\frac{1}{64}'' = .0156''$. It is in use by the Ford Motor Company and the values shown in Table II are some of those used in place of common fractions. The decimal equivalents of these common fractions are also shown rounded to 3 decimal places in accordance with American Standard Rules for Rounding off Numerical Values Z25.1-1940.

In the Ford system one and two-digit decimals carry the general tolerance of ± .010″. When greater accuracy is required three-place decimals are used to express a minimum and a maximum value.

The adoption of decimal dimensioning for all drawings prepared at Bell Telephone Laboratories is being actively considered. However, adoption

TABLE I

History of United States Dimensional Standards

Year	Action	Resulting Dimensional Relationships
1830–36	Adoption for Customs Service and for distribution to individual states of standards intended to be the English yard based on a certain portion of an 82 inch bar imported in 1813. The portion selected was supposed to be identical with the English yard.	
1856	Official copy of new British Imperial Yard accepted as standard	International Meter = 39.370147 British Inch
1866	Congress declared metric units lawful and established legal equivalents	Legal Meter in U. S. = 39.37 British Inch
1893	Mendenhall Order set up International meter as the fundamental standard	International Meter = 39.37 U. S. Inch
1933	American Standards Association (Representing Industry) adopts 1 inch = 2.54 centimeters	International Meter = 39.370078 U. S. Inch
1937–41	Bill before Congress but held in committee for amendments	International Meter = 39.370078 U. S. Inch

TABLE II

Examples of Ford Decimals Compared to Common Fractions

Ford Decimal	Common Fraction	Decimal of Existing Common Fraction	American Standard Decimal Equivalents (3 Place)
.02	1/64	.015625	.016
.03	1/32	.03125	.031
.05	3/64	.046875	.047
.06	1/16	.0625	.062
.08	5/64	.078125	.078
.3	7/32	.21875	.219
.46	15/32	.46875	.469

of decimal dimensioning would not of itself result in any changes in our system for establishing tolerance values.

Raw Material Sizes

In contrast to this continued trend toward simplification and rationalization of our systems of dimensional units raw material supply is still complicated by a multitude of obsolete systems of gauge sizes in every day use.

Many in industry have probably grown used to the standard gauges in particular fields but though gauge numbers were undoubtedly initiated as a simplified identification the variety of gauges and the variety of names for the same gauge now merely increases confusion. Sheet metals are handled in terms of a number of gauges such as B&S gauge, U. S. standard gauge and BWG gauge; and sheet soft rubber is even designated in decimals of $\frac{1}{64}$ such as $\frac{4.3}{64}''$. It has become good practice to specify sizes by decimal dimension values and not by gauge numbers and holes by actual decimal size rather than by drill numbers. The actual sizes used, however, are determined in many cases by the values corresponding to old gauge numbers long used commercially, though in large running items mills will and do manufacture to any specified decimal size. For some time it has been the practice of material manufacturers and other large industries thus to discontinue the use of gauge numbers though still using the decimal values of gauge sizes.

There is now under way an effort, organized under committee B32 of the American Standards Association, to eliminate the old wire and sheet metal gauge systems entirely and set up a rational series of American standard thicknesses for all metal sheets and preferred diameters for wire, and insure availability in these sizes. The basic conception of a rational series of sizes is that a uniform degree of choice should be presented between successive sizes. Therefore each size should differ from the next by a fixed percentage. The series should therefore be geometric. A variety of geometric series could be used but in order to permit extending the series indefinitely by shifting the decimal point, the particular series based on the root of 10 has been established internationally as the Preferred Numbers Series for standard sizes. The 5 series is one having 5 numbers between 1 and 10 (or between 10 and 100) and is produced by using as the multiplier the fifth root of 10; the 10 series is produced by multiplying by the 10th root of 10; the 20 series by multiplying by the 20th root of 10 etc. The complete Preferred Numbers Series is explained and listed in various forms in American Standard Z17.1-1936.

The subcommittee working on the sheet metal sizes has recently issued a proposed American Standard of preferred thicknesses for all uncoated flat metals thinner than .250″. These thicknesses are all decimals based on the 20 series of preferred numbers rounded in the standard manner to 3 decimal places. The Preferred Numbers and the proposed thicknesses are shown by Table III. It happens that this series closely approximates the Brown and Sharp gauge used in the nonferrous metals which simplifies that portion of the changeover. If this proposed American Standard is generally approved, as now appears most promising, we will be able to choose thicknesses of any metal interchangeably without the restrictions of ancient gauge sizes es-

TABLE III

Decimal Series of Preferred Numbers 10–100			Proposed Preferred American Standard Thicknesses		
5 Series $\sqrt[5]{10} = 1.6$	10 Series $\sqrt[10]{10} = 1.25$	20 Series $\sqrt[20]{10} = 1.12$	Under .010	.010 to .100	.1120 to .250
10	10	10		.010	
		11.2		.011	.112
	12.5	12.5		.012	.125
		14		.014	.140
16	16	16		.016	.160
		18		.018	.180
	20	20		.020	.200
		22.4		.022	.224
25	25	25		.025	
		28		.028	
	31.5	31.5		.032	
		35.5		.036	
40	40	40	.004	.040	
		45		.045	
	50	50	.005	.050	
		56		.056	
63	63	63	.006	.063	
		71	.007	.071	
	80	80	.008	.080	
		90	.009	.090	
100	100	100		.100	

tablished for reasons which were possibly good and sufficient but which certainly have long been forgotten. Meanwhile, another subcommittee is investigating the possibility of applying a similar series to the diameters of wire. Probably diameters to 4 decimal places will be required.

Dimensional Tolerances

Part Tolerances

Regardless of the dimension decided upon in a design it is obvious that it cannot be regularly manufactured to the exact size. Certain manufacturing variations or tolerances must be expected and these introduce a large share of our dimensional problems.

The usual statement on tolerances is that the larger the tolerance allowed the cheaper the part is to manufacture and, therefore, the tolerance specified should be the widest that will permit functioning. However, this is generally true only of overall tolerances which define the manufacturing methods that may be used. It is true in the sense that apparatus is inexpensive to manufacture if it can be so designed that its functioning is largely independent of variations in dimensions. However, such design is not usually achieved and in much apparatus fairly good overall accuracy of dimensions and fit is necessary for uniform functioning. The problem of setting tolerances then becomes one of distributing certain tolerances over various dimensions and different parts. This is a very difficult problem and in the case of any individual tolerance a larger value does not necessarily mean lower apparatus cost and may even mean the reverse.

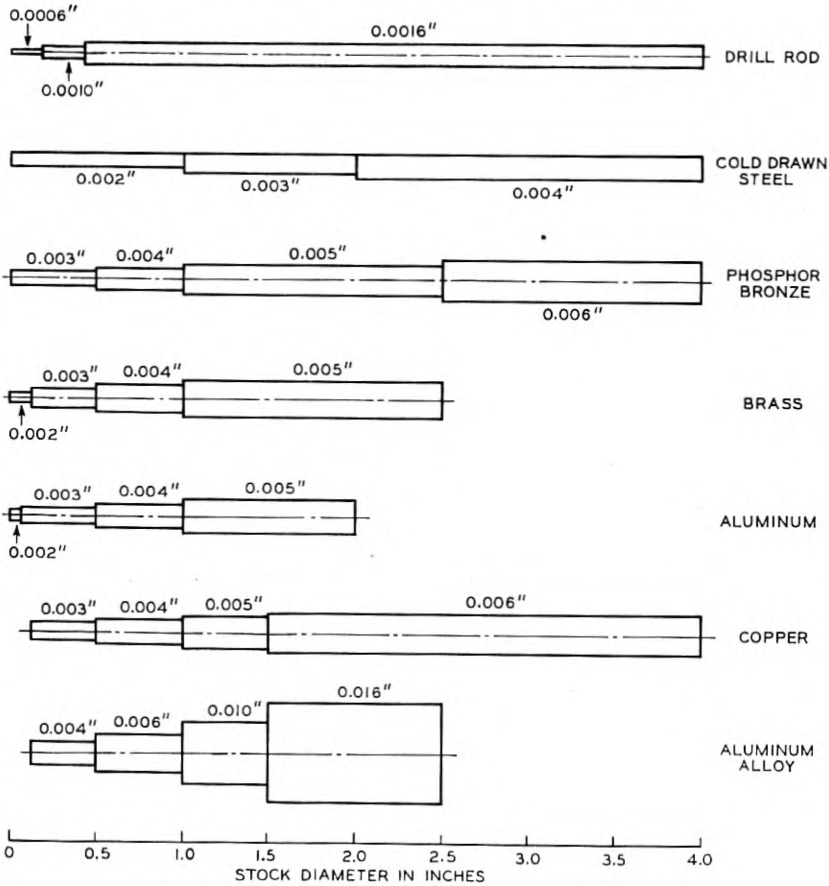

Fig. 1—Total diameter tolerances of commercial round stock

This is easily demonstrated in the case of part tolerances on dimensions which correspond to the dimensions of raw materials. Figure 1 shows the tolerances of commercial grades of round stock. If, for example, engineering requirements dictate the use of a particular material there is no gain in specifying larger tolerances than those to which it is regularly furnished and doing so may require greater accuracy in the mating part. There may even be economy in the use of higher priced material produced to closer tolerances, as for example, drill rod instead of cold drawn steel through economy in the manufacture of associated parts. Similarly manufacture of cantilever springs from sheet stock produced to closer tolerances may reduce the cost of subsequent adjustments. Therefore, when individual part tolerances are involved consideration must always be given to the size tolerances of raw materials.

The same situation exists in the case of tolerances on dimensions produced by a manufacturer's own tools. While close overall limits will require greater overall accuracy of the tools provided and greater frequency of set-ups the most economical distribution of tolerances will be that based upon the normal tolerances that can be expected from various manufacturing operations. Certain degrees of accuracy are inherent in certain types of machines and tools and allowing variations not in proportion to these values serves little if any purpose. Also there are types of combination tools and automatic machines, familiar in mass production practice with which wide tolerances are not an economy because accuracy is required for locating or nesting the part for subsequent operations. Since the distribution of tolerances involves such complex factors of manufacturing method and cost as these, it is desirable for the designing engineer to determine and to indicate unmistakably the effect of tolerances upon functioning and, where interchangeability of individual parts in service is not involved, to allow manufacturing considerations to determine the distribution of tolerances in an assembly.

It is apparent that considerable study of the requirements for functioning of the design, of available materials and the limitations of manufacturing process are required to establish the most economic balance between performance of the apparatus and the required tolerances. Consideration should be given to these tolerance factors in cooperation with manufacturing engineers in an early stage of a design problem so that they may influence the trend of design. This step may avoid the necessity for slow and costly manufacturing developments and delays in starting production. However, completely rigid adherence to the status quo of tolerances is not necessary in long range planning of major design projects. In such cases the trend of progress in materials and manufacture should be determined and anticipated. For example, some cantilever spring design requiring narrow control has been based on sheet material produced to tolerances not commercially available at the time but made so by the time it was needed for production. The extent of progress in this direction is shown by Fig. 2.

Similar progress in manufacturing technique can also be expected. For example, the development of broaching from a comparatively crude operation to the precision method it is today is recent and outstanding.

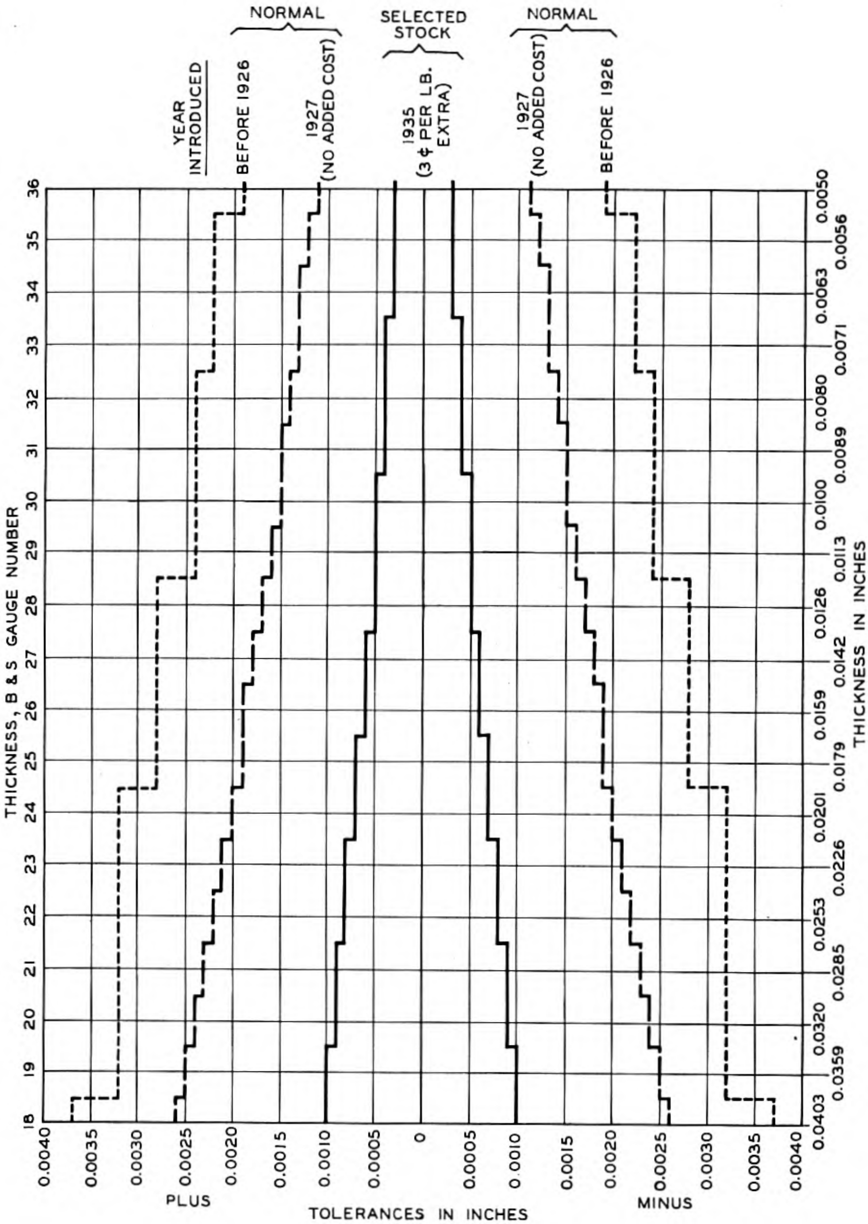

Fig. 2—Improvement in thickness tolerances for brass sheet 1926–1939

Cumulative Assembly Tolerances

Another problem in choice of tolerances is in those cases where a considerable number of parts are additively assembled into a unit as in the case of "spring pileups" used on electrical contacting apparatus such as relays and switches. These consist of considerable numbers of sheet metal springs and insulators alternating and clamped by screws. If the overall tolerance on such an assembly must be taken as the sum of the tolerances of the individual parts various courses of action are presented, the extremes of which are:

1. Very small tolerances must be maintained on the individual parts or
2. Adequate space must be provided in the apparatus for extremely large variations in the assembly.

Small tolerances on the individual parts may be extremely expensive and large space allowances and provisions in associated parts for variations in the assembly may be a serious design handicap.

However, it is recognized that there is obviously small probability that all minimum or all maximum parts will appear in any one assembly. It has been found satisfactory in certain types of such pileups to assume that the maximum dimensional variation that will actually be encountered in an assembly will not be greater than 70% of the sum of the part tolerances. A similar situation exists in many kinds of assemblies or associations of tolerances.

The statistical relationships involved in this problem are indicated by Fig. 3. The curves show the percentage of the cumulative part tolerances within which 99.7% of the assemblies may be expected to be found with different numbers of similar units in the assembly. The solid line is deduced from theoretical relationships. It assumes that the parts are all of one kind, that the parts going to assembly are controlled, of normal distribution and the limits are rationally set to represent the actual conditions. The dotted curves have been deduced from relationships which have been proposed as representing rectangular and triangular distributions of individual part tolerances. The curves may not be truly representative of specific cases because of inconsistent selection of limits or erratic distributions. However, they indicate that the 70% rule on pileups is probably on the safe side in most cases and that closer design of assembly or less restrictive tolerances and cheaper manufacture of piece-parts might be readily possible either (1) by better control, (2) by actual mixing of lots of piece parts or (3) even merely by knowledge of the actual statistical distribution of part dimensions.

The three points indicated in Fig. 3 show the results of a limited experiment in which pileups were assembled from 2083 individual insulators of $\frac{1}{32}''$ phenol fiber taken from factory stock. The establishment of curves by this type of experiment using a sufficiently large and representative sample would be practicable and would permit considerable condensation

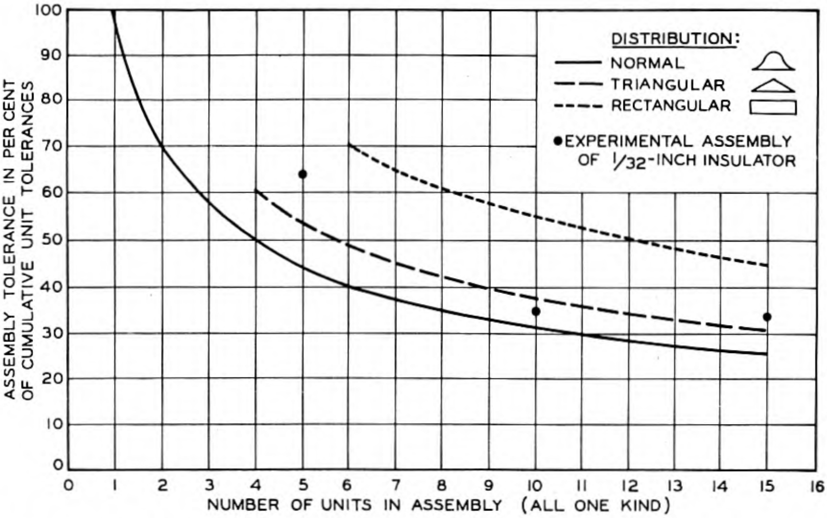

Fig. 3—Statistical relationship of overall tolerance on an assembly and the sum of the individual part tolerances

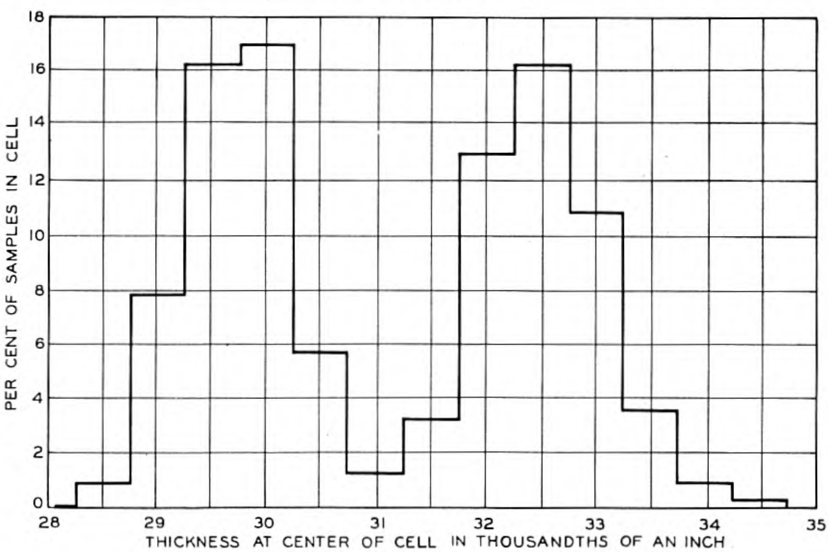

Fig. 4—Distribution of thickness in 2083 pileup insulators

of design on a sound basis. In this particular case the parts used apparently came from only two different sheets of fiber as indicated by the distribution of thickness of the individual parts shown by Fig. 4.

Further statistical analysis of this type of situation is needed together with experimental determination of the distribution of dimensions actually encountered in specific cases.

The distribution of dimensions in a product arises from a variety of causes. One type of cause is the variations such as those in the dimensions and physical properties of raw material which may produce different product dimensions even from a particular tool. A different and more systematic type of cause is the change in the dimensions of tools as a result of wear. The practice followed in establishing tool wear allowances will therefore affect the limits and statistical distribution of part dimensions during the life of the tool. Some designers and some tool makers consider that the specification of a nominal value with plus and minus variations requires a different handling of initial tool dimensions and wear allowances than does the specification only of minimum and maximum limits for a part dimension. Equally good authority maintains that a manufacturer recognizes no difference. Establishment of standard practices in such matters is a needed step in determining the distribution of dimensions to be expected in machined parts. In the present absence of standards or of any consistent attitude on the subject it is necessary for designing and manufacturing engineers to reach an agreement in specific cases where this factor is important.

Such are the factors which determine the tolerances which can be obtained economically or which perhaps will be unavoidably encountered. It is necessary for a designer to keep informed of the interaction of these factors as his design crystallizes and he must also determine the effect of such tolerances upon functioning in order to complete a design which will function properly when assembled in quantity production.

Functional Dimensioning

Effect of Tolerances

If apparatus parts are minute or have complicated relative motions it is recognized that manufacturing drawings to the usual scale have serious limitations to their usefulness in the analysis of the effects of combinations of tolerances. In such cases designers frequently make layouts to larger scales or large scale adjustable models to investigate the effect of variations on functioning. Illustrations of this practice are numerous in the experience of most designers of small apparatus.

Even in large parts which are stationary in use the application of tolerances, in effect, establishes several possible positions for each element and may present problems similar to those involving motion. These are not easily recognized because of a curious limitation inherent in small scale drawings. This limitation is probably well known to most engineers but it is worthwhile to analyze it because it is important to be always aware of it.

This limitation is the fact that in drawings the shape of the part and the effect of all nominal dimensions are actually shown graphically whereas, it is possible to indicate tolerances numerically but not graphically. We are therefore apt to visualize the part as it is graphically shown, that is, without tolerances and to think of the numerical tolerances one at a time rather than in combinations as they affect each other and the shape of the part.

If any dimension, significantly affecting the design of a part, is changed the drawing is immediately corrected so that its meaning will be clear and the functioning of the part can be checked. This obviously facilitates design and manufacture. Yet because they cannot be shown directly by regular drawing methods, we have grown accustomed to not being shown the effect of tolerances or changes in tolerances upon the shape of the part. Nevertheless it is obvious that these effects are critical in the functioning of the part or tolerances would not be set. The fact that these critical features of the design are not actually graphically shown and therefore are not easily seen and understood on the drafting board is a serious detriment in working out a design and in all later analysis of it. The full effect of interrelated variations particularly if in three dimensional space may appear only after tools are in process or the first parts produced and this may be rather late for economy.

Originally this difficult analysis of the effect of tolerances upon functioning probably involved only the designer. The manufacturer tried to make the part as nearly as possible to the nominal values shown and variations from them were accidental. Tolerances were looked upon as an indication of the care required and as a means of inspection for acceptance or rejection. With increasingly complex manufacturing tools the permitted tolerances are utilized more and more in the design of tools to allow the greatest possible wear before defective parts are produced and the tools must be replaced. For mass production parts progressive step type tools are used in which a continuous strip of stock advances by various stages from blank sheet to finished part. Tools of this type are extremely expensive and in order to obtain maximum life full use of allowed variations is made in their design. Design of such tools and the gauges required to maintain quality in mass production therefore also requires analysis of the effect of combinations of variables upon the desired part. As the designer has presumably already made this analysis, and incidentally is best qualified to do it, economy and accuracy dictate that his analysis be transmitted to the manufacturing engineer. The problem is to find means by which he can indicate unmistakably on the drawing his analysis of the required functioning of the part and the manner in which he intends the tolerances to apply, in the event that there is any possibility of misunderstanding.

The essence of this problem and some of the possibilities of solution can best be seen by reference to drawings which illustrate the major points.

Figure 5 shows the drawing of a flat plate dimensioned from center lines but without any tolerances whatever. Some minor dimensions not involved in this discussion are omitted in the interest of simplification but the part shown is in every way a normal one. The meaning of the drawing is completely clear and can be interpreted in but one way no matter from

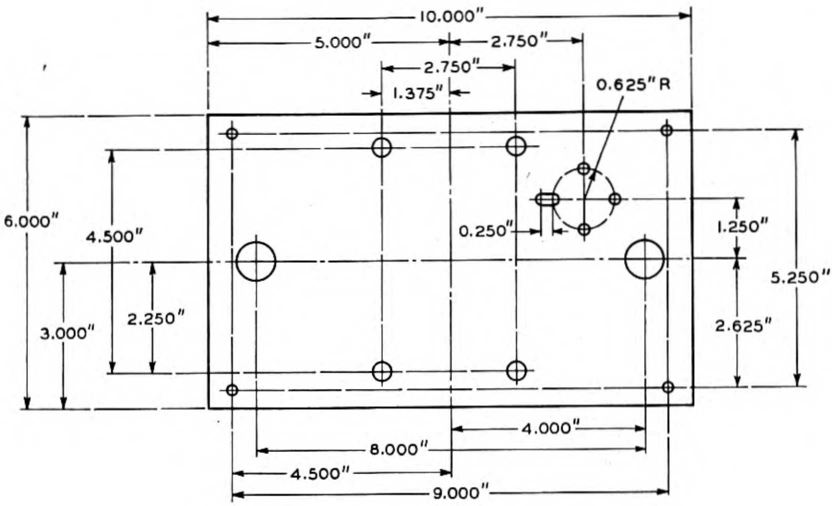

Fig. 5—Flat plate dimensioned without tolerances

what standpoint the analysis is made. The reason for this is obviously that but one value is shown for every dimension.

Figure 6 shows this same drawing dimensioned in exactly the same way with the exception that tolerances are shown for most of the dimensions. To the uninitiated it might appear to present no more problem than the previous drawing without tolerances because of the tendency to visualize the drawing in terms of the nominal dimensions only.

When the engineer analyzes the effect of the combinations of the various tolerances shown, interesting questions immediately arise. In the first place the combination of holes dimensioned 1.25″ ± .002″ from the center line appears to be definitely located because on the drawing the center line is shown in a definite position. Yet when the tolerances are considered the center line of this drawing could actually be shown in several different places as, for example:

1. It may be a line through the centers of the two large holes.
2. It may be a line anywhere from 2.992″ to 3.008″ from the outside edges.
3. It may be 2.247″ to 2.253″ from the small holes in the center of the plate.
4. It may be 2.615″ to 2.635″ from the holes numbered 2 and 4.

In brief, the center line which appears so definitely located on the drawing may actually be rather an indefinite location on the part when the various

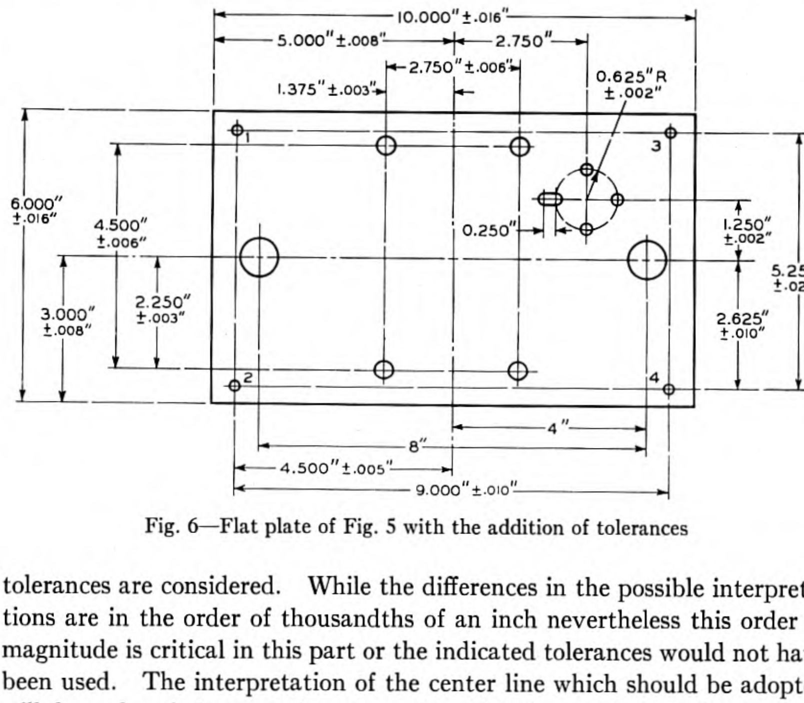

Fig. 6—Flat plate of Fig. 5 with the addition of tolerances

tolerances are considered. While the differences in the possible interpretations are in the order of thousandths of an inch nevertheless this order of magnitude is critical in this part or the indicated tolerances would not have been used. The interpretation of the center line which should be adopted will depend entirely upon the manner in which the part is intended to function and therefore should be indicated by the designing engineer. Obviously, not all designs or all dimensioning will present this difficulty but all should be studied from this viewpoint to determine whether or not they do.

Functional Datum Positions

When the type of uncertainty illustrated exists, it is necessary to indicate clearly the effect of tolerances on functioning by establishing the functional positions to which dimensions should refer. It may be difficult to do this graphically, in which case it is necessary to indicate by notes the particular interpretation which the designer intends. As an example, if the part of Figs. 5 and 6 functions by being located in position by means of the four holes numbered 1, 2, 3 and 4, the intentions of the designer are readily indicated by the following notes:

1. Functional datum line I is midway between the centers of holes 1 and 2 and the centers of holes 3 and 4.
2. Functional datum line II is perpendicular to datum line I at a point midway between the centers of holes 2 and 4.

These notes establish both horizontal and vertical center lines specifically in terms of the center of the one set of dimensions between the holes marked

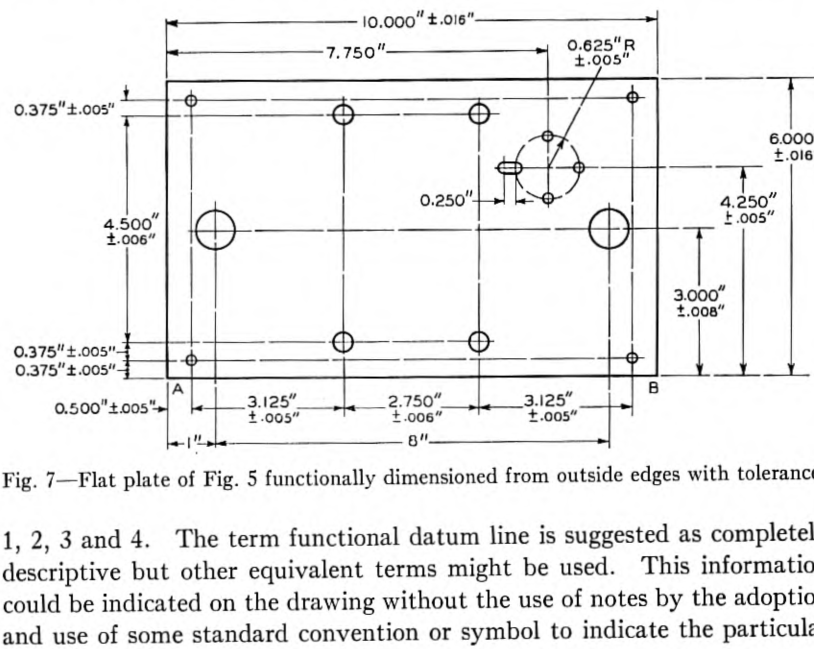

Fig. 7—Flat plate of Fig. 5 functionally dimensioned from outside edges with tolerances

1, 2, 3 and 4. The term functional datum line is suggested as completely descriptive but other equivalent terms might be used. This information could be indicated on the drawing without the use of notes by the adoption and use of some standard convention or symbol to indicate the particular dimension bisected by the center line.

If the functioning of this part were determined by location against the outside edges, this could be readily indicated by dimensioning the part as shown by Fig. 7 and using notes establishing the line A–B as one datum line and the perpendicular to it through A as the other.

In either of these cases the drawing becomes completely definite and subject to only one interpretation. In drawings of this type no change in the method of dimensioning may be required and the problem is solved simply by the addition of suitable notes or symbols indicating the intention of the designer as to functional datum lines.

It is sufficient to establish datum lines in the case of parts which are practically flat pieces with little depth but when a part has substantial depth it will be noted that center lines or other datum lines on a drawing really represent planes in space. In such parts it becomes necessary to establish datum planes rather than lines and three planes at right angles to each other are required.

Figure 8 illustrates such a part which might be an armature such as is used in many pieces of electrical contacting apparatus. In the typical operation of such a part its functioning is determined by the relation of its

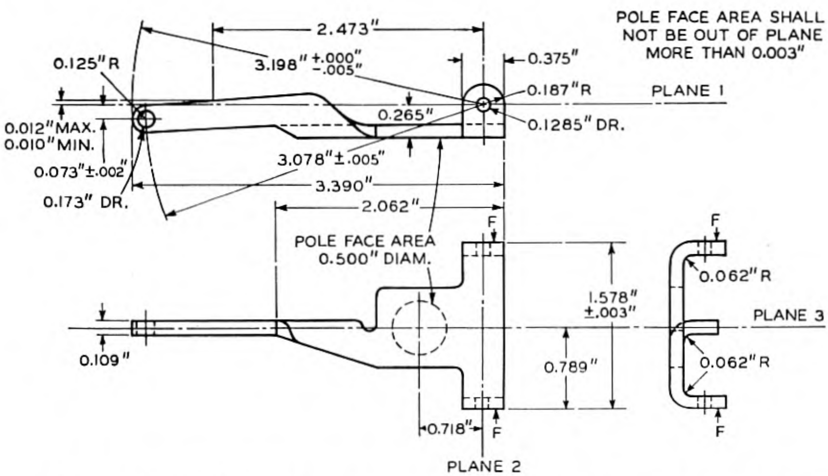

Fig. 8—Functional datum plane dimensioning of magnetic armature type of part

Functional datum plane I passes through the common axis of the two .1285 in. diameter holes and .265 in. above the pole face gauge position.

Functional datum plane II is perpendicular to plane I and passes through the common axis of the two .1285 in. diameter holes.

Functional datum plane III is perpendicular to planes I and II and passes midway between the finished surfaces which are 1.578 ± .003 in. apart.

various dimensions to the position of the pole face and the axis support. In order to indicate this on the drawing it is necessary to establish dimensioning as shown and add to the drawing the notes shown.

These notes establish three functional datum planes, the first through the axis at the point of support and a distance .265″ from the pole face area; the second at right angles to the first through the axis of support and the third at right angles to both the first and second and halfway between the finished surfaces 1.578″ apart. With these planes established the application of all the limits and tolerances shown is based on the operating position and analysis of the design is simplified. The drawing and the intentions of the design engineer cannot be misunderstood.

The clear expression of the designer's intentions by datum plane dimensioning will be appreciated by all concerned with the drawing or the resulting part. Inspection of the part is expedited no less than production. The inspector can usually by means of gauge blocks or simple fixtures set the part up on a surface plate as indicated by the drawings datum planes and positions. He can then establish the conformance of the part with the drawing by simple measurements to the indicated horizontal and vertical planes. When production quantities justify special gauges the required design of the gauge is established clearly by the datum planes.

Invariable or Gauge Dimensions

The drawing of Fig. 8 just described illustrates the use of gauge dimensions. The dimensions .265″ and .718″ and the indicated half-inch diameter for the pole face are all gauge dimensions without tolerances and some statement must be made or understanding reached that they are considered invariable and tolerances not permitted. They represent, it might be said, theoretical dimensions, on the drawing, or in practice they represent tools or gauging apparatus made to the highest standards of accuracy. These invariable dimensions are necessary in order to establish a starting point for the dimensioning of the part. It may appear at first that stating that a dimension has no manufacturing tolerance or variation is a hardship upon the manufacturer but this is not really so because the dimensions are not ones which are actually manufactured in the part. They represent usually dimensions built into tools or gauging equipment which are made to a precision greatly superior to that represented by part tolerances.

Invariable dimensions, or better, gauging dimensions or whatever it is proposed to call them are really not a new invention and it is possible to cite easily recognized examples. For instance, the dimension 2.473″ on Fig. 8 is an invariable gauging dimension not associated with the setting up of datum planes but typical of long standing use of invariable dimensions. We all can recall also the use of the term "theoretically correct position" and it is present practice in the case of vacuum tube bases and similar apparatus to designate the location of the contact studs in terms of a gauge having holes located on "true centers." Last but not least a minimum or maximum limit in its application is itself an invariable dimension.

In effect, datum lines or planes established when necessary by use of invariable or gauging dimensions remove the uncertainty as to the designer's intentions and prevent misunderstandings between design, production and inspecting engineers. Admittedly they do not completely solve all problems of dimensions as probably nothing will. They do, however, transfer whatever problems remain from the field of tolerances on finished product to the realm of tool making tolerances and gauging tolerances. The problem of how invariable is "invariable" remains but we are obviously then considering differences of an order of magnitude not usually vitally significant in the functioning of product parts. Theoretically, all "invariable" dimensions should be taken to the best accuracy of good gauging methods which means that any differences of opinion will be reduced at least to one-fifth and probably to one-tenth of the order of magnitude of those where tolerances themselves are involved.

It will be necessary to specially identify gauging dimensions on drawings to distinguish them from ordinary unlimited dimensions and to indicate that they are dimensions for gauges to which only gauge tolerances apply.

Practical Use of Datum Lines and Planes

It is not usual to establish datum lines on all drawings but if their use is necessary in the layout and design of the part they need to be permanently identified. This use of datum lines and planes on drawings, where necessary, may require somewhat greater drafting effort in the actual production of the drawing but their use results in a simplification of design and of the work of those subsequently using the drawings. It reduces the effort expended in analysis of drawings preparatory to the construction of tools and minimizes the possibility of misunderstandings or errors in tools. In products manufactured only intermittently it is particularly valuable as it minimizes the need for understandings and instructions supplementary to the drawings which may be forgotten between production periods or lost through shifts in personnel.

The overall economy in engineering effort and the reduction of the numerous possibilities of error more than compensate for the increase in the actual work of indicating datum positions, lines or planes upon drawings. In addition the choice of design of punches and dies and similar tools by production engineers is better guided by the designer's requirements if functional datum lines are clearly identified. An obvious example is the use of either the inside or outside of a punched and formed part as the starting point. In brief datum plane dimensioning is a more explicit expression on the drawing, of the designers "end point requirements".

When establishing datum planes, it is important to consider them in terms of the actual physical part rather than in terms of the drawing. Lines which appear as definite points on a drawing may not be actually part of the product when it is completed or may be on surfaces shown as a line on the drawing but rough or unfinished in the part. It is difficult to establish any set of rules covering what shall or shall not be done because each drawing and each part must be considered practically as an individual case. That this is so will be amply demonstrated by a serious study of even one part. However, there are obvious generalities which can be established and Fig. 9 shows some of them.

An example of functional datum plane analysis and dimensioning in three dimensions of a complicated part is shown by Fig. 10. This is the die cast frame for a special selector switch. It is the base upon which many interrelated parts and subassemblies are mounted. The proper functioning of the completely assembled switch depends in large measure on proper manufacture of this casting. In effect, the switch is designed around a vertical shaft passing through points P and Q and planes 1 and 2 are, therefore, established through the axis of this shaft. The production planning engineers intend to design the die and withdraw die plugs from such directions that the mounting surfaces will be smooth, flat and without

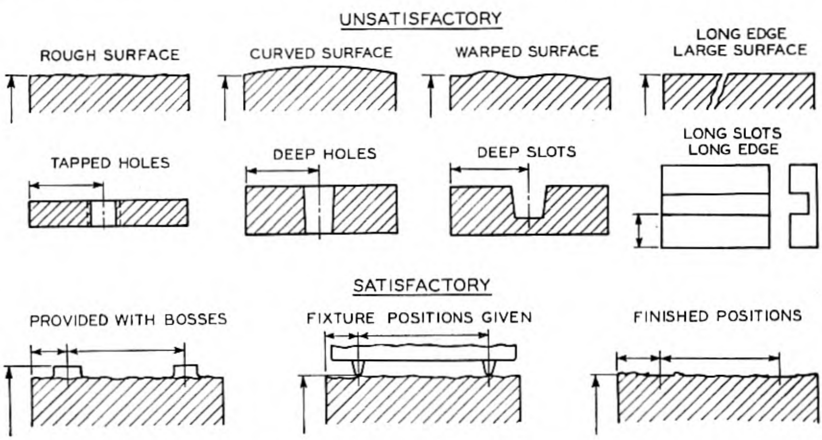

Fig. 9—Types of datum positions

any taper and they intend to use these surfaces as guiding points for their jigs and fixtures. It is for this reason that Plane 1 is established parallel to these mounting surfaces and an indicated distance from them. The other planes are established as shown on this drawing and described by the notes. With this arrangement of planes the designer's analysis in terms of Plane 1 is easily worked out and the reference of Plane 1 to the mounting surfaces permits the production or tool engineer to translate the design of the part into the design of his tools without necessity for further analysis and without the possibility of different interpretations. It will be noted that invariable or gauge dimensions are again used. The complete drawing of this part is very complicated and occupies a drawing practically 4 ft. x 6 ft. The perspective sketch shown and the accompanying notes are incorporated in the drawing as a separate view.

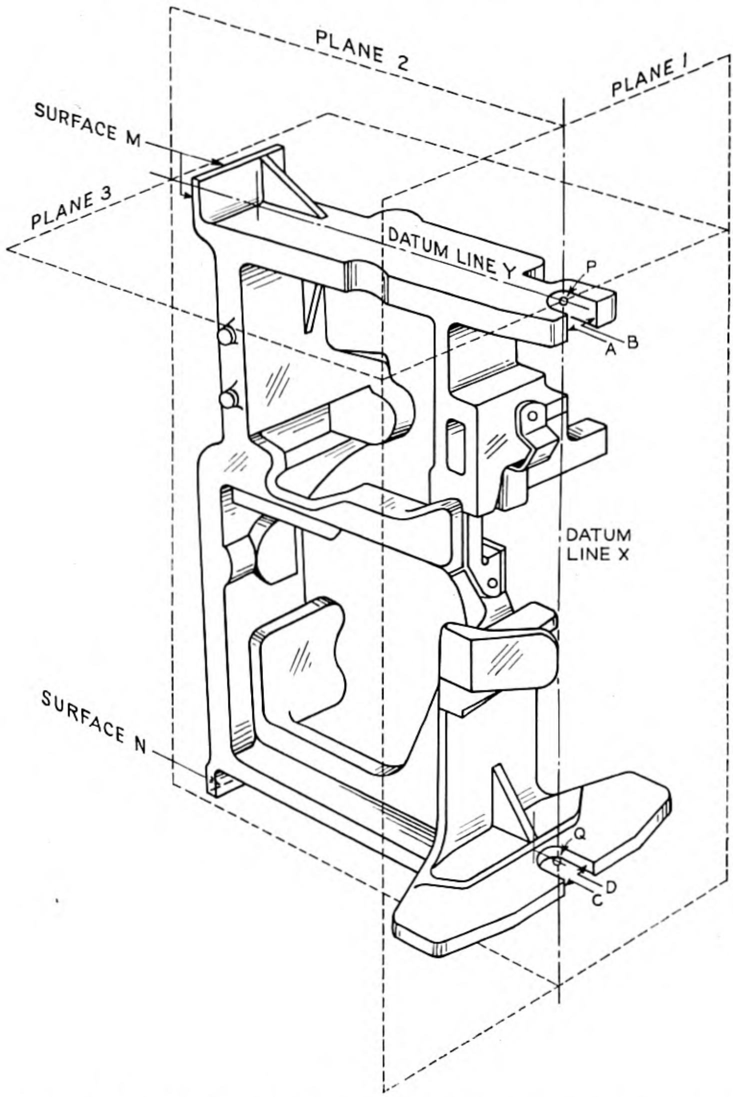

Fig. 10—Functional datum planes of complicated switch frame

Dimensions to datum line "X" or "Y" of the drawing of the frame refer to functional datum planes 1, 2 or 3 described below. Points "P" and "Q" are gauge points used in establishing these datum lines and planes. Points "P" and "Q" shall be half-way between the surfaces "A" and "B" and "C" and "D" respectively and 4.358 in. from the plane of surfaces "M" and "N" on the mounting lugs.

Datum line "X" shall pass through the points "P" and "Q".

Plane 1 shall be parallel to surfaces "M" and "N" and shall include datum line "X".

Plane 2 shall be perpendicular to plane 1 and shall also include datum line "X".

Plane 3 shall be perpendicular to plane 1 and to plane 2 at the point "P".

Datum line "Y" passes through point "P" and is the intersection of planes 2 and 3.

Required Standardization

It is not suggested that the drawings shown and the notes referred to represent a final practice on datum planes. A standard practice in designation of planes and standard terminology and understanding on gauge points and gauge dimensions is required. It will probably be desirable to adopt some symbol or designation for use on drawings to distinguish gauge dimensions which are invariable from ordinary unlimited dimensions to which manufacturing engineers for their own purposes usually add shop tolerances. One thing is certain and that is that datum planes, dimensions and tolerances when established should be primarily in terms of the required functioning of the apparatus. When that is done no one using the drawing in any capacity will have any doubts as to the designer's intention and this results in a great reduction in the discussions and analysis which might otherwise be necessary.

Summary

In summary it may be said that the whole approach to these problems in dimensions and tolerances should be on the basis of functioning. However, good engineering of dimensions and tolerances requires knowledge of what can reasonably be produced and the sources of reasonable tolerance values are:

1. Raw material limits including some knowledge of future trends and developments.
2. The normal accuracy of manufacture, also including anticipation of future improvement.
3. Discussion of trend of design with manufacturing engineers.

Solution of tolerance problems in the final design may involve all of the following steps:

1. Study of the effect of combinations of tolerances on functioning, allowing for statistical effects in accumulations of tolerances.
2. Discussion of this analysis with the production planning engineer because the analysis of tolerance combinations is important in the design of long life tools.
3. Indication of the results of such an analysis by the method of dimensioning drawings.
4. Indication on drawings of functional datum positions, lines or planes established on geometrically correct principles to permanently and unmistakably record the intentions of the designer regarding combinations of variations wherever this is necessary.

Time Division Multiplex Systems

By W. R. BENNETT

Introduction

THE idea of transmitting and receiving independent signals over a common line by means of synchronized switches at the terminals is quite old and has been used in multiplex telegraphy for many years. In general if N signal channels are to be provided over one line, the switching cycle includes N equal time intervals, one of which is allotted to each channel. Each channel is connected to the line throughout a part of its particular time interval and is disconnected throughout the remainder of the cycle. Absence of interference between the channels depends upon the fact that the channels are connected to the line throughout mutually exclusive time intervals. It is thus possible to avoid the use of channel band filters such as are necessary in carrier systems employing frequency as the basis of separation.

Application of time division multiplex methods to telephone channels has been proposed from time to time and some experiments have been made.[1,2,3,4,5,6] It is fairly evident that the concept of simple on-and-off switching giving alternately transmission and complete suppression for the signal from a particular channel on the line is inadequate for speech waves in actual telephone circuits. Imperfections in the transmission properties of the line tend to distort the wave form of the successive signal components and prolong the contribution of one signal into the time allotted for a different channel. It is the object of this paper to present a general quantitative discussion of the factors which enter into the transmission of any type of signal by a system of this kind. It has been found possible to arrive at definite criteria for such matters as the required switching frequency, the conditions to be imposed on contact time for good crosstalk suppression with economy of frequency band, and the transmission requirements which must be met by the intervening circuit to hold the interference between channels to tolerable values. The analysis leads directly to a physical viewpoint of the whole process which, to those familiar with the carrier and

[1] Patten and Minor, *U. S. Patent* 745,734, 1903.
[2] *Electrical World*, Dec. 5, 1903.
[3] Goldschmidt, *U. S. Patent* 1,087,113, Feb. 17, 1914.
[4] Poirson, *Soc. Fr. El.*, Apr. 1920.
[5] Marro, *L'Onde Electrique*, Oct. 1938.
[6] M. Cornilleau, *Revue de Telephones, Telegraphes et T. S. F.*, 13 (1935), pp. 625–643.

sideband philosophy of signal transmission, illuminates the manner in which departures from ideal amplitude and phase characteristics cause crosstalk between the several message channels. It further leads directly to other physical methods for producing and detecting a transmitted signal identical with the essential components derived in time division or switching processes.

A first step in the theoretical solution of the problem was taken by Dr. J. R. Carson, who, in an unpublished memorandum of May 25, 1920, derived quantitative relations between band width and interchannel interference in time division multiplex transmission. Applying Fourier series analysis to on-and-off switching, he showed that if the transmission medium had constant attenuation and linear phase shift for all frequencies below cutoff and no transmission of frequencies above the cutoff, the band width required for satisfactory multichannel telephony would be much wider than needed in conventional carrier methods. A further step was taken by Dr. H. Nyquist, who, in unpublished memoranda of August 24, 1936 and November 12, 1936,[7] showed that the width of band necessary may be reduced by providing a specially devised type of non-uniform transmission characteristic. In the following discussion, we shall see that a similar result can be obtained by control of the switching, and specific switching processes will be described which allow a flat transmission band of minimum width to be used.

In order to arrive at requirements which must be imposed on the various components of the system, we shall first give a theory of time division multiplex transmission in which both the switching processes and the transmission characteristic are completely general. Specific forms which give crosstalk suppression will then be discussed and effects of small departures estimated.

General Theory

We shall assume an N-channel system with a sinusoidal signal impressed on the j^{th} channel. An illustrative arrangement is shown in Fig. 1. Since the system is linear, we may represent currents and voltages by complex quantities with the understanding that the actual currents and voltages are the real components of the expressions used. Accordingly, let the signal voltage impressed on the j^{th} channel be

$$E_j(t) = E_j e^{i\omega_j t} \tag{1}$$

[7] Basic concepts used in Nyquist's analysis were included in his paper, "Certain Topics in Telegraph Transmission Theory," *A. I. E. E., Trans.*, April, 1928, pp. 617–644. Mention is also there made of the equivalence of signal shaping and equalizing in effect on reception of telegraph signals.

and let the switching between the j^{th} channel and the line at the sending end be represented by:

$$I_{sj}(t) = F_j(t)E_j(t), \tag{2}$$

where $I_{sj}(t)$ is the current flowing into the line from the j^{th} channel. The function $F_j(t)$ has the dimensions of an admittance and, in the arrangement shown in Fig. 1, is periodic in time with fundamental frequency $q = 2\pi/T$ radians per second, where T is the time occupied by one cycle of the switching operation. In the interests of economy of analysis, it is preferable for our purposes to assume for $F_j(t)$ a somewhat more general function of time

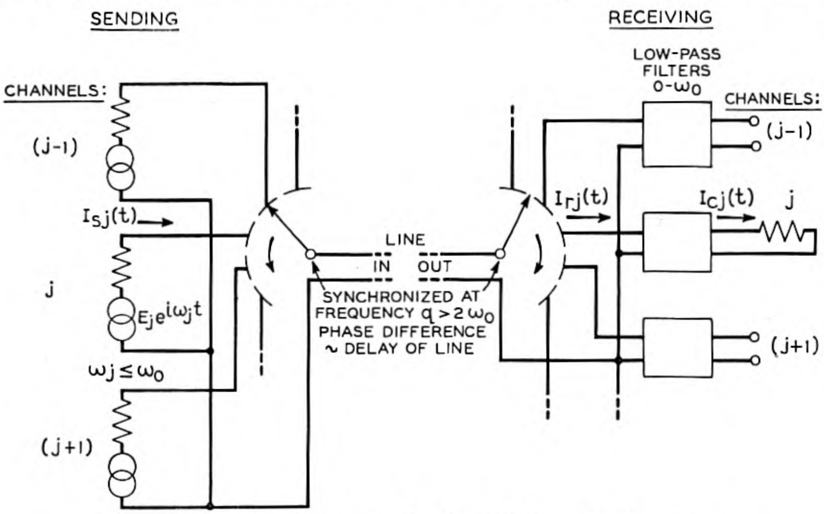

Fig. 1—Elementary arrangement for time division multiplex system

than is directly obtainable with the elementary arrangement of Fig. 1. We shall let

$$F_j(t) = \sum_{m=0}^{\infty} A_{mj} \cos [(\nu + mq)t - \theta_{mj}]. \tag{3}$$

To make the results applicable to Fig. 1, we merely let $\nu = 0$; then by the usual Fourier series analysis,

$$\left.\begin{aligned} A_{0j} &= a_0/2, & \left.\begin{aligned} A_{mj}^2 &= a_{mj}^2 + b_{mj}^2 \\ \tan \theta_{mj} &= b_{mj}/a_{mj} \end{aligned}\right\} m > 0 \\ \theta_{0j} &= 0, & \\ a_{mj} &= \frac{2}{T} \int_{t_1}^{T+t_1} F_j(t) \cos mqt\, dt & \\ b_{mj} &= \frac{2}{T} \int_{t_1}^{T+t_1} F_j(t) \sin mqt\, dt,\ t_1 \text{ arbitrary} & \end{aligned}\right\} \tag{4}$$

The wave (3) consists of the output of the circuit of Fig. 1 with all frequencies shifted by a constant amount ν radians per second; various means of accomplishing this result in the switching process will be discussed later. It is sufficient to point out here that such a shift in frequency is often desirable for optimum utilization of the transmission medium. Combining (2) and (3), we then have:

$$I_{sj}(t) = \frac{E_j}{2} \sum_{m=0}^{\infty} A_{mj} [e^{i(\nu+mq+\omega_j)t - i\theta_{mj}} + e^{-i(\nu+mq-\omega_j)t+i\theta_{mj}}] \tag{5}$$

It is clear from (5) that the result of the switching process is the production of upper and lower side frequencies from the signal on each harmonic of the switching frequency. It is also evident that if more than one signal component is superimposed, the resulting side frequencies constitute sidebands of the same nature as used in amplitude modulation systems. A significant difference between time division and amplitude modulation appears in that in the latter only one sideband or at most one pair of sidebands is transmitted, while the essential character of time division depends on the transmission of a plurality of sidebands. Thus if one pair of sidebands were selected from the output (5) by filtering, the time division process would merely be a particular way of generating the sidebands required in an amplitude modulation system.

The next step in a time division system is the transmission of the wave (5) over a line. The properties of the line in general may be specified by a complex transfer impedance, which we may express here by the ratio of open-circuit output voltage to input current:

$$E_r/I_s = Z(i\omega) \tag{6}$$

The result of applying the wave (5) to the line is then the open-circuit voltage:

$$\begin{aligned} E_{rj}(t) = \frac{E_j}{2} \sum_{m=0}^{\infty} A_{mj} Z[i(\nu + mq + \omega_j)] e^{i(\nu+mq+\omega_j)t - i\theta_{mj}} \\ + \frac{E_j}{2} \sum_{m=0}^{\infty} A_{mj} Z^*[i(\nu + mq - \omega_j)] e^{i(\nu+mq-\omega_j)t+i\theta_{mj}} \end{aligned} \tag{7}$$

In the above we have adopted the notation $Z^*(i\omega)$ to represent the conjugate of $Z(i\omega)$ and have made use of the fact that the response of a network to the applied wave $e^{-i\omega t}$ is the conjugate of the response to $e^{i\omega t}$.

At the receiving end another switching process takes place synchronously with that at the transmitting end. We shall assume that the switching process between the k^{th} channel and the line is represented by the relation

$$I_{rk}(t) = G_k(t) E_{rj}(t), \tag{8}$$

where $I_{rk}(t)$ is the current received in the k^{th} channel and $G_k(t)$ is a periodic function of time with fundamental frequency q. It is understood that j and k may be any two of the N channels. We shall express $G_k(t)$ in a manner analogous to the corresponding switching function at the transmitter, i.e.,

$$G_k(t) = \sum_{n=0}^{\infty} B_{nk} \cos [(\nu + nq)t - \Phi_{nk}] \tag{9}$$

Combining (7), (8), and (9), we find

$$I_{rk}(t) = \frac{E_j}{4} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} A_{mj} B_{nk} Z[i(\nu + mq + \omega_j)]$$

$$(e^{i[2\nu+(m+n)q+\omega_j]t-i(\theta_{mj}+\Phi_{nk})} + e^{i[(m-n)q+\omega_j]t-i(\theta_{mj}-\Phi_{nk})})$$

$$+ \frac{E_j}{4} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} A_{mj} B_{nk} Z^*[i(\nu + mq - \omega_j)]$$

$$(e^{-i[2\nu+(m+n)q-\omega_j]t+i(\theta_{mj}+\Phi_{nk})} + e^{i[(n-m)q+\omega_j]t+i(\theta_{mj}-\Phi_{nk})}) \tag{10}$$

The received wave thus consists of a doubly infinite set of side frequencies involving harmonics of q. It is, however, possible to set up conditions under which the original signal may be selected and the frequencies involving the switching rate may be suppressed by filtering. If $\nu = 0$, such separation is possible provided

$$\omega_j < q/2, \tag{11}$$

for it then follows that a low-pass filter with cutoff frequency at $q/2$ will not pass any of the components with frequencies dependent on q. The condition (11) follows from the fact that the lowest frequency of (10) dependent on q is $q - \omega_j$, and hence we must make $q - \omega_j > \omega_j$ in order to separate ω_j from $q - \omega_j$. In other words the sidebands on adjacent harmonics must not overlap. If $\nu > 0$, the condition (11) also suffices as far as suppression of terms dependent on q are concerned, but an additional condition is required to suppress frequencies dependent on ν in the special case in which $\nu < q/2$, i.e., the case of ν less than the maximum allowable value of ω_j. For in the latter case the frequency $2\nu + (m + n)q - \omega_j$ is less than $q - \omega_j$ in the special case of $m = n = 0$. The additional condition needed is evidently either $A_{0j} = 0$ or $B_{0k} = 0$. If $\nu = 0$ or if $\nu > q/2$, this condition is unnecessary.

Assuming then that (11) is fulfilled, and that a low-pass filter with cutoff at $q/2$ is inserted in the output of each channel, we calculate for the typical channel output current:

$$I_{ck}(t) = Y_{jk} E_j e^{i\omega_j t}, \tag{12}$$

where the value of Y_{jk} is as follows:

Case 1, $\nu = 0$

$$Y_{jk} = A_{0j}B_{0k}Z(i\omega_j) + \frac{1}{4}\sum_{m=1}^{\infty} A_{mj}B_{mk}Z[i(mq + \omega_j)]e^{-i(\theta_{mj}-\Phi_{nk})} + \frac{1}{4}\sum_{m=1}^{\infty} A_{mj}B_{mk}Z^*[i(mq - \omega_j)]e^{i(\theta_{mj}-\Phi_{nk})} \quad (13)$$

Case 2, $\nu > 0$; A_{0j} or $B_{0k} = 0$ if $\nu < q/2$

$$Y_{jk} = \frac{1}{4}\sum_{m=0}^{\infty} A_{mj}B_{mk}Z[i(\nu + mq + \omega_j)]e^{-i(\theta_{mj}-\Phi_{mk})} + \frac{1}{4}\sum_{m=0}^{\infty} A_{mj}B_{mk}Z^*[i(\nu + mq - \omega_j)]e^{i(\theta_{mj}-\Phi_{mk})} \quad (14)$$

The combination of an N-channel time division multiplex system with low-pass filters in the receiving branches is thus found to be equivalent to a linear network having N pairs of input and output terminals with the transfer admittance from the j^{th} pair of input terminals to the k^{th} pair of output terminals given by Y_{jk} in (13) or (14). The transfer admittance is calculated by summing the contributions of upper and lower sidebands on harmonics of the switching frequency and is affected directly by the transmitting properties of the medium at the side band frequencies. The result we have obtained is of sufficient generality to include all cases we shall treat in this paper. We shall now proceed to specific examples.

On-and-Off Switching with Commutator

When an ideal commutator is used as a switching means, the switching functions for the N channels are identical except for a time displacement which is the same between all pairs of consecutive channels. This condition is expressed by:

$$F_j(t) = F_1[t - (j - 1)T/N] \quad (15)$$

Thus $F_1(t)$, the switching function for the first channel becomes a reference function, $F_2(t)$ is the same except for a time delay of T/N, $F_3(t)$ is delayed by $2T/N$, etc. Substitution of (15) in (3) gives the relations:

$$\left.\begin{aligned} A_{mj} &= A_{m1} \\ \theta_{mj} &= \theta_{m1} + (j - 1)2m\pi/N \end{aligned}\right\} \quad (16)$$

If we further suppose that the commutator makes contact between the typical channel and the line throughout a fraction x of the time interval

T/N allotted to that channel and breaks contact throughout the remainder of the switching cycle, we may write the reference switching function as:

$$F_1(t) = \begin{pmatrix} A, & -xT/2n < t < xT/2N \\ 0, & xT/2N < t < (2N - x)T/2N \end{pmatrix} \tag{17}$$

Hence from (4)

$$\left.\begin{aligned} A_{01} &= Ax/N, \\ A_{m1} &= \frac{2A}{m\pi} \sin \frac{mx\pi}{N}, \; m > 0 \\ \theta_{m1} &= 0 \end{aligned}\right\} \tag{18}$$

In the receiving device, the corresponding switching process should be delayed with respect to the transmitter by a time interval t_0 equal to the time of transmission of the line. Hence we write

$$\left.\begin{aligned} B_{01} &= Bx/N \\ B_{m1} &= \frac{2B}{m\pi} \sin \frac{mx\pi}{N}, \; m > 0 \\ \Phi_{m1} &= mqt_0 \end{aligned}\right\} \tag{19}$$

B_{mj} and θ_{mj} are related to B_{m1} and Φ_{m1} in a manner analogous to (16).

The time of transmission of a distorting line is not precisely definable, but may be represented for our purpose by a linear phase component of $Z(i\omega)$. That is, we write

$$Z(i\omega) = Z_0(i\omega)e^{-it_0\omega}, \tag{20}$$

where t_0 is the slope of a straight line giving the best linear approximation to the phase vs. frequency curve, and $Z_0(i\omega)$ is the impedance function remaining after the subtraction of $t_0\omega$ from the actual phase shift ordinates. Substituting (15)–(19) in (12), we find

$$\begin{aligned} Y_{jk} = ABe^{-it_0\omega_j}\Bigg(&\frac{x^2}{N^2} Z_0(i\omega_j) \\ &+ \sum_{m=1}^{\infty} \frac{\sin^2 mx\pi/N}{m^2\pi^2} Z_0[i(mq + \omega_j)]e^{-i(j-k)2m\pi/N} \\ &+ \sum_{m=1}^{\infty} \frac{\sin^2 mx\pi/N}{m^2\pi^2} Z_0^*[i(mq - \omega_j)]e^{i(j-k)2m\pi/N}\Bigg) \end{aligned} \tag{21}$$

If the attenuation of the line is constant throughout the range ω_j to $Mq + \omega_j$ and all frequencies above the latter value are suppressed, (21) becomes

$$Y_{jk} = \frac{ABx^2 Z_0 e^{-it_0\omega_j}}{N^2}\left[1 + 2\sum_{m=1}^{M}\left(\frac{\sin mx\pi/N}{mx\pi/N}\right)^2 \cos\,(j-k)2m\pi/N\right] \quad (22)$$

The crosstalk ratio or ratio of amplitude of signal received in the k^{th} channel to that received in the j^{th} channel when signal is transmitted in the j^{th} channel is, therefore,

$$\frac{Y_{jk}}{Y_{jj}} = \frac{1 + 2\sum_{m=1}^{M}\left(\frac{\sin m\pi\, x/N}{m\pi\, x/N}\right)^2 \cos 2m\pi(k-j)/N}{1 + 2\sum_{m=1}^{M}\left(\frac{\sin m\pi\, x/N}{m\pi\, x/N}\right)^2} \quad (23)$$

Results of calculations made for a 10-channel system from (23) for $x = 1$ and $x = .5$, corresponding to no lost time and half lost time respectively in switching are shown in Fig. 2. It may be noted that adjacent channel crosstalk with half lost time is equivalent to alternate channel crosstalk with no lost time. Examination of the curves reveals a number of significant facts, among which are:

1. Crosstalk is quite imperfectly suppressed when the band width of the line is smaller than the theoretical minimum—the width of one sideband multiplied by the number of channels.

2. As the band width of the line is increased above the theoretical minimum, improvement in crosstalk suppression increases slowly, so that in general the use of frequency range on the line is uneconomical compared with other systems. For example, with no lost time in switching, the band width of the line must be increased tenfold to suppress adjacent channel crosstalk by 40 *db*. This conclusion is, however, to be qualified as follows:

3. When the duration of contact is decreased (less of the available channel time used) definite optimum transmission band widths appear which give better crosstalk suppression than bands somewhat wider or narrower. This suggests the possibility of critical phase relations existing between the contributions from the various sidebands such that if the right number having proper amplitudes and phases can be combined, complete suppression of crosstalk may occur even when the transmitted band width is finite.

When x, the fraction of contact time used, is made to approach zero, the limit of the amplitude factor (18) for the typical harmonic of the switching function is $A_{m1} = 2Ax/N$, which is independent of m. This is consistent with the known fact that a wave consisting of periodically repeated sharp pulses is composed of a large number of harmonics of nearly equal amplitude. If we use very short contact durations in time division, we should accordingly expect a large number of sidebands of nearly equal amplitude. The combination of proper numbers and phases of these sidebands offers a key to the realization of a time division multiplex system giving good crosstalk suppression with economy of frequency band.

Suppose that the duration of contact time is made sufficiently small to realize approximately the limiting values $A_{m1} = 2Ax/N$, $B_{m1} = 2Bx/N$ in transmitting and receiving respectively for the first $2M + 1$ of the sidebands and that by means of a low-pass filter with linear phase shift and

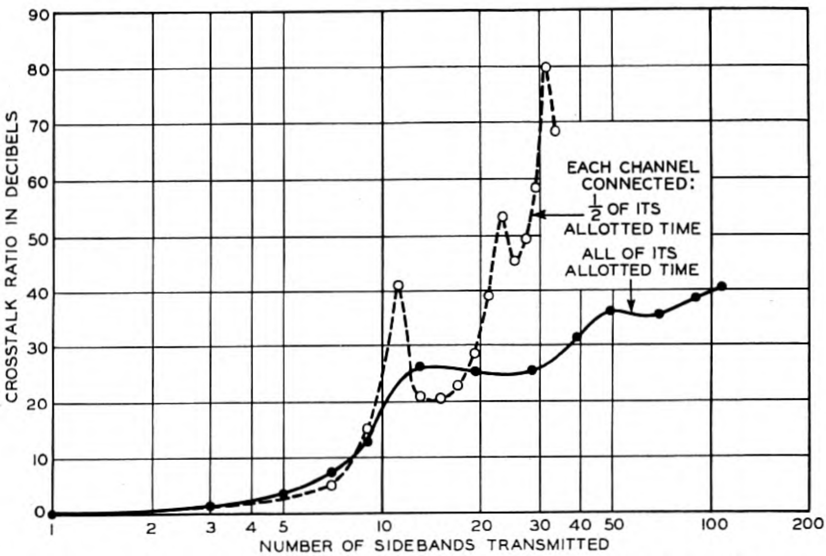

Fig. 2—Crosstalk between adjacent channels of ten-channel time division multiplex system with on-and-off switching. No attenuation or phase distortion in pass band of line

uniform attenuation in its pass-band all other sidebands are removed from the line. The expression (23) then becomes:

$$\frac{Y_{jk}}{Y_{jj}} = \frac{1 + 2\sum_{m=1}^{M} \cos 2m\pi(k - j)/N}{1 + 2M} = \begin{cases} 1, & k = j \\ \dfrac{\sin (2M + 1)\pi(k - j)/N}{(2M + 1) \sin \pi(k - j)/N}, & k \neq j \end{cases} \tag{24}$$

In particular, if

it follows that

$$\left.\begin{aligned} M &= (N - 1)/2, \\ Y_{jk}/Y_{jj} &= 0,\ k \neq j \end{aligned}\right\} \tag{25}$$

Thus there exists in theory a system employing sidebands on zero frequency and the first $(N - 1)/2$ harmonics of the switching frequency, in which multichannel transmission is possible without interchannel interference. Since the required condition (25) may also be written $N = 2M + 1$, an odd number of channels is obtained. Since N sidebands are transmitted, the band width used is the same as the minimum required for N single sideband amplitude modulation channels on a frequency discrimination basis. Sidebands produced on higher harmonics in the time division process must be removed by filtering.

It is to be noted that since it is equality of the N sideband contributions which is important and the amount of each contribution is determined by the transmission characteristic of the line as well as the transmitting and receiving switching processes, it would be theoretically possible to make up for sideband irregularities by equalizing the line. However, the equalization required in the line would be of "stairstep" type rather than smoothly varying with frequency since an error in the value of one harmonic of the switching function produces the same error throughout the entire range occupied by the pair of sidebands associated with that harmonic.

General Switching Functions with Crosstalk Suppression and Minimum Band Width

The above discussion based on the properties of a commutator has led us to an ideal switching function which is, except for an unimportant proportionality factor,

$$F_j(t) = 1 + 2 \sum_{m=1}^{(N-1)/2} \cos m[qt - (j - 1)2\pi/N],\ N \text{ odd} \tag{26}$$

This type of switching is approximately realizable with synchronized commutators having contact widths very narrow in comparison with the spacing between contacts. For a 3000-cycle speech band, the minimum switching rate would be 6000 cycles per second. Such a speed would be difficult to obtain with ordinary mechanical means but would be feasible with rotating electron beams.

The concept of combining detected contributions from a number of sidebands in proper phase to give in-phase addition of desired components and cancellation of unwanted ones leads to a generalization of the switching processes over those possible with synchronized commutators. We note that the switching functions $F_j(t)$ of (2) and $G_k(t)$ of (9) are analogous to carrier waves applied to a product modulator, and an electrical analogue of time division may be realized therefore by applying signal and a suitable carrier to a product modulator. Phase shifts in the carrier supply circuit may be made to serve the same purpose as the angular displacements between commutator segments. It is thus of interest to examine various other possible forms of the function $F_j(t)$ which are suitable for multiplex transmission and investigate methods by which they can be realized.

We note that (26) is suitable for an odd number of channels because it makes use of the direct signal component (or sideband on zero frequency) in addition to the paired sidebands on harmonics of the switching frequency. It seems reasonable to expect that systems for even numbers of channels can be devised using only upper and lower sidebands on harmonics and omitting the signal itself. Complete information for the separation of N channels should be contained in any set of N sidebands; hence we should not be forced to start with the sidebands of lowest frequency, but be able to use other sets with a more suitable place in the spectrum or with better equalization of amplitudes.

We shall derive an expression for a quite general switching function meeting the desired conditions of freedom from crosstalk and economy of band width for an even number of channels by assuming the following forms for A_{mj} and θ_{mj} in (3),

$$A_{mj} = \begin{pmatrix} A, \; n \leqq m \leqq n + N/2 - 1 \\ 0, \; m < n, \text{ or } m > n + N/2 - 1 \end{pmatrix} \tag{27}$$

$$\theta_{mj} = (j - 1)\,(m + h)\,\psi + \alpha \tag{28}$$

The switching function assumed contains $N/2$ harmonics and hence will produce N sidebands. The values of n, h, α and ψ are first assumed to be arbitrary. At the receiving end, a switching function similar except for a time displacement t_0 will be assumed. That is, in (9), we take

$$B_{mk} = \begin{pmatrix} B, \; n \leqq m \leqq n + N/2 - 1 \\ 0, \; m < n \text{ or } m > n + N/2 - 1 \end{pmatrix} \tag{29}$$

$$\Phi_{mk} = (k - 1)\,(m + h)\psi + (mq + \nu)\tau + \alpha \tag{30}$$

Transmission over the line is assumed to be of the distortionless form obtained by setting $Z_0(i\omega) = Z_0$, a constant, in (20). Substituting (27)–(30) in (14), we then calculate

$$Y_{jk} = \frac{NABZ_0 e^{-it_0\omega_j}}{4} \begin{cases} 1, & j = k \\ \dfrac{2 \sin N(j-k)\psi/4 \cos (4n + 4h + N - 2)(j-k)\psi/4}{N \sin (j-k)\psi/2}, & j \neq k \end{cases} \tag{31}$$

Figure 3 shows the curve of admittance vs. frequency required for on-and-off and plus-and-minus switching. Referred to the mid-band admittance as unity, the admittance is reduced to one-half (six *db* loss) at the frequencies $rNq/2$ and $(r + 2)Nq/2$ which are the nominal upper and lower cutoffs. N is an even integer and r is zero or any positive integer. The admittance curve has odd-symmetry about the cutoff frequencies—that is, if at a frequency x cycles below a cutoff frequency, the admittance has the value a, it must be $1 - a$ at a frequency x cycles above the cutoff. The nominal

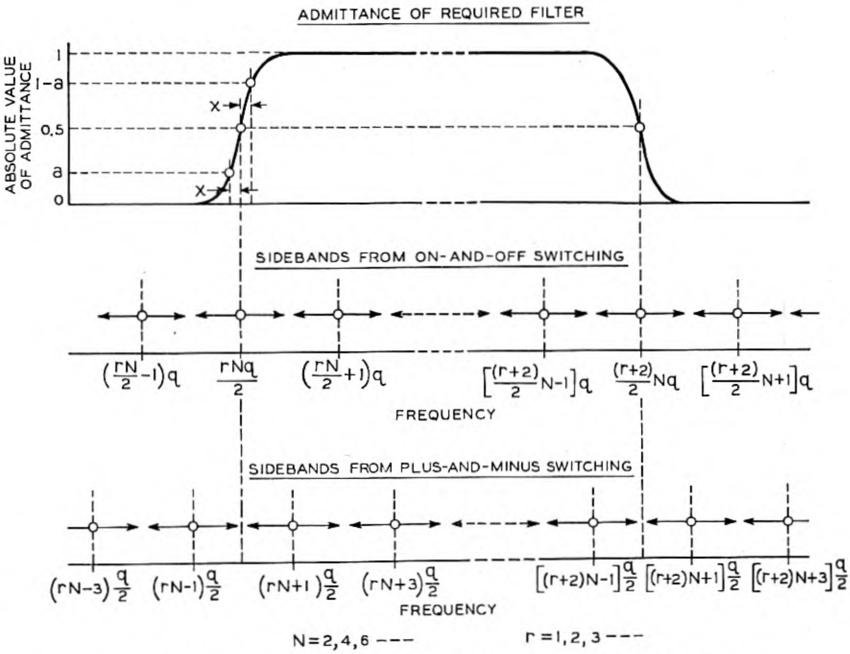

Fig. 3—Vestigial sideband transmission in time division multiplex systems

band transmitted in the case of on-and-off switching consists of the upper sideband on the harmonic $rNq/2$, the lower sideband on the harmonic $(r + 2)Nq/2$, and all intervening sidebands. In the case of plus-and-minus switching the upper and lower sidebands on frequencies $(rN + 1)q/2$ to $[(r + 2)N - 1]q/2$ inclusive are transmitted. Impairment of the nominal band by the filter is made up by transmitting the appropriate parts of sidebands outside the nominal range. It is easily verified that either of the systems depicted in Fig. 3 satisfies the required conditions for multiplex transmission without interchannel interference when the sidebands produced by a given signal have equal amplitudes over the range utilized.

The vestigial method is required only when strong sidebands very near the desired ones must be removed. Modifications of the time division process exist in which vestigial filters are unnecessary because very little energy appears at unwanted side frequencies. It is clear that if we regard the problem as one of producing certain sidebands on carriers of definite phases, we are not restricted to commutating devices only but may make use of general modulator technique. Further details concerning specific circuit arrangements are described in *U. S. Patent* 2,213,938, W. R. Bennett; and *U. S. Patent* 2,213,941, E. Peterson. As a general guide the following table of carrier phases for an N-channel system (N even) is furnished:

TABLE OF PHASE SHIFTS FOR N-CHANNEL SYSTEM
($N/2$ Carrier Frequencies Required)

Carrier Frequency		ν	$\nu + q$	$\nu + 2q$	$\nu + 3q$	$\cdots$
Phase Shift In Carrier	Channel 1	0	0	0	0	$\cdots$
	2	π/N	$3\pi/N$	$5\pi/N$	$7\pi/N$	$\cdots$
	3	$2\pi/N$	$6\pi/N$	$10\pi/N$	$14\pi/N$	$\cdots$
	4	$3\pi/N$	$9\pi/N$	$15\pi/N$	$21\pi/N$	$\cdots$

TRANSMISSION REQUIREMENTS

Practical success of a time division multiplex system requires the maintenance of a satisfactory ratio of wanted signal to crosstalk. In order to accomplish this, the transmission link must preserve the amplitude and phase relations of a group of sidebands. A physical picture of the relations involved may be obtained from Fig. 4, which is drawn for the particular case of a 5-channel system of the on-and-off switching type. For this example the theory previously developed shows that five sidebands of equal amplitude are sufficient, namely—the signal itself (which may be regarded as a sideband on a carrier of zero frequency), the upper and lower sidebands on the switching frequency and on the second harmonic of the switching frequency. If we take the phases of the switching fundamental and its second harmonic as applied to the first channel as a reference, the proper phases of fundamental and second harmonic respectively for the other four channels are given by the following table:

Channel Number	Fundamental Phase	Second Harmonic Phase
2	72°	144°
3	144°	288°
4	216°	432°
5	288°	576°

In Fig. 4, we have assumed a single-frequency signal component as input to the first channel. If the line has distortionless attenuation and phase characteristics, the five resulting side frequencies are received in the first channel as the five in-phase vectors of equal amplitude shown in (*a*). Reception in the second channel is shown by (*b*) in which the vector 1 represents the directly transmitted signal component (or sideband on *d-c*), 2 and 3

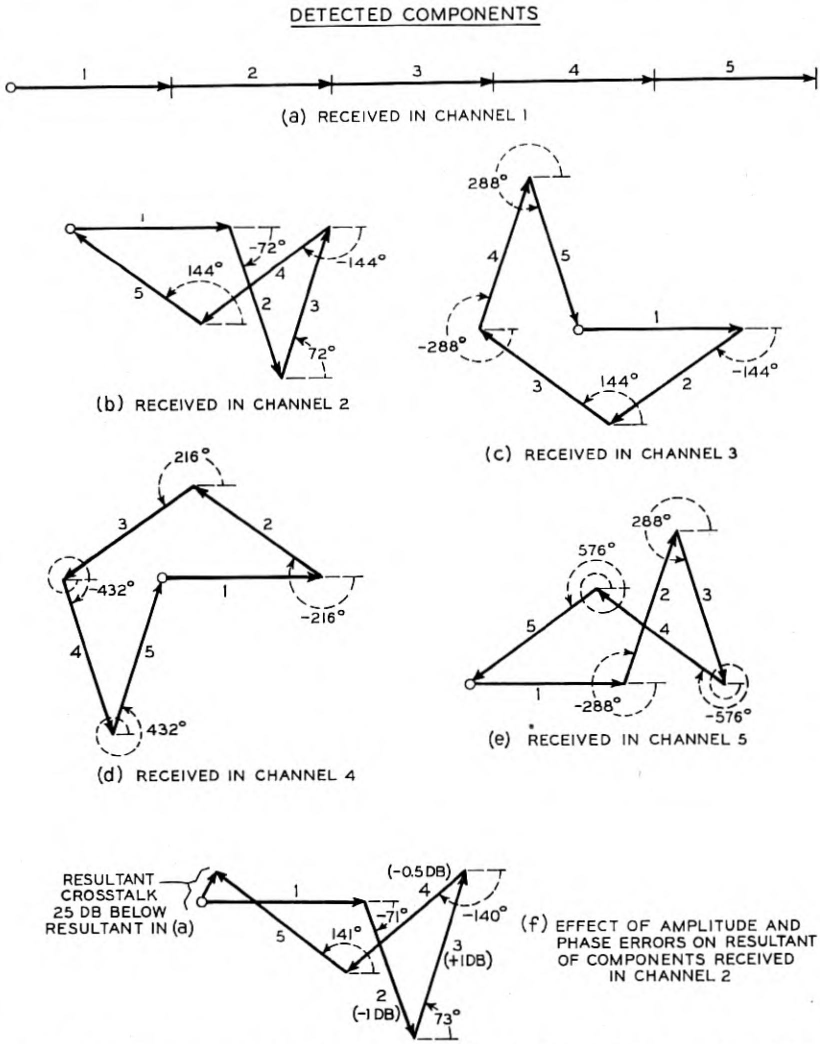

Fig. 4—Graphical representation of operation of time division multiplex system. Signal transmitted in channel 1 of 5-channel 5-sideband system

represent the detected components from the upper and lower side frequencies associated with the fundamental switching frequency, and 4 and 5 the components resulting from the upper and lower side frequencies of the second harmonic. Components 2 and 3 are shifted $-72°$ and $+72°$ respectively and 4 and 5 are shifted $-144°$ and $+144°$ in relation to the phase of component 1. As shown in (*b*), the five vectors combine in the form of a closed polygon giving a resultant of zero amplitude. Similar vector diagrams for reception in the third, fourth, and fifth channels are shown in (*c*), (*d*), and (*e*). The appropriate diagrams for transmission in channels 2, 3, 4, and 5 and receiving in any channel can be obtained from (*a*) – (*e*) by cyclic permutation of the channel numbers, i.e., transmission in 1 and reception in 2 corresponds to transmission in 2 and reception in 3, etc.

Production of crosstalk by phase and amplitude distortion in the transmission medium is illustrated by (*f*), Fig. (4), which shows the resultant component received in channel 2 when signal is transmitted in channel 1 and an imperfect line is used to connect the transmitting and receiving terminals. The vector 1 is taken as the reference amplitude and phase. The gain characteristic of the line is assumed to be one *db* low at the side frequency producing vector 2, one *db* too high for vector 3, 0.5 *db* low for vector 4, and with no error for vector 5. The phase curve is assumed to depart from a straight line by $-1°$, $-1°$, $-4°$, $+3°$ at the side frequencies from which components 2, 3, 4, 5 respectively are derived. The vector polygon fails to close and the resultant represents an unwanted signal received in channel 2 at a level 25 *db* below the wanted signal received in channel 1.

We may make an estimate of the accuracy of the equalization required in the general case by writing the transfer impedance $Z(i\omega)$ in the form:

$$Z(i\omega) = \rho(\omega)Z_0e^{-it_0\omega - i\beta(\omega)} \tag{39}$$

where $\beta(\omega)$ represents the departure of the phase shift from a straight line and the variation from flat gain is given by

$$g(\omega) = 20 \log_{10} \rho(\omega) \tag{40}$$

The expression (39) may be rewritten as:

$$Z(i\omega) = [1 + z(i\omega)]Z_0e^{-it_0\omega}, \tag{41}$$

where

$$z(i\omega) = \rho(\omega)e^{-i\beta(\omega)} - 1 \tag{42}$$

If we assume that the switching function is of the general form (34), we calculate from (14) the general relation:

$$Y_{jk} = K \left\{ \begin{array}{l} N + \sum_{m=n}^{n+\frac{N}{2}-1} (z[i(\nu + mq + \omega_j)] + z^*[i(\nu + mq - \omega_j)]), \; j = k \\ \sum_{m=n}^{n+\frac{N}{2}-1} (z[i(\nu + mq + \omega_j)]e^{-i(j-k)(2m-2n+1)\pi/N} \\ \qquad + z^*[i(\nu + mq - \omega_j)]e^{i(j-k)(2m-2n+1)\pi/N}), \; j \neq k \end{array} \right\} \tag{43}$$

where K is a complex constant of proportionality. The case of $j = k$ which gives transmission within the channel contains a variation with signal frequency caused by the summation of the departures from ideal transmission at the N sideband frequencies. This term presumably will be unimportant if the transmission characteristic is sufficiently good to meet crosstalk requirements; hence we may neglect the z and z^* terms in the case of $j = k$ and write the ratio of interference to desired signal as:

$$C_{jk} = \frac{Y_{jk}}{Y_{jj}} = \frac{1}{N} \sum_{m=n}^{n+\frac{N}{2}-1} (z[i(\nu + mq + \omega_j)]e^{-i(j-k)(2m-2n+1)\pi/N} + z^*[i(\nu + mq - \omega_j)]e^{i(j-k)(2m-2n+1)\pi/N}) \tag{44}$$

The crosstalk ratio will in general vary with the signal frequency. The requirement would logically be based on the total interference power weighted in accordance with the interfering effect at individual frequencies. Thus we might set

$$X_{jk} = S_j \int_{\omega_a}^{\omega_b} W_{jk}(\omega_j) \mid C_{jk} \mid^2 d\omega_j, \tag{45}$$

where X_{jk} represents the weighted interference power received in the k^{th} channel when a reference signal wave of mean power S_j is applied to the j^{th} channel. The limits of integration ω_a and ω_b are the lowest and highest signal frequencies used. The function $W_{jk}(\omega_j)$ represents the proper weighting with frequency of the interference and takes into account the distribution of the interfering signal and the relative importance of the different interfering frequencies.

Equation (45) is sufficient for computation of interchannel interference introduced by the line when the transmission characteristics of the line are known. A more valuable result, however, would be the expression of the required line characteristics in terms of the allowable interference. In general this would require some specification of the nature of the departures from the ideal characteristic. Except perhaps for systems with very few channels, it seems reasonable to assume that the departures are distributed fairly uniformly throughout the frequency range transmitted by the line, and hence that for purposes of estimating requirements we may replace $| C_{jk} |^2$ in (45) by its average value over the band. We may then write (45) in the form:

$$\overline{| C_{jk} |^2} = \frac{X_{jk}}{S_j \int_{\omega_a}^{\omega_b} W_{jk}(\omega_j)\, d\omega_j} \equiv U_{jk} \tag{46}$$

The value of the right-hand member, which we have designated by the symbol U_{jk}, is either known or can be determined for the particular type of signal. Hence our problem is reduced to finding the allowable departures in transmission which keep the mean square absolute value of C_{jk} from exceeding a prescribed maximum value.

We note that C_{jk} is the sum of N complex quantities, each of which is restricted to a range of values determined by the transfer impedance of the line in an individual band of frequencies. A convenient simplification may be made by regarding the N complex quantities as N independent chance variables. This is tantamount to assuming that departures in one band do not affect departures in any other band; the assumption is not strictly true, but should lead to no important error. We may then make use of the following theorem[8]: If

$$\zeta = b_1 z_1 + b_2 z_2 + \cdots + b_n z_n , \tag{47}$$

where $z_1, z_2, \cdots z_n$ are n independent complex chance variables and $b_1, b_2, \cdots b_n$ are complex constants,

$$\overline{| \zeta |^2} = \overline{| b_1 |^2}\, \overline{| z_1 |^2} + \cdots + \overline{| b_n |^2}\, \overline{| z_n |^2} \tag{48}$$

Application of this theorem to (44) gives

$$\begin{aligned} \overline{| C_{jk} |^2} &= \frac{1}{N^2} \sum_{m=n}^{n+\frac{N}{2}-1} \left(\overline{| z[i(\nu + mq + \omega_j)] |^2} + \overline{| z^*[i(\nu + mq - \omega_j)] |^2} \right) \\ &= \overline{| z |^2}/N, \end{aligned} \tag{49}$$

if the average square of the absolute value of the departure is the same in all bands and is equal to $\overline{| z |^2}$, which we shall define as the average squared absolute value of the departure for the entire line band used.

From (42) and (40),

$$\begin{aligned} | z(i\omega) |^2 &= 1 - 2\rho(\omega) \cos \theta(\omega) + \rho^2(\omega) \\ &= 1 - 2 \cdot 10^{g(\omega)/20} \cos \theta(\omega) + 10^{g(\omega)/10} \end{aligned} \tag{50}$$

[8] R. S. Hoyt, *B. S. T. J.*, Vol. XII, No. 1, Jan. 1933, p. 64.

Since it seems certain that g and θ must remain small to make the system operative, we investigate the nature of (50) when expanded in powers of g and θ. The leading terms are:

$$|z(i\omega)|^2 = \frac{(\log_e 10)^2}{400} g^2(\omega) + \theta^2(\omega) + \cdots \tag{51}$$

Hence for g and θ small, we have independent of any correlation which may exist between g and θ,[9]

$$\overline{|z(i\omega)|^2} = \left(\frac{\log_e 10}{20}\right)^2 \overline{g^2(\omega)} + \overline{\theta^2(\omega)} \tag{52}$$

Let

$$\sigma_1^2 = \overline{g^2(\omega)}, \qquad \sigma_2^2 = \overline{\theta^2(\omega)} \tag{53}$$

Then from (46), (49), (52),

$$U_{jk} = \left[\left(\frac{\log_e 10}{20}\right)^2 \sigma_1^2 + \sigma_2^2\right] \Big/ N \tag{54}$$

In (54) σ_1 is the r.m.s. departure of the gain in *db* from a constant and σ_2 is the r.m.s. departure of the phase shift in radians from a straight line. If σ_2 is expressed in degress instead of radians, (54) becomes

$$U_{jk} = 10^{-3}\,(13.25\,\sigma_1^2 + .3046\,\sigma_2^2)/N \tag{55}$$

The total interference received in any one channel is the sum of the individual contributions from the other $N - 1$ channels. The addition factor required to express the total in terms of the interference from one channel depends on the nature of the individual loads. Thus if the probability that any one channel is transmitting a signal wave is τ, the average total interference power received in one channel is

$$X = \tau(N - 1)X_{jk} = US_j \int_{\omega_a}^{\omega_b} W_{jk}(\omega_j)\, d\omega_j, \tag{56}$$

where

$$U = (N - 1)\tau U_{jk} = \frac{(N - 1)\tau}{N} 10^{-3}(13.25\sigma_1^2 + .3046\sigma_2^2) \tag{57}$$

For large values of N, the ratio $(N - 1)/N$ approaches unity, and the average interference becomes independent of the number of channels. The average interference may not be the most significant quantity, however. For example, if there is a considerable probability that all channels are

[9] This method of avoiding any assumption concerning correlation of attenuation and phase was suggested by Dr. T. C. Fry.

carrying energy simultaneously, as would be the case if the channels were subdivisions of a common signal band, the peak value of interference would probably be of more significance than the average value.

It is convenient to let

$$H = 10 \log_{10} \frac{S_j}{X} \tag{58}$$

$$F = -10 \log_{10} \int_{\omega_a}^{\omega_b} W_{jk}(\omega_j)\, d\omega_j \tag{59}$$

H is the ratio expressed in *db* of mean signal power in one channel to the total interference power received in one channel, and F is the weighting factor expressed in *db*. From (56),

$$U = 10^{-(H-F)/10} \tag{60}$$

Equation (57) may be written in the form,

$$\frac{\sigma_1^2}{a^2} + \frac{\sigma_2^2}{b^2} = 1, \tag{61}$$

where

$$\left.\begin{aligned} a &= 8.69 \sqrt{\frac{NU}{(N-1)\tau}}\ db \\ b &= 57.3 \sqrt{\frac{NU}{(N-1)\tau}}\ \text{degrees} \end{aligned}\right\} \tag{62}$$

Without the numerical factors, a and b are expressed in nepers and radians respectively.

If we regard σ_1 and σ_2 as variables, (61) determines a family of ellipses in which a and b are the semi-axes. By assigning values to N, τ, and $H - F$ we may thus represent the requirements on gain and phase variation by elliptical boundaries in the $\sigma_1\sigma_2$-plane. Figure 5 shows such a diagram constructed for a large number of channels each active one-fourth of the time and with flat weighting. In terms of the symbols above, we have set $N/(N-1)$ equal to unity, $\tau = 1/4$, and $F = 0$. Gain and phase variations included within a particular ellipse produce average interference power less than the amount designated on the boundary in terms of *db* down on mean power in one channel. The requirements imposed on both gain and phase variation are considerably more stringent than the corresponding requirements for carrier systems using frequency discrimination and employing comparable band widths.

Requirements on linear transmission of the line are, of course, not the only considerations involved in a comparison of time division multiplex

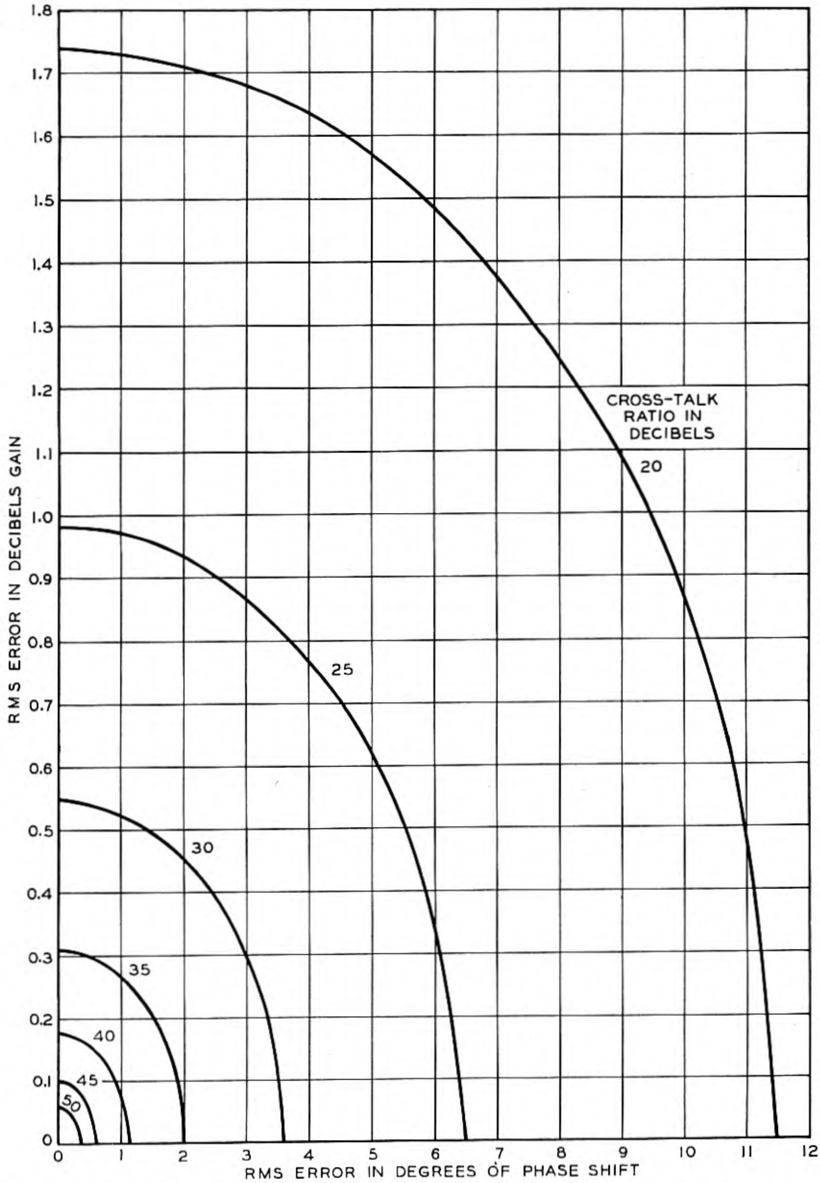

FIG. 5—Gain and phase requirements for transmission of time division multiplex signals. Each channel active 25% of time

methods with competing methods of superimposing channels. Other aspects to be considered are the synchronization of transmitting and receiving switching processes, the effects of non-linearities in the line, and the sensitivity of the system to external interference. It is thought, however, that the severe restrictions imposed on phase and attenuation characteristics when economy of band width is required form the weakest feature of the method and will in many cases provide the primary criterion for judging its availability in the solution of particular problems. Conversely, if the crosstalk requirements of the system are sufficiently mild to enable the transmission problem to be solved, the other problems also become relatively simplified.

Further References

A. M. Nicolson, *U. S. Patents*, 1,962,610; 2,021,743; 1,951,524; 1,956,397; 2,007,809.

Callahan, Mathes, and Kahn, "Time Division Multiplex in Radio-Telegraphic Practice," *Proc. I. R. E.*, 26 (1938), 1, pp. 55–74.

H. Raabe, "Untersuchungen an der wechselzeitigen Mehrfachübertragung (Multiplexübertragung)," *E. N. T.*, 16 (1939), 8, pp. 213–228.

Steady State Delay as Related to Aperiodic Signals

By R. V. L. HARTLEY

The concepts of phase and envelope delay, as applied to any linear system, rather than only to a medium, are discussed. Criteria are set up for the time of occurrence of that part of an aperiodic signal which corresponds to a small segment of the spectrum. The original spectrum of the signal gives the time of entry and this spectrum as modified by the phase characteristic of the system gives the time of exit.

If the amplitude is constant over the segment, it is shown that when the criterion is the time of maximum envelope of the disturbance, the aperiodic delay is identical with the envelope delay. When it is the time of maximum absolute value, the delay depends on the signal spectrum, the phase shift of the system, and the envelope delay, but not on the phase delay.

If the amplitude vaies rapidly with frequency, the component of an aperiodic disturbance which corresponds to a narrow segment of the spectrum persists so long that the resulting over-lapping of neighboring segments makes their interpretation difficult.

IN THE earlier applications of steady state theory to transmission problems the emphasis was placed on the variation of amplitude with frequency. The use of long loaded lines made it necessary to take account of phase distortion[1] as well. With the development of telephotography and television[2], the phase characteristic was found to provide a useful index for predicting the overlapping of adjacent picture elements. For these purposes it has been found convenient to express the phase characteristic in terms of phase or envelope delay. These may be called "steady state delays" since they are defined and measured in terms of sinusoidal disturbances of adjustable frequency. However, the signals for which they are intended to furnish an index are aperiodic in nature. It seemed worthwhile, therefore, to examine more closely the relations existing between "aperiodic delays," defined in terms of such signals, and steady state delays.

Let us first review the development of the concepts of steady state delay. Early in the study of the propagation of sinusoidal waves a distinction was made between phase and group velocity. If we fix on a particular distance of transmission the ratio of this distance to each of these two velocities may be interpreted as a delay associated with the transmission. In the

[1] For discussion and references see "Phase Distortion and Phase Distortion Correction," S. P. Mead, *B. S. T. J.*, Vol. VII, p. 195, 1928.
[2] Symposium on Television, *B. S. T. J.*, Vol. VI, p. 551.

communication art, these delays have been called phase and envelope delay, respectively. If the medium exhibits dispersion they vary with frequency. Let us fix our attention on the conditions throughout the medium at a particular instant during the transmission of a sinusoidal disturbance. We may determine the total change of phase in passing from the input to the output. This may be more than a single cycle. If now we divide this phase shift by the frequency, expressed in the same angular units, we get the time which will be required for the phase at the input to progress to the output, or the phase delay. Also it may readily be shown that the derivative of this phase shift with respect to frequency is equal to the envelope delay as defined above in terms of the group velocity. The simplest treatment of this is based on the consideration of two sinusoidal waves of equal amplitude and slightly different frequencies.

While these delays can be easily interpreted for most media, difficulties arise in the case of those substances which exhibit anomalous dispersion. Here, in the neighborhood of certain frequencies, the phase shift varies rapidly with frequency, and often appears to be discontinuous. Actually the apparent discontinuity is a region of very rapid decrease of phase with frequency, which leads to a negative value of envelope delay. In the same region the transmission varies rapidly with frequency, and selective reflection occurs at the entering boundary. This effect can be explained in terms of resonance in the elements which make up the fine structure of the medium.

The next step was to dissociate the idea of delay from that of velocity in a medium, and associate it with a steady state transfer characteristic between any two points of a linear system. This would permit its application to all sorts of complicated networks in which uniform propagation cannot be readily visualized. Here two types of characteristic are to be distinguished. One, which is associated with what might be called "damped" systems, exhibits a reasonably gradual variation of both phase shift and attenuation with frequency. This is the analog of a medium having normal dispersion. The other, which is associated with "resonant" systems, exhibits the phenomena associated with anomalous dispersion. In the case of filters and hollow wave guides these resonances give rise to regions of high attenuation and reactive impedance, which are the analogs of the regions of high absorption and selective reflection at the boundary of a medium. In applying the idea of delay to networks then, we can expect the results to agree with our intuitions only so long as we keep away from the critical frequencies of resonant systems.

In computing or measuring the phase shift of a system, at a single frequency, the result is indeterminate so far as the addition of multiples of 2π is concerned. This does not affect the envelope delay, which depends only on the derivative, and so this type of delay can be generalized directly to include the transfer characteristics of arbitrary networks. To give an exact meaning to phase delay some convention would have to be adopted for determining what, if any, multiple of 2π is to be added to the computed phase for the frequency in question. Apparently no such convention has been agreed upon which is of general application. For damped networks which transmit frequencies down to zero, it is customary to assume the phase shift to be zero at zero frequency, and, for higher frequencies, to add multiples of 2π so that the phase shift varies continuously with frequency. If, then, B is the computed phase shift, between $-\pi$ and π, we may represent the continuously varying phase shift by $B + 2m\pi$, where m is the number of discontinuities in B which have been eliminated in passing from zero to the frequency in question. The phase delay may then be defined as

$$D_p = \frac{B + 2m\pi}{\omega}. \tag{1}$$

Any similar convention for resonant systems would be less simple, and since, as will appear below, phase delay has little bearing on aperiodic signals, it seems unwise to attempt to formulate such a convention here.

In contrast with steady state delay, let us now examine the delay of an aperiodic signal. If the signal is transmitted without distortion the concept of delay of the signal as a whole is simple. If, because of distortion, the sent and received signals are different we may still agree upon some recognizable feature of each as determining its time of occurrence. If the distortion is considerable the delay may vary greatly with the distinguishing characteristic chosen. For example, if it depends on the behavior of components of high frequency the delay may be quite different from what it is if it depends on those of low. In the first case the result would be little affected if, before transmission, the signal were sent through a high-pass filter and, in the second, if it went through a low-pass filter. In each case we measure a delay associated with a disturbance which comprises only those Fourier components of the signal which occupy a particular limited range of frequency. We may carry this idea farther and make use of a very narrow band-pass filter. By varying the mid-frequency of this band we obtain a delay which is a function of frequency. Its value, at any frequency, is the delay, as defined by our convention, of a disturbance which corresponds to that part of the spectrum of the signal which is in the immediate neighborhood of the frequency in question. Our problem then is to find recognizable features of a disturbance of this kind such that, when they are used as criteria of delay, the result can be related directly to the phase or envelope delay as defined in terms of periodic disturbances.

Compared with the pair of equal sinusoids used in the derivation of envelope delay, this disturbance differs in that, in any finite range of frequency, there are an infinity of sinusoids, the amplitudes of which need not all be the same. For simplicity, we assume the actual filter to be replaced by an idealized one in which there is no distortion within the band and no transmission outside it. If the signal as a whole be represented by a Fourier integral, we may obtain the desired disturbance, for an angular frequency, ω_1, by integrating from $\omega_1 - \delta$ to $\omega_1 + \delta$. The disturbance may be represented by

$$f(t) = \text{real part of } M \int_{\omega_1-\delta}^{\omega_1+\delta} \exp\left[-\alpha + i(\omega t - \theta)\right] d\omega, \tag{2}$$

where M is a constant dependent on the magnitude of the signal and α and θ are functions of frequency and position which describe the spectrum of the signal at various points in the system.

The first step is to perform the indicated integration and express the resulting function of time in a convenient form. For this we let

$$\epsilon = \omega - \omega_1 .$$

Since we are interested only in small values of ϵ we may replace α by

$$\alpha = \alpha_1 + \alpha_1'\epsilon,$$

where α_1 and α_1' are the values of α and $\dfrac{\partial\alpha}{\partial\omega}$ at ω_1. Similarly,

$$\theta = \theta_1 + \theta_1'\epsilon.$$

We define an instant, T_e, by

$$T_e = \theta_1', \tag{3}$$

and a time, τ, by

$$\tau = t - T_e . \tag{4}$$

Substituting these in (2) and performing the integration, we get

$$f(t) = \text{real part of } 2M \exp\left[-\alpha_1 + i(\omega_1\tau - (\theta_1 - \omega_1\theta_1')\right] \frac{\sinh\,(-\alpha_1' + i\tau)\delta}{(-\alpha_1' + i\tau)} .$$

If we introduce the angles,

$$\beta = \text{arc tan}\, \frac{-\alpha_1'}{\tau},$$

and

$$\gamma = \text{arc tan}\, \frac{\tanh\,(-\delta\alpha_1')}{\tan\,\delta\tau},$$

and take the real part, we get

$$f(t) = 2M \exp(-\alpha_1) \frac{[(\cosh \delta\alpha_1' \sin \delta\tau)^2 + (\sinh \delta\alpha_1' \cos \delta\tau)^2]^{\frac{1}{2}}}{(\alpha_1'^2 + \tau^2)^{\frac{1}{2}}}$$

$$\cos(\omega_1\tau - (\theta_1 - \omega_1\theta_1') + \beta - \gamma). \qquad (5)$$

Let us consider first the extreme case where the spectrum of the signal is

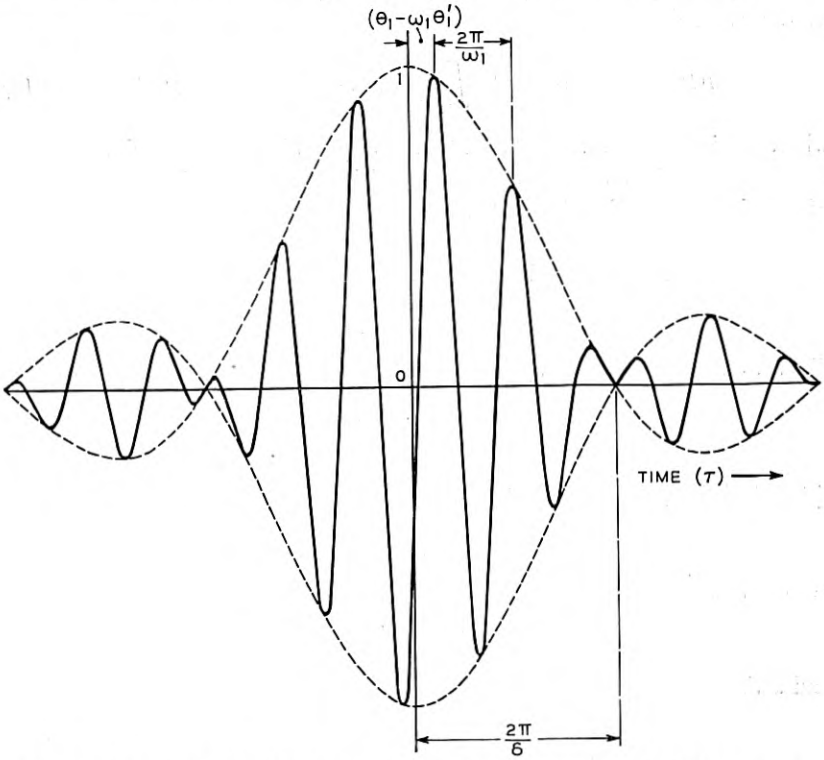

Fig. 1—Elementary disturbance corresponding to a narrow segment of the spectrum

uniform in amplitude in the neighborhood of ω_1, so that α_1' is zero. Then

$$f(t) = 2\delta M \exp(-\alpha_1) \frac{\sin \delta\tau}{\delta\tau} \cos(\omega_1\tau - (\theta_1 - \omega_1\theta_1')). \qquad (6)$$

Here the amplitude includes a constant factor which is proportional to the bandwidth, 2δ, and to the magnitude, $M \exp(-\alpha_1)$, at the frequency, ω_1, and a function of time, a plot of which is shown in Fig. 1. This function consists of a sinusoidal wave of frequency, ω_1, the amplitude of which varies with time, the envelope being symmetrical about the instant, $T_e = \theta_1'$, at which it is a maximum. T_e, the time of maximum envelope, is then a unique instant which is suitable for defining the time at which the disturbance occurs. It is determined solely by the slope of the phase frequency curve for the spectrum.

The instant, T_e, may be interpreted, in accordance with the principle of stationary phase, as the one at which the sinusoidal components of (2) are most nearly in the same phase, and so have the least destructive inter-

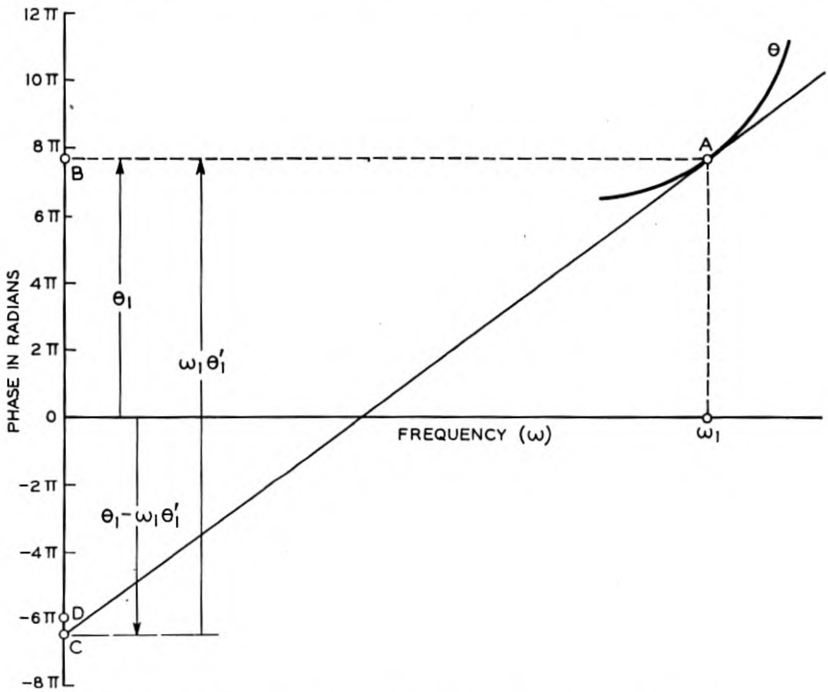

Fig. 2—Graphical representation of the phase of an elementary disturbance

ference. This condition will hold when the instantaneous phase angle is changing least rapidly with frequency, that is, when

$$\frac{\partial}{\partial \omega}(\omega t - \theta) = 0,$$

from which

$$t = \theta_1'.$$

The angle, $\theta_1 - \omega_1\theta_1'$, in (6), gives the phase of the wave at the instant, T_e, when its envelope is a maximum. The interpretation of this angle will be aided by the geometrical construction of Fig. 2 which is similar to that employed for phase and group velocity[3]. The abscissae are values of ω and the ordinates are values of phase in radians. A portion of the function, θ, in the neighborhood of ω_1 is shown. The distance, OB, is θ_1. The slope of the tangent, CA, to the curve at A is θ_1'. The distance, CB, is $\omega_1\theta_1'$. Consequently, OC, or the intercept of this tangent on the phase axis, is $\theta_1 - \omega_1\theta_1'$. If, as shown in the figure, the absolute value of this intercept is greater than π, we may transform (6) to a form in which the angle is less than π, by the substitution

$$\varphi = \theta_1 - \omega_1\theta_1' + 2n\pi, \tag{7}$$

where n is an integer and

$$|\varphi| < \pi.$$

In Fig. 2, n is 3, and φ is the distance DC. (6) then becomes

$$f(t) = 2\delta M \exp(-\alpha_1) \frac{\sin \delta\tau}{\delta\tau} \cos(\omega_1\tau - \varphi),$$

and φ is the ordinary phase lag of the sinusoid, relative to an origin of time given by the instant of maximum envelope.

We may choose as the instant at which the disturbance occurs, not T_e, at which the envelope is a maximum, but T_a, at which the instantaneous value of the function has its maximum absolute value. Since δ is small compared with ω_1, this will occur very nearly at the smallest absolute value of τ for which $\cos(\omega_1\tau - \varphi)$ is ± 1. This will occur for

$$\tau = \frac{\varphi}{\omega_1}, \quad \text{when} \quad -\frac{\pi}{2} < \varphi < \frac{\pi}{2},$$

and for

$$\tau = \frac{\varphi \pm \pi}{\omega_1} \quad \text{when} \quad -\pi < \varphi < -\frac{\pi}{2} \quad \text{or} \quad \frac{\pi}{2} < \varphi < \pi.$$

From (4), (3) and (7),

$$T_a = \frac{\theta_1 + k\pi}{\omega_1},$$

where k is an integer such that

$$-\frac{\pi}{2} < \Psi = \theta_1 - \omega_1\theta_1' + k\pi < \frac{\pi}{2}.$$

The significance of this can be seen from Fig. 3. Here, in addition to the θ curve of Fig. 2, there are plotted a series of curves whose ordinates differ

[3] Lamb, "Hydrodynamics," Cambridge U. Press 1916, p. 371.

from it by multiples of π. In so far as any one purely sinusoidal component of the disturbance is concerned, values of phase determined by those curves which differ by an even multiple of π would be indistinguishable. Those differing by an odd multiple would represent a reversal of sign. Let us

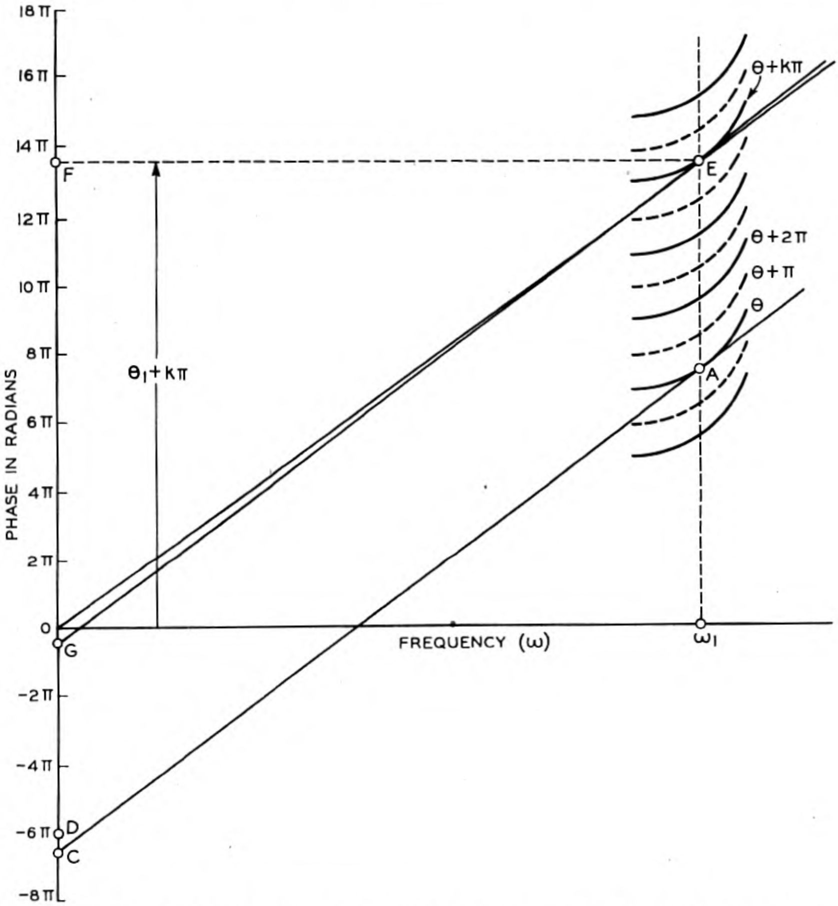

Fig. 3—Graphical representation of the time of maximum absolute value

now select that curve for which the tangent at ω_1 intersects the phase axis nearest the origin, and call it $\theta + k\pi$. Since, for the case drawn,

$$|DC| < \frac{\pi}{2},$$

$$k = 2n.$$

If it were greater we should have

$$k = 2n \pm 1.$$

It is then obvious that the time of maximum absolute value, T_a, is given by the slope of the line OE. It differs from T_e by the difference in slope of the lines OE and GE.

We have then deduced from the spectrum of the disturbance its time of occurrence in terms of two definitions of the latter. The next step is to compare these times for the input and output and determine the corresponding delays. Let us consider first the case where the attenuation is independent of frequency, so that α_1' is zero in the output signal also. We may then confine our attention to the phase, θ. Let us represent its value at the input by b, and the phase shift of the system by B. Then at the output θ will be equal to $b + B$. If we take the time of occurrence as determined by the maximum envelope, these times at the input and output are

$$T_{e_0} = b_1',$$
$$T_{e_1} = b_1' + B_1'.$$

The delay is then

$$D_e = T_{e_1} - T_{e_0} = B_1',$$

which is by definition the envelope delay of the system.

If we take the time of occurrence based on the maximum absolute value, we have, at the input,

$$T_{a_0} = \frac{b_1 + k_0\pi}{\omega_1},$$

where

$$-\frac{\pi}{2} < \Psi_0 = b_1 - \omega_1 b_1' + k_0\pi < \frac{\pi}{2}.$$

At the output,

$$T_{a_1} = \frac{b_1 + B_1 + (k_0 + k_3)\pi}{\omega_1},$$

where

$$-\frac{\pi}{2} < \Psi_3 = b_1 + B_1 - \omega_1(b_1' + B_1') + (k_0 + k_3)\pi < \frac{\pi}{2}.$$

The delay,

$$D_a = T_{a_1} - T_{a_0} = \frac{B_1 + k_3\pi}{\omega_1}.$$

While there is a superficial similarity between this and the phase delay (1), it is of little real significance; m, in (1), is determined by the aggregate increase in phase shift with frequency, while k, is determined mainly by the rate of increase at ω_1. An example of a situation in which the two delays are very different, is furnished by a wave guide when the frequency only just exceeds the cutoff. The phase delay is then almost zero while the rate of change of phase shift with frequency is very large.

Thus the delay based on maximum absolute value depends on both the envelope delay and the phase shift of the system, but not on the phase delay. There remains to examine this dependence in more detail. The value of k_3 depends on the spectrum of the signal as well as the characteristic of the system. It is of interest to see if it can be replaced by a quantity derived from the system characteristic alone. The most obvious thing to try is a delay which is derived from the phase shift of the system in the same way that the time of absolute maximum is derived from that of the signal spectrum. This would be

$$D_S = \frac{B_1 + k_2\pi}{\omega_1},$$

where

$$-\frac{\pi}{2} < \Psi_2 = B_1 - \omega_1 B_1' + k_2\pi < \frac{\pi}{2}.$$

The difference between this and the aperiodic delay based on absolute value is

$$D_S - D_A = \frac{\pi}{\omega_1}(k_2 - k_3),$$

$$= \frac{1}{\omega_1}(\Psi_0 + \Psi_2 - \Psi_3).$$

Since $k_2 - k_3$ is either zero or an integer and $|\Psi_3|$ is less than $\frac{\pi}{2}$, if

$$-\frac{\pi}{2} < \Psi_0 + \Psi_2 < \frac{\pi}{2},$$

$$D_S - D_A = 0.$$

If

$$-\pi < \Psi_0 + \Psi_2 < -\frac{\pi}{2},$$

$$D_S - D_A = -\frac{\pi}{\omega_1}.$$

If

$$\frac{\pi}{2} < \Psi_0 + \Psi_2 < \pi,$$

$$D_S - D_A = \frac{\pi}{\omega_1}.$$

Thus the delay as derived from the system characteristic alone may be identical with the aperiodic delay based on maximum absolute value or it may differ from it by $\pm\frac{\pi}{\omega_1}$, that is by half a period. Which condition holds depends on the interrelation of the phase functions which characterize the signal spectrum at the input and the transmission of the system, and not on either of these functions alone.

If the attenuation is not uniform, α_1' cannot be neglected and the expression for the output signal becomes more complicated. Both the amplitude and phase in (5) vary with time in a manner which depends on the value chosen for δ. The expression becomes fairly simple, however, for the case where α_1' is very large, as in anomalous dispersion and in highly resonant systems. Then, even when δ is small, we may assume that

$$\cosh(\delta\alpha_1') = \exp(\pm\,\delta\alpha_1'),$$
$$\sinh(\delta\alpha_1') = \pm\exp(\pm\,\delta\alpha_1'),$$

according as $\alpha_1' \gtrless 0$.

The amplitude factor in (5) then becomes

$$\frac{M\exp(-\alpha_1 \pm \delta\alpha_1')}{(\alpha_1'^2 + \tau^2)^{\frac{1}{2}}}.$$

Here the exponent is equal to the value of α at that edge of the segment of the spectrum where the amplitude is greatest. The amplitude is symmetrical about $\tau = 0$, that is, about $t = \theta_1'$, at which point it has its maximum value. Hence the instant of maximum envelope is still given by the slope of the phase, frequency curve, as when α_1' is small. However, the maximum is now extremely flat and its sharpness no longer depends directly on δ. Over the range of values of τ for which $\tau^2 << \alpha_1'^2$, the amplitude is sensibly constant. When $\tau = \pm\alpha_1'$, it is reduced to $\frac{1}{\sqrt{2}}$ times its maximum. For $\tau^2 >> \alpha_1'^2$, it varies inversely as $|\tau|$.

To investigate the oscillating factor of (5) we note that now

$$\gamma = \pm\delta\tau \pm \frac{\pi}{2},$$

where the sign of $\delta\tau$ depends on that of α_1' and that of $\frac{\pi}{2}$ does not. The oscillating factor then is

$$\cos [(\omega_1 \mp \delta)\tau - (\theta_1 - \omega_1\theta_1') - \eta],$$

where

$$\eta = \arctan \frac{\alpha_1'}{\tau} \pm \frac{\pi}{2}. \tag{8}$$

The frequency, $(\omega_1 \mp \delta)$, is that of the edge of the segment of the spectrum where the amplitude is relatively very large. The phase differs from that for small values of α_1' by a quantity η which is an ambiguous function of the time τ. This ambiguity may be removed if we assume that the phase varies continuously and that, for very small values of τ, the amplitude has the same sign as the spectrum component corresponding to an infinitesimal value of δ. As τ increases through zero, $\arctan \frac{\alpha_1'}{\tau}$ changes discontinuously from $\mp \frac{\pi}{2}$ to $\pm\frac{\pi}{2}$ according as $\alpha_1' \lessgtr 0$. To avoid a similar discontinuity, in η we say that the sign of $\frac{\pi}{2}$ in (8) is to be taken opposite for positive and negative values of τ. If we make it $\pm$ for $\tau < 0$, and $\mp$ for $\tau > 0$, according as $\alpha_1' \gtrless 0$, then η is zero in the neighborhood of $\tau = 0$. Since the amplitude factor is always positive, this corresponds to a spectral component of positive amplitude. If we make the sign of $\frac{\pi}{2}$ $\mp$ for $\tau < 0$, and $\pm$ for $\tau > 0$, η becomes $\pm\ \pi$, which is the equivalent of a negative amplitude. Hence a knowledge of the spectral component of frequency ω_1 enables us to determine the sign in (8). For large values of (τ), η reduces to $\pm\frac{\pi}{2}$.

Here we have assumed the amplitude of the input signal to be independent of frequency. If this is not the case the same conditions hold at the input as have just been discussed for the output of a resonant system.

The main conclusion to be drawn from the foregoing is that when the amplitude is changing rapidly with frequency, the component of an aperiodic disturbance which corresponds to a narrow segment of the spectrum persists for a considerable period so that there is much overlapping of the contributions of neighboring segments. It is therefore difficult to deduce the nature of the disturbance at any particular time from any narrow region of its spectrum. For the same reason it is difficult to associate the delay experienced by an aperiodic signal with the steady state characteristic of a network when the attenuation of the latter is changing rapidly with frequency.

The net result of our study then is that steady state phase delay has no direct relation to the particular types of delay of an aperiodic signal which we have chosen to investigate. When the amplitude does not change rapidly with frequency, envelope delay is identical with the delay produced in the maximum value of the envelope of a disturbance corresponding to that part of the signal spectrum which is in the immediate neighborhood of the frequency in question. The envelope delay, together with the phase shift, determines the delay in the maximum absolute value of this disturbance, subject to an uncertainty of half a period. This uncertainty depends on the particular combination of signal spectrum and system characteristic. When the amplitude does change rapidly with frequency, the envelope delay still gives the delay in the maximum value of the envelope. However, this maximum is so flat that the interpretation of the results is very difficult.

Engineering Requirements for Program Transmission Circuits*

By F. A. COWAN, R. G. McCURDY and I. E. LATTIMER

Present-day program networks are reviewed from the standpoints of engineering, design, and operation as developed to meet the needs of the broadcasters. The factors requiring consideration in the further development of program networks in anticipation of future needs are also discussed. The presentation of the paper is supplemented by a demonstration of the quality obtainable by transmission over various types of telephone facilities.

Introduction

THE growth of radio broadcasting to the magnitude of a major national industry within the last twenty years has been accompanied by the development of a nation-wide system of wire-line networks interconnecting hundreds of broadcasting stations. Papers have been presented before this Institute from time to time[1,2,3] describing the types of plant used for these networks and discussing important features of their design and operation. With these twenty years of experience as a background, it should now be of interest to review how the various requirements of broadcasting have influenced the development of the networks and to consider some of the factors which have determined the point to which transmission and operating features have so far been carried.

Simply stated, broadcasting is a means by which sounds originated at one place are reproduced simultaneously to large numbers of listeners distributed over wide areas. The simplest possible radio broadcasting system would consist of a microphone, a radio broadcast transmitter and some radio receiving sets. Such a system could serve only the listeners within the comparatively limited service area of the transmitter. To serve the whole nation many transmitters must be established about the country. Furthermore the most desirable sources of program are not usually in the neighborhood of the transmitter to which a particular listener can tune, since talent tends to be concentrated in certain parts of the country, and special events of interest may occur anywhere. To give a true country-wide service so that every listener can hear the programs he enjoys wherever they may

* Presented at A.I.E.E. Winter Convention, Philadelphia, Pa., January 27–31, 1941. Published in *Electrical Engineering*, Transactions section, April 1941.

[1] For all numbered references, see list at end of paper.

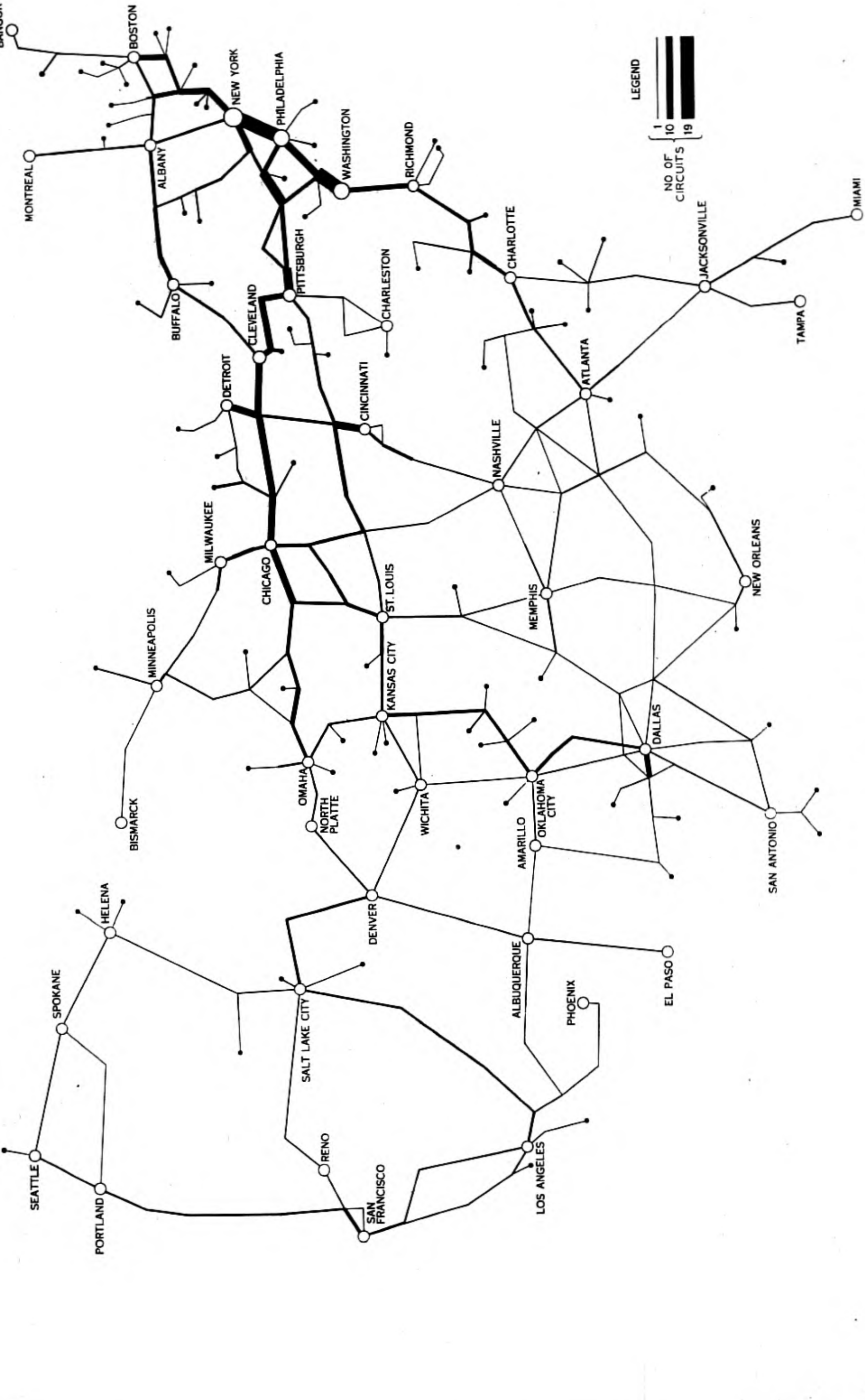

Fig. 1—Major network circuits in the United States

originate, a supplementary transmission system must be provided, interconnecting the many studios and broadcasting stations. The wire networks that perform this function comprise the subject matter of this paper.

The present extent of the wire-line facilities which are associated with the major portion of these networks is indicated in Fig. 1. The width of the lines on this chart has been made proportional to the numbers of circuits in the various sections. The total length of these circuits is in excess of 110,000 miles, and it is not unusual for a program originating at some point on a network to traverse more than 7,000 miles of circuit before being broadcast by the most remote station.

The requirements which the program networks must meet are in the final analysis determined largely by the needs of the broadcasters. The objective of a program network service is to meet these needs in as complete and prompt a manner as possible consistent with reasonable cost. With this objective in mind, it is necessary in planning the plant to consider not only the day-to-day needs, but the possible future needs as well. The importance of this may be appreciated when it is considered that plant provided today for program transmission service will need to be adaptable to the service requirements ten or twenty years hence. As a result of such planning, cables and equipment installed five, ten, and fifteen years ago meet present-day requirements, and, with some rearrangements, will take care of those likely to develop tomorrow.

The detailed planning of program transmission circuits requires consideration of:

1. The numbers of circuits likely to be required, section by section, over each route;
2. The provisions for reliability, flexibility, operation, and supervision essential to a high-grade network service;
3. The transmission requirements, or electrical characteristics, necessary to achieve a natural reproduction of the program.

These three general classes of requirements will be considered in order.

Number of Circuits Required

The circuits which have been established on a full-time basis for continuing use form the backbone of the program networks. Even for these circuits, however, permanence is relative since frequent extensions and rearrangements are made to meet changing requirements of the broadcasters. Aside from these fulltime circuits there are intermittent requirements occasioned by special events and other short-period needs of the broadcasters, some of which involve networks almost as extensive as the full-time networks. In addition reliability of service requires provision for rerouting the networks in the event of trouble. Figure 2 shows the year-by-year growth in the operated mileage of program circuits for the period 1926 to 1940. Of the more than 110,000 miles of circuits shown for 1940, about 45,000 miles have been provided for the short-period services and as stand-by facilities for protection. In addition to these, there are still other circuits, normally assigned to other services, which are arranged to be readily adaptable to program service to supplement the reserve facilities maintained on a full-time basis.

The time interval necessarily accompanying any extensive construction project makes it necessary to engineer plant considerably in advance of actual service requirements to meet, not only the expected growth, but also the changes in network routing. Figure 3 shows for two typical sections

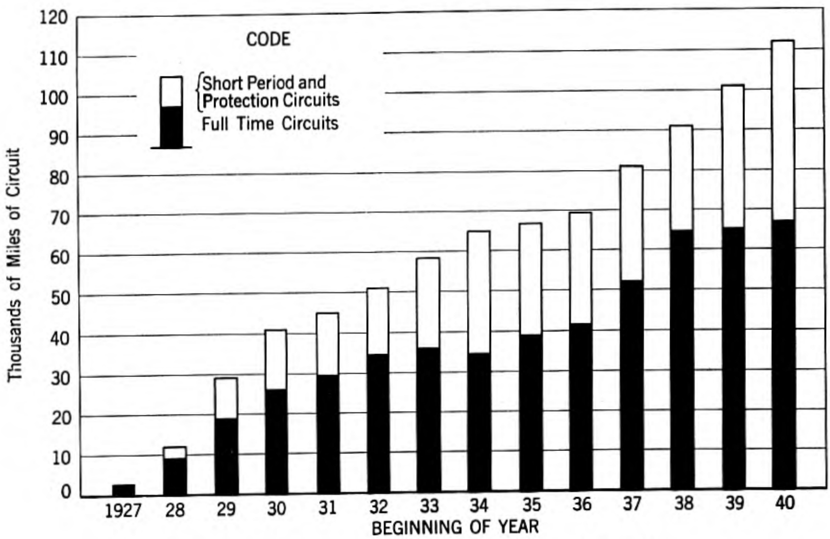

Fig. 2—Growth in mileage of major network circuits

along major routes the variations in requirements for full-time network circuits resulting from growth and rearrangements required by the broadcasters. While, in planning to meet these rapid variations in circuit requirements, advantage can be taken of some latitude which exists in the choice of routes for occasional services and protection facilities, the task of balancing the provision of circuits against requirements is an entertaining and at times difficult one for the circuit engineer.

Operating Requirements

Considering for a moment the variety of programs originating at many different points that can be heard on any home radio set in the course of an evening without once changing the tuning, it will be apparent that minute-to-minute rearrangements of an established interconnecting network must be possible. For example, studios have to be changed from receiving to originating, sections of the network have to be made to transmit first in

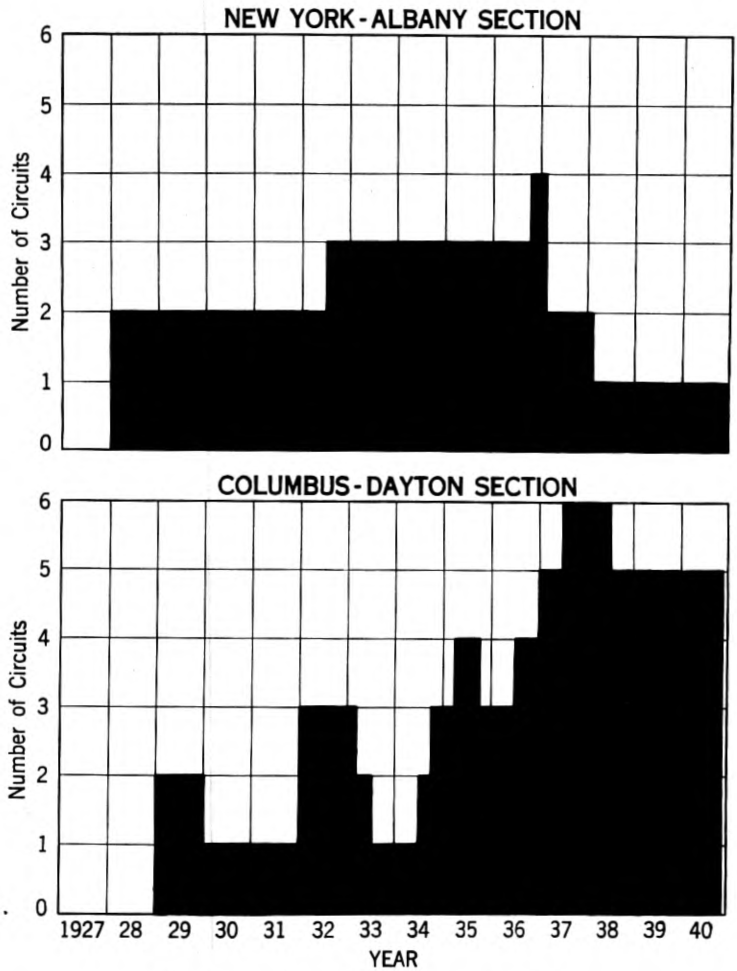

Fig. 3—Variations in full-time program circuits

one direction and then in the other, and branches have to be connected and disconnected. These changes in the network have to be made in the few seconds elapsing between the close of one program and the start of the next on the receipt of selected cue words or sounds. Even during the course of a single program, switches or reversals may have to be made to change the originating point temporarily. To provide for these rapid changes, special operation and special switching and reversing equipment are required at many points along the network. Much of this equipment is under remote control from selected points.

The greater portion of the switching of program circuits is done at about 25 points throughout the country on the major networks. On the average more than 25,000 switching operations per month are performed at these 25 points. During the busy hours of any typical evening there may be something over 500 men on duty at all of the offices about the networks.

At points where switching requirements are simple, the switching equipment consists merely of a few keys. At the larger points where the switching requirements are complex, the switching equipment consists of elaborate relay and control arrangements. These are so designed that it is possible to set up in advance the circuit combinations required for the ensuing program period without disturbing the programs in progress. The actual switching operation takes place at the instant the monitoring attendants signal the the receipt of the last of selected cues, and not before then. This type of arrangement affords a maximum of protection against error, as it is possible to check the presetting for the next switch or make a last minute change if necessary any time before the switch has been made.

Figure 4 shows a picture of such a switching arrangement in use at Omaha, Nebraska for one broadcasting company. At this point 13 circuits used in various trunk and branch sections of two networks are connected to the switching equipment. These are grouped in various combinations to take care of as many as five simultaneous programs. A maximum of five cues might, therefore, be involved in a switch at this point.

The operation and maintenance of the networks are carried out by a special organization under centralized authority and trained in the application of uniform methods and procedures found by experience to be productive of best results. Transmission is monitored continuously at strategic points about the networks. In order to facilitate the activities of this group many thousands of miles of intercommunicating telephone and telegraph circuits are provided full time for their use.

A picture of a monitoring position in the program transmission office at Washington, D. C. is shown in Fig. 5. It will be noted that the monitoring attendant is using an individual headset. This is of a special high fidelity type and is used to avoid the confusion that would result from attempting to monitor a number of different programs simultaneously with loud-speakers. Loud-speakers are available, however, for supplementary checks of quality whenever required.

Accurate transmission measuring equipment is necessary at the various operating points about the networks to insure satisfactory transmission maintenance results.

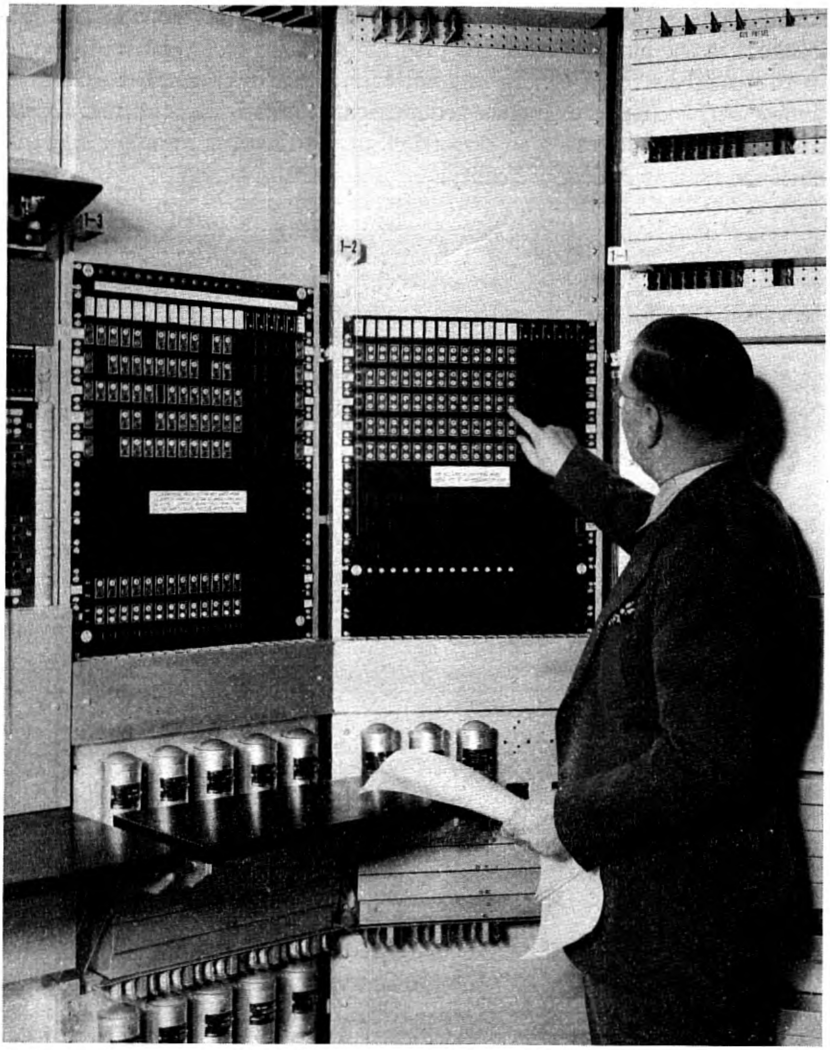

Fig. 4—Switching panel at Omaha, Nebraska

TRANSMISSION REQUIREMENTS

The general transmission requirement for a broadcasting system is that the program material be transmitted with a high degree of naturalness. Although the exact determination of the transmission characteristics which would accomplish this involves many considerations it will be assumed in this discussion that an ideal transmittion system is one in which the sound waves impressed on the listeners' ears in the home are an exact replica of the sound waves striking the microphone in the distant studio. Limitations inherent in the human ear, in the program material to be transmitted, and in the usual listening conditions, however, make such ideal transmission unnecessary. In expressing the requirements for satisfactory transmission, frequency range, attenuation distortion, delay distortion, nonlinearity and noise are used as indices of quality.

Fig. 5—Monitoring position at Washington, D. C.

Before taking up the transmission requirements of a program circuit, it is important to consider further the fundamental factors that are involved in fixing the characteristics considered desirable for the entire system. According to Harvey Fletcher,[4] the zone of audibility of the average normal human ear for pure or single frequency sounds is the area within the curve of Fig. 6. The abscissas represent frequency and the ordinates show the range of intensity recognizable as sound, between the lower limit or threshold of audibility and the upper limit where the sensation of pain is felt. It is seen that the extreme frequency range shown on the chart is from about 20 to 20,000 cycles per second. This range is for young people. It is considerably less for middle-aged and elderly people, and varies with individuals.

In addition to the limitation of the ear there is the fact that there is little energy present in most program material in the extremes of this range, particularly in the upper frequencies. The energy versus frequency spectra of music and other forms of program have been published elsewhere.[5] Figure 7 shows the frequency range which must be transmitted for a number of instruments, speech, and certain noises, so that competent observers cannot detect any impairment.[6] For whole orchestras, experiment has

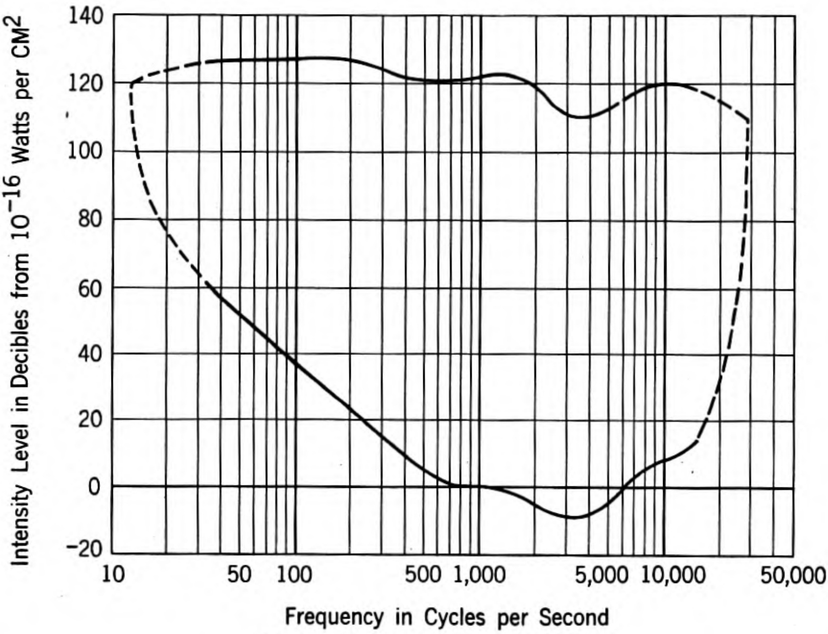

Fig. 6—Limits of audible sound

shown that the elimination of frequencies below 40 and above 15,000 cycles per second is undetectable.[4] It the upper limit of the transmitted frequencies is lowered from 15,000 cycles, the impairment is at first barely detectable but increases at an accelerating rate. When the limit is materially lower than 8,000 cycles, the loss is readily apparent to many people.

Another important consideration is volume range—that is the difference between the maximum and minimum levels of the program. The ordinates of Fig. 6 show that for part of the frequency range, the ear can respond to a range of intensities of more than 120 decibels, with perhaps 100 decibels as a mean. However, the following considerations show that the volume range which the transmission system needs to accommodate is considerably narrower than the intensity range to which the ear can respond.

In the first place, the range of program volumes to which the ear can respond is much less than the range of single-frequency intensities shown by the curve. Program waves are in general very irregular in shape, and even at constant volume contain large and small peaks differing in amplitude by many decibels. The range between the volume at which the highst peaks reach the maximum instantaneous intensity which the ear can tolerate and

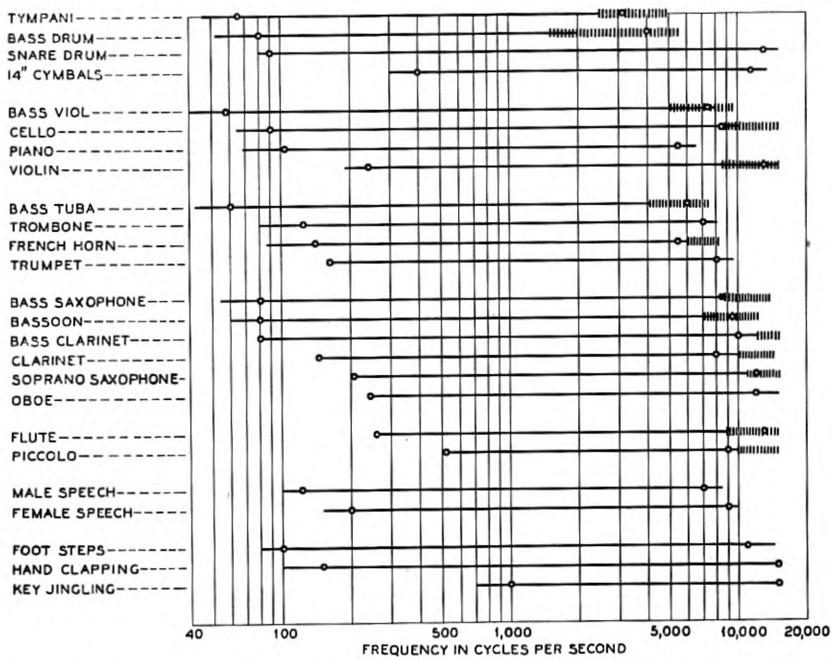

Fig. 7—Audible frequency ranges of music, speech, and noise

Observers voted for entire band by ratio of 60 to 40 over band shown by extremities of lines, and by ratio of 80 to 20 over band indicated by circles. Broken lines show range of noise accompanying music.

the volume at which the smallest peaks are just above the threshold of audibility is therefore less by a number of decibels than the intensity range of the ear as measured by single frequencies.

In the second place, the volume range of the usual program material has definite limits. Measurements have shown that a large symphony orchestra produces a maximum volume range of about 70 decibels.[4] The volume range of most other types of program is considerably less than this, for example, being only about 25 to 30 decibles for dance music and as little as about 15 decibels for much of the dialogue of actors in radio drama.

In the third place, the usual listening conditions impose a definite limit on the useful volume range. The loudest passages in the music of a symphony orchestra correspond to a sound level of about +95 decibels at a point, say one-third the way back in an auditorium, but most people in their homes prefer a level which is lower than this by 5 to 10 decibels. Figure 8

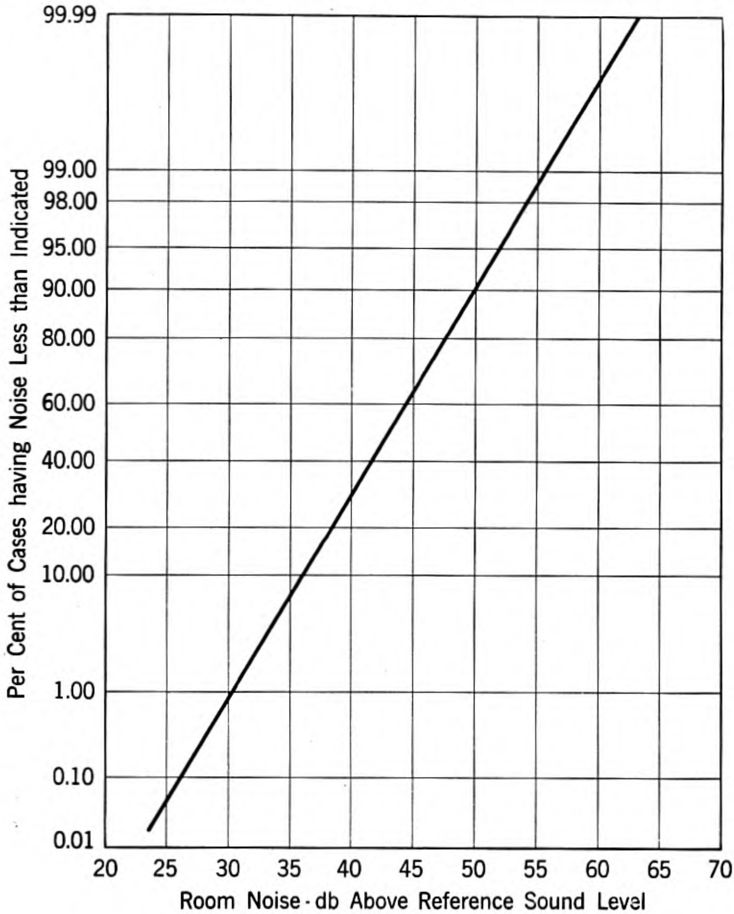

Fig. 8—Residence room noise with radio sets silent
Average 43 db. Standard deviation 5.5 db.

shows the results of an extensive survey[7] of acoustic room noise in homes. It will be noted that the average noise level is +43 decibels on the sound-level scale, and that few homes are quieter than +30 decibels even in the suburbs. The signal-to-noise range inherent in the listening conditions, and allowing nothing for the noise contributed by the transmission system or the room noise where the program is being produced, is therefore seen to be somewhere between 45 and 65 decibels. There is, therefore, no advantage to the listener in providing a permissible volume range materially wider than this in the transmission system.

The above discussion applies primarily to the transmission of symphonic and similar high-grade program material. Much program material is less exacting in its requirements, but on the other hand some sound effects such as the tearing of paper require the reproduction of higher frequencies for complete naturalness.

With these broad considerations in mind, the requirements of high-quality program circuits may now be taken up. As has been noted, the program network is but one part of the over-all broadcasting system which, in addition, includes microphones and studio equipment, radio transmitters, and the home receivers with their loud-speakers. It may be taken as a goal for the program networks that their transmission be nearly enough distortionless so that the over-all performance in regard to naturalness of reproduction will not be limited by them.

To meet such a requirement for short program circuits having only one or two sections is not difficult technically and does not in general require costly types of plant. However, the vast country-wide program networks are made up of many sections of circuits in tandem, which as mentioned before may total in some cases as much as 7,000 miles. This makes it necessary to design and operate the individual circuits to very close limits so that the cumulative discrepancies in the whole network will not exceed tolerable values; and to consider carefully the types of plant employed lest by virtue of sheer numbers of units involved, the total cost be out of line with the over-all grade of service being given the listner. These two conflicting factors are important ones in the consideration of transmission requirements for networks. The determination of the practical working characteristics of program networks involves a consideration not only of the physical and cost factors discussed above but also of such other factors as cost of studios, broadcast transmitters and receivers, and the limitations of the frequency allocations of broadcast stations.

From the standpoint of frequency band the consideration of all factors has resulted in the major present-day program networks being set up to transmit a frequency band with an upper limit of about 5,000 cycles. All program facilities installed in the last ten years or so, however, have been designed to be adaptable to the future transmission of frequencies up to 8,000 cycles. Operation on an 8,000-cycle basis, however, requires the release of additional frequency space now occupied by other services in much of the plant and a general readjustment of the program-circuit characteristics. In 1933, experimental wire circuits were set up between Philadelphia and Washington to transmit frequency bands up to 15,000 cycles. These were employed in a demonstration of stereophonic transmission and reproduction of music.[4] Studio-transmitter loops have been provided to transmit wider frequency bands than the 5,000 cycles currently provided on the nation-wide networks. At the present time, many of the studio-transmitter loops are being set up to transmit bands up to 15,000 cycles. A demonstration will be given at the close of the paper of the transmission of programs over cable circuits about 1,200 miles in length with frequency bands extending to 15,000, 8,000, and 5,000 cycles. The 5,000-cycle circuit is of the type in present commercial use. The 8,000-cycle circuit is of a type to which much of the present program plant can readily be modified. The 15,000-cycle circuit consists of a standard carrier system to which has been added program terminal equipment now under development.

In the consideration of transmission requirements for program circuits other than nominal frequency band, the variation in performance with length and type of circuit is important, since the factors tending to impair transmission are in the nature of small amounts of distortion or noise which accumulate over the length of the circuit. If these effects varied in some definite manner with length, transmission requirements could be fixed on that basis. However, good engineering practice frequently requires choosing for the various sections of a long circuit, different types of facilities whose contributions to the total effects are not in proportion to their length. Even the determination of the maximum permissible distortion and noise on a circuit is influenced by outside factors such as are involved in the broadcasters weighing operating flexibility and cost against the frequency of occurrence of unfavorable network routings and the number of stations affected. For example, in order to secure operating flexibility with a minimum of total network mileage, most of the networks employ the so-called "round robin" principle for a part of the network. In this arrangement the circuit follows a route from station to station forming a continuous loop which returns to its starting point. This naturally results in increased circuit mileage between the program source and the more distant listeners with an attendant increase in undesired transmission effects. For these reasons no exact or specific transmission requirements can be stated for even the over-all performance of program transmission service.

Volume Range

The permissible volume range for a program circuit is determined by the maximum volume which can be transmitted as limited by nonlinear distortion or crosstalk, and the minimum volume which can be transmitted without impairment from the noise present on the circuit.

In connection with their design the various types of program circuit are subjected to listening tests in which the transmission of program over a long loop of the circuit is compared with transmission over a local distortionless circuit. Each type of circuit thus is rated as to the maximum volume it can transmit without noticeable distortion. The highest volume which can be permitted without excessive crosstalk into other program or message telephone circuits is also investigated, and whichever limit is the lower determines the maximum allowable working volume for service. The range between the maximum permissible volume and the noise level on very long lengths of the present program circuits is about 45 or 50 decibels, except under some conditions on certain open-wire sections. On the individual links making up the long circuit, the range is 10 or 20 decibels greater than this.

Attenuation and Delay Distortion

Another important consideration is the amount of attenuation and delay (or phase) distortion to be permitted within the transmitted frequency band. It is the practice to equip program circuits with adjustable attentuation equalizers. By means of these once the desired frequency band has been chosen the deviation in attenuation at any frequency within that band, compared with that at 1,000 cycles, can be adjusted within close limits. On very long circuits, however, experience has shown that even with automatic regulating features and careful operation residual variations which may amount to several decibels may develop as a result of changing temperature and other conditions. These variations are kept within tolerable limits by readjustment of the equalizers from time to time.

Associated with the attenuation distortion is another effect detrimental to program quality, namely, differences in time of transmission for different frequency components of the signal. In practice, circuits tend to have a lower velocity of transmission near the edges of the frequency band than in the middle portions. This results in frequency components near the edges being delayed as compared to the middle portions of the band. This difference in time of transmission is called delay distortion of the circuit. Careful listening tests have shown that it becomes noticeable if, at the highest transmitted frequency, the delay is more than eight milliseconds greater than at 1,000 cycles, and if, at 100 cycles, it is more than about 15 milliseconds greater than at 1,000 cycles. It is controlled by careful attention to the design of loading systems, amplifiers, repeating coils, and all other elements of the circuit. Since such small amounts of over-all delay distortion are detectable and since networks frequently have 100 or more amplifiers in tandem between an originating point and the broadcasting stations on the more distant portions of the networks, it is necessary that the delay distortion of all individual components of a network be held to exceedingly close limits. Accumulations of residual delay distortion which cannot be entirely eliminated in design are reduced by the use of delay equalizers along the circuits when they are set up.

Conclusion

From this discussion it is seen that the program networks are comprised of many parts, each of which must meet exacting requirements in order that over-all results will be satisfactory. It is seen that equally important with transmission are the requirements for plant flexibility, adequate reserves, uniform practices, and centralized supervision of the networks.

The features discussed have been those found desirable for present-day network service. As indicated earlier, consideration of the needs of the future as well as those of the present is an essential feature of the design and engineering of the plant for program-network service. As a result of having done this it will be possible to provide with present plant, and with new plant currently being installed, adequate network facilities as the broadcasting art develops toward higher standards of performance. With the past experience as a guide, it appears that there should be no fundamental difficulty in meeting all reasonable requirements, always remembering that in the long run, requirements and costs bear definite relations to each other.

Acknowledgment

The authors of this paper gratefully acknowledge the assistance in its preparation of many of their associates, especially Mr. W. E. Bloecker, Mr. D. K. Gannett, and Mr. G. S. Bibbins.

References

1. "Telephone Circuits for Program Transmission," F. A. Cowan, *A.I.E.E. Transactions*, Vol. 48, no. 3, July 1929, pages 1045–9.
2. "Long Distance Cable Circuit for Program Transmission," A. B. Clark and C. W. Green, *Bell System Technical Journal*, Vol. 9, no. 3, July 1930, pages 567–94.
3. "Wide Band Open Wire Program System," H. S. Hamilton, *Electrical Engineering*, Vol. 53, no. 4, April 1934, pages 550–62.
4. "Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective," *Electrical Engineering*, Vol. 53, no. 1, January 1934, pages 9–32, 214–19; *Bell System Technical Journal*, Vol. 13, no. 2, April 1934, pages 239–308:
   "Basic Requirements," Harvey Fletcher.
   "Physical Factors," J. C. Steinberg and W. B. Snow.
   "Loud Speakers and Microphones," E. C. Wente and A. L. Thuras.
   "Amplifiers," E. O. Scriven.
   "Transmission Lines," H. A. Affel, R. W. Chesnut and R. N. Mills.
   "System Adaptation," E. H. Bedell and Iden Kerney.
5. "Absolute Amplitude and Spectra of Certain Musical Instruments and Orchestras," L. J. Sivian, H. K. Dunn, and S. D. White. *Journal Acoustical Society of America*, Vol. 2, January 1931, pages 330–71.
6. "Audible Frequency Range of Music, Speech and Noise," W. B. Snow, *Journal Acoustical Society of America*, Vol. 3, July 1934, pages 155–6.
7. Two papers: "Room Noise at Telephone Locations," D. F. Seacord, *Electrical Engineering*, Vol. 58, no. 6, June 1939, pages 255–7, and Vol. 59, no. 6, June 1940, pages 232–4.

Abstracts of Technical Articles by Bell System Authors

Notes on the Time Relation between Solar Emission and Terrestrial Disturbances.[1] CLIFFORD N. ANDERSON. Although the correlation between general solar activity and terrestrial disturbances is quite evident, the association of individual storms with specific sunspot groups has never been very satisfactory. Disturbances sometimes have occurred when no sunspots were visible and at other times large sunspots have been unaccompanied by any abnormal disturbances. A possible explanation of such anomalies may lie in longer transit times for the disturbing solar emission than is usually assumed. Some indication is given in this paper that these transit times may range from periods as short as only one or two days to as much as three months. The corresponding velocities for the above transit times are of the order of 2000 and 20 kilometers per second.

Curves show the approximate relation between the angle of emission, velocity, day of emission, and the days intervening between the passage of a spot through the central meridian of the sun and the corpuscular encounter with the earth.

The Effect of the Earth's Curvature on Ground-Wave Propagation.[2] CHARLES R. BURROWS and MARION C. GRAY. Curves are presented for the rapid calculation of the ground wave for radio propagation over a spherical earth of arbitrary ground constants, antenna heights, and polarization.

Based on the pioneering work of G. N. Watson, a rigorous theory of the propagation of electromagnetic waves round a spherical earth has been developed in the past twenty years. Watson developed his method in detail only in the limiting case of an earth of infinite conductivity, but his work has since been extended by various authors to cover other values of the earth's conductivity. Theoretically, therefore, solutions are available for any values of the earth's constants (dielectric constant and conductivity) and for either vertically polarized or horizontally polarized waves. In practice, unfortunately, the computations required are lengthy and involved, and for the most part the recent theoretical papers have confined their calculations to a few specific values of the earth's constants. The present paper attempts to summarize the results so far obtained in a manner

[1] *Proc. I.R.E.*, November 1940.
[2] *Proc. I.R.E.*, January 1941.

that will make them more easily available to the practical engineer, and to fill the gaps in these results by developing a series of curves from which the field for any values of the earth's constants may be read, with all the accuracy that could be expected in engineering practice.

Electrical Breakdown of Anodically Oxidized Coatings on Aluminum: A Means of Checking Thickness of Anodized Finishes.[3] K. G. COMPTON and A. MENDIZZA. The existing methods for determining the thickness of anodically produced oxide coatings on aluminum are relatively few and are almost entirely of a destructive nature. It is a fairly well established fact that, within the thickness limits normally encountered in practice, the voltage breakdown is a linear function of thickness of oxide film. The authors have endeavored to utilize this fact in developing a test method for determining the thickness of coatings produced under known and controlled conditions with practically no injury to the finish. Data are given which show the relationship between breakdown resistance, anodizing time, thickness of coating, current density and sealing of anodically oxidized polished commercially pure aluminum. Statistical data for the values obtained are also given, indicating the good reproducibility of the breakdown values. By calibrating a particular anodic process, satisfactory results may be obtained in a relatively short time and often without destroying or marring the article. Since the oxide coating is not entirely homogeneous it is necessary to obtain a fairly large number of readings for every test condition. The authors have found that approximately twenty-five readings are usually sufficient and can be made in a relatively short time. Although only one of the many anodizing possibilities has been investigated, the applicability of this method of evaluating the thickness of oxide coatings may be extended to all commercial treatments.

Ultrasonic Absorption and Velocity Measurements in Numerous Liquids.[4] GERALD W. WILLARD. By means of ultrasonic light-diffraction phenomena the velocity and absorption of sound in some forty transparent liquids were measured in the frequency range of 6 to 30 Mc. Among the list of materials studied are mixtures of liquids in varying proportions, several solutions of solids in liquids, and a non-liquid jell. A novel-construction glass-to-metal-to-quartz cell made possible the study of highly solvent liquids. Velocity values were obtained from measurements of the diffraction spectra spacing. Absorption values were obtained by measurement of the sound radiator voltages required to produce certain color transmission effects at measured distances from the sound radiator. The use of a mercury arc light-source

[3] *A.S.T.M. Proc.*, Vol. 40, 1940.
[4] *Jour. Acous. Soc. Amer.*, January 1941.

enhanced the necessary color effects. The relation between sound beam width (in the optical direction) and light transmission was studied. In general, the values of velocity obtained were found to be independent of frequency, and the absorption to be proportional to frequency squared and unrelated to calculated viscous and thermal losses. A simple calculation is proposed for estimating absorption errors caused by sound beam diffraction and spreading. These apply as well to absorption measured in other methods than here used.

Contributors to this Issue

W. R. Bennett, B.S., Oregon State College, 1925; A.M., Columbia University, 1928. Bell Telephone Laboratories, 1925–. Mr. Bennett has been engaged in the study of the electrical transmission problems of communication.

F. A. Cowan, B.S. in Electrical Engineering, Georgia School of Technology, 1919. American Telephone and Telegraph Company at Atlanta, Ga., 1920; Long Lines Engineering Department, Special Service Group, New York, 1922; Division Transmission Engineer, Division No. 1, New York, 1926; Engineer of Transmission, Long Lines Department, 1928; Engineering Department, American Telephone and Telegraph Company, New York, 1940–. Mr. Cowan is Transmission Engineer and as such is responsible for assisting the Associated Companies in connection with telephone and telegraph transmission and protection matters.

R. V. L. Hartley, A.B., Utah, 1909; B.A., Oxford, 1912; B.Sc., 1913; Instructor in Physics, Nevada, 1909–10. Engineering Department, Bell Telephone Laboratories, 1913–. Mr. Hartley took part in the early radio telephone experiments and has since been associated with research on telephony and telegraphy at voice and carrier frequencies. As Research Consultant he is concerned with general circuit problems.

Irving E. Lattimer, B.E.E., University of Michigan, 1913. Central Union Telephone Company, Engineering Department, Chicago, 1913–14; Chicago Telephone Company, Engineering Department, 1914–18. American Telephone and Telegraph Company, Long Lines Engineering Department, New York, 1918–. Mr. Lattimer is Transmission Methods Engineer.

R. G. McCurdy, B.S., University of California, 1913. Pacific Gas and Electric Company, 1913; Joint Committee on Inductive Interference (California), 1913–1916. American Telephone and Telegraph Company, 1916–1934; Bell Telephone Laboratories, Inc., 1934–. Assistant Director of Transmission Development, 1937–1939; Director of Transmission Engineering, 1939–.

S. O. Rice, B.S. in Electrical Engineering, Oregon State College, 1929; California Institute of Technology, 1929–30, 1934–35. Bell Telephone Laboratories, 1930–. Mr. Rice has been concerned with various theoretical investigations relating to telephone transmission theory.

W. W. Werring, M.E., Cornell University, 1922; Instructor, Mechanical Engineering Laboratory at Cornell, 1920–22. Engineering Department, Western Electric Company, 1922–25; Bell Telephone Laboratories, 1925–. Mr. Werring was at first associated with the Apparatus Analysis Laboratory and went with the Materials Department on its formation. His earlier work there was in the field of insulating materials and plastic molding. More recently he has been concerned with the application of welding, x-rays, optical and photographic methods, gauging and precision measurements and the engineering features of standardization of dimensions and tolerances.