Convention Record

of the I·R·E

1953 NATIONAL CONVENTION

Part 5—Circuit Theory

SESSIONS ON . . .

Circuits I — Network Theory

Circuits II — Symposium: Panel Discussion on Wide-Band

Amplifiers

Circuits III — Time Domain Networks — Delay Lines

Circuits IV — Active Networks — Transistors

SPONSORED BY

IRE PROFESSIONAL GROUP ON . . .

Circuit Theory

The Institute of Radio Engineers

CONVENTION RECORD OF THE I.R.E.

1953 NATIONAL CONVENTION

PART 5 - CIRCUIT THEORY

TABLE OF CONTENTS

A GENERAL RLC SYNTHESIS PROCEDURE

By Louis Weinberg
Hughes Aircraft Company
Research and Development Laboratories
Culver City, California

I. Introduction

There is a wide variety of existing synthesis procedures, but as anyone conversant with the synthesis field fully realizes, much remains to be done. The inadequacy of available procedures shows up particularly in a broad field of communications, namely, synthesis for prescribed transient response[1]. In this synthesis both magnitude and phase are important so that the methods for realizing a prescribed magnitude of transfer function are inapplicable. Up to the present time the only procedure that could be used for the realization of both minimum-phase and nonminimum-phase transfer functions has been the one that yields a constant-resistance lattice. This type of lattice suffers from many disadvantages. In general each of the arms requires close coupled or lossless coils. An important disadvantage, too, for those cases in which an unbalanced form of network is definitely preferable, is that the series and cross arms are so complicated relative to each other that without the use of ideal transformers reduction to an unbalanced form is virtually impossible.

The lattice synthesis procedures treated in this paper realize a given transfer function within a multiplicative constant. No restriction other than physical realizability is placed on the function to be realized. Among the advantages claimed for the final lattice are that it contains no mutual inductance and all its coils are lossy, i.e., every inductance may be associated with a series resistance. In addition, the arms are of so simple a form as to render the lattice amenable to reduction to an unbalanced network. For the case of a transfer admittance, moreover, reduction can always be achieved with the use at most of real transformers, i.e., transformers with winding resistance, finite magnetizing inductance, and a coupling coefficient smaller than one.

The dimensions of the transfer function to be realized depend, of course, on the type of system in which the synthesized network is to be used. If the driving force, for example, is a pentode which approximates a current source, a transfer impedance is needed. On the other hand, a transfer admittance is necessary for a voltage-source drive like a cathode follower and a current output. Finally, a ratio of output to input voltage calls for the synthesis of a dimensionless transfer function. The procedures show how each of these functions may be realized with the practical terminations of an open circuit, a resistance, or a parallel resistance-capacitance combination.

II. The Synthesis Problem Presented

In Fig. 1 is shown a general passive, two terminal-pair network which may be completely characterized by the pair of simultaneous equations

$$I_1 = y_{11}E_1 + y_{12}E_2$$
$$I_2 = y_{21}E_1 + y_{22}E_2 \tag{1}$$

or by the inverse set

$$E_1 = z_{11}I_1 + z_{12}I_2$$
$$E_2 = z_{21}I_1 + z_{22}I_2 \tag{2}$$

where $z_{12} = z_{21}$ and $y_{12} = y_{21}$ by reciprocity, and the y's and z's are the familiar short-circuit admittances and open-circuit impedances, respectively. Useful relations may be derived from the above basic equations. For example, for a pure resistance load of one ohm, since numerically $E_2 = -I_2$ the second of Eqs. 1 yields

$$Y_{12} \equiv \frac{I_2}{E_1} = \frac{y_{12}}{1 + y_{22}}, \tag{3}$$

an analogous relation being derived for Z_{12} from the second of Eqs. 2.

For the load an open circuit, that is with $I_2 = 0$, Eqs. 2 yield

$$K \equiv \frac{E_2}{E_1} = \frac{z_{12}}{z_{11}} \tag{4}$$

while for a short-circuit load, from Eqs. 1 may be obtained

$$\frac{I_2}{I_1} = \frac{y_{12}}{y_{11}}. \tag{5}$$

In this paper the two terminal-pair network is particularized to the lattice shown in Fig. 2 for which

2

$$z_{11} = \frac{1}{2}(Z_b + Z_a)$$

$$z_{12} = \frac{1}{2}(Z_b - Z_a), \tag{6}$$

an analogous set of relations holding true for the y's. Thus, with the substitution of Eqs. 6, Eq. 4 becomes

$$K = \frac{Z_b - Z_a}{Z_b + Z_a} \tag{7}$$

whereas substitution of the y's for the lattice transforms Eq. 3 to

$$Y_{12} = \frac{\frac{1}{2}(Y_b - Y_a)}{1 + \frac{1}{2}(Y_b + Y_a)} . \tag{8}$$

It is well known, furthermore, that the general form for any of the transfer functions is a rational function given by the quotient of polynomials

$$\frac{a_n s^n + a_{n-1} s^{n-1} + \ldots + a_o}{b_m s^m + b_{m-1} s^{m-1} + \ldots + b_o}$$

$$= \frac{(s-s_1)(s-s_3) \ldots (s-s_n)}{\frac{b_m}{a_n}(s-s_2)(s-s_4) \ldots (s-s_m)} \quad n \le m \tag{9}$$

$$= \frac{p(s)}{Hq(s)} \quad ,$$

where q must be a Hurwitz polynomial, that is, have all its zeros in the left half-plane, and H is a positive constant.

III. Open-Circuited Lattice Realization for $K = \frac{E_2}{E_1}$

It is desired to realize the given transfer function

$$K = \frac{E_2}{E_1} = \frac{p}{Hq} \tag{10}$$

as an open-circuited lattice with as large a gain (that is, as small an H) as possible. The given function may therefore be set equal to Eq. 7; this yields

$$K = \frac{p}{Hq} = \frac{Z_b - Z_a}{Z_b + Z_a} . \tag{11}$$

We break q into the sum of two polynomials[2], so that

$$q = q_1 + Aq_1' , \tag{12}$$

where q_1 is a Hurwitz polynomial, A is a positive real constant, and q_1' is the derivative of q_1. This can always be done as demonstrated in the reference cited. After dividing the numerator and denominator of the resulting K by q_1 to obtain

$$K = \frac{\dfrac{p}{q_1}}{H(1 + \dfrac{Aq_1'}{q_1})} , \tag{13}$$

we expand p/q_1 into partial fractions. Its residues are in general positive or negative real for real poles, complex for complex poles. A similar expansion of Aq_1'/q_1 makes the total denominator of K

$$H(1 + \frac{Aq_1'}{q_1}) = H(k_o^{(d)} + \frac{k_1^{(d)}}{s - s_1} + \frac{k_2^{(d)}}{s - s_2} + \ldots) \tag{14}$$

where $k_o^{(d)} = 1$, and all the residues $k_\mu^{(d)}$ for $\mu \ne 0$ are equal to A. If $(Z_b - Z_a)$ and $(Z_b + Z_a)$ are also thought of as expanded in partial fractions, the residues of like terms of p/q_1 and $(Z_b - Z_a)$ may be equated as may those of $H(1 + Aq_1'/q_1)$ and $(Z_b + Z_a)$. We thus obtain

$$k_\mu^{(b)} - k_\mu^{(a)} = k_\mu^{(n)}$$

$$k_\mu^{(b)} + k_\mu^{(a)} = Hk_\mu^{(d)} \quad \begin{array}{l} (\mu = 0, 1, 2, \ldots m, \\ \text{where m is degree of q)} \end{array} \tag{15}$$

where the superscripts a, b, n, and d refer respectively to Z_a, Z_b, and the numerator and denominator of K, while the subscript μ designates the poles $s_\mu = -\sigma_\mu + j\omega_\mu$, or $s_\mu = -\sigma_\mu - j\omega_\mu$, or $s_\mu = -\sigma_\mu$ in which σ_μ and ω_μ are positive.

Solving Eqs. 15 for the unknown Z_a and Z_b residues as indicated in Eqs. 16 below

$$k_\mu^{(b)} = \frac{1}{2}(Hk_\mu^{(d)} + k_\mu^{(n)})$$

$$\alpha_\mu^{(b)} + j\beta_\mu^{(b)} = \frac{1}{2}(Hk_\mu^{(d)} + \alpha_\mu^{(n)} + j\beta_\mu^{(n)})$$

$$k_\mu^{(a)} = \frac{1}{2}(Hk_\mu^{(d)} - k_\mu^{(n)}) \tag{16}$$

$$\alpha_\mu^{(a)} + j\beta_\mu^{(a)} = \frac{1}{2}(Hk_\mu^{(d)} - \alpha_\mu^{(n)} - j\beta_\mu^{(n)})$$

finally yields

3

$$\mu \neq 0$$
$$\alpha_\mu^{(b)} = \frac{1}{2}(HA + \alpha_\mu^{(n)}) \qquad \alpha_\mu^{(a)} = \frac{1}{2}(HA - \alpha_\mu^{(n)})$$

$$\beta_\mu^{(b)} = \frac{1}{2}\beta_\mu^{(n)} \qquad \beta_\mu^{(a)} = -\frac{1}{2}\beta_\mu^{(n)}$$

$$\tag{17}$$

$$\mu = 0$$
$$\alpha_0^{(b)} = \frac{1}{2}(H + \alpha_0^{(n)}) \qquad \alpha_0^{(a)} = \frac{1}{2}(H - \alpha_0^{(n)}).$$

For negative real poles the requirement that the residues, $\alpha_\mu^{(a)}$ and $\alpha_\mu^{(b)}$, be real and positive, when used in conjunction with the above Eqs. 17, gives as the condition to be satisfied for $\mu \neq 0$

$$-1 \le \frac{\alpha_\mu^{(n)}}{HA} \le 1 , \tag{18}$$

and for $\mu = 0$ we substitute the constant one for A. This, of course, is the same as the condition that arises in the Bower-Ordung RC synthesis[3], since for an RC lattice the poles must all be real. In the general synthesis considered here, however, the complex poles must also be provided for. The real parts of the residues in these complex poles must not only be positive, but must also be equal to or greater than a positive constant c_μ which is defined below. This is seen by application to the residues of Z_a of the condition for realizability that is derived in Appendix I (with a similar result holding for Z_b):

$$\frac{\left|\beta_\mu^{(a)}\right|}{\alpha_\mu^{(a)}} \le \frac{\sigma_\mu}{\omega_\mu} \tag{19}$$

or

$$\alpha_\mu^{(a)} \ge \frac{\left|\beta_\mu^{(a)}\right|\omega_\mu}{\sigma_\mu} = c_\mu \text{ (positive constant)}.$$

$$\ge \frac{\frac{1}{2}\left|\beta_\mu^{(n)}\right|\omega_\mu}{\sigma_\mu} = c_\mu. \tag{20}$$

When the constant c_μ is substituted in those relations of Eqs. 17 for which $\mu \neq 0$, the conditions to be satisfied become

$$1 \ge \frac{2c_\mu - \alpha_\mu^{(n)}}{HA} \tag{21}$$

and

$$1 \ge \frac{2c_\mu + \alpha_\mu^{(n)}}{HA} . \tag{22}$$

We need satisfy only the stronger of the above two inequalities for any specific complex

pole. If $\alpha_\mu^{(n)}$ is positive, Eq. 22 is the stronger and must be used to determine the minimum value of H; if $\alpha_\mu^{(n)}$ is negative, we use Eq. 21. Therefore, to summarize the two steps for the complex poles, we must determine first the c_μ for each pole and then the value of H necessary to satisfy the stronger of Eqs. 21 and 22.

By satisfaction also of Eq. 18 for the real poles we may thus tabulate the necessary value of H for each pole. In doing this we may use the equals sign in Eqs. 18, 21, and 22; then we choose a value of H greater than the largest required value, which automatically guarantees the satisfaction of the condition for each pole with the inequality sign. This is necessary, as is pointed out in Appendix I, in order that every inductance appear with an associated resistance and that each of the partial fraction components (complex conjugate poles taken in pairs) of

$$Z_a = k_0^{(a)} + \frac{k_1^{(a)}}{s - s_1} + \frac{k_2^{(a)}}{s - s_2} + \dots + \frac{k_m^{(a)}}{s - s_m}$$

and

$$\tag{23}$$

$$Z_b = k_0^{(b)} + \frac{k_1^{(b)}}{s - s_1} + \frac{k_2^{(b)}}{s - s_2} + \dots + \frac{k_m^{(b)}}{s - s_m}$$

be positive real. Then Z_a and Z_b may be realized by inspection in the Foster manner for two-element kind networks; and as Appendix I demonstrates, every inductance will have an associated series resistance.

One final point, useful in the subsequent procedures, is made regarding the constant term in Eqs. 23. As is obvious from Eqs. 17 for $\mu = 0$, we always obtain this constant term. Furthermore, if the degree of p is less than that of q, then

$$\alpha_0^{(b)} = \alpha_0^{(a)} = \frac{1}{2}H, \tag{24}$$

while if the degrees of p and q are equal,

$$\alpha_0^{(b)} = \alpha_0^{(a)} + 1 = \frac{1}{2}(H + 1). \tag{25}$$

IV. Procedure Using the Integral of q

An alternate procedure is obtained by making use of the integral of q, designated by $q^{(-1)}$. This method has the drawback that it may not always work; it works only when $q^{(-1)}$ is Hurwitz. It can be easily shown that the integral of a Hurwitz polynomial, unlike the derivative, is not always Hurwitz. It is a simple matter in any particular problem, however, to form the integral, choosing the arbitrary constant conveniently, and then check for Hurwitz character[4]. The advantage of this method is that it generally yields a network with a higher gain and coils of lower Q.

4

The steps in the procedure for synthesizing an open-circuited lattice begin with

$$K = \frac{E_2}{E_1} = \frac{Z_b - Z_a}{Z_b + Z_a}$$

$$= \frac{p}{Hq} \tag{26}$$

$$= \frac{\dfrac{p}{q^{(-1)}}}{\dfrac{Hq}{q^{(-1)}}} \; .$$

Now $q/q^{(-1)}$ is obviously of the same form as q_1'/q_1, if $q^{(-1)}$ is Hurwitz; and the procedure follows along the same lines as the previous one except for a few minor differences. The differences to be noted are that the constant term in the partial fraction expansions, i.e., for $\mu = 0$, is nonexistent and the constant A is equal to one.

For this method the useful equations that correspond to Eqs. 17 are

$\mu \neq 0$ $\mu \neq 0$

$$\alpha_\mu^{(b)} = \frac{1}{2}(H + \alpha_\mu^{(n)}) \qquad \alpha_\mu^{(a)} = \frac{1}{2}(H - \alpha_\mu^{(n)})$$

$$\beta_\mu^{(b)} = \frac{1}{2}\beta_\mu^{(n)} \qquad\qquad \beta_\mu^{(a)} = -\frac{1}{2}\beta_\mu^{(n)} \tag{27}$$

which yield for the real poles the condition corresponding to Eq. 18,

$$-1 \leq \frac{\alpha_\mu^{(n)}}{H} \leq 1, \tag{28}$$

and for the complex poles yield the inequalities that correspond respectively to Eqs. 21 and 22,

$$1 \geq \frac{2c_\mu - \alpha_\mu^{(n)}}{H} \tag{29}$$

and

$$1 \geq \frac{2c_\mu + \alpha_\mu^{(n)}}{H} \; . \tag{30}$$

The use of the above equations, along with the definition of c_μ given in Eq. 20 allows a synthesis to be carried out.

V. Realization of Transfer Impedance in Form of Terminated Lattice

A. Resistance Termination

To obtain the transfer impedance

$$z_{12} = \frac{E_2}{I_1} = \frac{p}{Hq} \tag{31}$$

as a resistance-terminated lattice, it is possible to make use of the dual form of Eq. 3,

$$z_{12} = \frac{z_{12}}{1 + z_{22}} = \frac{\frac{1}{2}(Z_b - Z_a)}{1 + \frac{1}{2}(Z_b + Z_a)} \; . \tag{32}$$

By proceeding in a manner similar to that in Section III, we can then make the necessary identifications for direct synthesis of the lattice. The network obtained, however, is the same as the one obtained by application of the reciprocity theorem and well-known lattice equivalents to the open-circuited lattice of Section III. For simplicity of demonstration we will therefore consider the method of synthesis of that section as the basic one from which the other desirable forms of network are easily derived.

Since, as observed previously in Section III, a series resistance is always present in each arm of the open-circuited lattice, we derive an equivalent lattice[3,5] by removing one ohm from each arm, then convert to a current source by use of Norton's theorem, and finally by use of the reciprocity theorem obtain the desired network. The sequence of steps beginning with the previously realized open-circuited lattice for which

$$K = \frac{E_2}{E_1} = \frac{p}{Hq} \tag{33}$$

is illustrated in Fig. 3. The one-ohm series resistance at the output terminals is omitted in Fig. 3(c) because the output is open-circuited.

An improvement in gain can be effected by removing more than one ohm from each of the arms. Suppose, for the sake of illustration, that p is of lower degree than q so that, as noted in Eq. 24 $R_a = R_b = 1/2H$. Then we can remove R_a and follow the same sequence of steps as in Fig. 3 to obtain a network terminated in $R_a = 1/2H$ with the transfer function

$$z_{12} = \frac{E_2}{I_1} = \frac{1}{2}\frac{p}{q} \; . \tag{34}$$

It is pointed out, finally, that if we stop at the step given by Fig. 3(c), we realize a transfer impedance in the form of an open-circuited lattice, where the shunt resistance is useful in the instrumentation of the network since it may represent the finite internal resistance of the current source.

B. Parallel RC-Termination

In the instrumentation of a practical circuit it is often useful to have a shunt capacitance at the input or output of a network. The immediately preceding method showed how to obtain a resistance termination for the lattice network; this part demonstrates the realization of a parallel RC termination.

5

The following artifice is restricted in that it can only be used for those transfer functions in which the degree of p is less than the degree of q. Suppose that the q_1 given by the $q = q_1 + Aq_1'$ breakdown (similar remarks applying to the $q^{(-1)}$ procedure) has at least one negative real zero given by $(s + a)$. (If it does not and we wish to employ this technique we multiply numerator and denominator of the transfer function by a linear term $(s + b)$ to obtain a new denominator from which q_1 is determined.) Letting the impedance of the desired termination be $k/(s + a)$, we may then write for the RC-terminated lattice (using the additional constant k for convenience):

$$Z_{12} = \frac{E_2}{I_1} = \frac{kp}{Hq}$$

$$= \frac{kp}{H(q_1 + Aq_1')} = \frac{\dfrac{kp}{q_1}}{H(1 + \dfrac{Aq_1'}{q_1})}$$

(35)

$$= \frac{\dfrac{kp}{(s + a)q_2}}{H(1 + \dfrac{Aq_1'}{q_1})} .$$

Multiplying the transfer function by $(s + a)/k$ yields

$$K = (s + a)Z_{12}/k$$

(36)

$$= \frac{\dfrac{p}{q_2}}{H(1 + \dfrac{Aq_1'}{q_1})} .$$

Now this function K is realized as the voltage ratio of an open-circuited lattice by the method of Section III. Since the numerator residue in the pole s = -a is zero, it is clear by inspection of Eqs. 17 that the residues for Z_a and Z_b in this pole are each equal to HA/2. The network which thus has the form shown in Fig. 4(a) can be transformed by removal of the RC combination from each arm, after which the successive applications of Norton's Theorem and reciprocity give

$$Z_{12} = \frac{E_2}{I_1} = \frac{kp}{Hq} = \frac{1}{2}\frac{Ap}{q}$$

(37)

for the final desired form of network. The pertinent equations and steps are indicated in Fig. 4.

To summarize the procedure, we consider the given function as $Z_{12} = kp/Hq$ (this is what we finally achieve, where $k = 1/2(HA)$). We then multiply this by $(s + a)/k$ where $(s + a)$ is a

factor of q_1 (or of $q^{(-1)}$, if we are using the integral method). We then synthesize the resulting function as the voltage ratio for an open-circuited lattice, after which the steps illustrated in Fig. 4 will give Ap/(2q) as the desired transfer impedance for the RC-terminated lattice.

VI. Realization of Transfer Admittance as a Resistance-Terminated Lattice

As pointed out previously in Section V for the dual case, it is possible by use of the equation

$$Y_{12} = \frac{y_{12}}{1 + y_{22}} = \frac{\frac{1}{2}(Y_b - Y_a)}{1 + \frac{1}{2}(Y_b + Y_a)}$$

(38)

to develop the proper identifications for the synthesis procedure. The simpler approach, however, is to consider the entire procedure of Section III carried over to the dual problem; that is, instead of first synthesizing for a voltage ratio by means of an open-circuited lattice, we now realize

$$\frac{I_2}{I_1} = \frac{y_{12}}{y_{11}}$$

(39)

in the form of a short-circuited lattice. The dual of the remarks in Section V(A) now apply so that we obtain a final transfer function

$$Y_{12} = \frac{I_2}{E_1} = \frac{1}{2}\frac{p}{q}$$

(40)

for the same problem treated in that section. The final lattice with a resistance termination will be of the form shown in Fig. 5.

In Appendix II, where the general problem of the reduction of lattices to unbalanced forms is taken up, it is shown that the use of real transformers always allows the lattice obtained for Y_{12} to be converted to an unbalanced network.

VII. Illustrative Example

We desire to synthesize within a constant multiplier the given

$$Y_{12} = \frac{p}{q} = \frac{1}{(s^2 + 0.76536s + 1)(s^2 + 1.84776s + 1)}$$

as a resistance-terminated lattice. The given Y_{12} represents the Butterworth approximation to the normalized low-pass filter with n = 4, where the Butterworth functions are $B_{2n}(\omega^2) = 1 + \omega^{2n}$ and

$$\left. |Y_{12}|^2 \right|_{s=j\omega} = \left. \frac{1}{|q(s)|^2} \right|_{s=j\omega} = \frac{1}{1 + \omega^{2n}} .$$

6

We first realize

$$\frac{I_2}{I_1} = \frac{p}{q} = \frac{Y_b - Y_a}{Y_b + Y_a}$$

as a short-circuited lattice and then by transformations obtain the desired Y_{12}.

The breakdown $q = q_1 + Aq_1'$ for A=0.1 yields

$$q_1 = (s^2 + 1.61976s + 0.813459)(s^2 + 0.59336s + 0.975703)$$

and therefore

$$\frac{1}{q_1} = \frac{0.44509 - j1.0854}{s + 0.80988 - j0.39693} + \frac{-0.44509 + j0.21483}{s + 0.29668 - j0.94217}$$
$$+ \text{ conjugates.}$$

By inspection we see

$\alpha_1^{(n)} = 0.44509$ $\qquad$ $\beta_1^{(n)} = -1.0854$

$\alpha_2^{(n)} = -0.44509$ $\qquad$ $\beta_2^{(n)} = 0.21483$

$\sigma_1 = 0.80988$ $\qquad$ $\omega_1 = 0.39693$

$\sigma_2 = 0.29668$ $\qquad$ $\omega_2 = 0.94217.$

We find from

$$c_\mu = \frac{\frac{1}{2}\left|\beta_\mu^{(n)}\right|\omega_\mu}{\sigma_\mu}$$

that

$$c_1 = 0.2660$$
$$c_2 = 0.3411.$$

By use of Eqs. 21 and 22, we obtain

$$H_1 \geqq 9.8$$
$$H_2 \geqq 10.3$$

We choose H=25. Therefore

$\alpha_o^{(b)} = 12.5$ $\qquad$ $\alpha_o^{(a)} = 12.5$

$\alpha_1^{(b)} = 1.47255$ $\qquad$ $\alpha_1^{(a)} = 1.02746$

$\alpha_2^{(b)} = 1.02746$ $\qquad$ $\alpha_2^{(a)} = 1.47255$

$\beta_1^{(b)} = -0.5427$ $\qquad$ $\beta_1^{(a)} = 0.5427$

$\beta_2^{(b)} = 0.10742$ $\qquad$ $\beta_2^{(a)} = -0.10742.$

The lattice arms are then given by

$$Y_a = 12.5 + \frac{2.05492\,(s + 0.60022)}{s^2 + 1.61976s + 0.813459}$$

$$+ \frac{2.94510\,(s + 0.36541)}{s^2 + 0.59336s + 0.975703}$$

$$Y_b = 12.5 + \frac{2.94510\,(s + 0.95617)}{s^2 + 1.61976s + 0.813459}$$

$$+ \frac{2.05492\,(s + 0.19818)}{s^2 + 0.59336s + 0.975703}\,.$$

The application of the necessary transformations gives the lattice shown in Fig. 6 for which

$$Y_{12} = \frac{1}{2}\frac{p}{q}$$

$$= \frac{1}{2}\frac{1}{(s^2 + 0.76536s + 1)\,(s^2 + 1.84776s + 1)}.$$

VIII. Conclusion

A method has been demonstrated for realizing any transfer function by a lattice network. Each inductance used has an associated series resistance so that lossy coils may be used in building the network. General methods for reducing a lattice to an unbalanced form have been considered and it was shown that if real transformers are allowed, i.e., transformers with winding resistance, finite magnetizing inductance, and a coupling coefficient smaller than one, then the lattice realizing Y_{12} is always reducible to an unbalanced network.

Appendix I. Partial Fraction Expansion of a Positive Real RLC Function

For the driving-point functions of two-element kind networks the Foster method of synthesis can be successfully applied, that is, a partial-fraction expansion may be made to yield terms each of which is positive real and therefore realizable by inspection. Though the method breaks down for general RLC functions, it may work for a particular type of RLC function. In this appendix we determine the necessary conditions on the residues for the success of the method*; we determine, in addition, the conditions necessary so that a series resistance can be associated with every inductance;

---

*These conditions were pointed out by Dr. E. A. Guillemin in his Network Synthesis course at M. I. T.

and finally we indicate the application of the above to the synthesis treated in this paper

The partial-fraction expansion of a positive real RLC impedance

$$Z = \frac{k_1}{s - s_1} + \frac{k_2}{s - s_2} + \frac{k_3}{s - s_3} + \frac{k_4}{s - s_4} + \ldots \quad (I.1)$$

can be written more explicitly as

$$Z = \frac{\alpha_1 + j\beta_1}{s + \sigma_1 - j\omega_1} + \frac{\alpha_1 - j\beta_1}{s + \sigma_1 + j\omega_1} + \frac{R}{s + a} + \ldots$$

$$= \frac{2\alpha_1 \left(s + \sigma_1 - \dfrac{\beta_1 \omega_1}{\alpha_1}\right)}{s^2 + 2\sigma_1 s + \omega_o^2} + \frac{R}{s + a} + \ldots \quad (I.2)$$

$$= z_1 \qquad\qquad + z_2 \quad + \ldots$$

so that the typical terms z_1 and z_2 become apparent. To simplify exposition we have assumed only simple poles; we also have considered an impedance, the dual of the following remarks of course applying to an admittance. In Eq. I.2, σ_1 and ω_1 are real and positive constants, the constants R, α_1 and β_1 are real, and $\omega_o^2 \equiv \sigma_1^2 + \omega_1^2$. In order for terms like z_2 (with negative real poles) to be positive real and hence separately realizable as a simple driving-point impedance, it is clear that R must be positive. However, the condition on the residue of the partial-fraction component z_1 containing complex poles is not so obvious, except that it is necessary for α_1 to be positive. By application to z_1 of the well known test for positive real character[4], the additional condition is found to be

$$\frac{\left|\beta_1\right|}{\alpha_1} \leq \frac{\sigma_1}{\omega_1} . \quad (I.3)$$

In words, then, the condition for the existence of a positive real partial fraction component z_1 is that α_1 be positive and the angle formed by the imaginary axis and the radius from the origin to the pole be greater than or equal to the angle of the residue of the pole. This is illustrated in Fig. I.1 where angle $\varphi = \tan^{-1}(\beta_1/\alpha_1)$ must be less than or equal to the angle $\psi = \tan^{-1}(\sigma_1/\omega_1)$, or the residue may lie anywhere in the crosshatched portion of the plane.

It is desired that the inductance used in the synthesis of z_1 have a series resistance

associated with it. Since β_1 may be positive or negative, two cases arise. For $\beta_1 > 0$, if Eq. I.3 is satisfied with the equality sign, then $\beta_1\omega_1/\alpha_1 = \sigma_1$ and the quantity $(\sigma_1 - \beta_1\omega_1/\alpha_1)$ appearing in Eq. I.2 equals zero, so that a perfect coil is required. (For the admittance case all of the dissipation would be associated with the inductance.) On the other hand, if the inequality sign is used, $(\sigma_1 - \beta_1\omega_1/\alpha_1)$ is greater than zero, which guarantees a lossy coil. For the case of a negative β_1, it is possible to associate all of the dissipation with the coil by making $\sigma_1 - \beta_1\omega_1/\alpha_1 = 2\sigma$, with satisfaction of Eq. I.3 by the equals sign. Again, satisfaction with the inequality sign allows some dissipation to be associated with the coil. Thus, in both cases $(\beta > 0, \beta < 0)$, use of the inequality sign calls for an impedance containing a lossy coil; and since as is demonstrated in Section III, the lattice synthesis procedure requires that the lattice arms Z_a and Z_b have β's which are equal numerically but of opposite sign, it is mandatory that the inequality sign be used if it is desired to obtain lossy coils in both arms.

The synthesis procedure of this paper guarantees that the Z_a and Z_b have positive real partial-fraction components with lossy coils, and thus the arms may be realized by inspection in the Foster manner.

Appendix II. Reduction of Lattices to Unbalanced Networks

An unbalanced form of network, that is, one with a common ground from the input to the output terminals, is definitely preferred to a lattice. The problem of converting a lattice to an unbalanced network is therefore an important one. This appendix first considers general lattice equivalents, then passes on to reduction of the special types of lattices that may arise from the synthesis procedure of this paper and finally shows that the reduction of a network realizing a Y_{12} is always possible with at most real transformers.

Lattices may be transformed to unbalanced networks in the following ways[3,5]:

a) A series impedance may be removed from both Z_a and Z_b and placed in series with both the input and output terminals (see Fig. II.1(a)).

b) A shunt impedance may be removed from both Z_a and Z_b and placed in shunt with both the input and output terminals (see Fig. II.1(b)).

c) A shunt impedance may be removed from the series arm Z_a and considered as a bridge across the remainder of the lattice (see Fig. II.1(c)).

8

The ideal transformer that is necessary may be removed when the remainder of the lattice has been transformed to an unbalanced network.

   d) A series impedance may be removed from the cross arm as shown in Fig. II.1(d). Again the ideal transformer becomes unnecessary when the remainder of the lattice has been transformed to an unbalanced network.

   e) A lattice may be broken into a group of parallel lattices (see Fig. II.1(e)).

   In any specific problem, to reduce a lattice to an unbalanced form may require the application of a succession of the above methods, or it may first be necessary to resynthesize a lattice arm before one of the methods may be successfully applied. Thus considerable ingenuity is called for. However, the types of lattices that are realized in this paper, since they contain the same poles in both arms, are reducible in a large number of problems. We mention below a few of the forms which can be recognized as reducible in general.

   If all the residues in the real poles of one arm are larger than the corresponding residues in the other arm and, in addition, the coefficients of the numerators of the pairs of complex poles in the first arm are larger than the corresponding coefficients of the poles in the other arm, then the lattice is immediately reducible to an L network. Since this is a very restricted form of network, the residues will rarely have this desired distribution.

   A completely general form of unbalanced network is given if, at any stage of the lattice reduction process, the lattice arms can be resynthesized into the ladder forms shown in Fig. II.2, where G represents a conductance and Z an impedance. A two-element kind network is used for simplicity of illustration. How to bring this about in general is a matter requiring further investigation*, but when it is possible the lattice containing these arms can be reduced to the form shown in Fig. II.3.

   For the lattice obtained in the synthesis of Y_{12}, it is always possible to effect a reduction to an unbalanced network if one of the residues in a real pole is very large, specifically, large enough for the method to be applied. As an example of the reduction procedure when a large residue is present, consider the six-pole

---
   * After this had been written (in 1951), it came to the writer's attention that O. Aberth was conducting an investigation into this problem for RC networks as his Master's thesis research at M. I. T.

lattice shown in Fig. II.4, whose arms are given by

$$Y_a = y_{1a} + y_{2a} + y_{3a} + y_{4a} \qquad (II.1)$$

$$= \frac{2s + 1}{s^2 + 2s + 5} + \frac{3s + 4}{s^2 + 4s + 6} + \frac{8}{s + 2} + \frac{2}{s + 4}$$

and

$$Y_b = y_{1b} + y_{2b} + y_{3b} + y_{4b}$$

$$= \frac{3s + 2}{s^2 + 2s + 5} + \frac{4s + 7}{s^2 + 4s + 6} + \frac{2}{s + 2} \qquad (II.2)$$
$$+ \frac{1}{s + 4} .$$

   Because of the large residue in the admittance y_{3a}, the lattice can be reduced. First we remove from each arm the shunt branches y_{1a}, y_{2a}, y_{3b}, y_{4b} (see Fig. II.5(a)). Since the drive is a voltage source the shunt branches may be omitted from the input terminals. Then we split the remaining lattice into two parallel lattices with a bridging branch, as shown in Fig. II.5(b). Finally these may be transformed, as shown in Fig. II.5(c) to obtain a bridged twin-tee network with a complicated load.

   As a last resort, if all other methods fail, it is possible by the use of real transformers always to reduce any lattice obtained in the synthesis of Y_{12} by the methods presented in this paper. The procedure is explained below.

   As is well known, any lattice may be reduced to an unbalanced form by the use of an ideal transformer. The process calls for a rotation of the output terminals so that the cross arms become series arms. In order to compensate for this rotation an ideal transformer providing a phase reversal is necessary. The procedure is illustrated in Fig. II.6.

   However, we wish to use only real transformers, i.e., transformers possessing leakage inductance, winding resistance, finite magnetizing inductance, and core loss; the equivalent circuit of such a real transformer is an ideal one with a series resistance and inductance and a shunt resistance and inductance. We shall now show that the form of network realized in the synthesis of a transfer admittance automatically provides the necessary series and shunt branches.

   The general procedure is best explained by a simple example. Without loss of generality suppose we consider the lattice shown in Fig. II.7(a) where the numerator of y_{1a} is larger than

9

the numerator of the same pole in the cross arm, but the numerators of y_{2b} and y_{3b} are larger than the corresponding numerators of y_{2a} and y_{3a}. This is the form of lattice obtained in realizing a transfer admittance. We may remove y_{1b}, y_{2a}, and y_{3a} from both arms and divide the lattice into a group of lattices as shown in Fig. II.7(b). Then the final step of rotation and use of ideal transformers which now, moreover, have series and shunt inductances may be carried as in Fig. II.7 (c). Each of the networks within the broken lines represents the equivalent circuit of a real transformer.

If it is desired to use physically realizable mutual inductance without a core-loss resistance, we may use another method that is often applicable. The lattices that realize a transfer admittance may be divided into components similar to that shown in Fig. II.8(a). It is seen that the only requirement for the reduction of this component lattice is that R_b be greater than R_a. For, by removing R_a and L_a from each arm, we then obtain the network shown in Fig. II.8(b). If $(L_b - L_a)$ is negative, we can use the mutual inductance form of network shown in (c) of the figure for practical realization. All that remains is to show how to realize a sufficiently small R_a. This happy circumstance often comes about naturally in the realization of a transfer admittance. If it does not, it can often be brought about by the multiplication of numerator and denominator of Y_{12} by $(s + a)$ as the first step in the synthesis procedure, where $\underline{a}$ is a sufficiently small positive constant.
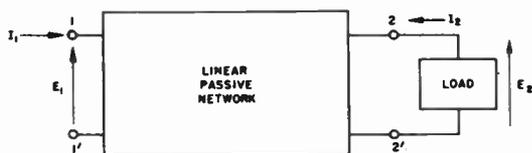
Thus, in summary, we see that we may often arrive at an unbalanced form of network by the methods presented in this appendix. Moreover, if we allow realizable mutual inductance in the synthesis of Y_{12}, we can <u>always</u> obtain an unbalanced network.

References

1. D. F. Tuttle, Jr.: "Network Synthesis for Prescribed Transient Response," Sc. D. Thesis in Electrical Engineering, M.I.T., 1948.

2. L. Weinberg: "Synthesis of Unbalanced RLC Networks," Proceedings of the National Electronics Conference, Vol. 8, 1952; also scheduled for publication, Journal of Applied Physics.

3. J. L. Bower and P.F. Ordung: "The Synthesis of Resistor-Capacitor Networks," Proc. I.R.E., March, 1950, pp 263-269.

4. E. A. Guillemin: "The Mathematics of Circuit Analysis," John Wiley, New York, 1949.

5. E. A. Guillemin: "RC-Coupling Networks," M.I.T. Rad. Lab. Report No. 43, Oct. 11, 1944.

Fig. 1 - Two terminal-pair network.
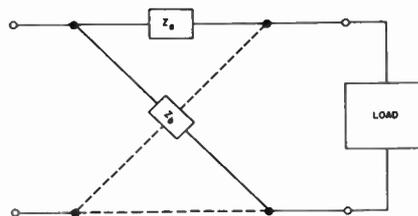
Fig. 2 - RLC lattice network.
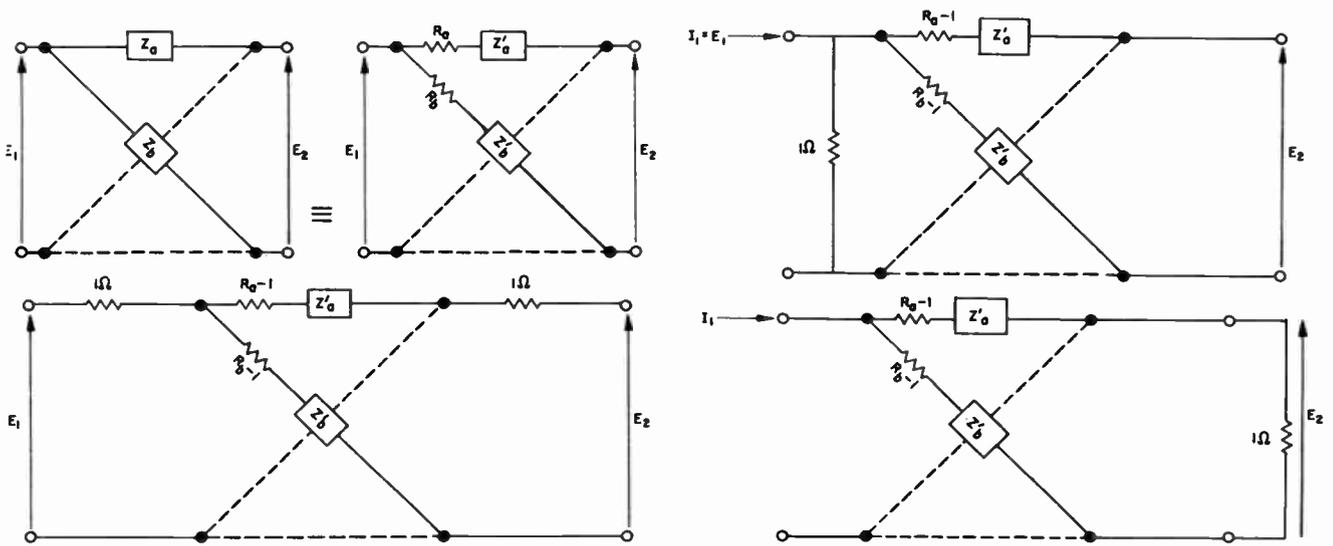
Fig. 3

Steps in the conversion of open-circuited lattice for which $K = E_2/E_1 = p/Hq$ to resistance-terminated lattice with $Z_{12} = E_2/I_1 = p/Hq$. (a) Open-circuited lattice for which $K = E_2/E_1 = p/Hq$. (b) Lattice equivalent to that in (a). (c) Lattice after application of Norton's Theorem. (d) Lattice given by application of reciprocity theorem, where $Z_{12} = E_2/I_1 = p/Hq$.
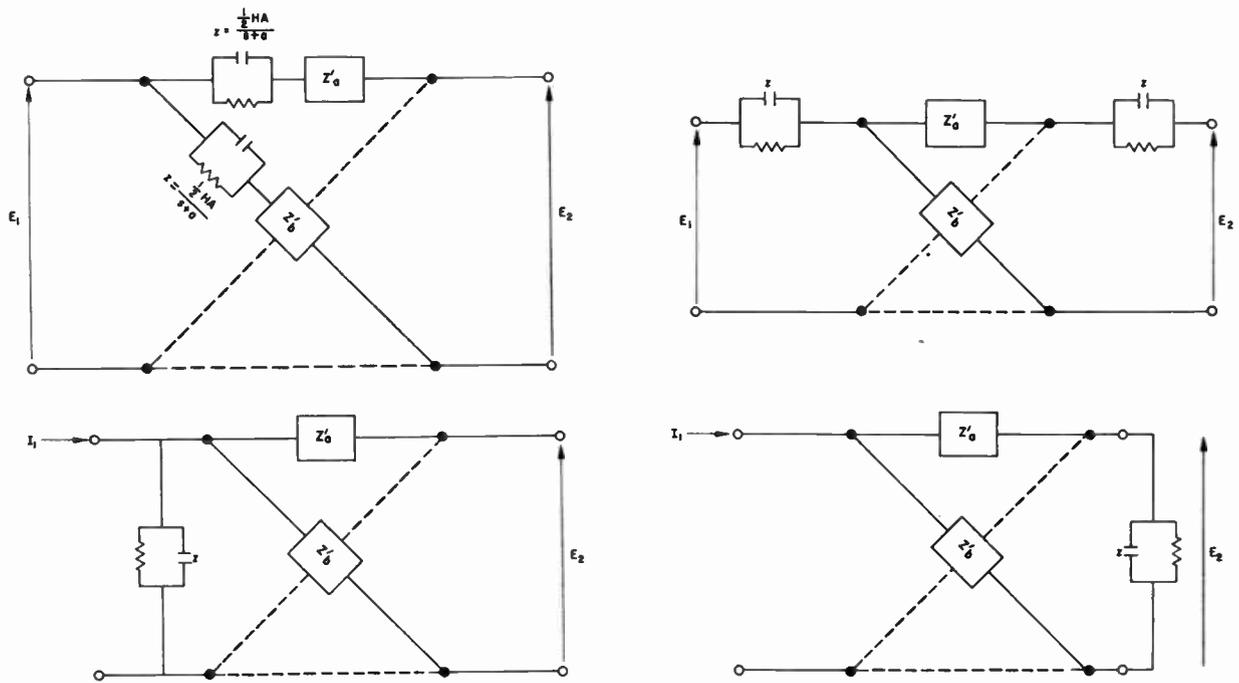
Fig. 4

Steps in the conversion of an open-circuited lattice for which $K = E_2/E_1 = (s+a)p/(Hq)$ to an RC-terminated lattice for which $Z = E_2/I_1 = Ap/(2q)$. (a) Open-circuited lattice for which $K = (s+a)p/(Hq) = E_2/E_1$. (b) Lattice equivalent to that in (a). (c) Lattice after application of Norton's Theorem, where $I_1 = 2(s+a)E_1/(HA)$. (d) Lattice after application of reciprocity theorem with $Z_{12} = E_2/I_1 = (1/2)HAE_2/(s+a)E_1 = (HA/2)K/(s+a) = Ap/(2q)$.

11
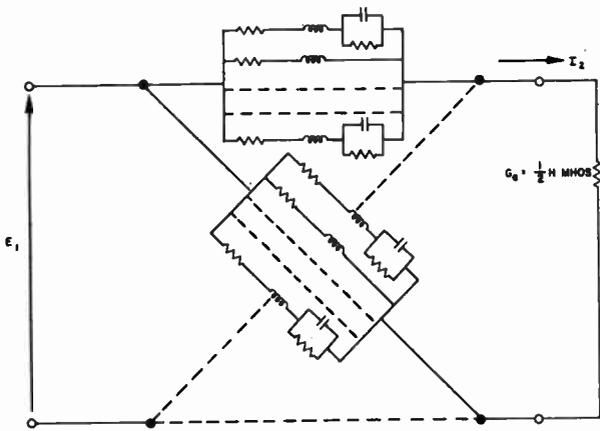
Fig. 5

Lattice for which $Y_{12}=1/2(p/q)$.
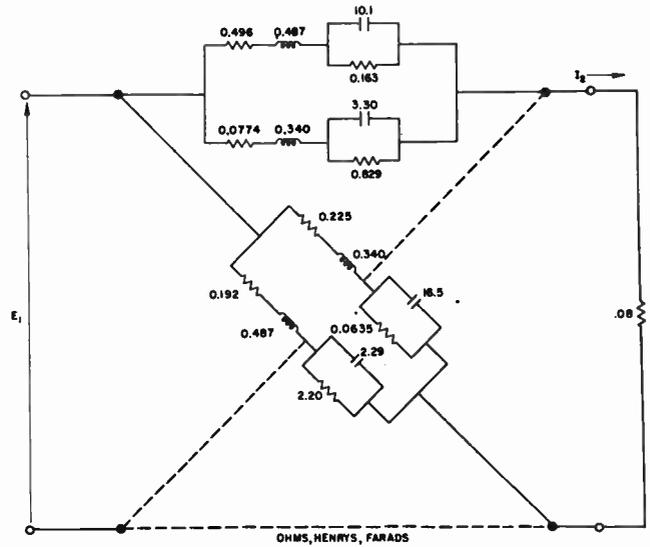
Fig. 6

Network realized in the illustrative example where $Y_{12}=1/2(p/q)$.
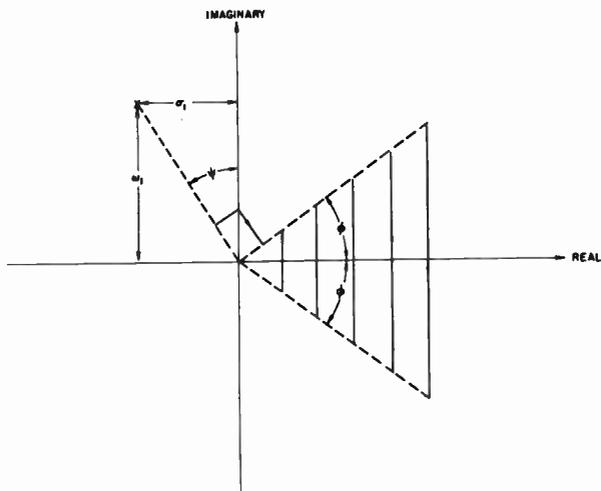
Fig. I.1

Graphical illustration of pole and residue relationship.
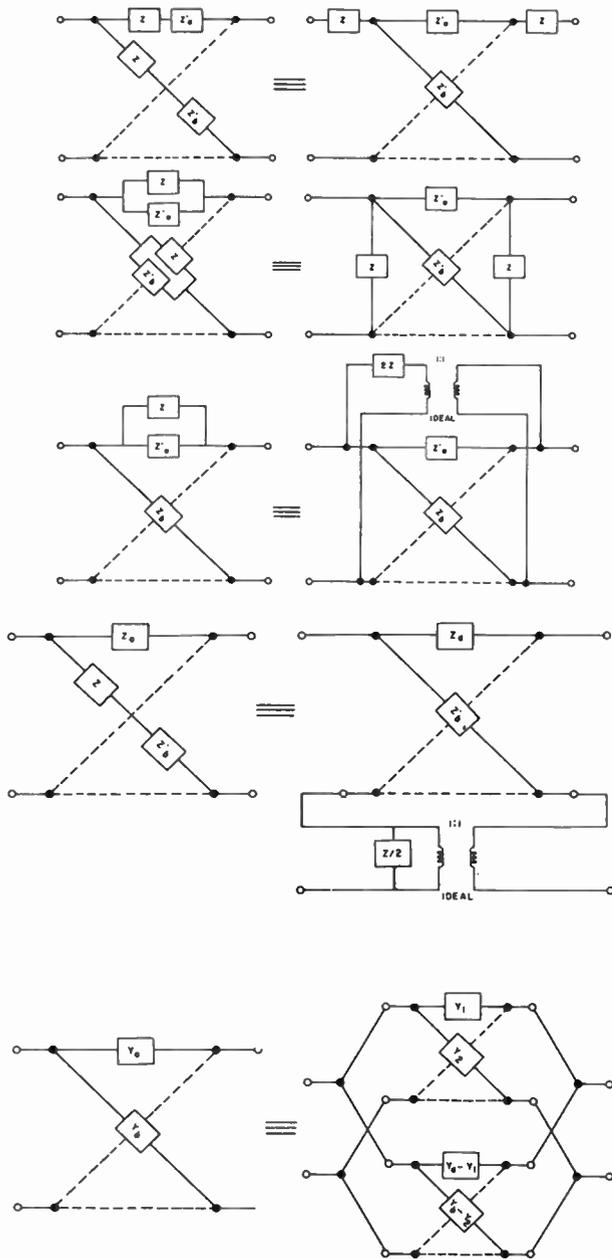
Fig. II.1

Methods for the conversion of lattices to
unbalanced networks.

a) Removal of a series impedance from each arm.

b) Removal of a shunt impedance from each arm.

c) Removal of a shunt impedance from series arm.

d) Removal of a series impedance from the cross
arm.

e) A single lattice decomposed into two lattices in
parallel.
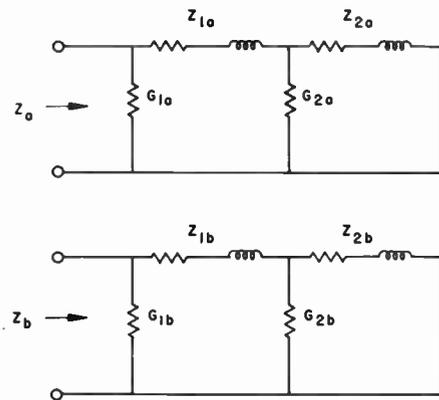
Fig. II.2

Desirable form of lattice arm impedances,
where $G_{na} \geq G_{nb}$ and $Z_{nb} \geq Z_{na}$.
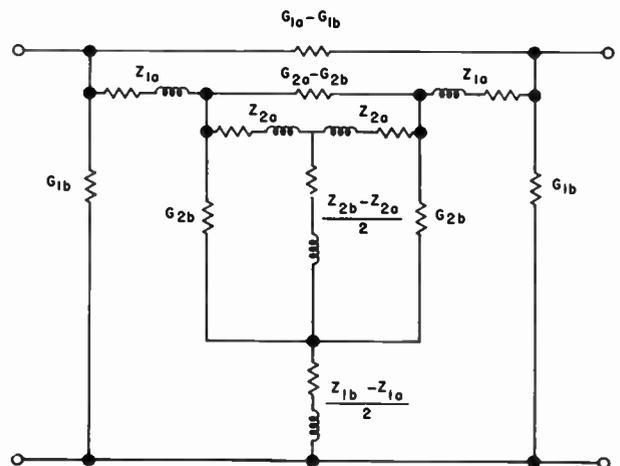
Fig. II.3

Unbalanced network corresponding to lattice
with arms given in Fig. II.2.
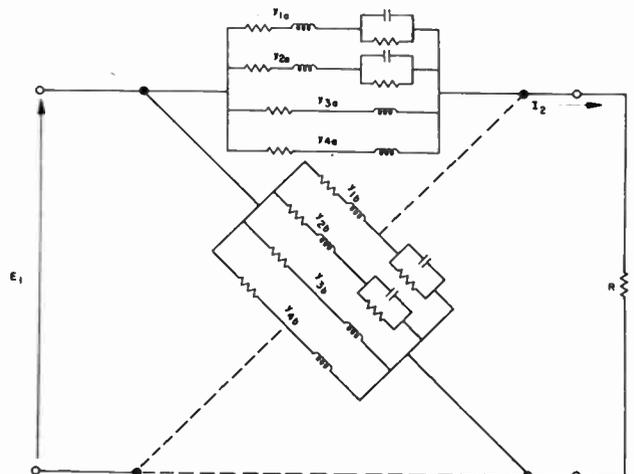
Fig. II.4

Six-pole lattice to be reduced to
unbalanced network.

13
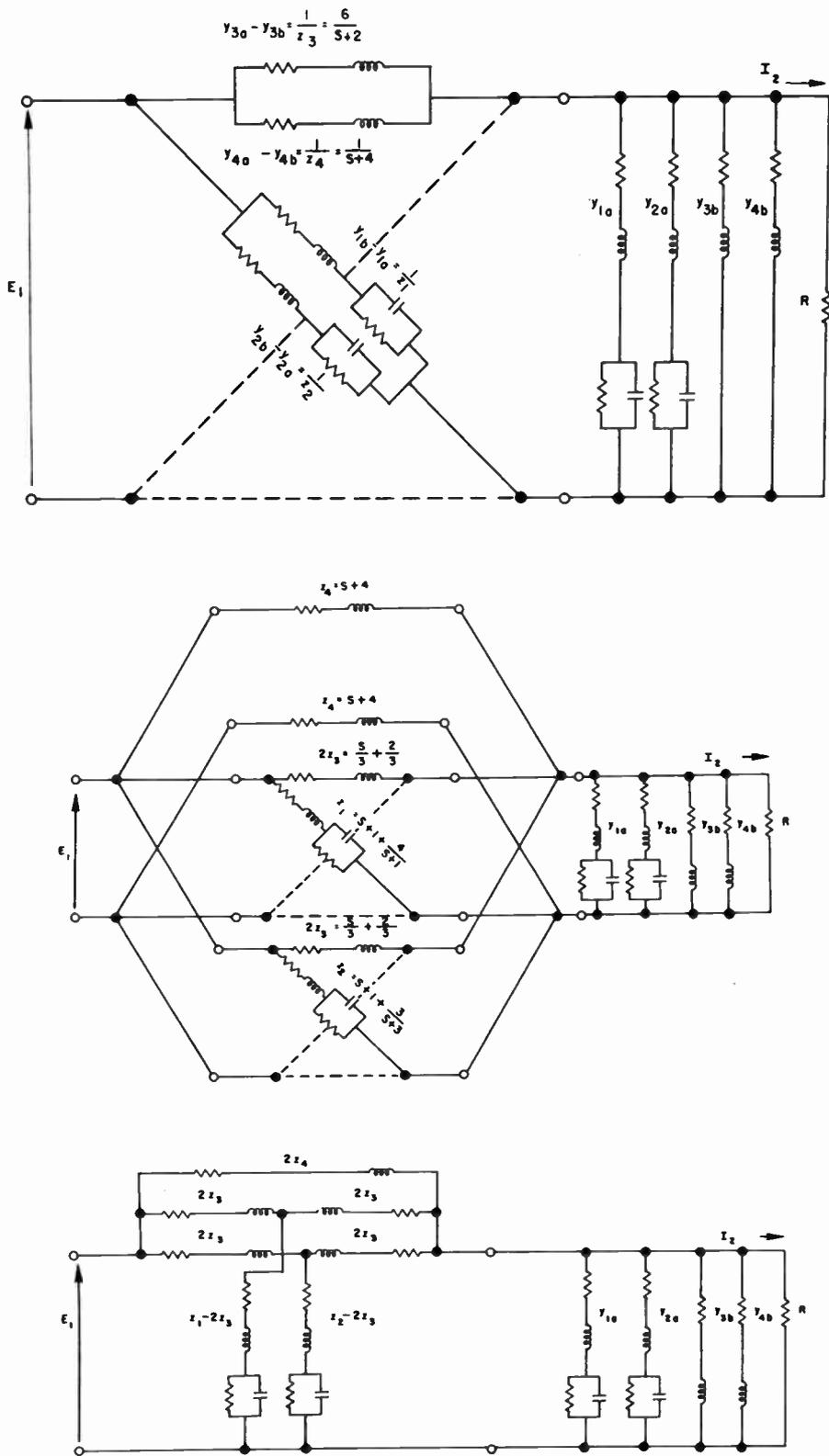
Fig. II.5

Steps in reduction of lattice that contains a sufficiently large residue.
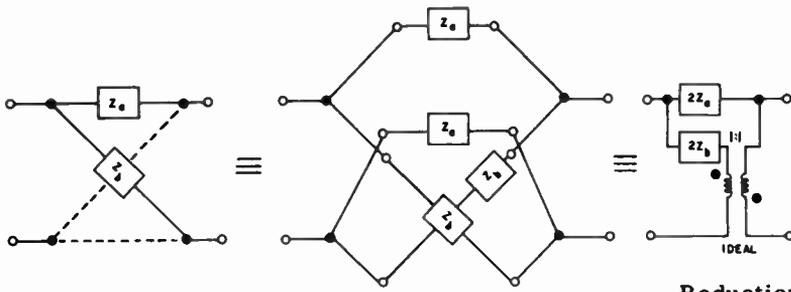
Fig. II. 6

Reduction of a lattice by use of ideal transformer.
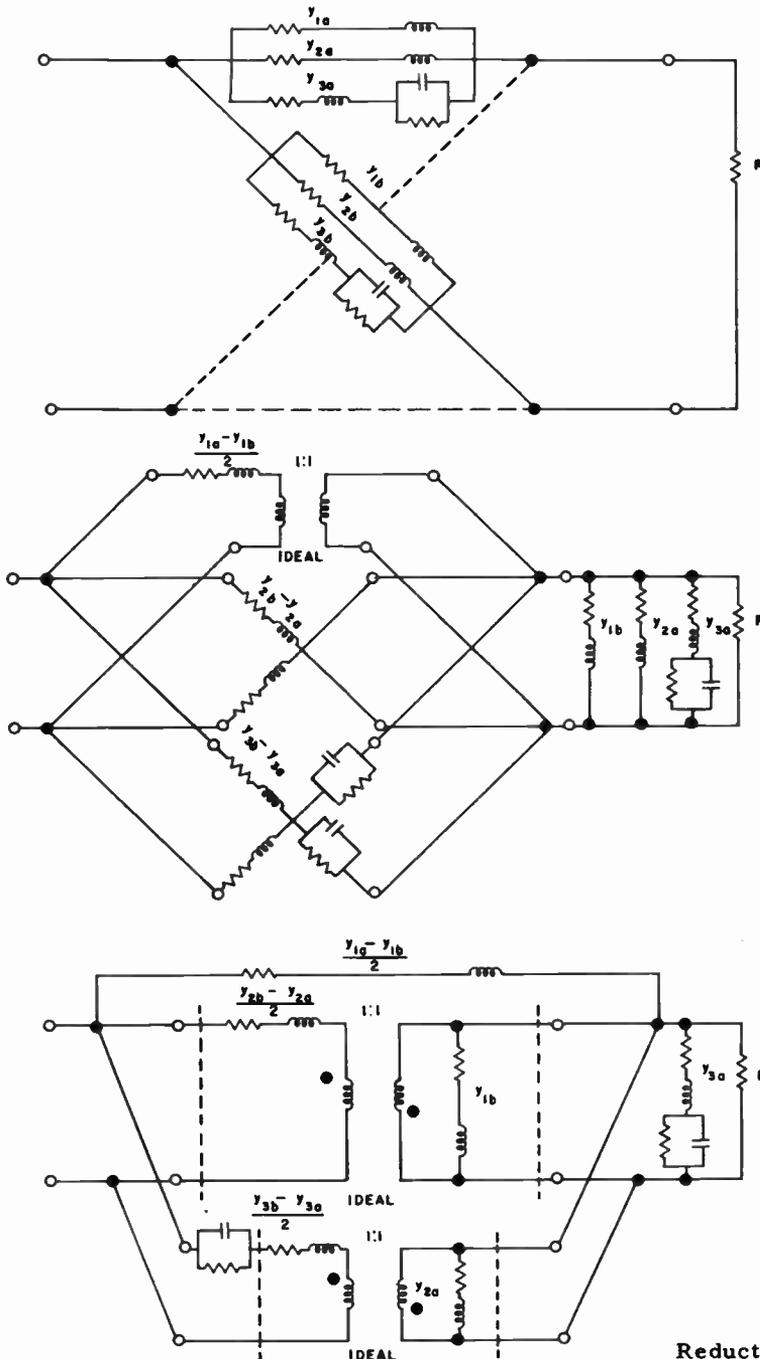
Fig. II. 7

Reduction of a lattice by use of real transformers.
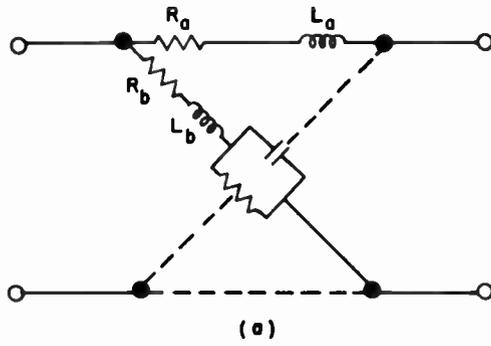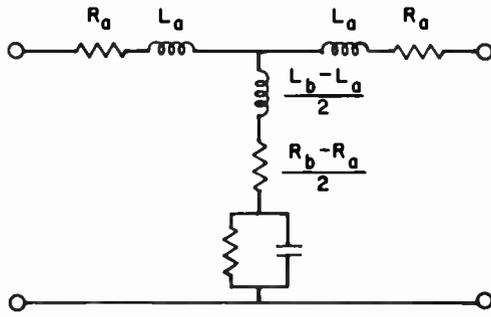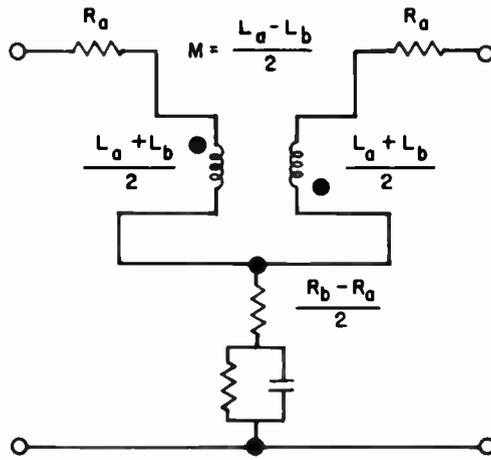
Fig. II. 8

Reduction of a lattice by use of mutual inductance.

A General Theory of Wide-band Matching

by

Richard La Rosa and Herbert J. Carlin

Microwave Research Institute
Polytechnic Institute of Brooklyn
Brooklyn, New York

It is often desirable to terminate a resistive generator in a resistance equal to the generator internal resistance. This is particularly true if the generator represents a length of nearly lossless transmission line because standing waves on the line can be avoided by terminating the line in its characteristic impedance.

When a resistive generator supplies power to an arbitrary complex load impedance, a two terminal pair matching network can be inserted between the generator and load to terminate the generator in an impedance which is almost equal to its internal resistance. When the matching is to be done over a broad band of frequencies, it is often not possible to obtain the desired quality of match with a lossless matching network; a lossy matching network must therefore be used. Lossy matching networks are also used to insert controlled amounts of attenuation for equalization purposes. When lossy matching networks are used, attenuation can be controlled independently of the matching function as long as the power-transfer efficiency is below a certain limit presented in this paper. The power-transfer efficiency is the ratio of power actually reaching the load to the available power of the generator.

The given load is represented by a lossless 4-pole E terminated in a unit resistor. The tandem combination of the matching network D and the lossless network E is a network S as shown in Fig. 1
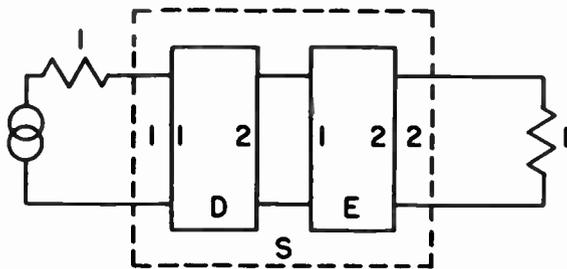
Fig. 1

Matching Network and Part of Load Representation
Combined to Form a Single Network

The network S is characterized by a scattering matrix (S) given by

$$[S] = \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix} \qquad (1)$$

The reflection coefficient of interest is $|S_{11}|$ and the power-transfer efficiency of the matching network is $|S_{12}|^2$. R. M. Fano[1] found the restrictions on these quantities for the case of a lossless matching network and H. J. Carlin and R. LaRosa[2] found in convenient form the limitations on $|S_{12}|^2$ when a lossy matching network makes S_{11} identically zero. R. LaRosa and H. J. Carlin have considered[4] the general case of a dissipative matching network where the input reflection factor may have any desired value.

The lossless network E imposes restrictions on the output reflection coefficient S_{22} which are best expressed as integral formulas involving log $|S_{22}|$ in the integrand. These integrals were tabulated by R. M. Fano[1] using a method shown by H. W. Bode[3]. The integral restrictions essentially establish the lower limit of $|S_{22}|$ over the entire frequency range.

The functions $|S_{11}|$ and $|S_{12}|^2$ are in turn limited by $|S_{22}|$ through the inequality

$$|S_{12}|^2 \leq (1+|S_{11}|)(1-|S_{22}|) \qquad (2)$$

as long as

$$|S_{22}| \geq |S_{11}| \qquad (3)$$

There is no advantage in violating inequality (3), so that inequality (2) controls the upper limit on $|S_{12}|^2$. Inequality (2) shows that for given $|S_{11}|$ and $|S_{22}|$ functions the greatest power-transfer efficiency is obtained when (2) is exactly satisfied as an equality. This condition exists when the matching network contains no more than a single resistor.

Inequality (2) shows that for minimum insertion loss networks (i.e. one resistor) increased input mismatch $|S_{11}|$ will allow increased power transfer $|S_{12}|^2$ up to the point where $|S_{11}|$ equals $|S_{22}|$; at this point the dis-

sipation in the matching network goes to zero and conditions for a lossless matching structure are defined. If $|S_{11}|$ is increased beyond this point another inequality becomes operative and $|S_{12}|^2$ decreases. The maximum power transfer is obtained with a lossless matching network.

The techniques used and results discussed above can also be applied to the matching of an arbitrary generator to a resistive load.

References

[1] R. M. Fano, "Theoretical Limitations on the Broadband Matching of Arbitrary Impedance". Journal of the Franklin Institute, Vol. 249, No. 1, p. 57, Jan. 1950, and Vol. 249, No. 2, p. 139, Feb. 1950.

[2] H. J. Carlin and R. LaRosa, "Broadband Reflectionless Matching with Minimum Insertion Loss". Proceedings of the Symposium on Modern Network Synthesis, Polytechnic Institute of Brooklyn, 1952, p. 161.

[3] H. W. Bode, "Network Analysis and Feedback Amplifier Design". D. Van Nostrand Company, Inc., New York, 1945.

[4] R. LaRosa and H. J. Carlin, "A General Theory of Wide-band Matching with Dissipative 4-Poles". Report No. R-308-53, PIB-247, Microwave Research Institute, Polytechnic Institute of Brooklyn, Feb. 17, 1953. Project Designation 075-215, Contract No. NOnr-292(00).

SYNTHESIS OF ELECTRIC FILTERS
WITH ARBITRARY PHASE CHARACTERISTICS

Byron J. Bennett
Stanford Research Institute
Stanford, California

I Introduction

This paper will deal with an insertion-loss method for the synthesis of electric filters, placing particular emphasis upon approximate realization of specified phase characteristics. It will be shown that:

1. A specified phase characteristic may be approximately realized in a filter network, provided that the approximating phase characteristic can also be realized in an all-pass network.

2. An attenuation characteristic approximating a constant pass-band value in Tschebyscheff equal-ripple fashion can also be realized in the same filter network.

Discussion will be confined primarily to the synthesis of low-pass filters. However, the same general method applies to other filter types.[1] The conformal transformation applied here in the low-pass case is not applicable to band-pass or high-pass filters, but the general philosophy involving the analogous potential problem still applies in these cases. The method to be described will lead directly to filters which have approximately the attenuation and phase characteristics desired, i.e., additional phase-correcting all-pass networks will not be required. A practical design procedure will be outlined first and then proof will be given that the procedure does lead to a network having approximately the attenuation and phase characteristics prescribed.

II Formulation of the Problem

A. The Insertion Voltage Ratio

It is required to design a network to be inserted between R_1 and R_2 in Fig. 1 such that the resulting output voltage (E_2 in Fig. 2) will vary with frequency in a prescribed manner. The network will be designed so that a given insertion ratio $\frac{E_2}{E_{20}}$ is obtained, in which E_2 and E_{20} are the output voltages with and without the network, respectively. (Figs. 1 and 2.) In order that the network may be composed of linear, lumped, and passive elements, the ratio $\frac{E_2}{E_{20}}$ must possess certain well-known properties:

1. It must be a rational function of p, where p is a complex frequency variable. ($p = \sigma + j\omega$ where ω is angular frequency.)

2. Its zeros and poles must occur in conjugate pairs.

3. Its poles must appear only in the left half of the p-plane since they are natural modes.

4. It must be equal to or less than the quantity $\frac{R_1 + R_2}{2\sqrt{R_1 R_2}}$ for all values of ω.

$(\frac{E_2}{E_{20}} = \frac{R_1 + R_2}{2\sqrt{R_1 R_2}}$ when the network matches R_1 and R_2 so that maximum power transfer is obtained.)
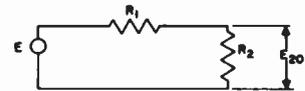
FIG. I
CIRCUIT WITHOUT FILTER NETWORK
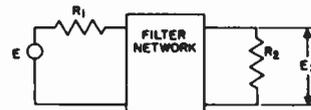
FIG. 2
CIRCUIT WITH FILTER NETWORK

B. An Example of Ideal Filter Characteristics

It is thus necessary to find an approximating function, to be called $W(p)$, which meets all the requirements of Section IIA so that it may be realized as $\frac{E_2}{E_{20}}$, and yet has characteristics which approximate desired transmission characteristics. The function having desired transmission characteristics will be termed $W'(p)$.

For purposes of illustration it will be assumed that $W'(p)$ has the characteristics of an ideal filter[2,3] (Fig. 3):

1. Its magnitude shall be a constant, unity, in the pass band which shall extend over the normalized frequency range, $-1 < \omega < 1$.

2. Its magnitude for all other values of ω shall approach zero.

3. The phase associated with $W'(p)$ shall be a linear function of ω over the frequency range $-1.43 < \omega < 1.43$.

19

It has long been known that characteristics such as those shown in Fig. 3 cannot be realized exactly by a $W(p)$ which meets the requirements enumerated in Section IIA. Accordingly, such characteristics are only approximated by a given $W(p)$, and then a network which realizes the given $W(p)$ is found by standard realization techniques.
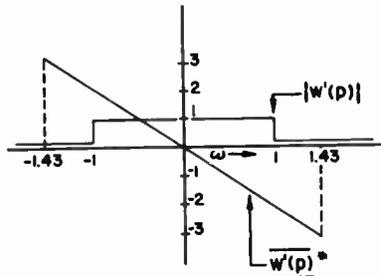
FIG. 3
MAGNITUDE AND PHASE OF W'(p) VERSUS ω
(σ = 0)

* Multiply vertical scale by 2 to get radians.

C. The Proposed Method

The method proposed for obtaining a proper $W(p)$ is as follows:

1. The phase-frequency characteristic desired is approximated by means of a rational fraction $F_{ap}(p)$, which possesses all the characteristics necessary for realization as $\frac{E_2}{E_{20}}$ of an all-pass network, i.e., $F_{ap}(p)$ meets all conditions of Section IIA, and its magnitude-frequency characteristic is constant for all ω. This approximation may be accomplished by any one of a number of methods. One of the best existing methods, making use of Tschebyscheff polynomial series, has been proposed by Darlington.[4]

2. The pass-band portion of the prescribed magnitude-frequency filter characteristic is then approximated in a Tschebyscheff equal-ripple fashion without altering the phase characteristic. The steps outlined in the following sections deal specifically with this aspect of the problem.

III Practical Procedure for Obtaining a Proper Approximating Rational Fraction W(p) from a Given All-Pass Type Rational Fraction $F_{ap}(p)$

A. Introduction

In this section the necessary steps for obtaining a given approximating filter function $W(p)$ will be outlined, assuming that an all-pass rational fraction $F_{ap}(p)$ has already been obtained. The procedure will be divided into two parts. The first part will deal with obtaining a rational fraction $G(p)$ whose magnitude is a Tschebyscheff

equal-ripple characteristic in the pass band; proof for the steps involved in this first part will be given in Section IV. The second part will deal with obtaining a $W(p)$ from a given $G(p)$.

Thus, the design procedure will begin with an all-pass rational fraction,

$$F_{ap}(p) = \frac{N(-p)}{N(p)} , \qquad (1)$$

which has a phase-frequency characteristic approximating the desired phase-frequency characteristic. In Eq. 1, $N(p)$ is a Hurwitz polynomial, i.e., its zeros are in the left half of the p-plane. The pole-zero arrangement for an $F_{ap}(p)$ having three poles and three zeros is shown in Fig. 4. The objective will be to find an approximating rational fraction, $W(p) = K \frac{M(p)}{[N(p)]^2}$, in which $M(p)$, because of quadrantal arrangement of its zeros in the p-plane, will have a constant phase-frequency characteristic in the pass band.*
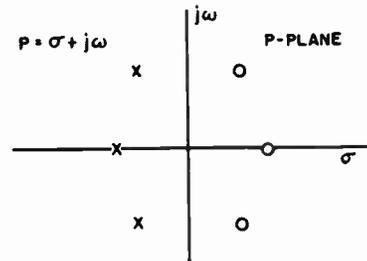
FIG. 4
POLES AND ZEROS OF AN ALL-PASS
RATIONAL FRACTION $F_{ap}(p)$

Thus, $W(p)$ will have the same phase characteristic as $F_{ap}(p)$, and if $M(p)$ is properly chosen the magnitude of $W(p)$ will be a Tschebyscheff equal-ripple approximation of the pass-band magnitude-frequency characteristic of $W'(p)$. The constant K is real and should be so chosen that
$$W(p)]_{p=j\omega} \leq \frac{R_1 + R_2}{2\sqrt{R_1 R_2}} \text{ for all } \omega. \text{ A prac-}$$
tical consideration sometimes further governs the choice of K. This will be discussed in Section V.

B. Steps in Obtaining G(p) from $F_{ap}(p)$

An initial objective will be the determination of a rational fraction

$$G(p) = \frac{[L(p)]^2}{AN(p)N(-p)} , \qquad (2)$$

---

* The pole-zero arrangement of a $W(p)$ obtained for the illustrative example of Section VI is shown in Fig. 7. For another $W(p)$, the zeros might be complex but they must always occur in quadrantal symmetry.

where $L(p)$ is a polynomial whose zeros occur in the pass band and the real constant A is chosen so that $G(p)]_{p=j\omega}$ varies between limits of 0 and 1 in the pass-band region. After $G(p)$ is determined, a few simple steps will lead to $W(p)$. The polynomial $L(p)$ may be considered to be a generalized Tschebyscheff polynomial whose coefficients depend upon the location of zeros of $N(p)$. It will be interesting to note that if $AN(p)N(-p)$ is a constant equal to unity, $L(p)$ will be a Tschebyscheff polynomial.[*]

The procedure for determining $G(p)$ is:

1. The quantities Z_1, Z_2, Z_3....Z_n are found by making use of the equation

$$Z_n = \frac{(p_n^2 + 1)^{\frac{1}{2}}}{p_n} , \qquad (3)$$

where the p_n's are the locations of the zeros of $N(p)$ [or the zeros of $N(-p)$]. In the solution of this equation only the Z_n's with the positive real parts are used.

2. The even part of the polynomial
$P(Z) = \prod_{i=1}^{i=n} (Z+Z_i)$ is transformed back into the p-plane, again using the transformation
$Z = \frac{(p^2 + 1)^{\frac{1}{2}}}{p}$. A rational fraction $U(p) = \frac{S(p)}{T(p)}$
will be obtained from this step.

3. The polynomial $S(p)$ will have either
 (a) the same number of zeros as $N(p)$ when n is even, or
 (b) one less than the number of zeros in $N(p)$ when n is odd.
The polynomial $L(p)$ must have the same number of zeros as $N(p)$. Thus if (a) applies, $L(p) = S(p)$, and if (b) applies, $L(p) = pS(p)$.

4. The square of $L(p)$ found in step No. 3 will form the numerator of $G(p)$, and $N(p)N(-p)$ will be its denominator. The real constant A is now chosen so that $G(p)$ will have a magnitude-frequency characteristic which oscillates between 0 and 1 in Tschebyscheff fashion in the pass band, i.e., $-1 < \omega < 1$.

In Fig. 5 the poles and zeros of a typical $G(p)$ are shown. If the $G(p)$ having these particular poles and zeros is obtained from the $F_{ap}(p)$ having poles and zeros shown in Fig. 4, then the pole locations of $G(p)$ are the same as the locations of the poles and zeros in $F_{ap}(p)$. The zeros of $G(p)$ are always located in the pass band on the jω-axis. Now $G(p)$ does not have all the characteristics which $\frac{E_2}{E_{20}}$ must have. It has properties No. 1 and No. 2 tabulated in Section IIA, but not property No. 3, and perhaps not property No. 4. The magnitude of $G(p)$ does vary between limits of 0 and 1 in the pass band, and it will be easy to obtain a $W(p)$ from $G(p)$. The next objective will be determination of $W(p)$, which will have:

1. The significant properties of $G(p)$,
2. Property No. 3 and property No. 4,
3. A magnitude-frequency characteristic which approximates a constant value in the pass band within a specified tolerance.
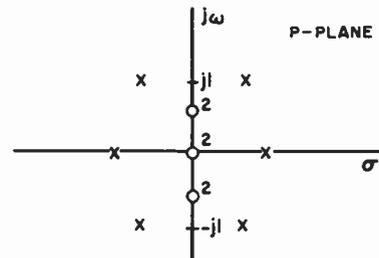
FIG. 5
POLES AND ZEROS OF THE
RATIONAL FRACTION $G(p)$

C. Steps in Obtaining $W(p)$ from $G(p)$

1. If the ratio of the maximum magnitude of $W(p)$ in the pass band to the minimum in the pass band is r, $G(p)$ may be modified to achieve this ratio by subtracting a positive real constant B from $F(p)$, i.e.,

$$H(p) = G(p) - B = \frac{M(p)}{N(p)N(-p)} , \qquad (4)$$

where the constant B is determined from the equation[**]

$$B = \frac{r}{r - 1} . \qquad (5)$$

Subtracting the constant B from $G(p)$ does not affect any of the necessary characteristics which $G(p)$ possesses but obtains the specified ratio of the maximum magnitude of $H(p)$ to the minimum magnitude of $H(p)$ in the pass band. This may be seen from Fig. 6 in which a magnitude-frequency plot of a typical $G(p)]_{p=j\omega}$ is shown. It may be noted that if the real positive constant B is greater than unity, then, when it is subtracted from $G(p)]_{p=j\omega}$ which is also real and positive for all ω, no zeros of the resulting $H(p)$ occur in the

---

[*] $AN(p)N(-p)$ would never be equal to unity in the the application discussed here since this would mean that all natural modes would occur at infinite frequency. The statement was made merely to indicate the relationship between $L(p)$ and Tschebyscheff polynomials.

[**] Alternatively, B may be chosen to obtain a given rate of cutoff.

pass band. The reason for subtracting B from $G(p)$ instead of adding is also easily seen from Fig. 6. Subtraction of B results in a low stop-band value for the absolute magnitude of $H(p)]_{p=j\omega}$, whereas addition of B would lead to a higher stop-band than pass-band value.
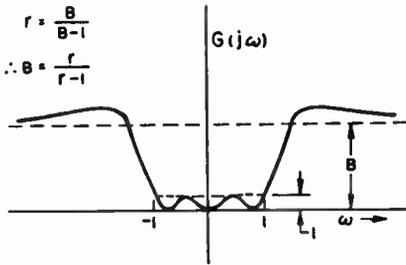
FIG. 6
CURVE OF $G(j\omega)$ VERSUS ω SHOWING EFFECT OF SUBTRACTION OF A REAL CONSTANT FROM $G(p)$

2. In order to obtain a function which possesses property No. 3, $H(p)$ is modified to form a new rational fraction $Q(p)$ by removing the poles in the right half of the p-plane and doubling the poles in the left half of the p-plane. Thus,

$$Q(p) = \frac{M(p)}{[N(p)]^2} . \quad (6)$$

Therefore, $Q(p)$ meets all conditions except No. 4 and, in addition, its magnitude has a specified r in the pass band.

3. In order to meet condition No. 4., i.e., $W(p)$ must be less than $\frac{R_1 + R_2}{2\sqrt{R_1 R_2}}$ for all ω, $Q(p)$ is merely multiplied by an appropriate constant K. Thus,

$$W(p) = K\frac{M(p)}{[N(p)]^2} . \quad (7)$$

In this manner, a $W(p)$ which approximates the characteristics desired for $\frac{E_2}{E_{20}}$ and, in addition, meets all necessary conditions of $\frac{E_2}{E_{20}}$ has been obtained (Fig. 7).

IV Proof That the Steps Outlined in Section III Produce a $G(p)$ Possessing an Equal-Ripple Magnitude Characteristic in the Pass Band

The procedure in this section is directed toward determination of a $G(p)$ from a given $F_{ap}(p)$ as outlined in Section III. However, the procedure is more detailed in order to show that $G(p)$ indeed has a Tschebyscheff characteristic in the pass band. The reason for inclusion of some of the steps may not be apparent until the end. At

each step a new function is formed using some of the properties of the function in the previous step.
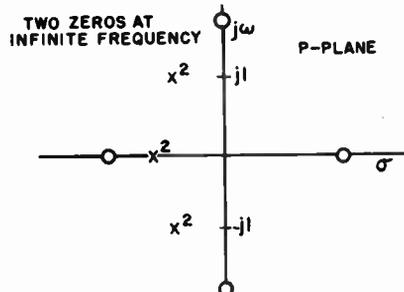
FIG. 7
POLES AND ZEROS OF A TYPICAL $W(p)$

A. The Rational Fraction $F_x(p)$

Suppose a rational fraction $F_x(p)$ is formed from a given $F_{ap}(p)$ by replacing the zeros of $F_{ap}(p)$ by poles. Thus,

$$F_x(p) = \frac{1}{N(p)N(-p)} . \quad (8)$$

The zeros of this function are located at infinite frequency. In Fig. 8 the pole locations of the $F_x(p)$ function corresponding to the all-pass pole and zero locations of Fig. 4 are shown.
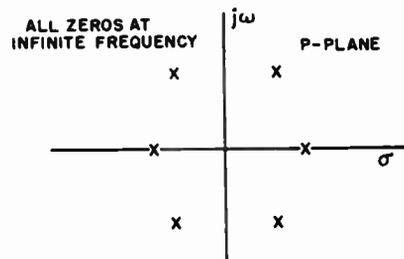
FIG. 8
POLES AND ZEROS OF $F_x(p)$

B. The Rational Fraction $\epsilon^{\varphi(Z)}$

Now a $G(p)$ which approximates a constant in the pass band, $-1 < \omega < 1$, is to be obtained. Therefore, some means should be taken to alter the p-plane of Fig. 8 so that a filter-like magnitude-frequency characteristic is obtained. A step in this direction is now attempted by introducing the condition that the potential be constant over the normalized pass-band region. Mathematically, this may be done by considering the analogous potential problem of Fig. 9 in which unit positive charges

22

are at the pole locations and all negative charge is distributed on a conducting surface in the pass band. The transformation $Z = \frac{(p^2 + 1)^{\frac{1}{2}}}{p}$ will transform the pass band and hence the conducting plate of Fig. 9 into the entire imaginary axis of the Z-plane of Fig. 10.
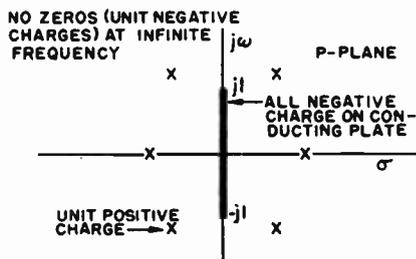
FIG. 9
ANALOGOUS POTENTIAL PROBLEM ASSUMING CONSTANT MAGNITUDE (SCALAR POTENTIAL) IN THE PASS BAND
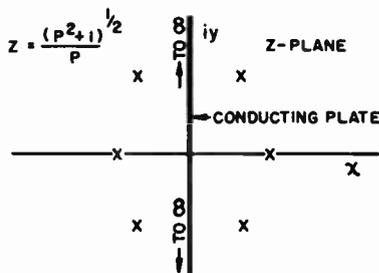
FIG. 10
POTENTIAL PROBLEM AS SEEN IN THE Z-PLANE

Since two values of Z exist for each value of p, two Riemann surfaces are needed to describe the Z-plane. Only one sheet in the Z-plane is essential to the potential problem, however. The complex Z-plane potential, $\varphi(Z)$, is now evaluated in the right half of the Z-plane using the method of images, i.e., the conducting plate is removed and the poles in the left half of the Z-plane are replaced by zeros. The rational fraction $\varepsilon^{\varphi(Z)}$ may be written:

$$\varepsilon^{\varphi(Z)} = \prod_{i=1}^{i=n} \left(\frac{Z + Z_i}{Z - Z_i} \right) , \tag{9}$$

where the real parts of Z_1, Z_2....Z_n are all positive and $Z_i = \frac{(p_i^2 + 1)^{\frac{1}{2}}}{p_i}$.

The entire imaginary axis of the Z-plane corresponds to the pass band of the p-plane. It is next convenient to perform a few mathematical manipulations in the Z-plane before returning to the p-plane.

C. The Rational Fraction f(Z)

The magnitude of the rational fraction $\varepsilon^{\varphi(Z)}$ is absolutely constant on the imaginary axis in the Z-plane, and thus if $\varepsilon^{\varphi(Z)}$ is transformed back into the p-plane, its magnitude in the pass band will be absolutely constant. However it will be impossible to obtain an absolutely constant magnitude for $\frac{E_2}{E_{20}}$ with a finite number of lumped circuit elements. Accordingly, steps will now be taken to modify $\varepsilon^{\varphi(Z)}$ in order to obtain a function whose magnitude on the imaginary axis varies about a constant value in equal-ripple fashion.[5]

Since the magnitude of $\varepsilon^{\varphi(Z)}$ is equal to unity on the imaginary axis of the Z-plane, and since its phase is a monotone function of the imaginary component of Z, then on the imaginary axis of the Z-plane the magnitude of the function $\varepsilon^{\varphi(Z)} + (-1)^n$ varies between 0 and 2. The magnitude of the function f(Z), defined by

$$f(Z) = \frac{\varepsilon^{\varphi(Z)} + (-1)^n}{2} , \tag{10}$$

varies between 0 and 1 in equal-ripple fashion. It should be noted that the numerator of f(Z) is the even part of the polynomial P(Z) needed in step No. 2 of Section IIIB.

D. The Rational Fraction g(Z)

In order to obtain a rational fraction in the p-plane, the following function of Z should be formed:

$$g(Z) = f(Z) f(-Z) . \tag{11}$$

The magnitude of this function also varies between 0 and 1 in equal-ripple fashion along the imaginary axis of the Z-plane. When

$$G(p) = g(Z)]_{Z = \frac{(p^2 + 1)^{\frac{1}{2}}}{p}} \tag{12}$$

is obtained, its magnitude will vary between the same limits in the pass band. The zeros of g(Z) appear on the imaginary axis of the Z-plane and are double. Therefore the zeros of G(p) appear in the pass band and are double. The poles of G(p) are, of course, the zeros of N(p)N(-p). In Fig. 5 the poles and zeros for the rational fraction G(p) corresponding to the all-pass pole-zero configuration of Fig. 4 are shown.

V Realization Methods

Once a W(p) which meets all necessary and sufficient conditions has been obtained, the network may be synthesized by any one of a variety of

methods. For instance, constant-resistance lattice sections in tandem connections might be used to realize a given $W(p)$. In some cases it may be desirable to isolate sections of the network by means of vacuum tubes. In many cases it may be convenient and desirable to realize the network as a reactance four-pole. If reactance four-pole realization is desired, it may always be obtained by methods introduced by Norton or Darlington.[6]

If the network is to consist of pure reactances, two properties in addition to the four properties of $\frac{E_2}{E_{20}}$ listed in Section IIA are necessary in order to avoid ideal transformers. These additional properties are:

1. If $\frac{E_2}{E_{20}} \neq 0$ at zero frequency, then $\frac{E_2}{E_{20}}$ must be equal to unity at zero frequency.

2. If $\frac{E_2}{E_{20}} \neq 0$ at infinite frequency, then $\frac{E_2}{E_{20}}$ must be equal to unity at infinite frequency. This means that in many low-pass filter cases, B must be chosen to produce a $W(p)$ which is zero at infinite frequency, rather than to produce a given r as discussed in Section IIIC.

If these adjustments can be tolerated, i.e., if r meets specifications, practical reactance four-pole realization without ideal transformers is in many cases possible even though zeros of $W(p)$ are complex.

VI A Design Example

A filter which has magnitude-frequency and phase-frequency characteristics which approximate those shown in Fig. 3 will now be designed. The reactance four-pole to be obtained will operate between a constant-current source and a 1-ohm resistance.

The zeros of $N(-p)$, obtained by the Tschebyscheff polynomial series method are:

$$p_1 = 0.57280 ,$$
$$p_2 = 0.51679 + j1.10250 ,$$
$$p_3 = \overline{p_2} = 0.51679 - j1.10250 .$$

The rational fraction $F_{ap}(p) = \frac{N(-p)}{N(p)}$ has the phase-frequency characteristic plotted in Fig. 11. The steps outlined in Section IIIB will now be followed:

1. From Eq. 3:

$$Z_1 = 2.01194 ,$$
$$Z_2 = 0.81789 - j0.31693 ,$$
$$Z_3 = \overline{Z_2} = 0.81789 + j0.31693 .$$

2. The even part of $P(Z)$ is $3.64772Z^2 +$

1.54797 and, when the transformation $Z = \frac{(p^2+1)^{\frac{1}{2}}}{p}$ is used, the corresponding p-plane rational fraction $U(p)$ is

$$U(p) = \frac{5.19569p^2 + 3.64772}{p^2} .$$

3. Thus,

$$S(p) = 5.19569\,p^2 + 3.64772 ,$$
$$L(p) = pS(p) = 5.19569p^3 + 3.64772\,p.$$

4. Finally,

$$G(p) = \frac{[L(p)]^2}{AN(p)N(-p)} ,$$
$$= \frac{26.995p^6 + 37.905p^4 + 13.306p^2}{A[p^6 + 1.5688p^4 + 1.5757p^2 - 0.72117]} .$$

If A is now chosen so that $G(j1) = 1$, then
$$G(p) = \frac{26.995p^6 + 37.905p^4 + 13.306p^2}{1.3866p^6 + 2.1753p^4 + 2.1849p^2 - 1}$$
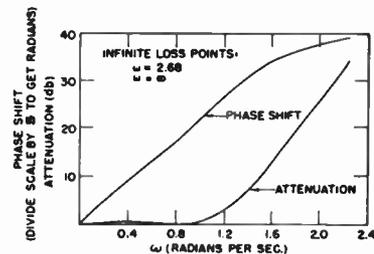
FIG. 11
PHASE AND ATTENUATION CHARACTERISTICS FOR A TYPICAL FILTER NETWORK

This $G(p)$ meets conditions No. 1 and No. 2. In addition its magnitude varies in Tschebyscheff manner between the limits of 0 and 1 in the pass band, i.e., $-1 < \omega < 1$. Steps outlined in Section IIIC for obtaining $W(p)$ from $G(p)$ will now be followed.

1. From Eq. 5, B might be obtained for a given r. However, as a step necessary in order to avoid use of an ideal transformer in this particular case, B is chosen so that $H(p) = 0$ at infinite frequency. Thus

$$B = 19.4685 ,$$

and from Eq. 5

$$r = 1.05415 .$$

From Eq. 4

$$H(p) = \frac{-4.445p^4 - 29.231p^2 + 19.4685}{1.3866p^6 + 2.1753p^4 + 2.1849p^2 - 1} .$$

2. From Eq. 6

$$W(p) = \frac{-4.445p^4 - 29.231 p^2 + 19.4685}{1.3866p^6 + 4.4549p^5 + 9.3316p^4 + 11.537p^3 + 9.75125p^2 + 4.8859p + 1}.$$

3. Another step necessary to avoid use of an ideal transformer is to choose k in Eq. 7 so that $W(o) = 1$. With k so chosen, condition No. 4 is still met since $\lim_{R_1 \to \infty} \frac{R_1 + R_2}{2\sqrt{R_1 R_2}} = \infty$, and $W(p)$ will certainly be less than infinity at all real frequencies. Thus, Eq. 7 becomes

$$W(p) = \frac{-0.22832p^4 - 1.50145p^2 + 1}{1.3866p^6 + 4.4549p^5 + 9.3316p^4 + 11.597p^3 + 9.75125p^2 + 4.8859p + 1}$$

In Fig. 11 curves of attenuation and phase shift vs. frequency for this $W(p)$ are shown. The attenuation characteristic was obtained from the formula:

$$\text{Attenuation (db)} = 20 \log_{10} \left| \frac{1}{W(p)} \right|.$$

The network which realizes $W(p)$ is shown in Fig. 12.
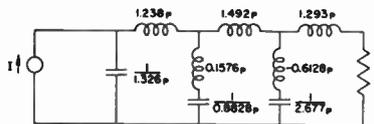
FIG.12
CIRCUIT OF LINEAR-PHASE ELECTRIC FILTER

The particular network shown is designed to operate between an infinite-impedance constant-current source and a 1-ohm resistance. However, a network which realizes this same $W(p)$ could have been designed to operate between finite resistances. The resistances might even have been equal, i.e., $\frac{R_1 + R_2}{2\sqrt{R_1 R_2}} = 1$, since for all ω the highest absolute magnitude of $W(p)$ is unity.

VII Design Variations

Discussion of design procedures has been formally restricted to low-pass filters. Although the same general approach may be used for design of other filter types, the conformal transformation is different for each type. For instance, for the high-pass filter the transformation is $Z = (p^2 + 1)^{\frac{1}{2}}$.

The design procedure outlined always results in a $W(p)$ having double poles, and if any zeros of $W(p)$ occur at infinite frequency, they must be double. Many practicable cases, notably many vacuum tube interstages, require an odd number of

zeros at infinite frequency. In order to obtain an approximating function $W_o(p)$ which has an odd number of zeros at infinite frequency, the following procedure, resulting in a $W_o(p)$ with a pass-band magnitude-frequency characteristic whose ripples about a constant value are nearly (but not exactly) equal, may be followed:

1. A proper $W(p)$ is obtained.

2. A polynomial formed from those zeros which occur in the finite p-plane is approximated by means of a polynomial having double zeros. This may be done by means of Tschebyscheff polynomial series. The approximating polynomial may have less zeros than the polynomial to be approximated. This will mean more zeros of $W_o(p)$ at infinite frequency. In any case, all zeros will be double after this approximation. Since the poles are already double, a rational fraction $W_o(p)$, having no double poles and an odd number of zeros at infinite frequency, will be obtained when the square root is taken. The function $W_o(p)$ will have one-half the phase of $W(p)$.

VIII CONCLUSION

Filters having the same phase characteristics as a given all-pass transfer function may be synthesized using the method outlined in this paper. The extent of the pass band may be arbitrarily chosen, and if no zeros of transmission are required at infinite frequency, then either an equal-ripple approximation to constant magnitude in the pass band may be made as close as desired, or a given rate of cutoff may be chosen. However, if transmission zeros are required at infinite frequency, the size of the pass-band ripples and the rate of cutoff of the magnitude-frequency characteristic depend upon the bandwidth of the filter, i.e., the size of the ripples and the rate of cutoff will be smaller for smaller bandwidths. The general method involving the analogous potential problem is applicable to filter types other than low-pass. The chief differences will be in the conformal transformation.

In order to obtain filters whose pass-band magnitude-frequency characteristic is an equal-ripple approximation of a constant value, double natural modes are required. Also, when these filters must have transmission zeros at infinite frequency, an even number of them is obtained. However, by means of a design variation described in Section VII, filters may be synthesized which have single natural modes and an odd number of transmission zeros at infinite frequency. The pass-band magnitude-frequency characteristic of these filters will be a close, although not exactly equal-ripple, approximation to a constant value.

References

1. Bennett, B.J., Linear-Phase Electric Filters, Stanford University Electronics Research Laboratory Report No. 43, ONR Contract N6 ONR-251 (NR-078-360) Task No. 7, (February 14, 1952).

2. Bode, H.W., and Dietzold, R.L., "Ideal Wave Filters." Bell System Technical Journal, Vol. XIV (April, 1935), 215-252.

3. Fredendall, G.L., and Kennedy, R.C., "Linear Phase Shift Video Amplifiers." RCA Review, Vol. XI (Sept. 1950), 418-430.

4. Darlington, S., "Network Synthesis Using Tschebyscheff Polynomial Series." Bell System Technical Journal, Vol. XXXI (July, 1952), 613-665.

5. Moore, A.D., Synthesis of Distributed Amplifiers for Prescribed Amplitude Response, Stanford University Electronics Research Laboratory Report No. 53, ONR Contract N6 ONR-251 (NR-078-360), Task No. 7, (September 1 1952).

6. Darlington, S., "Synthesis of Reactance 4-Poles." Journal of Mathematics and Physics. Vol. XVIII, (September, 1939), 257-353.

WIDE-BAND FILTER AMPLIFIERS
AT ULTRA-HIGH-FREQUENCIES

BY

W. A. Christopherson, IBM Research & Development
Laboratory, San Jose, California
D. O. Pederson, Electronics Research Laboratory,
Stanford University, Stanford, California
J. M. Pettit, Electronics Research Laboratory,
Stanford University, Stanford, California

Abstract

This paper describes the development of a stagger-tuned, band-pass amplifier at ultra-high-frequency using grounded-grid triodes and having a prescribed gain-magnitude response. For the first time, the concepts of stagger-tuning have been extended to grounded-grid cascades so that triodes designed for ultra-high-frequency operation can be effectively utilized to obtain amplifiers having large gain-bandwidth products and low noise-figures.

The amplifier described uses high performance, disc-seal triodes with special four-terminal interstages. The special interstages accomplish the impedance transformation needed for grounded-grid amplifiers and in addition have a frequency characteristic which is approximately that of a single-tuned circuit. Because of this frequency characteristic, stagger-tuning is employed to conserve gain-bandwidth product; however, the nature of the characteristic requires the use of novel low-pass to band-pass transformations to determine the correct interstage tunings for a "maximally-flat" type of response. The feedback effects occurring between adjacent interstages are incorporated in the alignment and tuning procedure.

The design and performance of an amplifier having a gain of 50 decibels, a bandwidth of 50 megacycles, and a maximally-flat gain characteristic will be presented.

Text

If one is to build high-gain band-pass amplifiers for receiver or carrier applications with center frequencies in the ultra-high-frequency region, the use of pentodes and conventional techniques of stagger-tuning is ruled out. This paper describes the use of high performance triodes in a modified stagger-tuning scheme, in order to achieve 50 decibels of gain at 400 megacycles with a filter-type amplitude-frequency response, 50 megacycles wide, together with a relatively low noise-figure of 4 decibels.

It is seen immediately that the frequency specifications make the task a difficult one. The center frequency requirement places the problem in a region between lumped elements for lower frequencies and transmission line or distributed elements for the higher frequencies. The bandwidth requirement calls for efficient interstage networks and high performance electron tubes. The traveling-wave tube, which is an excellent very-wide-band amplification component, is not well suited because of its relatively high noise-figure, and because of the lack of a controlled filter-type gain characteristic.

A review of conventional electron tubes eliminates the pentode from consideration because of the detrimental effects of transit-time and lead-inductance loading. Similarly, other tubes must be precluded with the exception of presently available disc-seal triodes. Transit time and lead-inductance effects are negligible with these triodes at the frequencies of interest; however, the structure of the triodes necessi-

tates the operation of the triodes in grounded-grid connections, in general. Of the available triodes of this type, the WE 416-A has by far the highest performance figure, a practical gain-bandwidth product of approximately 700 mega-cycles.[1,2]

A. Grounded-grid Triodes

Considerations in the grounded-grid triode can best be seen by examining the equivalent circuit shown in Figure 1. For modern triodes, the cathode-to-plate capacitance can be neglected. The output circuit of the triode consists of the output capacitance and the plate resistance in parallel. At the input, in addition to the input capacitance, there is the electronic loading and the transferred impedance from the interstage connected to the plate. The grounded-grid triode can be considered an impedance transformer having a step-up ratio of $(\mu + 1)$. Thus, the impedance transferred from the plate to the grid circuit, together with the electronic loading, is:

$$\frac{(r_p + Z_{in_{K+1}})}{\mu + 1}$$

where it should be noted that $Z_{in_{K+1}}$ includes C_{out} of the triode but not r_p. If the transferred impedance, $Z_{in_{K+1}}$, is for the moment ignored, the high input conductance, approximately equal to g_m, if $\mu + 1 \approx \mu$, necessitates the use of four-terminal interstages which provide an impedance transformation. An additional requirement on interstage K comes from the transferred impedance, $Z_{in_{K+1}}/\mu + 1$, which will in general exhibit a variation with frequency, and thus a two-fold appearance of the characteristics of an interstage is produced. Thus means must be provided either to incorporate this transferred impedance or to minimize it. A combination of both of these methods will be used in this paper. If these non-unilateral effects of grounded-grid triodes can be properly incorporated in the inter-stage network, the stages can be considered unilateral.

The gain function (output voltage / input voltage) of an individual stage will be:

$$\frac{E_{g_{K+1}}}{E_{g_K}} = \frac{\mu_K + 1}{\mu_K} \; g_{m_K} \; Z_{T_K} \qquad (1)$$

where E_{g_K} is the grid-to-cathode voltage of the

K^{th} triode, g_{m_K}, μ_K are the conventional trans-conductance and amplification factor of the Kth triode. Z_{T_K} is the transfer impedance of output-voltage to input-current for the interstage connected to the plate of the Kth triode:

$$Z_{T_K} = \frac{E_{out}}{I_{in}} = \frac{E_{g_{K+1}}}{I_K}$$

The total gain function of an amplifier of n tubes and n interstages will be:

$$\frac{E_{out}}{E_{in}} \approx \prod_{K=1}^{n} g_{m_K} \; Z_{T_K} \quad (\mu \gg 1) \quad (2)$$

The gain of a stage or an amplifier will refer to the magnitude of the gain function at the center frequency, ω_o, of the stage or the amplifier.

B. The Interstage

As implied previously, the center frequency specifications necessitate either the use of lumped elements of extremely small value or else the use of transmission-line elements which at these frequencies will be quite large and bulky. Lumped elements are usually preferred, and the interstage networks must be simple in configuration for the realization of these elements.

The amplifier requirements of high gain and a square-type gain characteristic lead one to consider stagger-tuning. Stagger-tuning provides a means of conserving gain-bandwidth product, and in addition, through stagger-tuning, existing approximation techniques can be used to obtain a desirable frequency charac-teristic, such as a maximally-flat, band-pass characteristic. If the criterion of network simplicity is kept in mind together with stagger-tuning techniques, the desirable interstage specifications considered to this point can be summarized as follows. The interstage must:

1. Provide an impedance transformation; hence, a four-terminal interstage must be used.
2. Be simple in configuration and consist of simply realizable lumped elements.
3. Incorporate the nodal admittances of the grounded-grid triode, and
4. Provide a single pole of the transfer impedance in proximity to the frequency region of interest.

28

An interstage which satisfies these requirements is shown in Figure 2.

In comparing Figures 1 and 2, it can be seen that the conductances and capacitances can be supplied at least in part by the input and output admittances of the grounded-grid triode. An extensive investigation of the properties of the interstage has made it apparent that no additional conductance should be introduced and that C_1 should not be increased, if possible, above the value supplied by the triode. Therefore, $G_1 = \frac{1}{r_p}$, $C_1 = C_{out}$, $C_2 = C_{in} +$ padding. In this section, the transferred impedance from the following interstage will be neglected; therefore, $G_2 = g_m$. The effects of the transferred impedance will be taken up in the next section.

Although this network appears to be a low-pass type, through a proper choice of the values of the network elements, this network will provide a band-pass gain characteristic which is approximately that of a single-tuned circuit. This can be seen from an investigation of the transfer impedance of the interstage.

$$Z_T = \frac{E_{out}}{I_{in}} = \frac{1}{LC_1C_2} \bigg/ s^3 + (\frac{G_1}{C_1} + \frac{G_2}{C_2}) s^2 + (\frac{1}{LC_s} + \frac{G_1G_2}{C_1C_2}) s + \frac{G_1 + G_2}{LC_1C_2} \quad (3)$$

where s = complex frequency variable, $\sigma + j\omega$

$$C_s = C_1C_2 / (C_1 + C_2)$$

Equation (3) can also be written in terms of the poles (points of infinite gain) of the transfer impedance:

$$Z_T = H / (s + \gamma)(s + \alpha + j\eta)(s + \alpha - j\eta)$$
$$= H/s^3 (2\alpha + \gamma) s^2 (2\alpha\gamma + \lambda^2) s + \gamma\lambda^2 \quad (4)$$
$$\text{where: } \lambda^2 = \alpha^2 + \eta^2$$

These poles can be plotted in the complex frequency plane as in Figure 3.

Also shown in Figure 3 is the pole-zero plot of a simple, single-tuned GLC circuit. If α of the complex pole pair, $-\alpha \pm j\eta$, can be made small in relation to η, the response of the network will be approximately that of a single-tuned circuit in the real frequency region, $\omega \approx \eta$. The center frequency of the interstage, ω_o, is defined as the angular frequency of maximum magnitude of the transfer function.

Methods will be advanced later in the paper which will permit the design to commence in a low-pass domain with subsequent transformation of the design data to the actual band-pass domain, thereby eliminating the necessity of closely approximating a single-tuned characteristic in order to utilize stagger-tuning concepts.

In the design procedure for an entire amplifier it is usual to specify that each interstage realize a particular complex pole pair, $-\alpha \pm j\eta$. Knowing the desired pole locations, the remaining elements of the interstage, L and C_2, can be obtained by equating the corresponding coefficients of (3) and (4). Three non-linear simultaneous equations result, which can be solved to provide the following design formulas.

$$C_2^2 + \xi C_2 - \frac{G_2}{2\alpha}\xi = 0 \quad (5)$$

$$\xi = 2\alpha \left(\frac{C_1}{2\alpha - \frac{G_1}{C_1}} - \frac{G_T}{\lambda^2} \right) \quad (6)$$

where $G_T = G_1 + G_2$

$$\gamma = \frac{G_1}{C_1} + \frac{G_2}{C_2} - 2\alpha \quad (7)$$

$$L = \frac{G_T}{\gamma \lambda^2 C_1 C_2} \quad (8)$$

$$\text{Gain} = \frac{E_{out}}{E_g} (\omega_o) \approx \frac{\gamma \omega_o}{2\alpha |\gamma + j\omega_o|} \quad (9)$$

In addition to the design of the interstage, attention must be given to the tuning of the interstage. Tuning adjustments must, in general, be employed because of the impossibility of producing at ultra-high-frequencies the exact values of the small network elements. For the interstage, at least two elements should be variable in order to provide adjustment of center frequency and bandwidth. Investigations have shown that L and C_2 are convenient elements to be varied to adjust center frequency and bandwidth, respectively. The adjustments of these two elements do not provide completely independent variations of center frequency and bandwidth; however, these two elements provide the greatest independent adjustment.

C. Feedback

The characteristics and formulas just

29

presented were obtained with the assumption that there was no feedback from the following interstage. That is, the effects of the transferred input impedance originating in the interstage connected to the plate of the triode were assumed negligible. It remains, then, to investigate these effects and to determine how to incorporate or to minimize them if they are pronounced. A thorough mathematical presentation of this interaction proves to be very cumbersome and quite devoid of practical worth. However, for the situation where bandwidths of the order of only 10-20% of center frequency are required and where high performance triodes are used, a first order knowledge of feedback can be provided through simple graphs.

For first order effects, the feedback from the $(K+2)$th triode to the Kth interstage can be assumed negligible. In addition, it can be shown that only the real part of the transferred impedance, $Z_{in_{K+1}}$, need be considered to obtain an accurate measure of the feedback. This quantity, Re $Z_{in_{K+1}}$, will have approximately the same frequency characteristic as the transfer impedance $Z_{T_{K+1}}$ of the parent interstage with the exception of being about twice as sharp. Therefore, only the input conductance of the triode (which is element G_2 in Figure 2) will exhibit the feedback. Since only a first order knowledge is desired, the plate resistance of the triode can be ignored, thus $G_1 = 0$. The transfer impedance of interstage K can then be expressed as:

$$Z_{T_K}(j\omega) = \frac{1}{LC_1C_2}\Bigg/\Bigg[\frac{G_2}{C_2}\left(\frac{1}{LC_1} - \omega^2\right) +$$

$$j\omega\left(\frac{1}{LC_s} - \omega^2\right)\Bigg] \tag{10}$$

$$= H\Big/\Big[M + jN\Big]$$

In (10), it is seen that only M contains G_2, while N does not; thus, M will contain the feedback. Graphs of M and N for a representative situation (as in Figure 4) provide a means of visualizing the effects.

The first case to investigate is where the center frequency of interstage $\underline{K+1}$ is equal to the center frequency of interstage $\underline{K}$, i.e., $\omega_{o_{K+1}} = \omega_{o_K} = \omega_o$. In Figure 4, the approximate location of ω_o is shown, and it is seen that in this region, the magnitude of N is small compared to M. Therefore, the effect of M is

predominant in Z_{T_K}, and a decrease of G_{2_K} because of the frequency peak of $Z_{in_{K+1}}$ at ω_o will produce an increase in the magnitude of Z_{T_K}. For the two stage combination, the overall magnitude of the transfer function in the frequency region near ω_o will be increased; however, the overall band-width of the combination will be decreased because the relative sharpness of the real part of the transferred impedance from interstage $K+1$ produces an increase of Z_{T_K} only near ω_o. See Figure 5.

Next, $\omega_{o_{K+1}}$ is assumed to be less than ω_{o_K}. If the difference of the two frequencies is not too great, $\omega_{o_{K+1}}$ will be located in the region of minimum M_K. Therefore, any variation of M_K, due to the transferred impedance, will produce little if any effect on the magnitude of Z_{T_K}.

Finally, for $\omega_{o_{K+1}} > \omega_{o_K}$, $\omega_{b_{K+1}}$ will probably lie in the minimum region of N_K, and any variation of M_K will appear directly in the magnitude of Z_{T_K}.

For these last two situations, a peak in the magnitude of Z_{T_K} will appear in greater or lesser degree at $\omega_{o_{K+1}}$. For the cascaded response of two stages it is convenient and informative to assign the feedback effects to the parent stage of the feedback, $\underline{K+1}$, rather than the actual stage K. Thus, stage $\underline{K+1}$, through feedback, effectively has a greater gain at its center frequency, $\omega_{o_{K+1}}$, and consequently, has a smaller effective bandwidth. Where feedback is present in an amplifier, it can be minimized and absorbed by increasing the original bandwidth of stage $K+1$, such that with the narrow-banding effects of feedback, the proper effective bandwidth for the stage is obtained.

The conclusion can be drawn that in a stagger-tuned arrangement the center frequencies of the interstages should be assigned in a decreasing order as one progresses from the amplifier input to the output. This arrangement will minimize the effective bandwidth shrinkage of the individual interstages. At the input to the first stage it will be seen that the feedback present at this point can be incorporated through a proper tuning adjustment technique.

With the proper interstage arrangement, a

30

grounded-grid amplifier can be assumed to be composed of unilateral interstages. Although the feedback can be ignored to a first order, if a desired response is to be realized very closely, it will probably be necessary to adjust some s stages empirically in order to incorporate and minimize the feedback that appears.

D. The Staggered-triple

In using conventional stagger-tuning with pentodes, one usually determines the number of stages in a staggered-arrangement from the desired selectivity needed in the pass-band shape or from the desired overall magnitude of gain. It is well known that the most efficient use of the gain-bandwidth limitations of the interstages and tubes is obtained in an n-uple arrangement where n is the needed number of single-tuned interstages. Of course, practical considerations often dictate a compromise between ease of realizability and efficiency.

For stagger-tuning with the triodes and interstages of this paper, it can be shown that ideally the most efficient arrangement is again an n-uple. However, as noted in the last section, a small amount of empiricism will probably be necessary in order to obtain the desired effective interstage bandwidths. In addition, it will be seen in the next section that it is not possible in general to find the exact tuning data to realize exactly a desired frequency characteristic. Therefore, the small empirical adjustments will also be needed to compensate for tuning inaccuracies. In order to realize as closely as possible a desired frequency characteristic, it will be necessary to determine which stages need adjustment. The larger the value of n (the number of staggered interstages), the smaller will be the bandwidth requirement for certain of the interstages. The feedback effects from the narrow stages will be greater than for broader stages, necessitating greater empirical adjustment. For a large value of n, it will not only be difficult to establish which stages need adjustment, but also how much adjustment is necessary. This follows because of the closeness of the center frequencies of the individual interstages. Thus, although a large number of stages is desired for most gain-bandwidth efficiency, a small number is desired for the feedback problem. These considerations point to the staggered triple as the best compromise arrangement for these grounded-grid amplifiers.

For a triple where the individual interstages have center frequencies at either the extremities or the center of the pass-band, the empirical realignment can be easily and accurately provided following an inspection of the overall triple response.

The staggered-triple will have a pole complement as shown in Figure 6.

The output of the last stage of the triple will usually be furnished with a load such as a terminated coaxial cable. If, for example, an amplifier is designed for use in a 50 ohm coaxial system, G_2 of the last stage will be supplied by terminated 50 ohm cable rather than the input conductance of another 416-A.

On the basis of the amplifier gain magnitude, the best stage for this decreased loading would be the wide stage of the triple. Placing this stage last is not in accordance with the arrangement for minimum feedback; however, the interchange of the low frequency and center frequency stages produces a negligible increase in feedback. Thus, the triple arrangement from input to output that will be used is: high-frequency stage, low-frequency stage, center-frequency stage.

E. Wide-Band Synthesis Procedure

Because of the required ratio of bandwidth to center frequency of only 10-20%, and because a small amount of empirical adjustment has been permitted, it might seem sufficient to employ narrow-band synthesis techniques. The usual narrow-band technique is to consider the region in the band-pass plane near the desired center frequency as a low-pass region, positioning the complex poles in this region in accordance with knowledge of the low-pass situation, i.e., ignoring the presence of poles outside of this region. If such a technique is used for the interstages of this report, a serious distortion of the frequency characteristic of the amplifier results primarily due to the presence of the real poles. Even for the modest bandwidth requirement herein, the empirical correction of this distortion to make a close realization of a desired characteristic is a very tedious and lengthy process.

A procedure to obtain wide-band, cas-

31

caded amplifiers of desired pass-band shape is to investigate the desired type of interstage, with respect to poles and zeros of the transfer function; to determine the poles in a low-pass situation which will realize the desired frequency characteristic; and, finally, to evolve a conformal low-pass to band-pass transformation which will transform the low-pass pole positions to actual band-pass positions which can then be realized by the desired interstage type.

This process can be used here except that it has not been possible to find a simple transformation which is exact with respect to the actual poles of the interstage. However, transformations have been found which evolve poles approximating closely the actual poles of the interstage. One such transformation is:

$$p = -s(s^2 + 1) \qquad (11)$$

The low-pass pole configuration which provides a maximally-flat transfer characteristic and which is compatible with a triple is shown in Figure 7. Also shown in Figure 7 are the poles in the band-pass plane that are evolved through use of (11). It is seen immediately that the band-pass poles of Figure 7 are not physically realizable because of the poles in the right half plane. However, if the negative real poles of the triple, shown in Figure 6, are close to the origin, say, $\sqrt[3]{\gamma_1 \, \gamma_2 \, \gamma_3} \, / \omega_o \leq 0.3$, (11) will provide a close approximation to desired performance.

If the mean coordinate of the real poles is greater than 0.3, a tilted pass-band results. For the 416-A triode, and with triple specifications for a 400 megacycle center frequency and 50 megacycle bandwidth, the real pole locations are $0.5 < \frac{\gamma}{\omega_o} < 1.0$. If for simplicity the mean value of $\gamma/\omega_o = 1$ is assumed, the error in the transformation is as indicated by the upper curve of Figure 8. This large error, incurred because the negative real poles are not close to the origin, leads one to consider another simple transformation which neglects the real poles entirely,

$$p = -j(s^2 + 1) \qquad (12)$$

The error due to this transformation, again for the 416-A, is shown by the lower curve of Figure 8.

The nearly equal and opposite nature of the errors from the two transformations indicates that approximately valid interstage locations can be obtained for the 416-A by an arithmetic comparison of the pole positions evolved by the two transformations. These mean locations are shown in Table 1. The validity of these locations can be established analytically; however, for the purpose of this paper, the results from an experimental amplifier may suffice.

The complex pole locations thus can be used with (5) - (9) to determine the network elements and gain of the triple. It is still necessary to establish the tuning data of the individual interstages of the triple. The approximate transformations, (11) and (12), can again be used to obtain this data. For the two necessary adjustments, it is convenient to adjust the center frequency and the frequency of the upper 3-decibel band-edge. The frequency characteristic of the interstage is then completely defined. These two frequencies can be obtained by noting for each pole in the low-pass plane, the low-pass frequencies corresponding to the interstage center frequency and upper band-edge.

These low-pass frequencies can then be substituted into (11) or (12) where the complex variables are now considered to be pure imaginary, and the real frequencies of alignment are obtained. For the 416-A amplifier, a mean comparison can again be used to obtain valid, accurate tuning data.

Rather than use an averaging process as above, a method has been devised which will provide accurate tuning data for any mean pole location. The method is derived for a criterion of zero tilt at band-center for the triple, and results in:

$$v = \frac{\omega + a}{1 + a} (\omega^2 - 1) \qquad (13)$$

where v is an alignment frequency in the low-pass plane (from the jv-axis), and a ($= \frac{\sqrt[3]{\gamma_1 \, \gamma_2 \, \gamma_3}}{\omega_o}$) is the mean real pole location.

Thus, if it is not desired to predict accurately the performance of the amplifier, either (11) or (12), together with design formulas, can be used to determine the order of

32

magnitudes of L and C_2. (13) can then be used to tune the amplifier properly to realize a desired frequency characteristic.

F. The Practical Amplifier

Using the techniques described to realize a desired band-pass gain characteristic with a staggered-triple, a complete amplifier will be available with the addition of a suitable input circuit. Also, attention must be given the problems of actual tuning procedures.

The input circuit should possess the following properties:

1. A low noise-figure for the amplifier
2. The incorporation of the feedback effects from the high-frequency stage of the triple
3. A broad-band transfer function when operated with a low-impedance source.

The simplest input circuit will be a direct connection of the signal source to the input of the first triode, together with shunt inductance to resonate the input capacitance of this triode. This connection is shown in Figure 9.

It can be shown that this circuit will provide a satisfactory transfer characteristic, in spite of a mismatch, for reasonably low levels of source impedance.

In a manner similar to that used for the interstages, a qualitative knowledge of the feedback effects on this circuit can be simply obtained. Such an investigation points out that the feedback effects are much greater here than are encountered in later interstages. However, the effect of feedback at the input can be incorporated through the use of the following tuning procedure. First, the single-tuned input circuit is resonated at the center frequency of the overall amplifier. The source is then connected and the output of the first stage of the triple is monitored. This first stage of the triple can now be adjusted for the proper center frequency and bandwidth, and the narrow-banding effects of the feedback will automatically be compensated.

It should be noted that this adjustment technique could theoretically be used for the individual interstages; however, it is difficult

in practice to introduce the source to the input of an internal triode without effecting a frequency distortion in the alignment.

In practice, high-impedance probes directly-connected into the input of each interstage, except the first, couple energy to the interstage; and the output of the interstage is monitored with a high impedance crystal detector circuit which is also permanently connected to the interstage. The permanent connection of these probes and monitors provides for a constant loading effect on the interstages.

The actual tuning of an amplifier should be started with the last stage of the triple, progressing toward the input. At the input, however, the input circuit should be resonated, as mentioned above, before tuning the first stage of the triple.

An estimate of the noise figure of this input circuit can be obtained through the use of conventional formulas. For the purposes of illustration, a WE 416-A triode will be used, and a center frequency of 400 megacycles and a source impedance of 50 ohms will be assumed. For this triode, the transit time effects are negligible at 400 megacycles, and for careful construction of the input stage, circuit losses can be neglected. The noise figure of the input (and consequently the amplifier, since the power gain of a 416-A is large) is then approximately:

$$F \approx 1 + \frac{2.5}{g_m R_{Source}} \qquad (14)$$

For a g_m of 0.04 mhos and a source impedance of 50 ohms, a noise figure of 4 decibels can be obtained. [3]

If it is desired to cascade these staggered-triples, a requirement of the last stage of the first triple is a very constant resistive load. That is, the second triple of the cascaded system must have an extrememly flat input impedance, constant, say, within 1% over a band of 15% of center frequency. To accomplish this it has been found necessary to employ an active input network for the second amplifier rather than a passive network. Such an active network is a grounded-cathode triode. Thus the input becomes the familiar cascode circuit, as shown in Figure 10. [4]

33

For triodes such as the 416-A, the real part of the input impedance of the grounded-cathode triode is sufficiently large to be negligible with respect to a resistance which is equal in value to the source impedance. The imaginary part of the input impedance is capacitive and can be resonated with an inductance. The feedback effects for the cascode will be similar to the direct connection because of the single-tuned circuits. Here again, then, the feedback can be incorporated through proper tuning adjustments.

With this cascode circuit an extremely flat input impedance of desired magnitude can be obtained. Because of this flat input-impedance magnitude, cascading of two or more amplifiers can be accomplished without any difficulties due to inter-action between the individual amplifiers. It should be noted that although the noise figure of a cascode circuit can be remarkably low, it is not so in this version because of the grid resistor added for broad-banding reasons. But since the cascode follows a first triple with high gain and low noise-figure, the second triple has almost no effect on the overall noise figure.

G. Practical Example

An amplifier has been constructed using the design procedures outlined in this paper. The amplifier consists of a cascode input circuit and a staggered-triple employing 416-A triodes. The amplifier specifications were for a center frequency of 400 megacycles, a 3-decibel bandwidth of 50 megacycles and a maximally-flat gain characteristic. The cascode input circuit was chosen to illustrate the triple used in second (and later) stages in a cascade of triples. The direct input circuit is even simpler to obtain. The complete amplifier is shown in Figure 11. A schematic diagram of the amplifier is shown in Figure 12, with only the r-f circuit shown and those network elements supplied by the triodes indicated by broken lines.

The tuning data for the amplifier were obtained in the manner indicated in Section E (see Table 1). After the initial tuning of the amplifier, a slight tilt is usually sustained in the gain characteristic because of the incomplete absorption of the feedback effects and because of the approximate nature of the transformation. Therefore, an inspection of the initial results

is needed to determine which stages need further trimming in order to realize the desired overall characteristic of the amplifier.

The actual performance of the amplifier is shown in Figure 13. The observed insertion gain of 50 decibels, for Curve "b" of the figure, agrees almost exactly with the calculated value, which value included the gain increase at the amplifier input due to the feedback from the first stage of the triple. The cascode input circuit of the amplifier provided an input impedance which was constant within $1/2\%$ over the pass-band.

TABLE 1

Complex Pole Locations and Tuning Frequencies for a Maximally-flat, Staggered-triple

Center frequency: 400   3-db bandwidth: 50

	Center Freq. stage	Hi Freq. stage	Lo Freq. stage
Pole location	-24.9 + j402.35	-11.35+ j421	-14.1 + j377.8
Tunings...f_o	400	420.6	377
3 db	423.6	431.6	390.7

All values in megacycles

References

[1]   Christopherson, W. A., "The analysis and synthesis of grounded-grid amplifier transfer functions," Tech. Report No. 46, Electronics Research Laboratory, Stanford University, California, May, 1952.

[2]   Pederson, D. O., "The synthesis and design of grounded-grid, staggered-triples at ultra-high-frequencies," Tech. Report No. 12, Electronics Research Laboratory, Stanford University, California, September, 1952.

[3]   Murakomi, T., "A study of grounded-grid, ultra-high-frequency amplifiers," RCA Review, Vol. 12, pp. 682-701, December, 1951.

[4]   Wallman, H., MacKee, A. B., and Gadsen, C. P., "A low-noise amplifier," Proceedings of the I.R.E., Vol. 36, pp. 700-708.
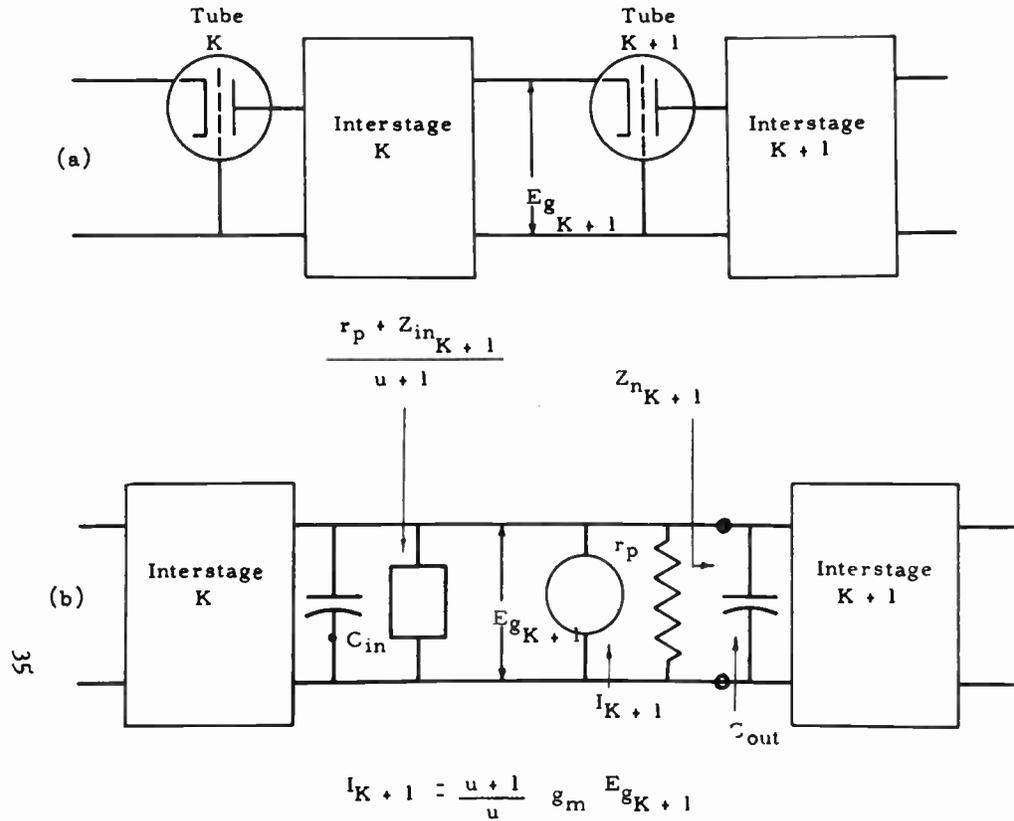
Fig. 1
The triode in grounded-grid operation.
(a) cascaded stages; (b) equivalent circuit of a stage.

$$I_{K+1} = \frac{u+1}{u} \; g_m \; E_{g_{K+1}}$$
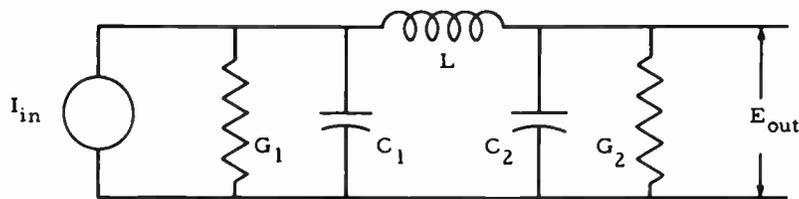
Fig. 2
The interstage.
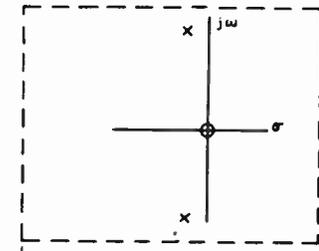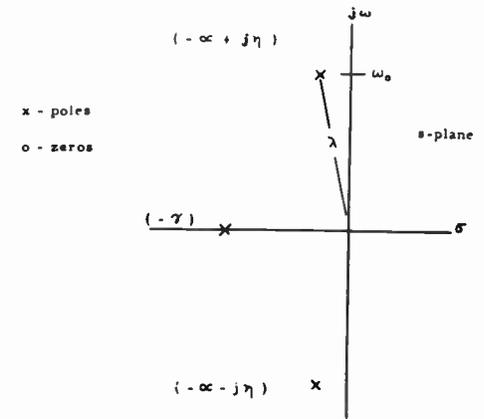
Fig. 3
The poles of the interstage transfer impedance, Z_T (with corresponding poles and zeros of a single-tuned circuit, a parallel GLC).
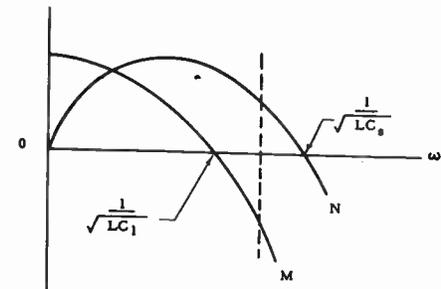
Fig. 4
Plots of M and N of Z_{T_K}.
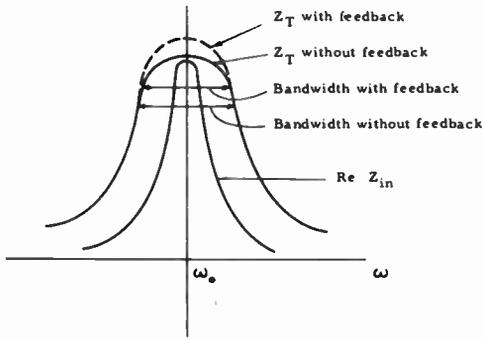
Fig. 5
Representative plots of Z_T and ReZ_{in}
illustrating the effects of feedback.
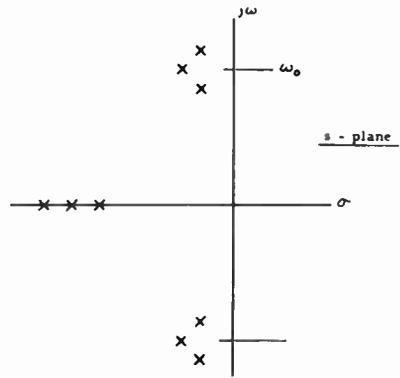
Fig. 6
The poles of a staggered-triple.
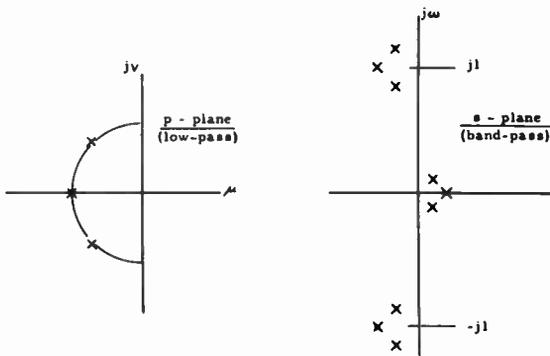
Fig. 7
Three-pole maximally-flat
pole locations.
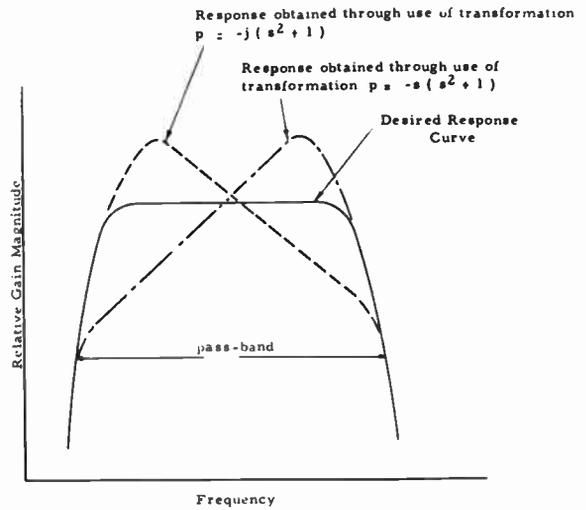
Fig. 8
Distortions of the gain response due to
the approximate nature of the
transformations.
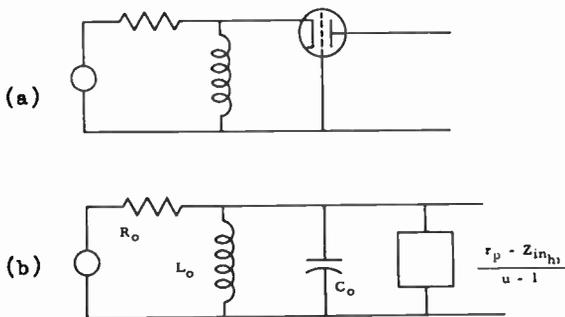
Fig. 9
The directly connected input circuit.
(a) the actual circuit;
(b) the equivalent circuit.

Fig. 10
The cascode input circuit.

STAGE NO.1.-OUTPUT MONITOR
STAGE NO.3.-INPUT PROBE
STAGE NO.3.-INDUCTANCE TUNER
STAGE NO.3.-OUTPUT CAPACITANCE PAD
OUTPUT TERMINATION & CRYSTAL DETECTOR
BLOWER

STAGE NO.1.-OUTPUT CAPACITANCE PAD
STAGE NO.1.-INDUCTANCE TUNER
CASCODE TUNING STUB
CASCODE INTERSTAGE MONITOR
AMPLIFIER INPUT

INPUT TUNING STUB
STAGE NO.2.-INPUT PROBE
STAGE NO.2.-INDUCTANCE TUNER

AIR DUCT
STAGE NO.2.-OUTPUT MONITOR
STAGE NO.2.-OUTPUT CAPACITANCE PAD

Fig. 11 - A 400-mc amplifier.

Fig. 12 - RF schematic diagram of the 400-mc amplifier.
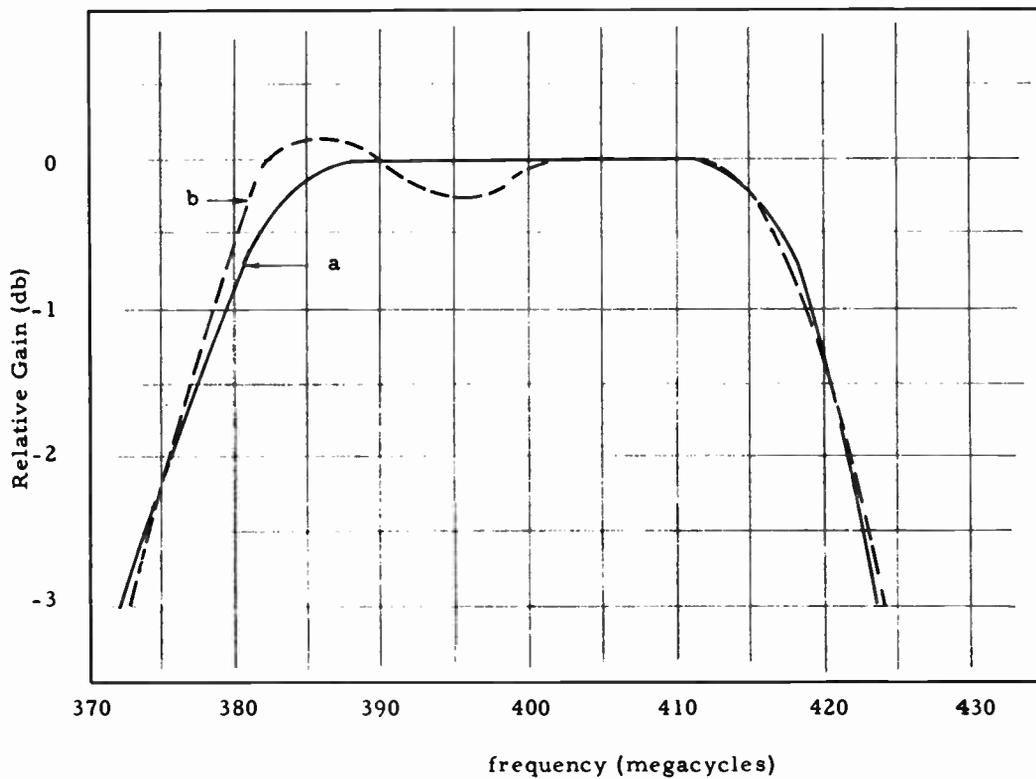
Fig. 13
Gain magnitude vs. frequency characteristics. (a) the exact maximally-flat
characteristic; (b) a representative amplifier characteristic with a
50-mc bandwidth and a 50-db gain.

NETWORK ANALYSIS WITH THE AID OF THE MATRIX GENERATING POLYNOMIAL[*]

by

Herbert Kurss

Microwave Research Institute
Polytechnic Institute of Brooklyn
Brooklyn 1, New York

I. Introduction

This investigation arose from an attempt to simplify the derivation of a long neglected formula published by G. A. Campbell.[1] Not only was this attempt successful, but from this new and physically evident derivation the important Sylvester identity[2] was an immediate by-product. The surprising feature is that the Campbell formula is the solution to a network problem while the Sylvester identity is regarded solely as a mathematical matrix identity. This feature was emphasized when it was soon apparent that the technique, implicit in this new derivation, allowed one to derive with remarkable simplicity the Jacobi ratio theorem[3] as well as theorems on equivalent networks. Here then is a technique which bridges the gap between the network analyst and the mathematician and whose ultimate range of applicability is yet to be charted. In order to accommodate the mathematical inclinations of various groups, the following equivalent epitomizations of the core of the technique can be given:

A. The characterization of a network by its input-output relation, or

B. The characterization of a matrix by its generating polynomial, or

C. The characterization of a (finite dimensional) operator by its spectral surface.

The following pages shall elaborate upon this point of view and indicate its fruitfulness as a means of simplifying and unifying derivations, correlating otherwise scattered formula, and extending known results.

II. The Network Element as an Operator

In this paper it is sufficient to define a multiterminal network as an operator which effects a transformation between two electrical vectors. As can be said of any operator, depending upon the choice of a vector basis, one can obtain a multitude of matrix representatives. In particular, the voltage and current vectors give rise to the impedance and admittance matrices while the incident and reflected waves result in the scatter-

ing matrix. (This definition of a network as an operator is strongly indicated when one deals with a multiterminal waveguide junction or a nonlinear network in the small signal region, especially if these networks are non-reciprocal as occurs for a gyrator.[4])

It should be remembered that an operator acts upon vectors in its domain and transforms them into vectors in its range, but the domain and range may by accident or intent be contained in a larger dimensional space. The question of equivalent networks is then identified with the question of those operators which have the same projection onto a given subspace. This shall be considered in later sections.

Although abstract operator viewpoint offers a convenient means for exhibiting the unity of the various topics, it shall be only sparsely and parenthetically used and a network shall, in general, be equated to a matrix representative.

III. The Generating Polynomial of a Matrix

Classical terminology describes the sum of a power series in a single variable as being the generating function of its coefficients[5] (For example, the Bessel functions of integral order can be so defined and G. N. Watson hence refers to them as Bessel coefficients.[6]) Analogously, a polynomial whose coefficients are the various minors (or subdeterminants) of a matrix can be said to generate the minors of the matrix. Such a polynomial is a generating polynomial of the matrix. When formed by the mathematical process of polarization[7] (described below), this shall be called the generating polynomial of the matrix. More precisely, given an arbitrary matrix $A = [a_{ij}]$, one introduces an auxiliary matrix $\Omega = [\lambda_{ij}]$ and defines the generating polynomial of A with respect to Ω as the polynomial in the λ's obtained from the expansion of the determinant $|A + \Omega|$. Essentially, this considers the nth order determinant $|A|$ as a homogeneous function of its n^2 elements a_{ij} and when each a_{ij} is augmented by an indeterminate λ_{ij}, one obtains the complete polarization of this determinant, i.e., $|A + \Omega|$.

The implications of this generating polynomial can be gleaned from the fact that the spectral theory of a finite operator is but the

special case which occurs when $\Omega = \lambda E$ where λ is a scalar and E is the unit matrix. The total set of values of the scalar λ for which $|A - \lambda E| = 0$ is then denoted as the spectrum of A. This spectrum consists of a finite number of complex points since a polynomial in a <u>single</u> variable of degree n has at most n distinct <u>roots</u>. Consequently, when viewed as a generalization of this spectral theory, the vanishing of the generating polynomial can be said to define the spectral surface.

In passing, one should note that the bilinear form or inner product often associated with a matrix A corresponds, essentially, to the generating polynomial of A^{-1} obtained when Ω has a rank of unity.

The motivation for the introduction of the generating polynomial into network analysis is clear from its physical interpretation. For definiteness, consider a network whose behavior can be stated by the matrix equation $V = ZI$ where $Z = [z_{ij}]$ and z_{ij} are <u>internal impedance</u> parameters which represent the network. The effective external couplings or terminations can also be characterized by a matrix, $\Omega = [\lambda_{ij}]$, which relates the same voltage V and current I by the equation $V = -\Omega I$. For example, $-\lambda_{11}$ is the input impedance at terminal 1 when terminating impedances λ_{ij} are connected between the remaining terminals. Combining these last two equations, one can eliminate the voltage to form $(Z + \Omega)I = 0$. It is now a well known result that a non-trivial current I can exist if and only if $|Z + \Omega| = 0$, i.e., if and only if the determinant of the matrix $(Z + \Omega)$ vanishes. This single equation, a polynomial in the λ's, is then the input-output relation of the network.

Two networks are called equivalent with respect to a class of terminations when each network presents the same input-output transformation for any set of terminations in the class. In other words, networks are called equivalent when their generating polynomials are equivalent, i.e., dependent in the mathematical sense. It is then a simple but vital fact that the corresponding polynomial coefficients are proportional, i.e., corresponding coefficients have a common ratio. Here then is the parent to a family of ratio theorems.

It is of interest to note that the number of terms in the generating polynomial $|A + \Omega|$ is equal to the number of minors of the matrix Ω. If Ω is non-singular and of order n, there are $_nC_m^2$ minors of order m where the binomial coefficient $_nC_m$ is defined by the generating polynomial $(1 + x)^n = \sum_n C_m x^m$. The total number of minors of all orders is then $\sum_n C_m^2 = _{2n}C_n = 2n!/n! \, n!$. (This last statement presumes the convention that the minor of order zero is unity.) Of course, since the matrix has only n^2 distinct elements only n^2 of these minors are independent!

IV. The Jacobi Ratio Theorem
(Functionally Related Operators)

The Jacobi ratio theorem relates the minors of a matrix to the minors of its inverse and applies directly to network theory where, for example, the inverse of an impedance matrix is an admittance matrix. (The theorem states[3] that every ratio of a minor of a non-singular matrix to the corresponding complementary cofactor of the inverse matrix has the same value.) From the physical equation, as given above,

$$V = ZI = -\Omega I \text{ , it follows that}$$
$$Z^{-1} V = I = -\Omega^{-1} V \text{ , and hence}$$
$$|Z + \Omega| = 0 \Longleftrightarrow |Z^{-1} + \Omega^{-1}| = 0.$$

(The symbol $\Longleftrightarrow$ reads "implies and is implied by" or "if and only if".) The equivalence of these generating polynomials already contains the Jacobi ratio theorem but the presence of Ω^{-1} is slightly awkward. This is not serious since Ω can choose with a simple form, for example a diagonal form, whence Ω^{-1} is evaluated directly.

It is illuminating to observe the absence or removal of Ω^{-1} in the mathematical derivation of the Jacobi theorem in this implicit form of equivalent generating polynomials. This derivation proceeds from the simple identities

$$Z + \Omega = Z(E + Z^{-1}\Omega) = (E + \Omega Z^{-1})Z$$

which merely postulate $|Z| \neq 0$. The associated determinants are related by the similar equation

$$|Z + \Omega| = |Z||E + Z^{-1}\Omega| = |E + \Omega Z^{-1}| \, |Z|.$$

Since the expansion of a determinant is not reducible into factors, no information is lost when this is put in the form of equivalent polynomials, namely,

$$|Z + \Omega| = 0 \Longleftrightarrow |E + Z^{-1}\Omega| = 0 \Longleftrightarrow |E + \Omega Z^{-1}| = 0.$$

From this the specific details of the Jacobi theorem are readily obtained. Upon accepting the additional hypothesis $|\Omega| \neq 0$, one can employ the identity

$$|Z + \Omega| = \Omega(\Omega^{-1} + Z^{-1}) \, Z$$

to obtain the result of the previous physical derivation, namely,

$$|Z + \Omega| = 0 \Longleftrightarrow |\Omega^{-1} + Z^{-1}| = 0.$$

From the above mathematical derivation, it is clear how one extends the results to any two functionally related matrices. Thus, if matrix Z is a known function of matrix S, $Z = f(S)$, it follows that

$$|Z + \Omega| = 0 \iff |f(S) + \Omega| = 0.$$

For example, the relative impedance matrix Z_R and the scattering matrix S are related by

$$Z_R = (S + E) / (S - E)$$

and hence

$$|Z_R + \Omega| = 0 \iff |(S + E) + \Omega(S - E)| = 0,$$

a total of $_{2n}C_n$ relations between the Z_R and S matrices when each has order n. It is significant to rewrite the above results as

$$|Z_R + \Omega| = 0 \iff |S + (E+\Omega)/(E-\Omega)| = 0.$$

This indicates the sufficiency of a single construction for a generating polynomial since a different construction can often be reduced to the question of functionally related operators.

V. "Elimination of Concealed Circuits"[1]
(The Projection of an Operator)

A familiar problem in analysis is the elimination of concealed circuits (linear constraints). (At microwave frequencies this corresponds to fixed terminations at some of the terminal planes of a multiwaveguide junction.) In matrix terms this permits the following typical illustration: given

$$\begin{bmatrix} y_1 \\ y_2 \\ -- \\ y_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ - - - & - & - - \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ -- \\ x_3 \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ - & - \\ A_3 & A_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ -- \\ x_3 \end{bmatrix}$$

when x_3 is eliminated, determine A' in

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a'_{11} & a'_{12} \\ a'_{21} & a'_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A' \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

(More succinctly, one seeks the projection of an operator onto a subspace which in this illustration is its range.)

One immediate solution of the problem exploits the above indicated partitioning of the matrix to obtain the reduction formula[8]

$$A' = A_1 - A_2 A_4^{-1} A_3 .$$

Alternatively, one can invoke Cramer's rule (a part of the Jacobi ratio theorem) to obtain

$$\begin{bmatrix} x_1 \\ x_2 \\ -- \\ x_3 \end{bmatrix} = \frac{1}{|a|} \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ - - - & - & - - \\ A_{13} & A_{23} & A_{33} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ -- \\ y_3 \end{bmatrix}$$

and hence

$$(A')^{-1} = \frac{1}{|a|} \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{bmatrix}$$

(A_{ij} denotes the cofactor of a_{ij}.)

The Campbell solution is more involved but the simplicity of its derivation by the polynomial technique shall now be indicated. The equivalent generating polynomials are:

$$\begin{vmatrix} a_{11} + \lambda_{11} & a_{12} + \lambda_{12} & a_{13} \\ a_{21} + \lambda_{21} & a_{22} + \lambda_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} =$$

$$\begin{cases} |\lambda| a_c + \lambda_{11} a_{22 \cdot c} - \lambda_{12} a_{12 \cdot c} + \ldots + |a| = 0 \text{ or} \\ |\lambda| A_\alpha + \lambda_{11} A_{\alpha/22} - \lambda_{12} A_{\alpha/21} + \ldots + A_0 = 0 \end{cases}$$

where $\alpha = 11 \cdot 22$ (accessible terminals)
$c = \bar{\alpha} = 33$ (concealed terminals) } and $A_{\alpha/ij} = A_{ij \cdot c}$

and

$$\begin{vmatrix} a'_{11} + \lambda_{11} & a'_{12} + \lambda_{12} \\ a'_{21} + \lambda_{21} & a'_{22} + \lambda_{22} \end{vmatrix} =$$

$$|\lambda| + \lambda_{11} a'_{22} - \lambda_{12} a'_{21} + \ldots + |a'| = 0.$$

(The symbol $a_{ij \cdot kl}$ denotes the minor which contains the product $a_{ij} a_{kl}$ while the cofactor of this minor is denoted[1] by $A_{ij \cdot kl}$. With the general case in mind $a_{ij \cdot c}$ can be described as the primary[9] superdeterminant of the concealed determinant a_c, this superdeterminant formed with respect to the accessible element a_{ij}.) The ratio of corresponding coefficients in these polynomials contains the Campbell formula as applied to the above illustration, namely,

$$\begin{bmatrix} a'_{11} & a'_{12} \\ a'_{21} & a'_{22} \end{bmatrix} = \frac{1}{a_c} \begin{bmatrix} a_{11 \cdot c} & a_{12 \cdot c} \\ a_{21 \cdot c} & a_{22 \cdot c} \end{bmatrix} = \frac{1}{A_\alpha} \begin{bmatrix} A_{\alpha/11} & A_{\alpha/12} \\ A_{\alpha/21} & A_{\alpha/22} \end{bmatrix}$$

(When the Campbell formula is equated to the previous partitioned matrix reduction formula, one obtains a cheap derivation of an identity noted by Cullis[9].) But, in addition to this Campbell formula, the equivalence of the generating poly-

nomials also contains the important Sylvester identity[2] which for this three dimensional case assumes the form of the single equation

$$|a'| = |a|/a_c$$

Both the Compbell formula and the Sylvester identity can be embodied in the single statement: When two matrices have the same accessible terminals and with respect to these they effect the same transformation then every ratio of the cofactor of an accessible minor of one of the matrices to the cofactor of the corresponding accessible minor of the other matrix has the same value.

VI. Alteration or Addition of Concealed Circuits

(Operators with the Same Projection)

A basic theorem on equivalent networks is the statement:

If $P\Omega Q = \Omega$, then $A \cong PAQ$
(where $\cong$ reads "is equivalent to")

The proof is immediate for from the hypothesis, $P\Omega Q = \Omega$, it follows that

$$|A+\Omega| = |P||A+\Omega||Q| = |PAQ+\Omega|$$

and, consequently, A and PAQ have equivalent generating polynomials.

A simple but important corollary occurs when $\Omega = -\lambda E$ for then $A \cong PAP^{-1}$ for all non-singular matrices P.

A corollary more directly applicable to network analysis (and indeed synthesis) is the following:

If $A = \begin{bmatrix} A_1 & | & A_2 \\ -- & | & -- \\ A_3 & | & A_4 \end{bmatrix}$ and $\Omega = \begin{bmatrix} \Omega_1 & | & 0 \\ -- & | & -- \\ 0 & | & 0 \end{bmatrix}$ then $A \cong PAQ$

where $P = \begin{bmatrix} E & | & P_1 \\ -- & | & -- \\ 0 & | & P_2 \end{bmatrix}$; $Q = \begin{bmatrix} E & | & 0 \\ -- & | & -- \\ Q_1 & | & Q_2 \end{bmatrix}$

and P_1, P_2, Q_1 and Q_2 are arbitrary.

This corollary is in no way limited to reciprocal networks. However, if the network represented by the matrix A is reciprocal, i.e., if the matrix is symmetrical and the equivalent matrix is to be symmetrical, then Q is the transpose of P ($Q = P^T$) and one obtains the familiar result[10] $A \cong PAP^T$.

The discussion is easily extended to equivalent networks with a different number of concealed circuits by the remark that

$$\begin{bmatrix} A & | & 0 \\ -- & - & -- \\ 0 & | & E \end{bmatrix} \cong A$$

The validity of this is evident since by the Laplace expansion of a determinant

$$\begin{vmatrix} A + \Omega_1 & | & 0 \\ ----- & | & - \\ 0 & | & E \end{vmatrix} = 0 \iff |A + \Omega_1| = 0$$

regardless of the dimension of E.

VII. Compound Networks in General

Before the interconnection of multiterminal networks is susceptible to treatment by the generating polynomial a bit of spade work must be done. Suppose two networks, distinguished respectively by subscripts K and L, are to be interconnected. For variety, the scattering formalism shall be used wherein the incident wave $\underline{a}$ is transformed into the scattered wave $\underline{b}$ by the scattering matrix S. It shall be presumed that the behavior of the networks are given by

$$\underline{b}_K = S_K \underline{a}_K$$
$$\underline{b}_L = S_L \underline{a}_L$$

i.e.,

$$\begin{bmatrix} \underline{b}_K \\ \underline{b}_L \end{bmatrix} = \begin{bmatrix} S_K & | & 0 \\ - & + & - \\ 0 & | & S_L \end{bmatrix} \begin{bmatrix} \underline{a}_K \\ \underline{a}_L \end{bmatrix} = S \begin{bmatrix} \underline{a}_K \\ \underline{a}_L \end{bmatrix}$$

both before and after compounding. (For definiteness, one might consider each of the two networks as a four-terminal pair magic tee and the networks are then interconnected with the aid of a gyrator so as to form a four-terminal pair circulator.[11]) The spade work alluded to above consists of eliminating the concealed b_i to obtain $\underline{b}' = S'\underline{a}'$ where all the b_i' are independent and the new matrix S' is square so that $\underline{b}'$ and $\underline{a}'$ are vectors of the same dimension. For example, from the connection constraints it is relatively easy to find matrices C_1 and C_2 such that $C_1 \underline{b} = \underline{b}'$ and $\underline{a} = C_2 \underline{a}'$ whence the goal is achieved by $S' = C_1 S C_2$.

It is at this point that the generating polynomial scheme is applicable. As expounded in the preceding sections, one can now obtain equivalent networks by the elimination, alteration, or addition of as many concealed a_i as desired.

VIII. Compound Networks Amenable to
Matrix Multiplication

If the networks are more special and the interconnections are of a more modest nature than envisaged above, it is possible to cast the results of the preceding section in the form of a

matrix multiplication scheme as commonly used for the tandem connection of quadripoles. The results, without a detailed derivation, shall be listed for the case of a quadripole connected to a three-terminal pair network where for each network the terminal observables are not V_i and I_i separately but rather their ratio. This result has special relevance for a microwave network since the maximum knowledge gained from slotted line impedance measurements is merely the set of coefficients in the generating polynomial which occur when the polarizing matrix Ω is diagonal. These coefficients are all the principal minors of the impedance matrix and no other minors. (Incidentally, since there are $\sum_n C_m = 2^n$ principal minors for a matrix of order n, including unity as the minor of order zero, it is readily shown that $2^n - 1$ independent impedance measurements suffice to determine these principal minors and additional such measurements add no new information. Thus, from the fact that the non-reciprocal or gyrator aspect of a network is not determinable from the principal minors, it follows that the gyrator effects can <u>not</u> be determined from measurements which depend only upon the ratio V_i/I_i at each terminal.)

Let the terminal pairs of the initial six pole be numbered 1, 2, and 3 while the quadripole terminal pairs are numbered 4 and 5. Furthermore, let terminal pairs 3 and 4 be directly connected. Then with primes to denote the elements of the equivalent three-terminal pair network and a use of subscripts explained in section 5, the equivalence can be stated by the following matrix equation:

$$
\begin{bmatrix}
z'_{11} & z'_{11\cdot33} \\
z'_{22} & z'_{22\cdot33} \\
z'_{11\cdot22} & z'_{11\cdot22\cdot33} \\
1 & z'_{33}
\end{bmatrix}
=
$$

$$
\begin{bmatrix}
z_{11} & z_{11\cdot33} \\
z_{22} & z_{22\cdot33} \\
z_{11\cdot22} & z_{11\cdot22\cdot33} \\
1 & z_{33}
\end{bmatrix}
\begin{bmatrix}
z_{44} & z_{44\cdot55} \\
1 & z_{55}
\end{bmatrix}
$$

(The horizontal lines within the matrices serve as a reminder that these are proportional or fractional[12] matrices which are only defined by the <u>ratio</u> of their elements.)

IX. Conclusion

The preceding sections have touched upon facets of multiterminal network analysis each of which could be expanded upon at great length. This was avoided here since the point which was to be stressed was the ease and directness in which basic results were obtained with the aid of the single technique of characterizing a matrix by its generating polynomial. Armed with a working knowledge of this method, one should be able to reproduce and extend the results contained herein. Moreover, the proceedings can be made palatable to the mathematician as well as to the engineer. However, for a deeper insight into the mechanism and a more co-ordinated view of the various topics the operator view of a network has much to recommend it.

References

1. Campbell, G. A., Cisoidal Oscillations, Trans. A.I.E.E., 30, 879-909 (1911).

2. Kowalewski, G., Einführung in die Determinantentheorie, Chap. 7, Chelsea, 1948.

3. Turnbull, H. W., The Theory of Determinants, Matrices and Invariants, p. 77, Blackie and Son Ltd., 2nd Ed., 1945.

4. Tellegen, B. D. H., The Gyrator, a New Circuit Element, Philips Res. Rep., 3, 81-101 (1948).

5. Doetsch, G., Theorie und Anwendung der Laplace-Transformation, p. 8, Dover Publications, 1943.

6. Watson, G. N., A Treatise on the Theory of Bessel Functions, Chap. 2, Cambridge Univ. Press, 1922.

7. Fricke, R., Lehrbuch der Algebra, Vol. I, Chap. 1 ⫛ 18, Friedr. Vieweg & Sohn, 1924.

8. Kron, G., Tensor Analysis of Networks, Chap. 10, John Wiley & Sons, Inc., 1939.

9. Cullis, C. E., Matrices and Determinoids, Vol. II, p. 17, 92, Cambridge Univ. Press, 1918.

10. Burington, R. S., Matrices in Electric Circuit Theory, Jl. Math. Phys., 14, 325-349 (1935).

11. Hogan, C. L., The Microwave Gyrator, B.S.T.J., 31, 1-31 (1952).

12. Bocher, M., Introduction to Higher Algebra, Macmillan, 1907.

43

CONDENSED VERSION OF

TWO NEW EQUATIONS FOR THE DESIGN OF FILTERS

By Milton Dishal

Federal Telecommunication Laboratories

Nutley, N. J.

I. INTRODUCTION

There are some situations where a selective circuit which is equivalent to an inverse arm, i.e., constant-K-configuration filter is to be designed, and the unloaded Q of the elements to be used is sufficiently high for them to be considered "non-dissipative". This paper presents two equations which, for the non-dissipative case, specify the exact element values required for the filter to produce that attenuation shape having the highest possible rate of cutoff, i.e., the Chebishev attenuation shape.

II. SOME EXAMPLES OF NON-DISSIPATIVE EQUIVALENT-INVERSE-ARM FILTERS AND THE THREE CIRCUIT CONSTANTS WHICH MUST BE CORRECTLY ADJUSTED

Because so many of the selective circuits now being used, or designed, seem physically so different from the basic inverse arm configurations, many engineers new to the field do not realize that the design equations for the constant K configuration can be applied.

It thus seems worthwhile calling attention to a few of these equivalent inverse arm filters to stress the wide applicability of the two design equations to be presented.

It will be noticed that with one exception the bandpass examples are from the U.H.F. and microwave region, because it is mainly in this region that the ratio of unloaded Q to fractional mid-frequency $\frac{Q_o}{f_o BW}$ is high enough for the elements to be considered non-dissipative.

Figure 1a shows a common direct-coupled waveguide bandpass filter using four resonators; in another language it would be called a quadruped tuned bandpass filter. The equivalence of this to the fundamental constant K configuration (either bandpass or low pass) has been excellently described in W. W. Mumford's paper. (Reference 1 to which the reader is referred.) In the "language" which this present paper will use, the design information which the engineer must possess (and which is required for all equivalent constant K configuration filters) is:

1. The required coefficient of coupling $K_{r(r+1)}$ between adjacent resonators. This fixes the size of the opening which must be made in the wall between adjacent resonators, and as is well known this opening can take the form of a slot parallel to the electric field vector which will give the equivalent of mutual-inductance coupling between resonators, or a slot perpendicular to the electric field vector which will give the equivalent of "low-side" capacitive coupling between resonators, or a post parallel to the electric field vector which will give the equivalent of self-inductance coupling between resonators, or in general any kind of opening which will allow some of the electric and/or magnetic field of one resonator to enter the adjacent resonator.

2. The required resonant frequency (f_o) of each resonator. This fixes the distance between the walls of each resonator. As is well known the coefficient of coupling mechanism must be correctly considered a part of each resonator to which it is connected; otherwise the passband mid-frequency will not coincide with the resonant frequency.

3. The required singly loaded $Q(Q_1)$ of the first resonator (produced by correctly coupling the generator to this first resonator); and the required singly loaded Q (Q_n) of the last resonator (produced by correctly coupling the load to this last resonator). If a terminated wave guide is used on each side of the filter, then this fixes the size of the opening in the first and last wall of the structure of Figure 1a; or if desired, these first and last walls can be completely closed off and, as Figure 1a attempts to show, the generator and load can be capacitively coupled to the first and last resonators by probes (or magnetically coupled by loops). Whatever the method used, this generator and load coupling must be adjusted until the first and last resonators respectively have the required singly loaded Q_1 and Q_n.

- - - - -

The above three well-known circuit constants have been discussed in a previous paper (Reference 2) and methods of measuring and adjusting them have also been presented (Reference 3).

Continuing with some other examples of equivalent-constant -K structures, Figure 1b and 1c show that by discarding the waveguide concept in favor of the coupled resonator concept, additional useful, and different-looking, filters can be built with the same four resonators. Figure 1b shows the four resonators of Figure 1a rotated by 90 degrees and placed together in such a way that the openings between adjacent resonators produce the equivalent of "high side" capacity coupling. Figure 1c shows the same four resonators arranged in yet another physical configuration which will still produce the same small-percentage bandwidth filtering action: there is equivalent "high-side" capacity coupling between resonators 1 and 2, mutual inductance coupling (due to a vertical slot) between resonators 2 and 3, high-side capacity between resonators 3 and 4 and the generator sets Q_1 by being inductively coupled to the first resonator and the load sets Q_4 by being inductively coupled to the last resonator.

Figure 2 is included to stress the fact that the "different-looking" filters produced by using coaxial resonators are also equivalent to constant-K configuration filters insofar as bandpass response and required circuit constants are concerned.

Figure 3 shows a triple tuned bandpass filter which while physically in no way resembling the classical inverse arm structure, still is described by exactly the same design constants as the inverse arm structure. It is the spherical resonator which is so designed that three of its resonant modes occur at the same frequency, i.e., are degenerate. The two screws shown project into the cavity and correctly adjust K_{12} (the coefficient of coupling between the first resonance and the second resonance), and K_{23} (the coefficient of coupling between the second resonance and the third resonance). The opening on the left is of the proper size to allow the terminated waveguide shown to load properly the first resonance, i.e., to set Q_1; and the opening on the right allows the terminated waveguide shown there to load properly the last resonance, i.e., to set Q_3.

Finally, Figure 4 shows a three resonator filter using mechanical resonators for the filter elements. Here, the coefficient of couplings K_{12} and K_{23} are set by the material and diameter and "tap" point used for the quarter-wave-long (approximately) thin rods which connect two adjacent resonators. Q_1 of the first resonator is correctly set by the thin low Q resonant rod connected to the first resonator, and the last resonator is similarly correctly loaded by the low Q rod connected to it. The coils, by magnetostrictive action convert the electric energy to mechanical energy and then vice versa and because of the unfortunately poor coupling produced by this phenomenon is usually negligible electrical loading coupled into the first and last resonators.

There are many other examples of filters which at first glance do not resemble the basic inverse arm configuration, but which actually are equivalent to it; and in all of these many filters the designing engineer must know the required numerical value for all the coefficient of couplings in the structure; the required numerical value of the singly loaded Q of the first resonator and that of the last resonator; and the proper element values or physical lengths to produce the proper midfrequency, (or design information exactly equivalent to these three quantities).

III. THE TWO PAIRS OF DESIGN EQUATIONS

Figure 5 gives the two pairs of equations which, when infinite Q elements are used, supply the engineer with the above described information for two cases which often arise in practice: (a) both the generator and load are resistive, and (b) one of these, either generator or load is reactive. An example of the first case would be the microwave preselectors designed to work from a 50 ohm generator and into a 300 or 400 ohm mixer crystal; while an example of the second case would be the loaded-on-one-side-only interstage networks used in IF amplifier strips.

As indicated at the top of the figure, the attenuation shape which will be obtained will be the optimum Chebishev attenuation shown in Figure 7C. It is important to realize that the design equations of Figure 5 are given in terms of the "valley bandwidth" (BW_v) which is the bandwidth between the points on the skirt which are down by the same number of db as the peak to valley ratio (V_p/V_v). The quantity Sn is a function of the number of resonators n used and the peak-to-valley ratio desired and the reader should note that as (V_p/V_v) approachs unit Sn becomes very large

and therefore the required K's will be a large number times the fractional valley bandwidth. However, a required bandwidth at some other db down rather than at the valley db down is very often specified and it is therefore necessary to get the numerical relationship between the valley-db-down-bandwidth and the specified-db-down-bandwidth by using the shape equation at the top of Figure 5.

IV. THE DESIGN EQUATIONS FOR THE BUTTERWORTH RESPONSE

In the limit, when the peak to valley ratio is made to be zero db, Sn will become infinite and simultaneously the valley bandwidth will become zero. By correctly approaching the limit we find that the following simple changes in the equations of Figure 5 will give the design equation for the Butterworth response.

(a) In place of the valley db down bandwidth (BW_v) in the equations, use the 3 db down bandwidth BW_{3db}.

(b) Use unity for the numerator of the coefficient of coupling equations and for the denominator of the Q equations.

V. APPLICATION OF THE DESIGN EQUATIONS TO THE LOW PASS LADDER

If one writes the transfer impedance equations for the low pass ladder and compares them to those obtained for the bandpass case and uses a suitable normalizing procedure it is found that in the low pass case the frequency variable ω (i.e., $2\pi f$) is exactly equivalent to the bandpass case frequency variable

$$\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right) \equiv \frac{BW}{f_o}$$

and in the low pass case the quantity

$$1/\sqrt{L_{series}\,C_{shunt}}$$

is exactly equivalent to the well-known coefficient of coupling of bandpass coupled circuit theory; and the quantity L/R in a series arm and RC for a shunt arm are exactly equivalent to the well-known resonant frequency Q of bandpass circuit theory. Thus to apply the equations of Figure 5 to a low pass ladder we do the following:

in place of BW_v/f_o use ω_v

in place of Q_1 use L_1/R_1 or $R_1 C_1$

in place of $K_{r(r+1)}$ use

$$1/\sqrt{L_r C_{r+1}} \quad \text{or}$$
$$1/\sqrt{C_r L_{(r+1)}}$$

- - - - -

2. M. DISHAL, "Design of Dissipative Bandpass Filters Producing Desired Exact Amplitude-Frequency Characteristics," Proc. IRE v.37 pp 1050-1069 Sept. 1949.

3. M. DISHAL, "Alignment and Adjustment of Synchronously Tuned Multiple Resonant Circuit Filters," Proc. IRE v 39, pp 1448-1455, Nov. 1951

1. W. W. MUMFORD, "Maximally-Flat Filters in Waveguide," B.S.T.J. v.27 pp 684-713, Oct. 1948
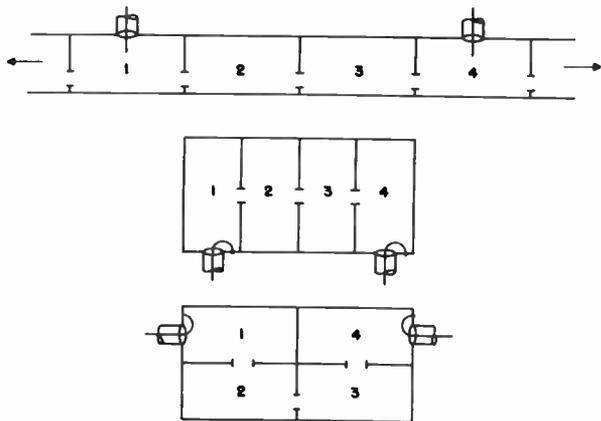
Fig. 1
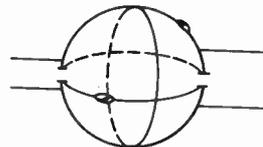
Fig. 3
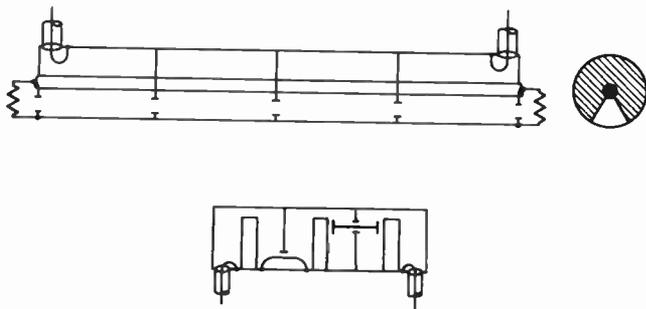
Fig. 2
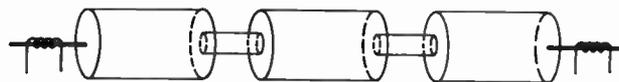
Fig. 4

TO OBTAIN THE SHAPE $\left(\dfrac{V_p}{V}\right)^2 = 1 + \left[\left(\dfrac{V_p}{V_v}\right)^2 - 1\right] \cos h^2 \left\{ n \cos h^{-1} \dfrac{\omega}{\omega_v} \right\}$

RESISTIVE GENERATOR AND RESISTIVE LOAD

$\dfrac{Q_{1,n}}{f_0/BW_v} = \dfrac{2 \sin\theta}{S_n}$ $\qquad\qquad Q_{2\rightarrow(n-1)} = \infty$

$\left(\dfrac{K_{r(r+1)}}{BW_v/f_0}\right)^2 = \dfrac{\left[S_n^2 + \sin^2 2r\theta\right]}{4\left\{\sin(2r-1)\theta\right\}\left\{\sin(2r+1)\theta\right\}}$

RESISTIVE GENERATOR AND REACTIVE LOAD (OR VICE VERSA)

$\dfrac{Q_1}{f_0/BW_v} = \dfrac{\sin\theta}{S_n}$ $\qquad\qquad Q_{2\rightarrow n} = \infty$

$\left(\dfrac{K_{r(r+1)}}{BW_v/f_0}\right)^2 = \dfrac{\left[S_n^2 + \sin^2 r\theta\right]}{\sec^2(r\theta)\left\{\sin(2r-1)\theta\right\}\left\{\sin(2r+1)\theta\right\}}$

$\theta = \dfrac{90°}{n}$ $\qquad\qquad S_n = \sin h\left[\dfrac{1}{n}\sinh^{-1}\dfrac{1}{\sqrt{(V_p/V_v)^2-1}}\right]$

Fig. 5

47

CONVENTIONAL AMPLIFIERS

William E. Bradley
Co-Director of Research
Philco Corporation
Phila., Pa.

For present purposes let us agree to define a conventional amplifier as one in which feedback is negligible so that the interstage coupling networks are effectively isolated from each other by the vacuum tubes of the amplifier. As is implied by the adjective "conventional", such amplifiers are at the present time almost universally used in electronic equipment. The vacuum tube used is almost always a pentode. Triodes driven from the grid require neutralization and are more difficult to utilize in such circuits. Grounded grid triodes however isolate the input from the output fairly effectively and are here considered to be conventional amplifiers. Most of the following discussion deals with amplifiers consisting of simple interstage coupling networks separated by pentodes.

In order to compare conventional amplifiers with other types, consider first their performance as a function of frequency. The over-all amplification of a set of such stages is the product of all of the tube transconductances multiplied by the product of the separate transfer impedances of the interstage coupling networks.
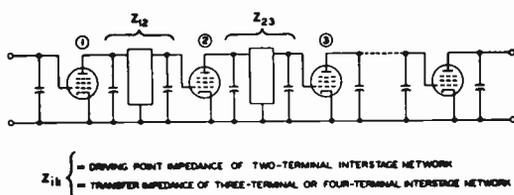
Fig. 1 - Conventional amplifier chain.

The limitations of conventional amplifiers in gain and bandwidth result from the shunt capacitances of the input and the output terminals of each tube to ground. These so-called "parasitic" capacitances dominate the design of such amplifiers for most purposes.

To obtain the highest stage gain, the transfer impedance of the interstage coupling network must be made as high as possible over the desired band, subject to the condition that the shunt capacitances at each end of the network must include the parasitic capacitances of the tubes. Network theory shows that there are limits to the transfer impedance which can be obtained under this condition; it also shows that the transfer impedance is likely to fall off rapidly outside of the band for which it is designed.

This leads to the most outstanding peculiarity of conventional wide-band amplifiers, which is that they are part amplifier, part filter. Their ability to reject unwanted frequencies close to their wide pass-band is often useful as, for example, in television receiver design. It is also true that the amplitude and phase characteristics through the pass-band are subject to convenient control with a minimum of design difficulty.

The extremely direct and flexible design of the frequency characteristics of conventional amplifiers is due to the fact that they are really chains of simple filters each of which is easy to design and adjust. The over-all transfer impedance of the amplifier is simply the product of the separate stage transfer impedances and therefore very complex over-all characteristics can be obtained reproducibly from such chains.

At this point two main branches to the subject at interstage coupling design can be distinguished. The first branch deals with the design of the best possible single stage. Here the ideal is usually to obtain flat frequency response and linear phase characteristic across the entire pass-band. The final video stage of a television receiver is often designed this way as are many of the video amplifiers re-

quired in television studio and broadcasting equipment.

$$R_0 = \frac{2}{C \Delta\omega} \; ; \; L = \frac{2R_0}{\Delta\omega} \; ; \; \Delta\omega = \text{RADIAN CUT-OFF FREQUENCY}$$

-a-
TWO-TERMINAL INTERSTAGE NETWORK

$$R_0' = \frac{2}{C'\Delta\omega} \; , \; L' = \frac{2R_0'}{\Delta\omega} \; ; \; \Delta\omega = \text{RADIAN CUT-OFF FREQUENCY}$$

-b-
FOUR-TERMINAL INTERSTAGE NETWORK
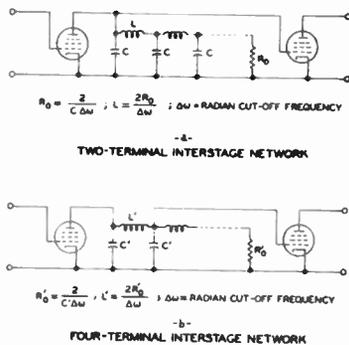
Fig. 2
Parasitic capacitances absorbed in
filter structure.

The interstage couplings for such amplifiers are wave filter sections which have shunt capacitances across each section. These shunt capacitances absorb the parasitic capacitances of the tubes while the filter properties of the network ensure equal transmission across the pass-band. Very good approach to ideal characteristics can be obtained with a single stage in this way if enough circuit elements are used in the structure.

Usually only a few circuit elements are used and the result is a compromise between ideal performance of the single stage and complexity of structure.

The voltage amplification per stage through the pass band obtained in this way with a four-terminal interstage coupling is

$$A_1 = \frac{2 \; g_m}{C \; \Delta\omega}$$

where C is the mean parasitic capacitance of the tube. The amplification is half of this for the same tube type with a two-terminal interstage coupling. Since then the shunt capacitance is the sum, not the mean of the two parasitic capacitances.

The second branch of the general subject of interstage coupling design takes as its objective optimum over-all performance of a set of cascaded stages. The frequency response of any one stage of the set may be very far from ideal. It is usual in such amplifiers to employ very simple interstage coupling networks,

seldom using more than two inductances per stage and usually only one. The designer of this form of amplifier is principally concerned with how the stage responses can be made to fit together to achieve the results which he desires. Because the interstage couplings are simple and non-interacting it is possible to control the complex resonant frequencies of these networks with great precision. These complex resonances correspond to poles and zeros of the over-all amplification of the set of stages. One of the greatest advantages of this type of amplifier is the ease with which these poles and zeros can be placed in prescribed patterns in the complex frequency plane to realize accurately design goals.

The gain obtained in this way is not greatly different from that obtainable with ideal filter interstage couplings and carries the advantage of simplicity of circuit adjustment and flexibility of design.

$$R = \frac{1}{C \Delta\omega} \qquad L = R^2 C$$

λ PLANE
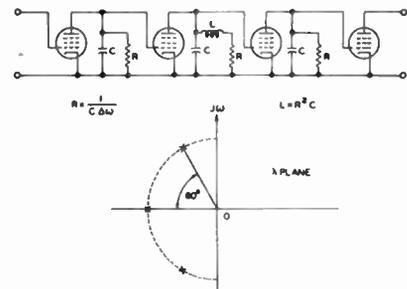
Fig. 3 - Three-stage maximally flat amplifier.

As a simple example consider the amplifier shown in Figure 3. Here a low pass amplifier achieves a maximally flat amplitude characteristic with only one "peaking coil". Its three poles are arranged on a semicircle. In spite of its simplicity the mean low frequency gain per stage of this amplifier is as high as can be obtained with any two-terminal interstage coupling networks. Since it has only three poles, however, the frequency response does not remain flat all of the way to the nominal band limit, $\Delta\omega$, but is down 3 db at that point.

49
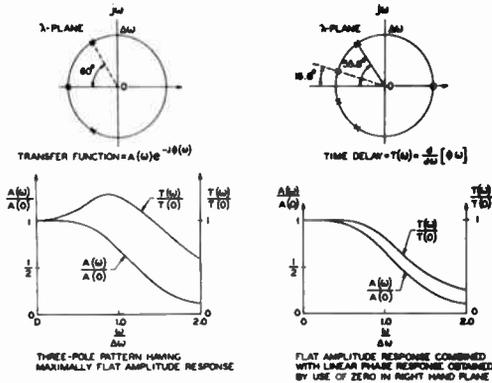
Fig. 4 - Use of zero in right-hand plane.

Fig. 5
Ideal pass band obtained by distribution
of poles and zeros on circumference of circle.

The design possibilities of so-called
conventional amplifiers are still not
common knowledge to design engineers. For
example, it is not well-known that not
only the amplitude but also the phase
characteristic may be controlled through
the pass-band to achieve equal amplifica-
tion and constant time delay for all pass-
band frequencies. A simple case is illus-
trated in Figure 4 where the gain and
phase shift for the three pole, maximally
flat response can be compared with those
resulting from four poles and one zero in
the right-half plane. The amplitude re-
sponses are not very different for the two
cases but the phase characteristic of the
second pattern is much straighter, corre-
sponding to improved transient response,
with no loss of selectivity. A general
recipe for obtaining perfect amplitude and
phase characteristics through the pass-
band is shown in Figure 5.

Interstage coupling networks providing
zeros in the right-half plane may be
readily constructed using principles
similar to those used in designing bridged
T equalizers or resistance compensated
filters. A commonly used example is the
resistance compensated trap circuit used
in television receivers which provides a
zero on the jω axis for adjacent channel
attenuation. Changing the value of the
compensating resistor in such a direction
as to increase its effect moves the zero
over into the right-half plane.

An important property of conventional
amplifiers in general is that the trans-
conductances of the tubes do not affect

the shapes of the gain and the phase
characteristics. This is not generally
true of other forms of amplifiers involv-
ing feedback or, more generally, in which
the pole and zero locations are functions
of transconductance. Hence conventional
amplifiers may have their gain controlled
over wide limits with no appreciable change
in transient response. This is important
in amplifiers intended for pulse communica-
tion or television use.
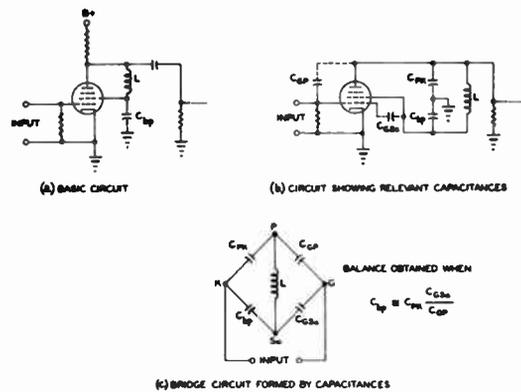
Fig. 6
Compensation for grid-to-plate
capacitance.

Even small traces of spurious feedback
from plate to grid can spoil the precision
of design of conventional pentode ampli-
fiers, especially if the gain per stage is
high. Figure 6 shows the screen neutral-
ization circuit used in a television re-
ceiver band-pass amplifier with type 6CB6

50

tubes. This circuit is able to substantially eliminate interaction between the interstage coupling networks used. The intermediate frequency amplifier pole and zero locations used in the design of a production receiver are shown in Figure 7.
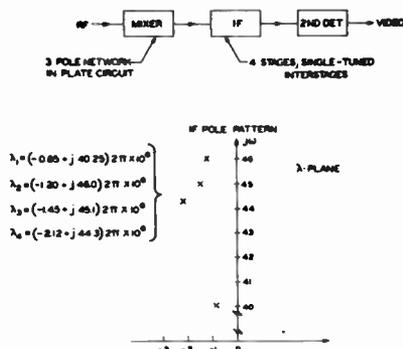
Fig. 7 - Pole pattern of the TV-90 IF.

Fig. 8 - Pentode figure of merit.

A pentode tube for a conventional wide-band amplifier has a figure of merit which varies directly with the transconductance and inversely as the parasitic capacitance of the tube. From a practical point of view, the figure of merit defined in this way is slightly misleading since the external circuits add some parasitic capacity but, of course, do not affect the transconductance. It follows that a high transconductance is a more valuable asset for many applications than a low shunt capacitance.

The figure of merit for a number of conventional pentodes is illustrated in Figure 8. It is interesting to notice that the figure of merit of the subminiature pentode 5702 is very high while the shunt capacitances of this tube are so low as to be very much affected by the capacitances of the associated wiring. When these capacitances are added the 5702, while still a very good pentode, is not greatly different in performance from several other pentodes.

The upper frequency limit to conventional tube operation is probably somewhere above four thousand megacycles, if grounded grid triodes are considered to be conventional. Extreme delicacy of construction of the tube elements and difficulty of connection to them discourage at present

the application of tubes of this type above the UHF band. The UHF band itself seems to be within reach of conventional structures although large scale commercial use of amplifiers in television tuners for this band is only beginning.

The good and bad points of conventional amplifiers can be summarized as follows. On the bad side it can be said that:

1. Grid to cathode capacitance is subject to some variation in tube production. This makes the frequency characteristics of conventional wide-band amplifiers change slightly when tubes are replaced without circuit readjustment.

2. They tend to be slightly non-linear.

3. Their absolute level of gain is subject to drift with change of g_m.

4. They have an amplification per stage which varies inversely as the bandwidth so that conventional amplifiers are of doubtful utility for bandwidths greater than about thirty megacycles using presently available commercial tubes.

The good points of conventional amplifiers are:

1. The ease and accuracy with which good frequency response and transient performance can be obtained in a predetermined frequency band together with rejection of frequencies outside this band.

2. Their relatively low cost.

3. The independence of their frequency characteristics and transient response with respect to changes of tube transconductance.

51

BROAD-BAND FEEDBACK AMPLIFIERS

Harold N. Beveridge
Raytheon Manufacturing Company
Newton, Mass.

The introduction of television and radar imposed the problem of designing wide-band i.f. and video amplifiers. Techniques using double-tuned circuits, stagger tuning, shunt peaking, and series peaking are fairly well known. The use of resistive feedback from plate to grid in i.f. and video amplifiers was pointed out by Harold A. Wheeler[1]. The practical design of this type amplifier was started at the National Research Council in Ottawa, Canada, by the author and A.J. Ferguson and later continued at the Naval Research Laboratory in Washington[2]. The purpose of this morning's paper is to indicate the simple basic circuits and their performance and to discuss some of the second order effects so important in amplifier design.
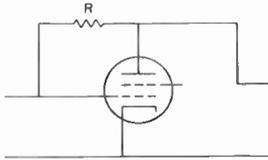
Fig. 1

Fig. 1 shows the basic feedback configuration. The coupling networks used between such stages may be very simple and still give effective gain bandwidth performance.
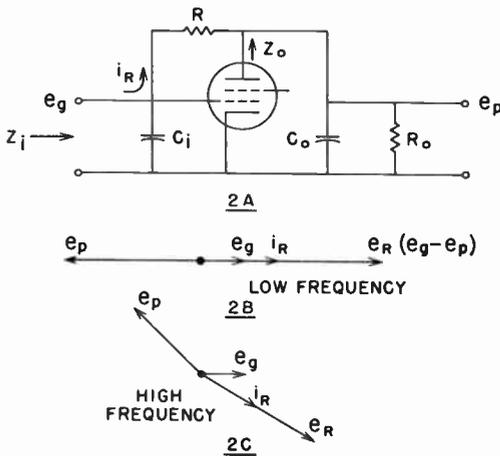
Fig. 2

A qualitative understanding of this type of feedback can be obtained by considering the behavior of a single video stage. This is shown in Fig. 2A. R_o is the plate load for the desired gain. Consider the input impedance Z_i due to the feedback resistance R.

Fig. 2B illustrates the low frequency case. It is apparent that the impedance presented is purely resistive (i_R in phase with e_g). In a manner similar to Miller effect, the impedance is given by

$$Z_i = \frac{R}{1 + A}$$

A is the absolute voltage gain of the stage.

Next consider the circuit behavior at a frequency corresponding to the 3 db point in the output circuit, Fig. 2C. The feedback current i_R is about 70% its previous value and it lags 45°. Thus, the input impedance Z_i shows an increased resistive component and a large inductive component. This will tend to "tune out" C_i. This type of feedback is most useful if a constant current generator is used to drive Z_i. This allows the rising impedance characteristics of Z_i at higher frequencies to result in an increase in gain to compensate the fall-off of gain due to the decreasing impedance in the output circuit. This is the basic mechanism by which this type of feedback produces broad-banding.
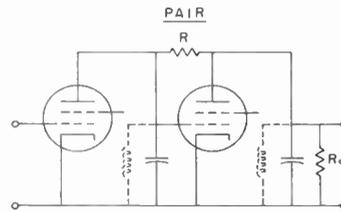
Fig. 3

Adding a tube to drive the circuit previously shown produces what has come to be called a "feedback pair". This is shown in Fig. 3. R is chosen so that it effectively loads the plate of the first tube with a value about equal to R_o. Thus, the low frequency voltage gain of each stage is about the same. By adding inductances to tune both circuits to the same frequency, the amplifier becomes an i.f. or band-pass amplifier.

There is an exact correspondence in these amplifiers between the video and i.f. case. The total bandwidths are identical for both. That is, the semi bandwidth of the i.f. is just one-half the video bandwidth. This follows from the normal situation of having reactance change at twice the rate when deviating from a center frequency (band-

pass case) than occurs when increasing from zero frequency (low-pass case).

The transmission characteristics and gain bandwidth product of a feedback i.f. pair and a stagger pair are nearly alike.

The pass-band shape of a pair can be adjusted by varying the value of Ro.

Bandwidth may be varried by varying R.
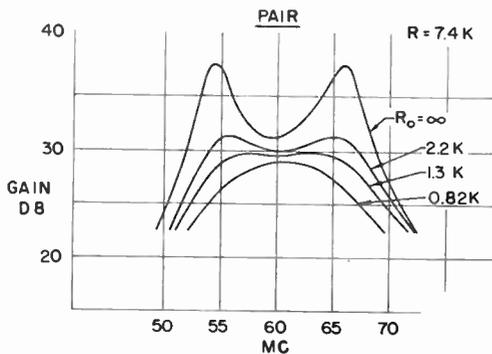
Fig. 4

The effects of the termination are shown for the i.f. case in Fig. 4. When Ro = ∞, the response is markedly doubled peaked. There is a value for Ro that gives a flat response. Lower values of Ro result in a single peaked response. It is not obvious, but equivalent results will be obtained by removing Ro and replacing it with an equivalent resistance across the first tuned circuit, or by placing 2Ro across each circuit.
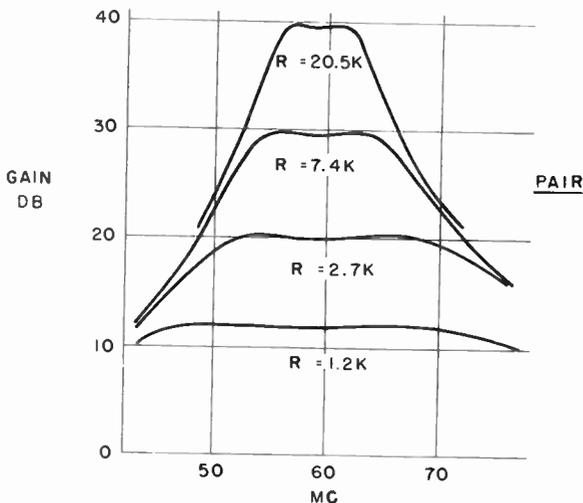
Fig. 5

Fig. 5 shows on an absolute scale the gain and measured bandwidth obtainable using 6AK5 tubes operated at a Gm of 4,500 umhos, a total circuit capacity of 9.5 UUF, feedback resistors as indi-

cated, and Ro adjusted for approximately flat response. Capacitance shunting the feedback resistance establishes a practical lower bandwidth limit of 5 mc. The greatest bandwidth shown is 30 mc at 1 db, corresponding to a gain of 12 db for two stages.
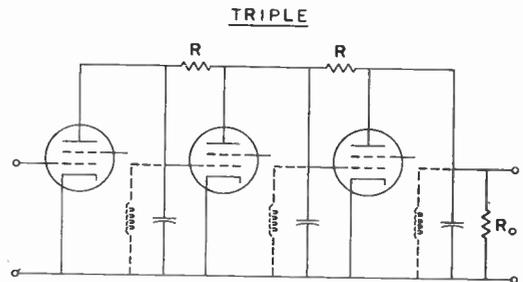
Fig. 6

Fig. 6 shows a feedback "triple". For the i.f. case, it is practically identical to a stagger triple. As in the case of a pair Ro in the triple may be placed across the first circuit instead of the last without altering the bandpass characteristics. However, unlike the pair, if 2Ro is placed across the first and last circuits, the amplitude response of the amplifier becomes unsatisfactory. It is difficult to compare exactly gain bandwidth performance of pairs and triples as the result is somewhat dependent upon the number of stages used. Results were measured on a six-stage amplifier using in one case three pairs and in the other case two triples. When adjusted for equal gain, the triples gave 20% more bandwidth.

Clearly the number of stages may be increased beyond three. The theoretical case for an infinite chain has been worked out by A.J.Ferguson[3]. The longer chain produces a small increase in gain bandwidth product compared with triples.

Gain control is a problem in feedback amplifiers. The bandwidth of these amplifiers is a function of the Gm of those tubes having resistance in shunt. Gain control is usually achieved by controlling the Gm of a tube around which there is no feedback. In the case of a pair or triple, the first tube is available for gain control purposes. In some cases, it is allowable or even desirable to have bandwidth increase when gain is decreased. The introduction of the germanium diode allows such control if placed across each tuned circuit (i.f. case). Varying the current through the diode varies the degree of loading across each circuit.

Inasmuch as the bandwidth of a feedback amplifier is a function of tube Gm, the question arises as to the practical problems with such amplifiers. Tests were conducted on two types of fixed tuned 60 mc i.f. amplifiers. One amplifier contained two feedback triples, the other two staggered triples. Measurements of gain and band-

width were made on each type amplifier with a
large number of 6AK5 tubes. The results are shown
in Fig. 7. It is seen that the spread in bandwidth
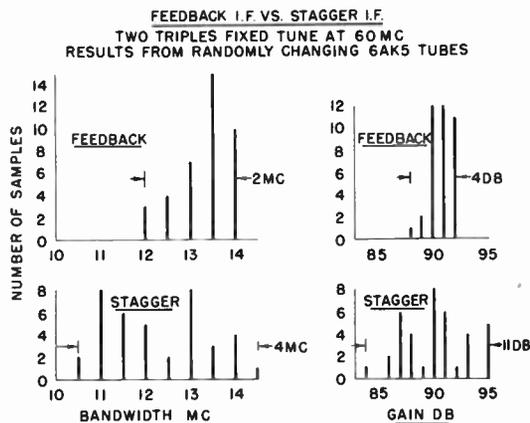
Fig. 7

was only 2 mc for the feedback amplifier compared
with 4 mc for the stagger tuned amplifier. The
feedback amplifier exhibits even better char-
acteristics with respect to gain stability showing
a change of only 4 db compared with 11 db in
stagger tuned. Feedback amplifiers, in general,
show good repeatability in production.

The amplifiers so far described have employed
the simplest possible interstage coupling networks.
It is possible to use more complex coupling net-
works. By splitting the circuit capacitance into
two parts, a further improvement in gain bandwidth
product may be achieved. The factor of improvement
may approach two.
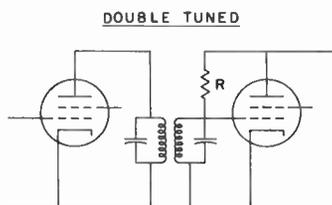
DOUBLE TUNED

Fig. 8

Fig. 8 illustrates a typical i.f. stage using
a double-tuned circuit as the interstage coupling.
R.Q. Twiss[4] has worked out the theoretical
analysis of such an amplifier and verified
experimentally the results. It should be pointed
out, however, that the more complex networks with
larger gain bandwidth factors have poorer transient
characteristics.

An improvement in gain bandwidth product can
be achieved in the video case by splitting the
capacitance with an inductance as shown in Fig. 9.
If the capacitances are equal, an improvement

VIDEO
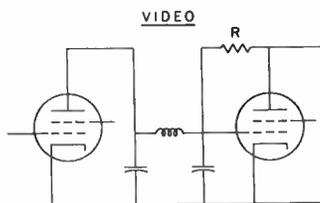
Fig. 9

factor approaching two is possible.

Second order effects play a very large part
in the practical design of wide-band amplifiers,
and particularly so in the case of feedback
amplifiers. The treatment of the impedance be-
tween cathode and ground, the transit angle in
the tube, and the capacitance shunting the feed-
back resistance are all important.

Grid input loading in pentodes at high
frequencies is generally appreciated as being
present. However, it is of interest to know how
much of this is due to transit time effects and
how much due to cathode lead inductance. This
has been measured.
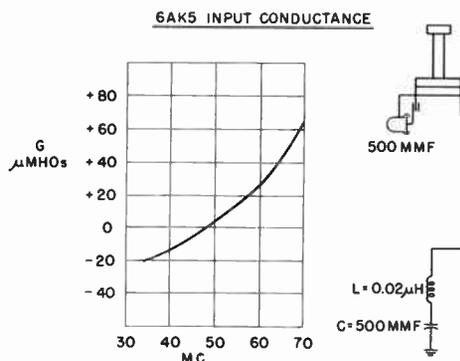
Fig. 10

Fig. 10 shows the input conductance of a
6AK5 tube as a function of frequency with a 500
UUF cathode by-pass capacitor. It should be
noted that the input conductance is zero at 47 mc.
The frequency at which the total cathode lead
inductance series resonated with the by-pass
capacitor was measured and found to be about
48 mc. At this frequency the impedance between
cathode and ground is practically zero, and the
effect of cathode lead inductance is zero. As
the input conductance is also practically zero,
it is concluded that transit time loading is small
compared with loading due to cathode lead in-
ductance. The change in input conductance would
tend to produce a small slope on the flat top of
an amplifier. Reactance between cathode and
ground may be used either to advance or retard the
plate current phase in a tube relative to the
applied voltage between grid and ground. It is
quite practical to retard or advance the plate

54

current by 4 or 5 degrees.

At i.f. frequencies in feedback amplifiers, there are two second order effects which complicate design. Fortunately, they are of opposite sign and can be used to cancel one another. They are transit angle effects in the tube and the unavoidable capacitance which shunts the feedback resistance.
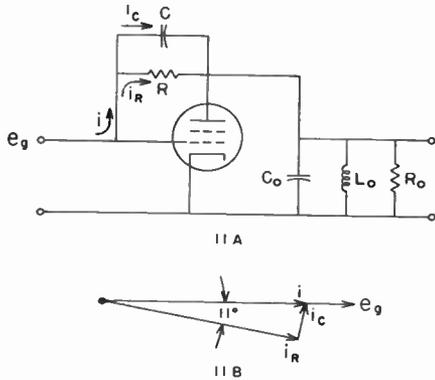
Fig. 11

First, let us consider transit angle. This has been measured on the 6AK5 at 60 mc and found to be 11°. Fig. 11B illustrates the relationships. If left uncorrected, this leads to an assymmetrical response. Actually capacity always shunts the feedback path. For a certain capacity, the current i_C will add to i_R to produce i in phase with e_g. An approximate correction of this sort is desirable.

It is worth while considering a numerical example. The highest practical value of R is about 20,000 ohms. The total capacity, which should shunt it, is 0.04 UUF (for 11°correction). The tube itself has approximately 0.02 UUF. This leaves 0.02 UUF for the end-to-end capacity of the feedback resistor. Actually most half watt carbon resistors have end-to-end capacities of the order of 0.3 UUF, about fifteen times too great.

A resistor suitable for feedback amplifiers was developed by the International Resistance Company. It is known as the MPM and has a capacity of 0.02 UUF. Probably there are now other deposited film resistors suitable for this purpose.

For very wide-band amplifiers, the value of the feedback resistor may be of the order of 1,000 ohms. In this case the capacity in the ordinary half watt carbon composition resistor is about right to compensate for transit angle. The foregoing may lead the designer to ask what can be done for intermediate values of feedback resistors. Actually, the practical answer to this problem is to select the most applicable of these two resistor types. Further adjustment of effective transit angle may be obtained by

choosing the cathode by-pass capacitor in such a way as to leave the total cathode path to ground either slightly inductive or capacitive.

Gain bandwidth product has been measured on feedback pairs and triples. The circuit capacity was 9.5 UUF and a tube Gm of 4,500 umhos. With voltage gains per stage of two or three the product for both pairs and triples was about 55 mc times. With voltage gains of eight per stage, the product for pairs was 65 mc times and for triples 75 mc times. All the above is based on 3 db bandwidths and 100 db of gain.

VOLTAGE GAIN FOR FEEDBACK COMBINATIONS
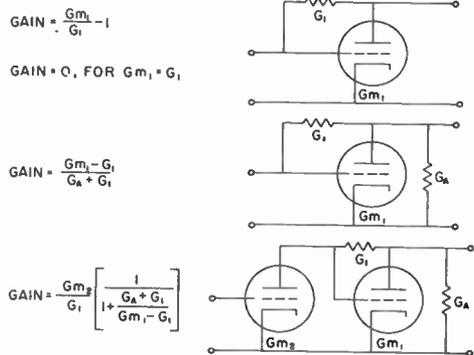
Fig. 12A
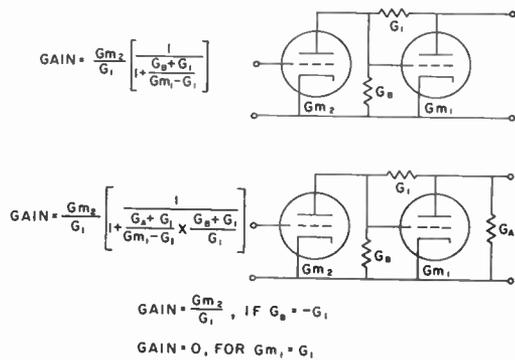
Fig. 12B

Two important parameters in amplifier design are, of course, overall bandwidth and voltage gain per stage. The product of these two quantities has been given above. Relatively simple expressions for voltage gains are given in Fig.12. These apply only for the low frequencies in video amplifiers and the center frequency in i.f. amplifiers.

The exact mathematical analysis of feedback amplifiers is cumbersome. However, with knowledge of the second order effects, gain bandwidth product, and the gain equations in this paper, the practical design of such amplifiers is rel-

55
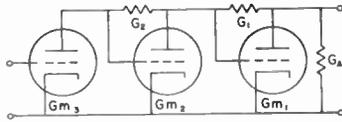
$$\text{GAIN} = \frac{Gm_1 Gm_3}{G_A G_2}\left[\frac{1-\dfrac{G_1}{Gm_1}-\dfrac{G_2}{Gm_2}+\dfrac{G_1 G_2}{Gm_1 Gm_2}}{1+\dfrac{G_1}{G_A}\left(1+\dfrac{Gm_1}{Gm_2}\right)+\dfrac{G_1}{Gm_2}}\right]$$

Fig. 12C

$$\text{GAIN} = A = \sqrt{\frac{Gm}{G}} - 1$$

FOR ONE STAGE OF INFINITE CHAIN, OR WHEN TERMINATED IN $G_0 = G(A+1)$

$$\text{GAIN} = A = \sqrt{\frac{Gm}{G}+\frac{G_1}{G}+\frac{G_1^2}{4G^2}} - \frac{G_1}{2G} - 1$$

FOR ONE STAGE OF INFINITE CHAIN, OR WHEN TERMINATED IN $G_0 = G(A+1) + G_1$
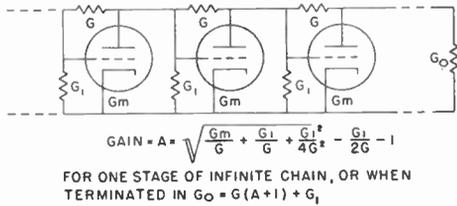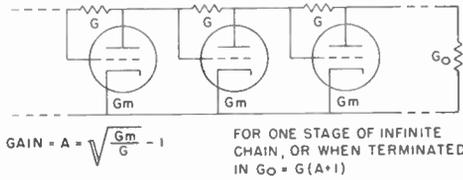
Fig. 12D

atively simple. These amplifiers, within limits, using resistive feedback eliminate the need for shunt peaking, the use of double-tuned circuits, and stagger tuning. Resistive plate to grid feedback amplifiers are inherently very stable and show good repeatability in production.

[1]Harold A. Wheeler, "Wide-Band Amplifiers for Television", Proceedings of the I.R.E., Vol. 27, pp 429-438, July, 1939.

[2]H.N. Beveridge, "Information on Broad-Band Feedback i.f. Amplifiers", Combined Research Group, Naval Research Laboratory Report CRG-93, Oct. 22, 1945.

[3]A.J. Ferguson, "The Theory of i.f. Amplifiers with Negative Feedback", Report No. PRA-59, National Research Council of Canada, Ottawa, Canada.

[4]R.Q. Twiss, "The Theoretical Design and Experimental Response of Single and Coupled Circuit Negative Feedback i.f. Amplifiers", Report T1649, Telecommunications Research Establishment, Malvern, England.

TRANSISTOR AMPLIFIERS

R. L. Wallace, Jr.
Bell Telephone Laboratories, Inc.
Murray Hill, N.J.

ABSTRACT

Mr. Wallace will discuss some of the properties of transistors which result in limitation of bandwidth and will indicate the order of performance which has been achieved.

DISTRIBUTED AMPLIFIERS

W. G. Tuller
Emmett H. Bradley
Melpar, Inc.
Alexandria, Virginia

With the advent of greater interest in milli-microsecond oscillography, the need for broadband amplifiers having increasingly high upper cut-off frequencies has become apparent. The conventional technique of cascading amplifier stages is inadequate to provide sufficient broadband operation since the maximum operating frequency of such a system is determined by its maximum "gain-band-width product".[1,2,3] However, the distributed amplifier is not limited by the conventional restrictions on bandwidth,[4,5,6] thus making possible the realization of amplifiers having a flat frequency response from d.c. to several hundred megacycles.

The basic philosophy of the distributed amplifier is not new; it was originally discovered by W. W. Percival[7] in 1935. Not until 1948 was this circuit recalled to the attention of an electronic age, hard pressed with the need for amplifiers having greater bandwidths. In their original paper,[8] Ginzton, Hewlett, Jasberg and Noe presented a comprehensive design procedure which ultimately led to the realization of this broadband amplifier.

Following the publication of their paper, considerable interest was created in this new circuit oddity, which applied traveling wave concepts to obtain amplification at video frequencies. Theoretically, such an amplifier could be designed to have as large a gain as desired over an arbitrarily wide bandwidth. It was soon found, however, that this was not the case since the upper cut-off frequency is limited by the high frequency characteristics of the tubes as well as by the geometric configuration of their envelopes.

In recent years, considerable work has gone into the development of wide band distributed amplifiers using multi-grid tubes.[9,10,11,12,13,14,15] At the present time, these amplifiers are commercially available with bandwidths in excess of 200 megacycles.[16,17] Several government agencies have developed distributed amplifiers whose upper cut-off frequencies were as high as 400 megacycles.[18,19] The men largely responsible for the development of these circuits have expressed their belief that this frequency represents the maximum upper cut-off frequency for distributed amplifiers using commercially available multi-grid tubes. Distributed amplifiers are limited to use at frequencies below 500 megacycles by the characteristically large effect of lead inductances, and grid loading in tetrode and pentode tubes.

The effects of grid and plate lead inductances can be largely overcome by using m-derived transmission lines having a negative mutual coupling. Grid loading resulting from the transit angle and the presence of cathode lead inductance in the tube can be partially eliminated by inductively loading the screen grids and by using tubes having twin cathode leads. Even when great care is given to the consideration of these effects, the attenuation along the plate and grid lines remains sufficiently large to limit the maximum operating frequency of these amplifiers. A representative group of distributed amplifiers along with their characteristics are tabulated in Table I.

In order to increase the operating frequency of the distributed amplifiers of the future, there are at least three alternatives.

1. A multi-grid tube can be designed having symmetrical lead pairs with small lead inductances, a large transconductance, small grid loading at high frequencies, and a large figure of merit (ratio of transconductance to input capacitance).

2. A tube development program can be initiated to develop a new type of vacuum tube similar to the transmission line tube proposed by V. J. Fowler.[20]

3. New techniques for employing triodes in distributed amplifiers can be developed.

The first and second proposals do not provide an immediate solution to this problem since tube development programs are time consuming and expensive. The latter approach appears more attractive for a number of reasons.

1. In recent years, triodes have been developed for operation at ultra high frequencies.

2. Many of these tubes have been designed to minimize lead inductances.

3. The effect of grid loading in these tubes is considerably less than in a multi-grid tube.

In the fall of 1950, Melpar, Inc., began the development of a distributed amplifier using ultra high frequency triodes. The problem is, however, more difficult than it might first appear. If the multi-grid tubes in a distributed amplifier are directly replaced with triodes, a considerable portion of the signal in the plate line is fed back into the grid line through the grid-plate capacitance of the tubes. This feedback becomes sufficiently large and of the proper phase to cause instability due to this inter-element coupling when more than a few tubes are utilized. For that reason a new technique had to be developed in order to eliminate this effect.

A circuit using paired triodes and called a "paraphase distributed amplifier" has been developed which does isolate the grid and plate lines. From the schematic diagram shown in Figure 1, it can be seen that the tube pair is comprised of a cathode follower driving a grounded grid amplifier. When a number of these tube pairs are spaced equidistantly along two transmission lines having an equal velocity of propagation, amplification is obtained in a manner similar to that in a conventional distributed amplifier. There are several advantages of this circuit:

1. The system uses triodes which minimize the effect of grid loading and transit time, and thereby make it possible for the amplifier to have a bandwidth of approximately two and one half times

that of any video or distributed amplifiers in use at the present time.

2. The problem of coupling between the grid and plate lines is avoided by this particular combination of tubes, thus reducing the possibility of instability in the system.

3. Since the input tube of the paraphase tube pair behaves very much like a cathode follower, the system provides amplification without inversion and operates satisfactorily with maximum input signals of approximately three times those used in conventional grounded cathode amplifiers.

In order to verify the theoretical investigations, a five tube pair paraphase distributed amplifier has been designed and tested. Preliminary tests on the five tube pair stage revealed a low frequency gain of 3 db. The voltage gain for the unit was flat within ±4 db up to 940 megacycles with the exception of a -7.8 db dip at 700 megacycles (Figure 2). The shape of the amplifier response appeared to be quite sensitive to changes in the plate voltage. By increasing the number of tube pairs, the stage gain can be increased proportionally without a serious reduction in bandwidth. The 5675 ultra high frequency triode was chosen for use in this amplifier because of its favorable figure of merit and geometric construction. Considerable care was required in the layout of the chassis in order to reduce the coupling effects between the transmission lines and to minimize the effect of stray capacitances. In Figure 3, the chassis layout for this breadboard amplifier is shown.

The 5675 pencil triode is not well suited for use in a driver or output stage. The maximum grid swing of this tube is small because of the relatively small operating voltage. Since the transconductance of the 5675 is only 6200 μmho, an impractical number of tubes would be required to develop an appreciable power gain. Furthermore, the large output signal traveling down the plate line would cause the instantaneous plate potential of the tubes to fluctuate greatly. Since the transconductance of these tubes is quite sensitive to changes in plate voltage, serious signal distortion would result. For these reasons, parallel work has been done on a paraphase distributed amplifier using 2C39A triodes. A stage using five tube pairs was designed similar to the previous model.

Considerable difficulty was encountered in the layout of this stage because the 2C39A has one side of the filament connected internally to the cathode. Since the triodes in the paraphase circuit operate with the cathodes above ground potential, a method of supplying the filaments from an isolated supply had to be developed before the unit could be tested. This problem was solved temporarily by the addition of a cathode line on which the needed filament voltage was supplied from an isolated storage battery. (The addition of a cathode line in the distributed amplifier should be avoided if possible because its low impedance reduces the gain per tube pair.) From experimental tests on

the unit, it has been found that the low-frequency voltage gain is 8.6 db, and the amplifier response is flat within ±2.5 db up to 470 megacycles (Figure 4). In addition to the investigation of its frequency response, the 2C39A stage was subjected to an impulse test. A test pulse, five millimicroseconds in duration and having a rise time of half a millimicrosecond, was fed into the amplifier while the output of the amplifier was applied to the plates of a high speed oscilloscope. Although the vertical deflection on the cathode-ray tube was extremely small, it was possible to determine with the aid of a cylindrical lens that the rise time of the pulse transmitted by the amplifier was not appreciably different from that of the original pulse.

There are several areas of uncertainty that must be resolved before success of the paraphase distributed amplifier can be predicted, and there are several difficulties to be overcome:

1. It can be shown that nearly thirty tube pairs (5675's) are required to give an optimum stage gain of e, the Naperian logarithmic base, into a load impedance of 50 ohms. It has been pointed out that in conventional distributed amplifiers using pentodes there is a limit to the number of tubes that can be used in one stage; beyond this number, stage gain decreases because of line losses. In an amplifier analysis made at the Allen B. DuMont Laboratories,[21] it was found that the maximum gain was obtained with twelve 6AN5 tubes per stage; larger numbers of tubes afforded less total gain. This particular number is not directly pertinent to the paraphase distributed amplifier, but it is important to realize that the number of tubes per stage cannot be increased without limit. Whether thirty pairs can be used will not be known until a large stage has been built and tested.

2. A serious problem is the inefficiency of coupling between stages. It can be shown that the total gain of several cascaded stages is proportional to $(1 + \rho)^m$, where ρ is the interstage reflection coefficient and m is the number of stages. For a plate line of characteristic impedance 120 ohms and a grid line of 50 ohms, $\rho = -0.412$. With five stages of 5675's, the quantity $(1 + \rho)^m$, which represents the effective coupling efficiency of the complete amplifier, is 0.07, or 7 per cent. If this coupling efficiency can be improved, a considerable reduction will result in the size of the amplifier, but there is no basis for optimism on this point. Although an awareness of the problem has existed from the very beginning, no satisfactory solution has been devised.

3. A satisfactory method must be found for supplying the heaters of the 2C39A's. The present technique of using an isolated battery or transformer is unattractive for a large stage. Moreover, the use of any cathode line is undesirable because it reduces gain by a factor of about one half. There is further question whether sufficiently high frequencies can be passed through a paraphase stage of 2C39A's because of the diff-

culty of maintaining sufficient impedance between cathodes and ground. A low capacitance transformer can be designed to supply each tube pair separately and eliminate the need for a cathode line.

4. Considerable difficulty has been encountered in obtaining line terminations that would be resistive and even nearly constant in value over the range of frequencies up to 1000 megacycles. A parallel program has been undertaken to develop terminations of a novel design for the needs of this amplifier. If it performs as anticipated, proper termination of the lines will no longer be a serious problem.

It is questionable whether the low "gain" limitation of the paraphase distributed amplifier can be resolved in the near future. It is of interest, however, to note that further work on triode distributed amplifiers may remove some of these difficulties. A "three dimensional distributed amplifier", which is now under study, shows considerable promise (Figure 5).

In conclusion, it can be said that distributed amplifiers have developed rapidly since 1948. There is reason to believe that bandwidths of over 1000 megacycles will be obtainable in the future with the introduction of new and more adaptable tubes. In the meantime, sufficient evidence has been uncovered to warrant a close investigation of the use of triodes in designing wide band distributed amplifiers.

References

1. H. A. Wheeler, "Wide-Band Amplifiers for Television," Proc. I.R.E. 27, pp. 429-38, 1939.

2. W. W. Hansen, "On Maximum Gain-Bandwidth Product in Amplifiers," J. App. Phys. 16, pp. 528-34, 1945.

3. H. W. Bode, Network Analysis and Feedback Amplifier Design, Ch. 17, Van Nostrand, New York, 1945.

4. W. C. Percival, "Thermionic Valve Circuits," British patent 460562, July 24, 1935 - January 25, 1937.

5. E. L. Ginzton, W. R. Hewlett, J. H. Jasberg, J. D. Noe, "Distributed Amplification," Proc. I.R.E. 36, pp. 956-69, 1948.

6. G. E. Valley, Jr., H. Wallman, Vacuum-Tube Amplifiers, p. 83, McGraw-Hill Book Co., New York, 1948.

7. W. C. Percival, loc. cit.

8. E. L. Ginzton, W. R. Hewlett, J. H. Jasberg, J. D. Noe, loc. cit.

9. F. Kennedy, H. G. Rudenberg, "200-Mc Traveling Wave Chain Amplifier," Electronics 22, p. 106, December, 1949.

10. W. W. Horton, J. H. Jasberg, J. D. Noe, "Distributed Amplifiers, Practical Considerations and Experimental Results," Proc. I.R.E. 38, pp. 748-53, 1950.

11. A. P. Copson, "A Distributed Power Amplifier," Elec. Engin., 69, pp. 893-98, October, 1950.

12. D. O. Pederson, "The Analysis and Synthesis of Distributed Amplifiers with Ladder Networks," Technical Report 34, Electronics Research Laboratory, Stanford University, 1951.

13. G. G. Kelley, "A High-Speed Synchroscope," Rev. Sci. Instr., 21, pp. 71-76, January, 1950.

14. J. M. Pettit, D. O. Pederson, "Distributed Amplifier Theory," Proc. Symposium Modern Network Synthesis, ONR, pp. 202-14, April 17, 1952.

15. D. O. Pederson, "The Distributed Pair," Trans. I.R.E., pp. 57-67, December, 1952.

16. Hewlett Packard Company, Laboratory Instruments, Catalog 20-A, pp. 62-63, 1950.

17. Spencer Kennedy Laboratories, Inc., "Chain Amplifier," Bulletins 212-3 and 202P-3.

18. J. Weber, "Distributed Amplification: Additional Considerations," unpublished - presented at the National Convention of the Institute of Radio Engineers, New York, March, 1951.

19. G. F. Myers, "Distributed Amplification for Pulses," unpublished - presented at the National Convention of the Institute of Radio Engineers, New York, March, 1951.

20. V. J. Fowler, "Transmission-Line Tubes," Technical Report 15, ONR Project 076 161, University of Illinois, September, 1951.

21. Y. P. Yu, H. E. Kallman, P. S. Christaldi, "Millimicrosecond Oscillography," Electronics 24, pp. 106-11, July, 1951.

Developed By	Tube Type	Band-width	Gain	Output Voltage (Peak)
Naval Research Laboratory	4X150	100 Mc	40 Db	160 V
Hewlett Packard Company	6AK5	140 Mc	20 Db	8 V
Spencer Kennedy Laboratories	6AK5	225 Mc	21 Db	6 V
Allen B. DuMont Laboratories	6AN5	250 Mc	16 Db	6.5 V
Naval Ordnance Laboratory	6AK5	400 Mc	9 Db	-

Table I
Typical distributed amplifier
characteristics.
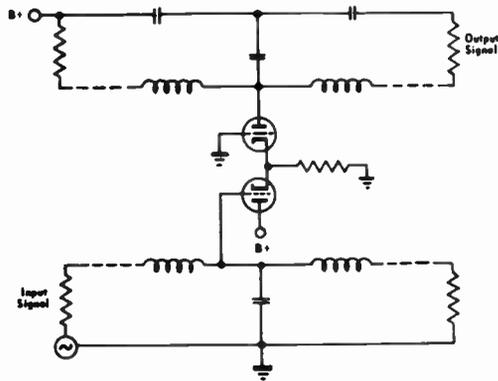
Fig. 1
Paraphase distributed amplifier.
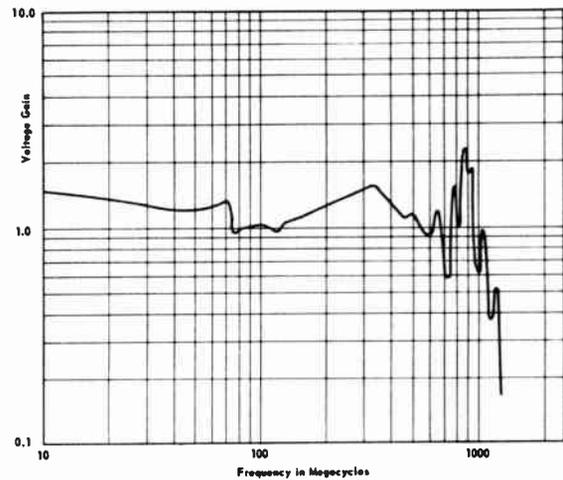
Fig. 2
Frequency response, paraphase
distributed amplifier using
5675 triodes.
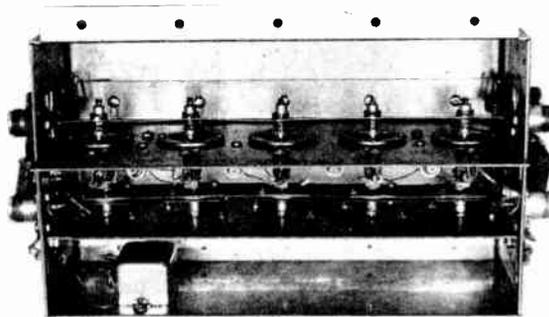
Fig. 3
Five-tube pair paraphase
amplifier (RCA 5675's).

60
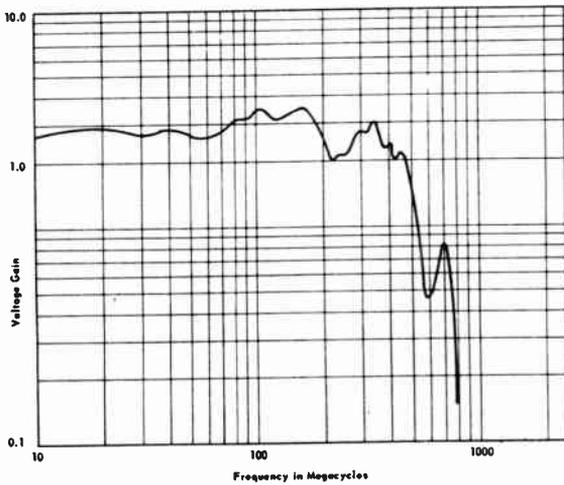
Fig. 4
Frequency response, paraphase
distributed amplifier using
2C39 triodes.
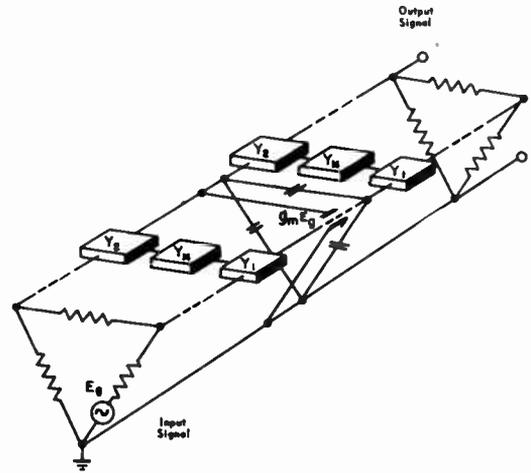
Fig. 5
Three-dimensional distributed amplifier.

TRAVELING-WAVE AND RELATED TUBES

L. M. Field
Stanford University
Stanford, Calif.

ABSTRACT

Circuit, electronic, and matching limitations on the wideband amplifying properties of helix-type tubes in the range of frequencies from 50 mc through the microwave range will be reviewed. Bandwidths of from one to several octaves in this range have been obtained and typical examples and design criteria will be given. Bandwidth and bandtuning of several related types of microwave tubes using other than a helix for wave propagation or nonpropagating structures will also be discussed.

CONTINUOUSLY VARIABLE PULSE SIGNAL DELAY SYSTEM

by

Carl Berkley
Allen B. Du Mont Laboratories, Inc.
Clifton, New Jersey

Summary

The technique of small artificial lines for the delaying of wideband signals with optimized amplitude and phase distortion is well known.[1,2,3] These lines are usually employed to produce fixed signal or pulse delays. Many applications would become possible with the availability of lines which could be adjusted over a large portion of their electrical length without appreciable change of the line constants. This has previously been accomplished experimentally with tapped delay lines[4], and with sliding contactors.[5] Because of the physically very small wires used in practical lines the taps and contacts are difficult to make without introducing undesired discontinuities. Also since the coil turns have to be exposed to the slider, difficulties arise in making good contact due to dirt and atmospheric corrosion of the surfaces. Commutation trouble and troubles with shorted turns are also encountered. A method is proposed and has been verified in principle experimentally which eliminates these difficulties.

Recently, continuously wound delay lines have become commercially available, which are generally similar to standard RG- cables, but with increased delay per unit length.[6] This is accomplished mostly by the use of powdered ferromagnetic materials in the insulated core of the cable which increase the inductance per unit length and thereby the delay. Because of the increased inductance, the characteristic impedance of the line is generally increased which is desirable for many purposes. If the core of the cable is made of magnetic material which is readily saturable, a magnetic field external to the cable may be used to alter the characteristics of the cable through the shielding without physical contact with the helix.

The characteristic impedance

$$Z_o = \sqrt{\frac{L}{C}}$$

Since we are dealing with a coil with a non-linear core, a type of analysis similar to that used in the case of magnetic amplifiers is applicable.[7] The inductance L, of a cored coil is given by, from the definition of inductance

$$L = N \frac{d\phi}{di}$$

where N = Number of turns

ϕ = the flux

i = the current

since $\phi = BA$ where B = the flux density and the magnetizing force is

$$H = \frac{Ni}{s} \qquad \text{therefore } i = \frac{Hs}{N}$$

then substituting

$$L = \frac{d(BA)}{d\left(\frac{Hs}{N}\right)} = \frac{AN^2}{s}\frac{dB}{dH}$$

Therefore it is seen that the inductance of a coil in the line is a function of the slope of the B/H curve. By adding a steady component to the field in the core either by a dc current or a superimposed external field from a magnet, we may change the slope that a signal current sees and therefore the characteristics of the line. If we use a core having a sharp shoulder in the B/H curve and use a field strength sufficient to change the operating point from one linear region to the other, then the point along the line where the inductance changes abruptly will appear to be a discontinuity to any incoming signal and the signal will be partly reflected. If the far end of the line is terminated properly, no signals will appear from this path and only the reflected signal will be returned to the input. Circuits exist and are described to separate the reflected from the ingoing signal.

The amplitude of the reflected signal depends, of course, on the ratios of the slopes of the different portions of the B/H curve. With a commercial line having a core of material not specifically intended to be saturable, we have obtained reflections attenuated approximately 40 db using a small horseshoe magnet against the cable. (See Figure 1) The reflected pulse can be continuously varied by moving the magnet. It is necessary when doing this to have the direction of the external field coincide, or be parallel, to the field in the cable solenoid, otherwise no reflections are observed. The cause of this effect is not clearly understood. It appears that even if the core is saturated in one direction by an external magnetizing force the component of this force in the direction at right angles is zero.

The delay of the reflected pulse with respect to the incident pulse is a linear function of the position of the saturating magnetic field. This makes possible the production of an accurately linear pulse delay.

By using the line as the frequency determining element in an oscillator as described by R. Palmer[8], and recirculating the pulses through an amplifier to compensate for line losses, a linear frequency scale may be obtained.

The sharp discontinuity produced may be thought of as a mirror or wall from which the waves are reflected. If we were to use a core ma-

terial such as Deltamax, which requires only a small field differential between the saturated and unsaturated regions and then produce a linearly increasing field along the cable perhaps by the use of a tapered helix wound external to the electrostatic shielding braid and concentric to it, we would produce a movable mirror by varying the helix current. By supplying an audio signal, for example, to the helix and pulses to the cable input, we could produce a PTM of the pulses.

By moving the mirror at speeds approaching the group velocity in the line we obtain a Doppler effect for the observer at the line input. The reflected beat frequencies which he sees depends on the rate at which the signal overtakes the mirror. This results in a lowering of the frequency of the input phenomenon. If the pulse signal just overtakes the mirror when it has reached the end of the line, the pulse has been "stretched" to the full length of the line.

For other applications, it may be desirable to use a core in which the inductance instead of varying abruptly, varies linearly with the external field. This requires the use of materials in which the second derivative of the B/H curve is a constant and the use of small signals. By using such materials and a series of adjustable external magnets along the line (see Figure 2), we can do a wide variety of waveshaping and transient generating functions in a manner similar to that proposed by Blumlein, Kallman, and Percival[9] with adjustable taps on the helix.

References

(1) R. A. Erickson and H. Sommer, "The Compensation of Delay Distortion in Video Delay Lines", Proc. I.R.E., vol. 38 #9, pp. 1036-1040; September 1950.

(2) M. J. Di Toro, "Phase and Amplitude Distortion in Linear Networks", Proc. I.R.E., vol. 36 #1, pp. 24-36; January 1946.

(3) J. P. Blewett and J. H. Rubel, "Video Delay Lines", Proc. I.R.E., vol. 35 #12, pp. 1580-1584; December 1947.

(4) H. E. Kallmann, "Transversal Filters", Proc. I.R.E., vol. 28 #7, pp. 302-310; July 1940.

(5) "Type 302 Delay Line", Advance Electronics Company, Passaic, New Jersey.

(6) "Type HH-1500 , Manufactured by Hackethal Wire and Cable Co., Hannover, W. Germany.

(7) H. J. Smith, "Magnetic Amplifiers", Rept. No. UA-514-P-1, Cornell Aero Lab., Buffalo, N. Y. August 23, 1949.

(8) R. C. Palmer, "Timing Marker Generator of High Precision", Presented at 1950 I.R.E. Annual Conv. Private Communication.

(9) A. D. Blumlein, H. E. Kallmann, W. S. Percival, "Improvements in or relating to Electric Wave Transmission Networks", British Patent Specification 517,516; February 1, 1940.

Fig. 1
Oscillograms of a pulse reflected from three positions along a delay line by use of a movable magnetic field.
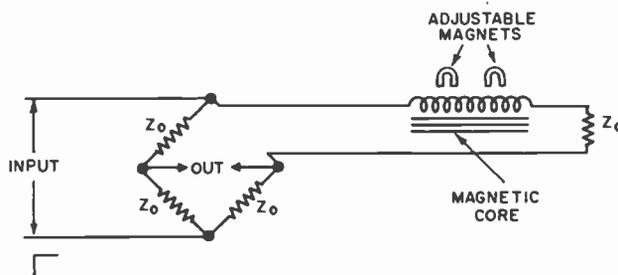
Fig. 2
Schematic circuit for generating an arbitrary waveshape with adjustable magnetic discontinuities.

63

GENERAL TRANSMISSION THEORY OF DISTRIBUTED
HELICAL DELAY LINES WITH BRIDGING CAPACITANCE

Dr. M. J. Di Toro
Allen B. Du Mont Laboratories, Inc.
Passaic, New Jersey

Abstract

Helical delay lines are dispersive trans-
mission systems because their group velocity of
propagation increases with increasing frequency.
The use of bridging capacitance to reduce this
dispersion is known, but no general analysis has
been presented previously for distributed (i.e.
non-lumped) lines. This paper derives the general
equations for the propagation constant and the
impedance of such distributed helical delay lines
with bridging capacitance. The problem requires
the solution of integral equations, which is
effected using Fourier transforms. The general
conditions for linear phase propagation are de-
rived, and design data are given for the bridging
capacitance geometry, along with the other delay
line parameters such as delay, usable bandwidth,
length, diameter, and the line's impulse response
overshoot or echo.

Introduction

Delay lines are becoming of increasing use-
fulness in various applications. An important
reason for this is their suitability in the time
domain synthesis of networks. In particular,
continuously wire-wound electrical delay lines
of the helical type, when suitably phase corrected
to eliminate dispersion, constitute a simple,
small network component which can displace the
bulkier delay lines comprising lumped induc-
tances and capacitances. As a result, consider-
able attention has been given to the design of
helical delay lines (Ref. 1 - 6).

Dispersion in Delay Lines

One of the major difficulties in the trans-
mission of signals along helical delay lines
arises because the line's inductance per unit
length L decreases with increase of frequency f.
As a result, the phase shift in radians per unit
length $B = \omega \sqrt{LC}$ (where C is the capacitance to
ground, per unit length) is not linear but is
concave downward with frequency. This is shown
in Fig. 1a (from Ref. 1 and 5). The group time
delay $dB/d\omega$ of a wave packet (or bundle of fre-
quencies) of width $df = d\omega/2\pi$ and centered at f
thus decreases with increase of frequency. Ac-
cordingly, if to the input of such a line is
applied a unit impulse (or delta) function,
whose spectrum comprises equally all frequencies,
the higher frequencies arrive at the output of
the line sooner than the lower frequencies. The
wave shape of the input impulse function thus
becomes changed. An example of this, in an
actual delay line, is given in Fig. 2a (from
Ref. 1; see also Ref. 7, Fig. 1).

Such transmission media, in which the group
time delay is not constant for all frequency
regions of finite applied signal spectrum, pro-
duce elongation in time, or dispersion, of the
output impulse response. It has been shown
(Ref. 8) that in a transmission system whose
transfer function is $A(\omega)$ exp $-jB(\omega)$, dispersion
is caused by large values of the mean squared
weighted group delay distortion $A(d\Delta B/d\omega)$ and
mean squared values of amplitude slope $dA/d\omega$.
Here $A(\omega)$ is the amplitude response vs frequency
of the transmission system, $B(\omega)$ is the phase
shift, and $\Delta B(\omega)$ is the phase distortion. The
latter is the deviation of B from the ideal value
given by ωt_d, where t_d, the delay, is defined as
the average value of $A^2 dB/d\omega$.

Phase Distortion Correction

Of these two sources of dispersion, only
that due to the phase distortion ΔB, of helical
delay lines, is considered here together with an
analysis of its correction. The use of bridging
capacitances to reduce the phase distortion ΔB
has been suggested previously (Ref. 1, 3, 4, 9).
Bridging capacitance in distributed helical delay
lines can be conveniently obtained by the use of
a reciprocating winding as in Fig. 1c (Ref. 1,
2, 5), by bridge capacitance strips (Ref. 3), or
by a conductive coating of the winding (Ref. 4).
The beneficial action of such bridge capacitance
is realized by comparing Fig. 2a with Fig. 2b
(from Ref. 5).

An elementary insight as to why bridging
capacitance works follows from Fig. 1b. One
observes that the effective inductance of a coil
with bridge capacitance increases with increase
of frequency. This annuls the decrease in L,
shown in Fig. 1a, providing a suitable value of
bridge capacitance is used (Ref. 2, 3, 6).

No general analysis of signal transmission
through such phase-corrected or dispersionless
delay lines has ever been presented previously
for distributed (i.e. non-lumped or smooth)
lines. It is the object of this paper to formu-
late the general problem, present the solution,
and ultimately show a set of design curves from
which the optimum bridging capacitance geometry
may be determined. Such design curves are given
below under Design Data, wherein are shown the
inter-relations between the various helical
delay line parameters such as delay, bandwidth,
length, diameter, and impulse response overshoot
or echo. Those not interested in the following
analysis should now go to that section below

wherein are given the design data.

General Transmission Theory

Consider the general formulation of the problem and its solution. The configuration of interest is that of Fig. 3. It comprises an infinite helix with capacitance (not indicated) to a nearby ground bus. It is assumed that between two elementary lengths of helix dx and dy, separated by the distance y, there exists (unintentionally) a mutual inductance $M(y)$ and (intentionally) a bridge capacitance $C_b(y)$. The (complex) current in the helix, at a distance x, from an arbitrary origin, is $I(x)$, while the (complex) voltage to ground at x is $V(x)$.

It is not difficult to show that the integral-differential equations between $I(x)$ and $V(x)$, in terms of the functions $M(y)$ and $C_b(y)$, and the fixed parameters L_o and C, are

$$- \frac{dV(x)}{dx} = j\omega L_o \int_{-\infty}^{\infty} m(y) \ I(x+y) \ dy \tag{1}$$

$$- \frac{dI(x)}{dx} = j\omega C \left[V(x) + \int_{-\infty}^{\infty} c(y) \left[V(x) - V(x+y) \right] dy \right] \tag{2}$$

The symbol meanings are:

$L_o =$ unit inductance, henry/meter, at zero frequency

$C =$ unit capacitance to ground, farads/meter

$M(y) =$ mutual inductance, henry/sq. meter, between two elementary sections of line length dx and dy, separated by y meters.

$C_b(y) =$ bridge capacitance, farads/sq. meter, between two elementary sections of line length dx and dy, separated by y meters.

$m(y) = M(y)/L_o$

$c(y) = C_b(y)/C$

$I(x) =$ complex current, amperes, in helix at x meters from an arbitrary origin.

$V(x) =$ complex voltage, volts, to ground at x.

Exact formulas for L_o and for m (in terms of Elliptic functions) are known (Ref. 11).

The solution of Eq. 1 and 2 is effected by the usual process of assuming

$$V(x) = A \ exp \ (-j\omega x/v) \tag{3}$$

$$I(x) = (A/Z) \ exp \ (-j\omega x/v) \tag{4}$$

where $A =$ arbitrary (complex) constant

$Z =$ characteristic impedance of the line

$v =$ phase velocity, meter/sec

$\omega/v =$ phase delay, radian/meter.

Insertion of Eqs. 3 and 4 into Eqs. 1 and 2 gives

$$z(Z/\omega L_o a) = 2\int_0^{\infty} m(y) \ cos \ (\omega y/v) \ dy$$
$$= 2 \ I_1(z) \ K_1(z) \tag{5}$$

$$z/(Z\omega C a) = 1 + 2\int_0^{\infty} c(y) \left[1 - cos(\omega y/v) \right] dy \tag{6}$$

where $z = \omega a/v =$ phase delay in 1 radius (a) of line length,

$a =$ radius of helix, meters.

The rather interesting thing from Eq. 5 and 6 is that the significant functions are not m nor c, but rather their Fourier (cosine) transforms. The transform of Eq. 5 has been determined by Blewett (Ref. 4.) The I_1 and K_1 functions are the modified Bessel functions of the first and second kind, which have been tabulated (Ref. 12). Accordingly, using $Y(z)$ as the Fourier cosine transform of $c(y)$, defined by

$$Y(z) = 2\int_0^{\infty} c(y) \ cos(\omega y/v) \ dy \tag{7}$$

it is possible to arrive at the following neat formulas for the phase velocity v and the characteristic impedance Z in terms of the radian frequency ω and the other line parameters:

$$(v_o/v)^2 = 2I_1(z)K_1(z) \left[1 + Y(0) - Y(z) \right] \tag{8}$$

$$(Z/Z_o)^2 = 2I_1(z)K_1(z) / \left[1 + Y(0) - Y(z) \right] \tag{9}$$

where $v_o = 1/\sqrt{CL_o} =$ phase velocity at zero frequency

$Z_o = \sqrt{L_o/C} =$ characteristic impedance at zero frequency.

The problem of phase distortion correction now reduces to that of selecting the bridge capacitance Fourier transform function $Y(z)$ such that $v_o/v = 1$ for the largest part of the pass band. It is also apparent that this is only done at the expense of getting a worse variation of the characteristic impedance Z.

It is now convenient to restate the function matching problem of Eq. 8 in another way which also gives the residual amount of phase distortion after correction from bridge capacitance.

The phase shift for a length of line equal to one radius(a) is $z = \omega a / v$. In an ideally phase corrected line, the phase velocity v does not change from its zero frequency value of v_0. The corresponding phase shift in one radius length of an ideally phase corrected line would be $z_0 = \omega a / v_0$. The phase deviation from this ideal phase shift, used here as the phase distortion, is thus $\Delta B_1 = z_0 - z$. From Eq. 8, one obtains

$$\Delta B_1 = z \left[\left[2 I_1(z) K_1(z) \right] \left[1 + Y(0) - Y(z) \right] \right]^{-\frac{1}{2}} - 1 \quad (10)$$

This can in turn be solved for the needed bridge capacitance transform function $Y(z)$, giving

$$Y(0) - Y(z) = \left[\frac{1}{2 I_1(z) K_1(z) \left[1 + (\Delta B_1/z)^2 \right]} \right] - 1 \quad (11)$$

On putting $\Delta B_1 = 0$ in this, one obtains the formula for the capacitance transform function resulting in no phase distortion as

$$\left[Y(0) - Y(z) \right]_1 = \frac{1}{2 I_1(z) K_1(z)} - 1 \quad (12)$$

Both Eq. 12 and Eq. 11, for various fixed values of ΔB_1, are shown in Fig. 4.

The problem now reduces to that of finding a suitable bridge capacitance c(y) function, and its corresponding Fourier transform Y(z), such that Y(z) comes within the supposedly tolerable and pre-assigned phase distortion limits of $+\Delta B$ and $-\Delta B$ of the curves of Fig. 4.

Synthesis of Bridge Capacitance Geometry

The synthesis problem of finding a suitable Y(z) does not have a unique solution. One method of solution will be indicated herein, which has both mathematical simplicity and is easily realizable physically. The bridge capacitance geometry in mind is that of Fig. 5. It is not difficult to show that the direct bridge capacitance function c(y) of Eq. 2, for the geometry of Fig. 5, is

$$c(y) = \frac{g}{w} \left[1 - \frac{|y|}{w} \right] = \frac{p C_p}{d^2 C} \left[1 - \frac{|y|}{w} \right] \quad (13)$$

where g = ratio of total helix capacitance to the bridge bus, to the helix capacitance to the ground bus, for a length d of helix.

w = width (along helix length) of bridge bus

p = width (along wire length) of bridge bus

d = winding pitch

C_p = capacitance between one helix wire and one bridge bus strip perpendicular to it (see Fig. 5).

The reason in choosing the linear function of equation (13) for c(y), and the corresponding geometry of Fig. 5, is that the Fourier transform of c(y) has the simple form

$$Y(0) - Y(z) = g \left[1 - \frac{\sin^2(wz/2a)}{(wz/2a)^2} \right] \quad (14)$$

The adjustable parameters of the bridge bus are the g and w defined above. In Eq. (14) these comprise scale changes in both the ordinates and the abscissa z. Accordingly, a convenient way to fit Eq. (14) to the ideal phase correction curve shown by $\Delta B_1 = 0$ in Fig. 4, is first to plot both Eq. 12 and Eq. 14 on log-log paper. The curve of Eq. 14 is then moved vertically to adjust g, and horizontally to adjust w, until a fit with the + and – tolerances of ΔB_1 is obtained.

As an example, Fig. 4 shows the suitable orientation of the (dash line) curve of Eq. 14 with the plots of Eq. 11, for the case where $\Delta B_1 = 0.05$ radians. It is noted that the proper location of the dash line curve is when it is tangent to both the $\Delta B_1 = 0.05$ and -0.05 radian phase distortion curves of Eq. 11.

The residual phase distortion is of the form indicated by Fig. 1d. As a result, some rather important parameters emerge which are of interest in the impulse response of the line. The first is the known relation (Ref. 13) between the echo, or impulse overshoot, of Fig. 1d and the residual (sinusoidal) phase distortion ΔB. This relation is

$$e = J_1(\Delta B)/J_0(\Delta B) \quad (15)$$

where J's are Bessel functions of the first kind.

The second important parameter is the value z_1 where the dash line curve of Eq. 14 intersects, for the last time, the curve for $\Delta B_1 = 0$. This is taken as the "useful" bandwidth of the delay line. (It should be recalled that the dimensionless variable z is related to the frequency f by the expression $z = 2\pi a f \sqrt{LC}$.)

Along with these data on overshoot or echo e and useful bandwidth f_1, one also obtains by this process the bridge capacitance geometrical parameters (g and w/2a) which, for $|\Delta B_1| = 0.05$,

66

are $g=3.5$ and $w/2a = 0.55$. Proceeding in this way for other values of ΔB_1, it is possible to arrive at other sets of values of the parameters ΔB_1, z_1, g and $w/2a$.

These relative parameters are related as follows to those of a line of length b. Since the phase distortion of a line of length a is ΔB_1, that for a line of length b is $\Delta B = b\Delta B_1/a = 2A\Delta B_1$, where A aspect ratio of the line = b/2a line length/line diameter. Hence

$$A = \Delta B/2\Delta B_1. \qquad (16)$$

The important parameter of delay-bandwidth product (Ref. 1) now can be obtained. The zero frequency delay t_{do} in a line of length b is $t_{do} = b\sqrt{L_oC}$. The useful bandwidth f_1 of the line is related to z_1 by the equation $f_1 = z_1/2\pi a\sqrt{L_oC}$. Hence, the delay-bandwidth product, in Cycles, is

$$\text{Delay-Bandwidth, Cycles,} \quad (z_1/2\pi\Delta B_1)\Delta B. \quad (17)$$

By the use of Eq. 15, relating the line's residual phase distortion ΔB to its impulse response echo e, and Eq. 16 and 17, together with the values of $z_1, \Delta B_1, g$ and $w/2a$ obtained from Fig. 4 in the manner described, it is possible to arrive at the useful design data described in the next section.

Design Data

In the previous section the general transmission theory is indicated of infinite length distributed helical transmission lines in which finite mutual inductance M and bridge capacitance C_b are present, (along with the usual capacitance C to ground) between all turns (Fig. 3). Linear phase transmission results when Eq. 12 is satisfied (see Ref. 10 for a special case).

The design of a suitable phase linearizing bridging capacitance geometry, to satisfy Eq. 12, is solved for the particular geometry of Fig. 5. The important parameters in this geometry are g (= helix capacitance to bridge bus/capacitance to ground bus) and $w/2a$ (= width, along helix length, of bridge bus/line diameter).

A residual oscillatory phase distortion of magnitude ΔB remains. This results in an echo or overshoot e of the line's impulse response (Eq. 15). The useful bandwidth of the line, defined as the frequency within which the line's phase distortion remains within a pre-assigned value of ΔB, is also found. In consequence, the important delay line parameter of delay-bandwidth product, expressed in Cycles, is determined (Eqs. 16 and 17). All of these design data are shown in Figs. 6 and 7.

Example of Bridge Capacitance Design

The use of the design data contained in Figs. 6 and 7 is best indicated by a specific example. Suppose it is desired to design a phase-corrective bridge capacitance geometry of the type of Fig. 5 for a line of 0.25 in diameter. Suppose that this line should have a useful bandwidth of at least 1 Mc., a delay of 10 microsec., and an impulse response overshoot which should not exceed 0.1.

The delay-bandwidth product is 10 cycles. From Fig. 6, one determines that the line's value of A(= length/diameter) must be 26. Thus the line must be at least 6.5 in. in length, and should be designed with a unit delay of not more than 1.5 microsec./in. From Fig. 7, the bridge capacitance parameters are found to be $g = 0.9$ and $w/2a = 1.6$. These data, along with those already presented (Refs. 1, 2 and 5) are sufficient to enable one to arrive at the other common parameters of the line such as L_o, C, Z_o, etc.

Symbol Definitions

$A(\omega)$ = amplitude vs frequency response of line.

A = line aspect ratio = line length/line diameter = b/2a.

a = radius of helical line (meters).

$B(\omega)$ = phase lag vs frequency response of line.

b = length of line (meters).

C = line unit capacitance to ground (farads/Meter).

$C_b(y)$ = bridge cpacitance(farads/sq.meter) between two elementary sections of line length dx and dy separated by y (meters).

$c(y) = C_b/C$.

d = winding pitch of helix(meters).

ΔB = phase distortion (radians).

$\Delta B_1 = z_o - z$ = phase distortion (radians) in one radius(a) length of line.

e = impulse response echo or overshoot

f = frequency (cps).

f_1 = useful bandwidth of line (cps).

g = ratio of total line capacitance to the bridge bus, to the line capacitance to the ground bus, for a length d of helix.

$I(x)$ = complex current (amperes) in helical line at x (meters) from an arbitrary origin.

$I_1(z)$ = modified Bessel function of the first kind.

J_o, J_1 Bessel functions of first kind.

K_1 = modified Bessel function of the second kind.

L_o = line unit inductance (henry/meter) at zero frequency.

L = Line unit inductance (henry/meter) at frequency f.

$M(y)$ = mutual inductance (henry/sq.meter) between two elementary sections of line length dx and dy, separated by y (meters).

$m(y) = M/L_o$.

p = width (meters), along line length, of bridge bus.

t_d = time delay (sec) of line of length b.

$V(x)$ = complex voltage (volts) of line to ground at x.

v = phase velocity of line (meters/sec) at frequency f.

v_o = phase velocity of line (meters/sec) at zero frequency.

ω = radian frequency = $2\pi f$.

w = width (meters), along line length, of bridge bus.

x = distance along helical line (meters).

$Y(z)$ = Fourier cosine transform of $c(y)$.

$z = \omega a/v$ = phase delay (radians) in one radius(a) of line length.

Z = characteristic impedance of line (ohms) at frequency f.

Z_0 = characteristic impedance of line (ohms) at zero frequency.

$z_0 = \omega a/v_0$ phase delay(radians) in one radius(a) length of ideal phase corrected line.

References

1) M. J. Di Toro "Phase Corrected Delay Lines" Hazeltine Electronics Corporation, Report No. 4000, 29 November 1945

2) M. J. Di Toro "Theory and Design of Delay Lines" Hazeltine Electronics Corporation, Report No. 4001, 11 December 1945

3) H. E. Kallmann "Equalized Delay Lines" IRE Proceedings, September 1946, pp. 646-657

4) J. P. Blewett, J. H. Rubel "Video Delay Lines" IRE Proceedings, December 1947, pp. 1580-1584

5) M. J. Di Toro "Phase Corrected Delay Lines" presented at 1948 IRE National Convention (see IRE Proceedings, March 1948, p. 374, abstract No. 68)

6) R. A. Erickson, H. Sommer "The Compensation of Delay Distortion in Video Delay Lines" IRE Proceedings, September 1950, pp. 1036-1040

7) M. J. Di Toro "Phase and Amplitude Distortion in Linear Networks" IRE Proceedings, January 1948, pp. 24-36

8) M. J. Di Toro "Dispersion in Transmission Systems" presented at 1952 IRE National Convention (being presented for publication in the IRE Proceedings)

9) M. J. E. Golay "The Ideal Low Pass Filter in the Form of a Dispersionless Lag Line" IRE Proceedings, March 1946, pp. 138P-144P

10) I. A. D. Lewis "Note on the Variations of Phase Velocity in Continuously-Wound Delay Lines at High Frequencies" Proceedings of the Institution of Electrical Engineers, Part III, July 1951, pp. 312-314.

11) W. R. Smythe "Static and Dynamic Electricity" 1939, McGraw-Hill Book Company, New York

12) G. N. Watson "Bessel Functions" 1944, The MacMillan Company

13) H. A. Wheeler "The Interpretation of Amplitude and Phase Distortion in Term of Paired Echoes" IRE Proceedings, June 1939, pp. 359-385.
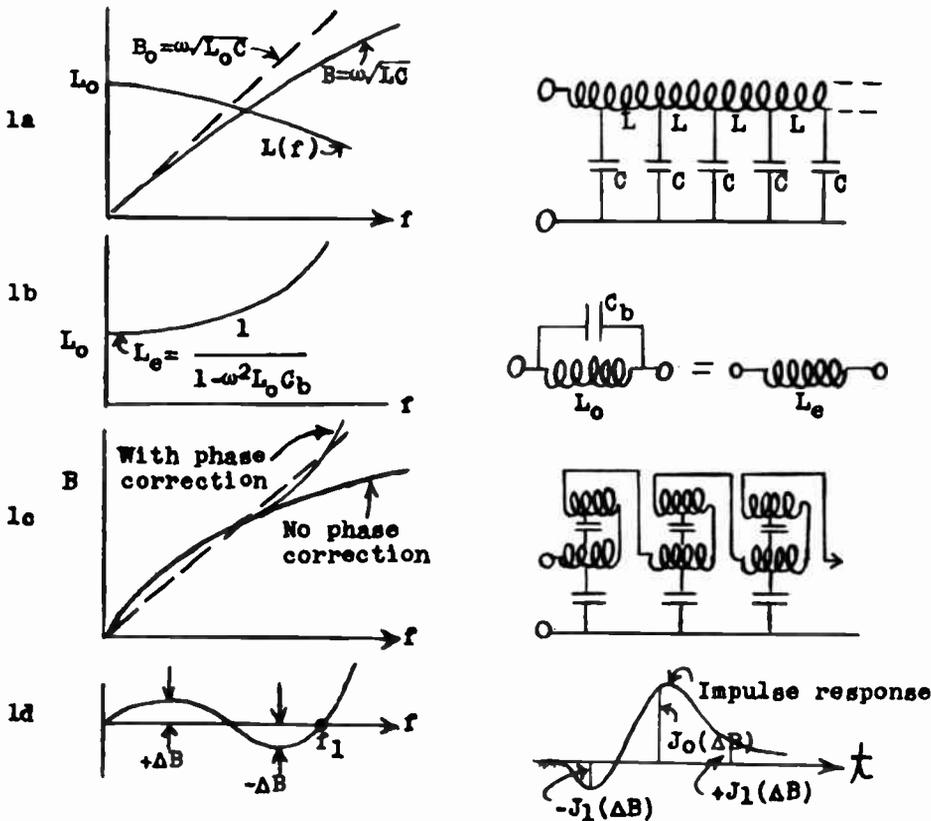
Fig. 1
Phase corrected helical delay line with bridging capacitance, and impulse response echo.

Impulse Response, Helical Delay Line, No Phase Correction

—|1 inch|—

Impulse Response, Helical Delay Line, Phase Corrected

—|1 inch|—

Fig. 2
Impulse response of helical delay lines, without
and with phase corrective bridging capacitance.
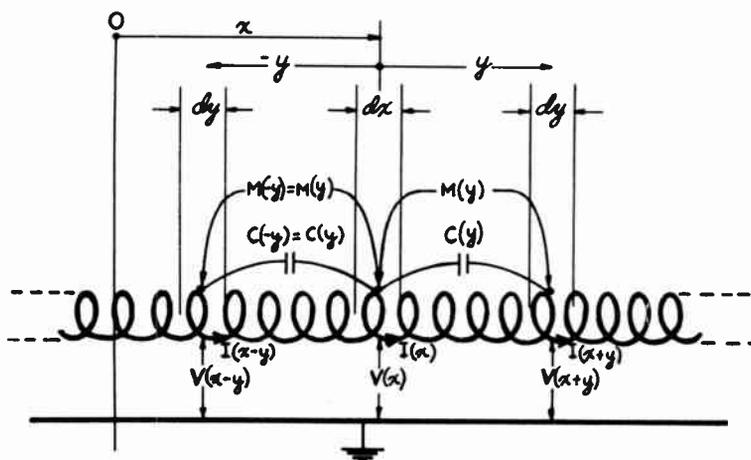
Fig. 3
Distributed helical delay line with bridging capacitance,
mutual inductance, and capacitance to ground.
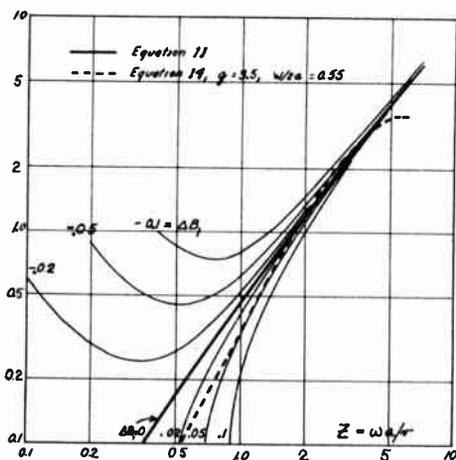
Fig. 4
Graphs of equations 11 and 14.

69
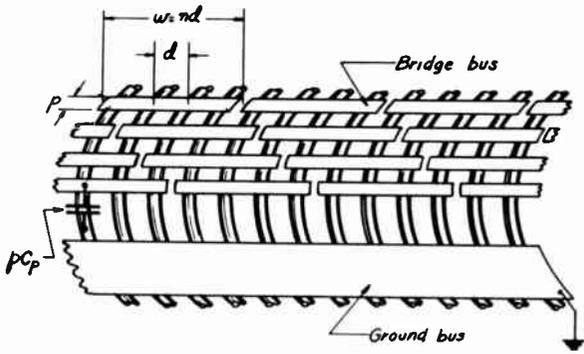
Fig. 5
Bridging and ground capacitance
geometry.
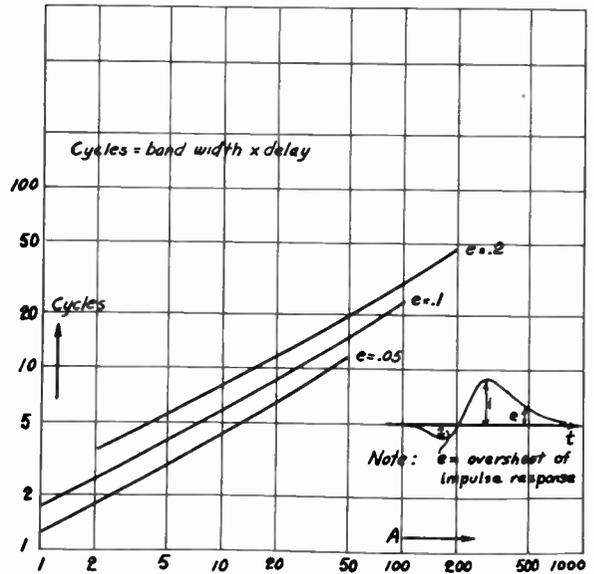
Fig. 6
Bandwidth-delay product vs.
echo and aspect ratio.
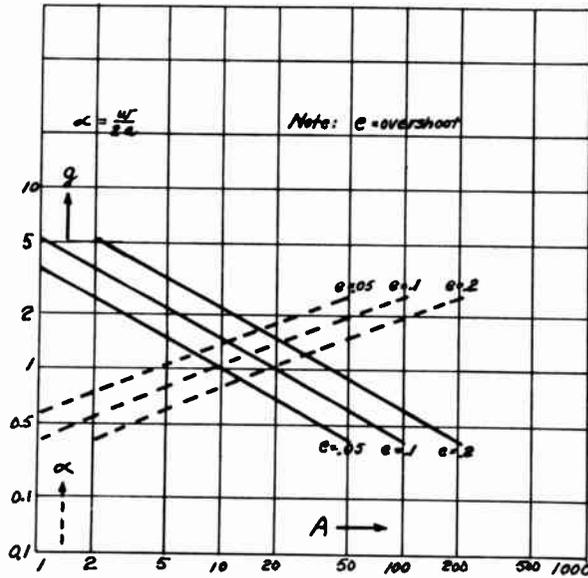
Fig. 7
Graphs of bridging capacitance parameters
g and w/2a, vs. echo and aspect ratio.

70

DISTRIBUTED CONSTANT DELAY LINES WITH
CHARACTERISTIC IMPEDANCES HIGHER THAN 5000 OHMS

William S. Carley
U. S. Naval Ordnance Laboratory
White Oak, Silver Spring, Maryland

Abstract

Artificial delay lines with characteristic impedances of greater than 5000 ohms have been developed for use with fractional microsecond pulses. These lines have delays ranging from 0.2 to 0.5 microseconds per axial inch. The attenuation of a one microsecond pulse may be as low as 0.2 db/microsecond of delay. Equations are given for the design of self-compensated multilayer delay lines. Experimental results are given for lines designed from these equations. Comparison of rise times, attenuations, time delays, and characteristic impedances of lines wound with various sizes of formex and teflon insulated wire is made. Photographs of the pulse response of these lines to 0.3, 0.4, 0.6 and 1.0 microsecond are shown.

Theory

Distributed constant electromagnetic delay lines are becoming increasingly useful circuit elements in present day electronic equipment. The characteristic impedance of these lines has been limited to values between 400 and 3,000 ohms.[1,2] Applications exist for lines with higher characteristic impedance. A brief discussion of the factors that determine the delay time and characteristic impedance will be given followed by a detailed description of methods for increasing the characteristic impedance.

The delay time, phase velocity, and characteristic impedance of a distributed constant delay line can be derived from the simplified equivalent circuit of Figure 1 where all losses have been neglected. These are

$$t_d = \sqrt{LC},\qquad (1)$$

$$\beta = \omega\sqrt{LC},\qquad (2)$$

and

$$Z_o = \sqrt{\frac{L}{C}},\qquad (3)$$

where L = inductance per unit length
and C = capacitance per unit length.
If R and G, the resistance and conductance per unit length, are present but $R \ll \omega L$

and $G \ll \omega C$ the following more general equations apply:[3]

$$t_d = \sqrt{LC}\left[1 + \frac{1}{2}\left(\frac{R}{2\omega L} - \frac{G}{2\omega C}\right)^2\right],\qquad (4)$$

$$\beta = \omega\sqrt{LC}\left[1 + \frac{1}{2}\left(\frac{R}{2\omega L} - \frac{G}{2\omega C}\right)^2\right],\qquad (5)$$

$$\text{and } Z_o = \sqrt{\frac{L}{C}}\left[1 + \frac{1}{2}\left(\frac{R^2}{4\omega^2 L^2} + \frac{RG}{2\omega^2 LC} - \frac{3G^2}{4\omega^2 C^2}\right) + j\left(\frac{G}{2\omega C} - \frac{R}{2\omega L}\right)\right]\qquad (6)$$

The design equations given in the following paragraphs have been derived on the assumption that the line is lossless.

It has been observed that the inductance of a delay line decreases at higher frequencies.[5,6,7] This is caused by the phase shift per turn increasing so that although the turns are still magnetically linked as the frequency increases they add less and less to each other's magnetic field. A plot[7] of normalized inductance $\frac{L}{L_o}$ and time delay $\frac{T}{T_o}$ vs. $\frac{d\,T_o\,f}{\ell}$ appears in Figure 2 where
 d = mean diameter of line,
 L_o = inductance at low frequencies,
 T_o = time delay for low frequencies,
 ℓ = length of line.
The actual expression[7] is

$$\frac{L}{L_o} = 2I_1(x)\,K_1(x),\qquad (7)$$

where $I_1(x)$ and $K_1(x)$ are modified Bessel Functions of the first and second kind,[8]

and $x = \frac{\pi d}{\lambda} = \frac{\pi d\,Tf}{\ell}$,

where x is the axial wavelength along the line.

The effect of turn to turn capacitance has been studied.[6,7] At low frequencies the effect of this capacitance is negligible as the phase of the voltage in each turn of the coil is the same. As the frequency increases the phase of the voltage in each turn changes. Thus the effect of the turn to turn capacitance increases with frequency.

The capacitance between turns is in parallel with the inductance of each turn. If this capacitance is C', the net impedance of the incremental series arm of our distributed constant line is

$$\frac{j\omega L_t}{1 - \omega^2 L_t C'} , \qquad (8)$$

where L_t = the self inductance per turn = $\frac{L}{N}$,
N = the total number of turns,

$$C' = \frac{\pi d K_e}{3.6 \log_e \left\{\frac{a+t}{a} + \sqrt{\left(\frac{a+t}{a}\right)^2 - 1}\right\}} \qquad (9)$$

where a = diameter of bare wire
t = wire separation.
K_e = dielectric constant of the insulation on the wire
For close wound coils

$$C' = \frac{\pi d K_e}{3.6 \sqrt{2}} \sqrt{\frac{a}{b-a}} , \qquad (10)$$

where b = over-all wire diameter with insulation.

From Figure 2 it is apparent that L_t decreases with increasing frequency. It is also evident from equation (8) that the effect of C' is at least in the direction of reducing the variation of the series arm impedance with frequency and thus equalizing the delay time.

Another method of equalization is the use of patches.[4,9] Patches are bridging capacitors over a number of turns, acting in similar fashion to our turn to turn capacitance except that it is a lumped type of compensation instead of a distributed type.

The purpose of the investigation, reported in this paper, was to produce self-equalized distributed constant lines having relatively large delays and high characteristic impedances. To achieve these goals both L and C were increased, but L was increased by a considerably larger factor than C.

In order to obtain as large a delay as possible it was decided to use the complete core as a ground. The capacitance per unit length can be varied by controlling the thickness and dielectric constant of the insulation material placed between the core and the winding. This large capacitance per unit length would necessitate a correspondingly large inductance per unit length in order to secure a high characteristic impedance.

The secret of success for the high characteristic impedance line is the

method of obtaining the high inductance.[10] First of all a small wire size was chosen. Most of the work was done with B & S gauge #46 and #44 wire. A bank winding with approximately 3 layers was found necessary to obtain the necessary inductance.

A derivation of the equations for compensation of a multilayer line will not be attempted in this paper. It is my anticipation to submit it for publication in the near future. Basically the effective capacitance between turns for a multilayer winding is derived assuming a bank winding, and is inserted in an expression similar to equation (8). A convenient expression for this approximate solution is

$$C_0 = \frac{K}{S} \frac{\sqrt{a}}{\sqrt{b-a}} \frac{4\pi \epsilon_0 K_e}{\sqrt{2} d} \left(\frac{T_0}{T}\right)^2 , \text{ farads/meter} (11)$$

where C_0 = capacitance to core per axial meter,
K = a factor depending on number of layers, (See Table 1)
T_0 = the time delay per meter axial length at low frequencies,
T = the minimum allowable time delay per meter axial length,
ϵ_0 = 8.854 x 10^{-12}
S = turns per meter per layer

Table 1

Winding layers	K
1	1
2	14
3	76
4	172
5	324

In case the number of layers does not come out even, successful results have been achieved by plotting number of layers against K and reading the value of K from the graph.

If the time delay per meter is high, the time delay variation over the desired pass band must be very small for good pulse reproduction. For the most part, in the lines described in this paper, the time delay was so high that T_0/T was made equal to 1.

Although it is not evident in the form given, equation (8) has a pole at $x = 1.3$. These lines then could give a reasonably flat delay time up to $x = 1$. From the definition of x (below equation (7)) better frequency response or larger time delays without excessive distortion can be achieved with a small diameter core.

72

It may be noted in passing that equation (11) does not reduce down to that given in reference (6) for a single layer winding. This is no doubt caused by several errors in reference (6). We know

$$C_o = \frac{2\pi\,\varepsilon_o}{\frac{1}{K_e}\,\ell n\,\frac{b}{a} + \frac{1}{K_{ed}}\,\ell n\,\frac{d}{b}} \qquad \text{farads/meter(12)}[11]$$

where K_{ed} = dielectric constant of insulation between core and winding. In many cases the first term in the denominator of equation (12) is negligible in comparison with the second.

For a long solenoid

$$L_o = \pi^2 N^2 d^2 \times 10^{-7} \qquad \text{henries/meter(13)}$$

The time delay per meter and the characteristic impedance are given by equations (1) and (3). With these equations and equations (11), (12) and (13) a self-compensated line can be designed.

Line Construction

Most of the lines were wound on 3/16 inch diameter polystyrene cores 12 inches long. These cores were given several coats of silver conducting paint to form the ground strip. Although the cores could be slotted after an overnight drying period, a much cleaner cut was made if the drying period was increased to several days. The cores were axially slotted forming 36 thin strips, each strip being about 0.015 inch wide. The slots were about 0.003 inch wide. A one inch length of the core was left unslotted to facilitate the connection of the external ground lead. A photograph of the slotted core and the slotting equipment is shown in Figure 3. The core was covered with a layer of insulating material to serve the dual purpose of insulating and controlling the winding-to-core capacitance. A piece of thin teflon tape was wound around the core. A number of small pieces of scotch cellophane tape held the teflon on the core until the line was wound. The scotch tape was removed piece by piece as the winding of the line progressed.

The winding was done on a lathe. In order to provide uniform wire tension, secure a good winding and prevent breakage, the wire feeding device shown in Figure 3 was used. The wire tension was adjustable over a range of about 10 to 70 grams. The tension was continuously indicated on a scale by a pointer.

A wire guide attached to the longitudinal feed of the lathe was placed about 1/16 inch from the core, which was chucked in the lathe. The longitudinal travel of the wire guide could be as low as 0.00066 inch per turn. As this distance is a fraction of the wire diameter, the result was a multiple layer coil approximately bank wound. The far end of the core was attached to a counter chucked in the tailstock. A steel drill rod was inserted through a hole in the core for rigidity. A 10 inch long winding was wound on the core. Lines were wound with speeds varying from about 200 to 800 rpm.

A short length of #26 wire was soldered to the ends of the winding and secured to the winding with polystyrene dope. A magnified view of the end of the line showing details of construction appears in Figure 4.

Measurements

The method for determining the characteristic impedance of these delay lines was based upon the fact that no reflections occur in an idealized delay line terminated in its characteristic impedance. The value of the characteristic impedance in a practical case involving complex waves must therefore be compromised for minimum reflections over the band of frequencies for which the line is designed to operate. The lines were terminated at the input as well as the output to minimize any possible secondary reflections at the input. A suitable means of determining the effective characteristic impedance when the line is used to delay rectangular pulses is to feed the pulse itself into the delay line and to adjust the terminating impedances for minimum reflections. A block diagram illustrating the experimental method for determining the characteristic impedance of these delay lines and for recording the response of the delay lines to rectangular pulses appears in Figure 5. The pulse generator was of the delay line type.[1] RG65U delay line cable was used. This pulse generator was triggered by a Lavoie type LA-592A pulse generator. A Hewlett Packard type 212A pulse generator was used on occasion in observing the response of the lines to pulses of greater than 1 microsecond duration. A Tektronix type 517 oscilloscope was used. During parts of the work a Dumont type 248 or Tektronix type 513D oscilloscope was used when the type 517 was not available. The oscilloscope sweep was triggered by the input pulse. A camera, mounted on the oscilloscope, was used to record the wave shapes of the pulse at the input and output of the delay line.

73

A video amplifier was placed between the delay line and the pulse generator. The load impedance of the video amplifier was made equal to the characteristic impedance of the line. A diagram of the video amplifier appears in Figure 6.

The pulse distortion and attenuation were also measured with the equipment connected as shown in Figure 5. The oscilloscope camera was used to record the wave forms of both the input and output pulses. Measurements were made directly from the photographs as the sweep of the type 517 oscilloscope is linear and the sweep time in milli-micro-seconds per centimeter is quite accurate. The vertical gain was kept constant for both input and output pulses so that attenuation measurements could be made from the photographs.

The delay time as well as the rise time and fall time was likewise measured on the oscilloscope. The delay time was defined as the time between the mid-point of the leading edge of the input and output pulses. The rise and fall times were defined as the time duration between the 10% and 90% values of the pulse amplitude. The pulse duration was defined as the time between the 10% values. The attenuation was measured by comparing the amplitudes of the input and output pulses.

Although the quality of the lines was determined from the pulse response, as a further check some data were taken with sinusoidal waves. In order to determine the validity of some of the equations a line was connected to a r.f. signal generator through a 10,000 ohm resistor. The Tektronix 517 oscilloscope was used as a VTVM. With the line short circuited, the frequencies at which the impedance of the line was a minimum was noted. The line was thus electrically $\lambda/2$, λ, $\frac{3\lambda}{2}$ etc. long. The measurements were repeated with the line open circuited, the line then being $\frac{\lambda}{4}$, $\frac{3\lambda}{4}$, $\frac{5\lambda}{4}$ etc. long electrically. From this data along with the measured values of L_0 and C_0 and the physical dimensions of the line, x could be calculated.

Results

The data on a particular line, typical of those wound, appears below:
Core diameter 0.188 inch
36 slots
dielectric: teflon 0.001" 6 layers
length 10 inches
winding 1305 turns per inch (0.00076 inch per turn)
#44 single formex wire 0.0020 inch in diameter (.0023" over-all)

The electrical characteristics of the line measured at 1000 cps were:
R = 1901 ohms
L = 17.6 millihenries
G = 0
C = 441 micromicrofarads
The following quantities were calculated from these measurements:
Z_0 = 6320 ohms
t_d = 284 microseconds
With a 1 microsecond pulse the experimental data obtained on this line were:
Z_0 = 6500 ohm resistance
t_d = 3 microseconds
t_{ri} = rise time of the input 1 microsecond pulse = 0.09 microsecond
t_r = rise time of the output 1 microsecond pulse = 0.14 microsecond

as
$$t_r = \sqrt{t_{ri}^2 + t_{rl}^2} \qquad (11)$$

where t_{rl} = rise time of output pulse if a perfect input pulse were applied to the line.
Thus t_{rl} = 0.1 microsecond.

Photographs of the input and output pulses of one #44 wire line appear in Figure 7 for pulse durations of 0.30, 0.37, 0.62, and 1.0 microseconds. Input and output waveforms superimposed to a larger scale are also included in Figure 7. (Note: If one defines the pulse duration as the time between 50% amplitude points, one should subtract 0.1 micro-second from the values of pulse duration given in this paper.) The reflections which appear between the input and output pulses no doubt occur at points where the spill over from true bank winding was particularly bad. All photographs were taken with the same value of terminating impedance which was the value obtained as the best impedance match with a 0.3 microsecond pulse applied. In some cases with longer pulse durations, slightly better waveforms can be secured by reterminating the line. An example will be shown later. In some cases a better termination was secured when a small choke was inserted in series with the resistor. A plot of $\frac{T}{T_0}$ vs x appears in Figure 8 for a typical self-compensated multilayer line along with the theoretical curve for an uncompensated line. It is observed that the variation in time delay for a compensated line is reasonable up to values of x in the neighborhood of 1. The series of points above the curve are from the open circuit data and the lower series from the short circuit data.

Lines with higher characteristic impedances have been obtained using a 1/4 inch diameter core and 2 3/8 layers of 3

TABLE 2

CHARACTERISTICS OF A NUMBER OF 3 LAYER BANK WOUND DELAY LINES

Wire	Dia. Core	Length	Z_o		td (usec)		rise times (usec)			atten- uation db/usec	t_d/" usec	f_o(Mc)
			meas.	calc.	meas.	calc.	t_{ri}	t_r	t_{rl}			
#36 F	1/2 "	4.94"	7,500	6,980	1.45	1.52	0.1	0.33	0.315	0.47	0.29	1.4
#44 F	1/4 "	9.3 "	10,000	9,500	4.5	4.3	0.08	0.24	0.22	0.2	0.48	2
#41 Teflon	1/4 "	2.7 "	7,540	7,000	0.53	0.58	0.09	0.18	0.15	0.45	0.20	2.9
#42 F	1/4 "	10 "	6,500	6,530	3.4	3.3	0.08	0.26	0.24	0.13	0.34	1.9
#44 F	3/16"	10 "	6,500	6,320	3.0	2.8	0.09	0.14	0.11	0.2	0.30	4.2
#44 Teflon	3/16"	8.1 "	9,640*	9,800	1.75	1.7	0.12	0.26	0.230	0.33	0.22	2.0
#46 HF	1/8 "	3.9 "	7,500**	6,420	0.79	0.78	0.08	0.125	0.096	0.3	0.20	4.6
#46 Teflon	1/8 "	6.2 "	9,060	9,050	1.1	1.1	0.08	0.15	0.123	0.42	0.18	3.6

* 50 microhenry choke in series with resistance

** 100 microhenry choke in series with resistance

mil teflon tape to separate the winding from the core. With a 0.3 microsecond pulse the characteristic impedance was 9000 ohms in series with a 400 microhenry choke. The input impedance (shunt impedance in output of video amplifier) was 7400 ohms in series with a choke of 400 microhenries. Photographs of the waveforms of this line appear in Figure 9. This particular line had only 14,165 turns and the winding was 9.3 inches long. It had a time delay of 4.5 microseconds, or a time delay of almost 0.5 microsecond per inch. When the line was reterminated for a 1.0 microsecond pulse, the terminating impedance turned out to be a 10,200 ohm resistor. The output impedance of the video amplifier was increased to 11,000 ohms. The waveforms of this termination also appear in Figure 9. This line was wound with #44 wire.

The characteristics of several 3 layer self-compensated lines are compared in Table 2 for the case of an input pulse of one microsecond duration. The cut off frequency f_o is computed on the basis of the rise time.[12]

$$f_o = \frac{0.445}{T_{rl}} \qquad (12)$$

It will be observed from Table 2 that the smaller the core diameter the higher f_o. Self-compensated lines wound with teflon wire have a higher characteristic impedance and lower time delay than those wound with formex wire due to the lower capacitance between turns. The attenua-

tion in db. per axial inch is about the same for both types of wire indicating that the copper losses are much greater than the dielectric losses in the frequency range of operation. The data on teflon insulated wire lines should be considered as preliminary data only and not necessarily indicative of its performance in this type of line. It is possible that the teflon lines may have some shorted turns. The teflon wire used was single teflon which has the same nominal insulation thickness as formex. It is apparently quite difficult to secure a continuous layer of teflon on the bare wire. It is thus possible, in multilayer construction, to get shorted turns. This possibility should be reduced for a single layer winding. Further experimental work is being performed in this area.

It will be observed that many of the lines in Table 2 are less than 10 inches long. This occurred due to wire breakage or some other mechanical failure. Some lines have been wound with more than 3 bank wound layers. Increasing the number of layers will increase the time delay and decrease the characteristic impedance because the effect of turn to turn capacitance will increase faster than the inductance increases. There are more minor reflections due to spill over in this type of line.

Conclusions

From the data presented, delay lines

with impedances higher than 5000 ohms and
reasonable attenuations for pulse widths
less than 1 microsecond can be obtained.
It appears that the attenuation and minor
reflections can be reduced if a better
winding technique can be developed to
reduce spill over. The preliminary
investigation of teflon insulated magnet
wire for multilayer delay lines has been
disappointing. Further work along these
lines is planned. It is possible that a
combination of self-compensation and
patches may yield higher time delays per
axial inch than self-compensation alone.

The author is deeply indebted to
Mr. Edward F. Seymour and Mr. J. F.
Peoples (Naval Ordnance Plant, Indiana-
polis, Indiana) for their general assist-
ance and for the designing of the wire
feeding device and the core slotting
device. The author wishes to acknowledge
the encouragement of Mr. M. F. Davis and
others in the Electricity and Magnetism
Research Division of the Physics Research
Department.

References

1. Britton Chance et al., "Waveforms",
   McGraw-Hill Book Co., Inc., New York,
   New York, p. 746, p. 240; 1949.
2. John F. Blackburn, "Components Hand-
   book," McGraw-Hill Book Co., New York,
   New York, Chapter 6; 1949.
3. George L. Ragan, "Microwave Trans-
   mission Circuits," McGraw-Hill Book
   Co. Inc., New York, New York, pp.
   20-23; 1949.
4. Heinz E. Kallman, "Equalized Delay
   Lines," Proc. IRE., Vol. 34, pp. 646-
   657; 1946 similar to report #550
   Massachusetts Institute of Technology;
   1944.
5. L. N. Brillouin, "Electromagnetic
   Delay Lines," Proceedings of a Sympo-
   sium on Large Scale Digital Calculat-
   ing Machinery, Harvard University
   Press, Cambridge, Massachusetts; 1948.
6. I. A. D. Lewis, "Notes on the Variation
   of Phase Velocity in Continuously-
   wound Delay Lines at High Frequencies,"
   Proceedings of the Institution of
   Electrical Engineers, Vol. 98, Part
   III, pp. 312-314; 1951.
7. T. F. Blewett et al., "Delay Lines,"
   General Electric Co. Report; May 31,
   1943. This report is essentially the
   same as T. P. Blewett and J. H. Rubel,
   "Video Delay Lines," Proc. IRE., Vol.
   35, pp. 1580-1584; 1947.
8. G. N. Watson, "Theory of Bessel
   Functions" Cambridge University Press,
   Cambridge, England; 1922.
9. R. A. Erickson and H. Sommer, "The
   Compensation of Delay Distortion in
   Video Delay Lines" Proc. IRE., Vol.
   38, pp. 1036-1040; 1950.
10. William S. Carley and Edward F.
    Seymour, "High Characteristic Imped-
    ance Delay Lines for Fractional
    Microsecond Pulses," Proc. National
    Electronics Conference, Vol. 8; 1952.
11. Ernst Weber, "Electromagnetic Fields,
    Theory and Application," Vol. 1,
    John Wiley & Sons, New York, New York,
    p. 150; 1950.
12. David K. Cheng, "A Note on the
    Reproduction of Pulses" Proc., IRE.,
    Vol. 40, p. 963; 1952.
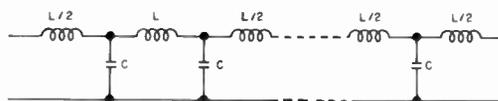
Fig. 1
Simplified equivalent circuit
of a distributed constant
line.
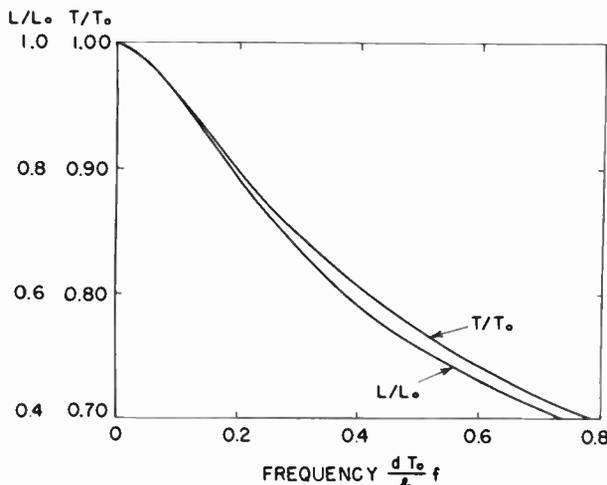
Fig. 2
Plot of $\frac{L}{L_0}$ and $\frac{T}{T_0}$ vs. $\frac{dT_0}{l} f$ (ref. 7).
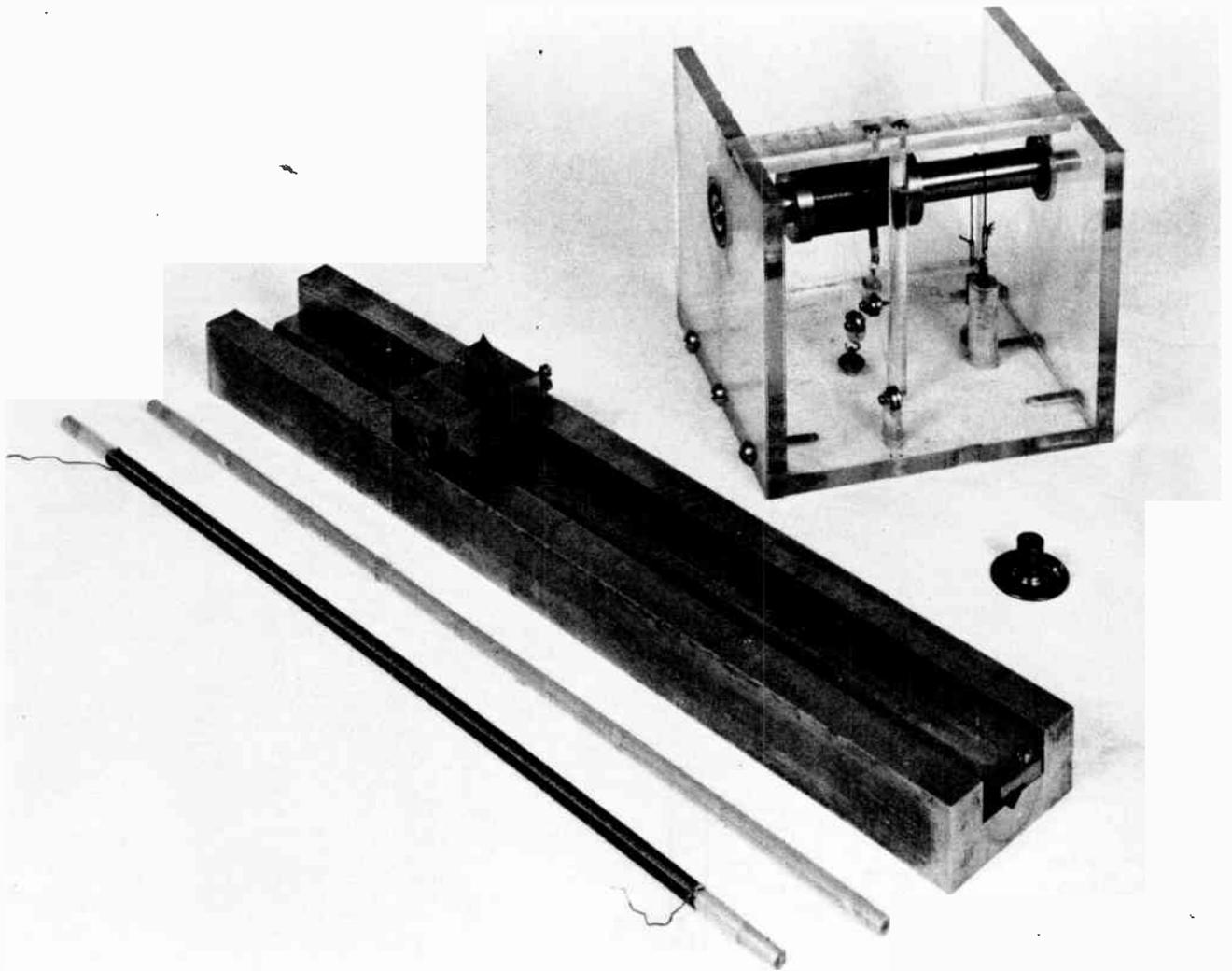
Fig. 3
Wound delay line, slotted core, slotting equipment and wire feeding device.
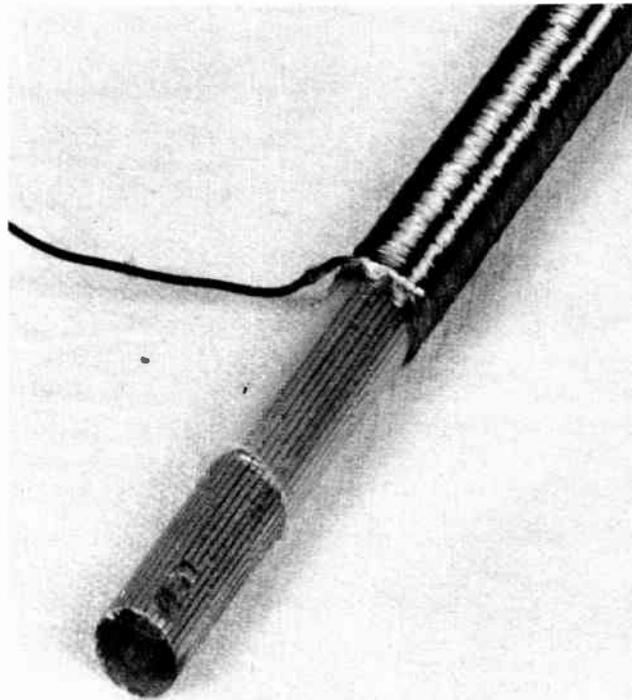
Fig. 4
End of a wound line illustrating core
details, dielectric and winding.
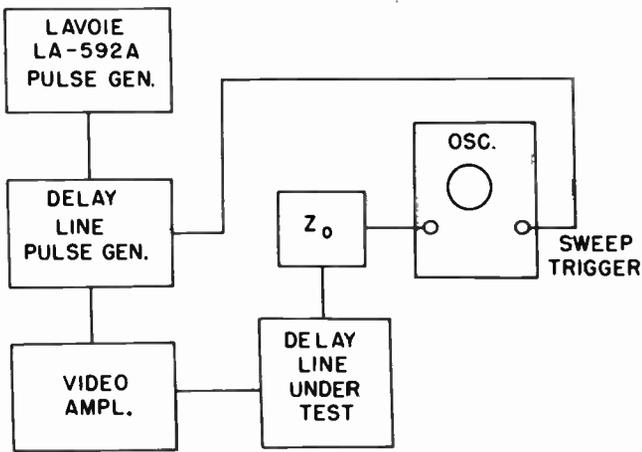
Fig. 5
Block diagram of measuring
equipment.
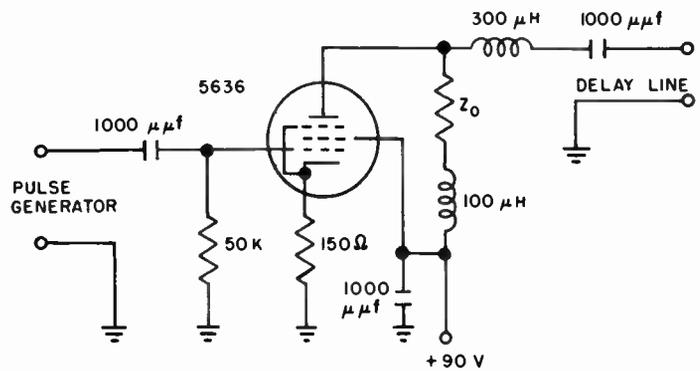
Fig. 6
Video amplifier circuit diagram.

78

INPUT   OUTPUT   INPUT AND OUTPUT

→| |←0.05 μ SEC	→| |←0.1 μ SEC	→| |← 0.5 μ SEC	0.3 μ SEC PULSE
→| |←0.05 μ SEC	→| |←0.1 μ SEC	→| |← 0.5 μ SEC	0.37 μ SEC PULSE
→| |←0.1 μ SEC	→| |←0.1 μ SEC	→| |←0.5 μ SEC	0.62 μ SEC PULSE
→| |←0.25 μ SEC	→| |←0.25 μ SEC	→| |← 0.5 μ SEC	1.0 μ SEC PULSE
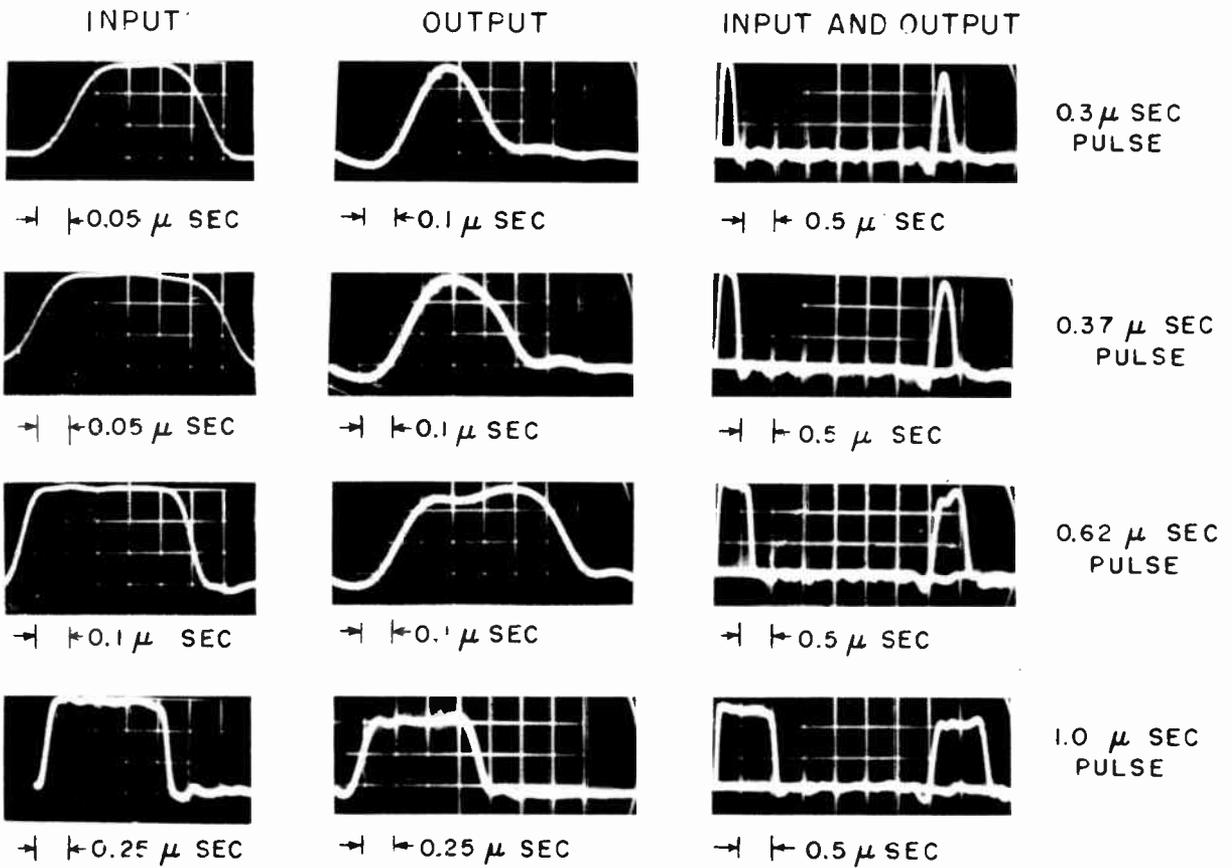
Fig. 7
Pulse response of a 5600-ohm line
wound with #44 wire.
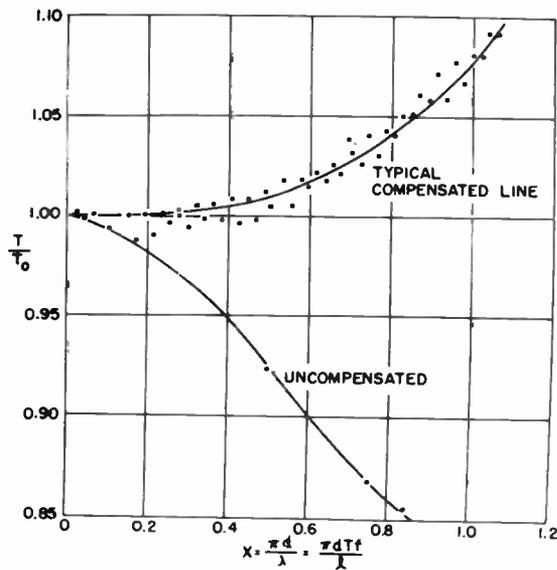
Fig. 8
Plot of T/T_c vs. x (for compensated
and uncompensated line).

INPUT
→| |← 0.2 μ SEC

OUTPUT
→| |← 0.2 μ SEC

INPUT AND
OUTPUT
→| |← 0.7 μ SEC

I μ SEC
PULSE
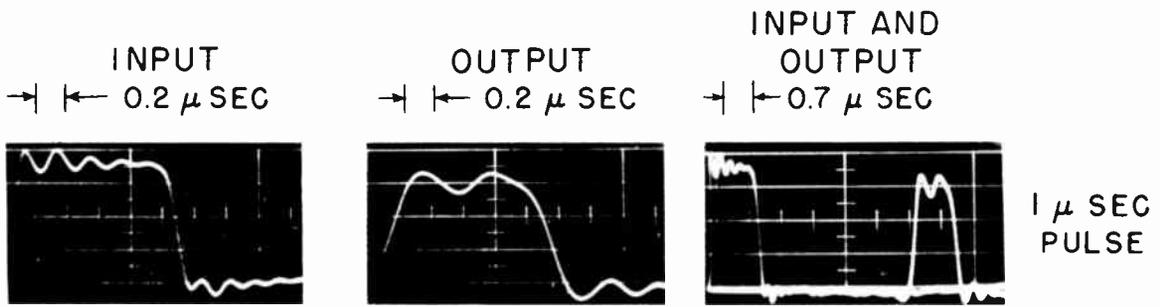
9000 OHM LINE WOUND WITH #44 WIRE AND TERMINATED
FOR A 0.3 μ SEC PULSE

I μ SEC
PULSE
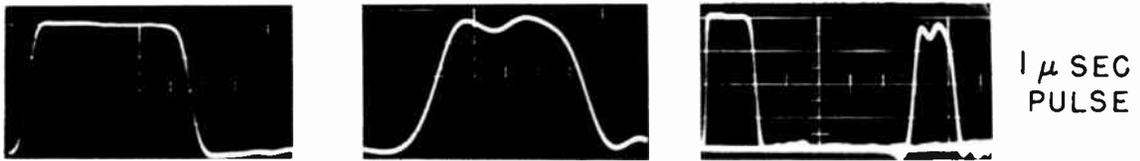
LINE ABOVE TERMINATED IN A 10,000 OHM RESISTOR

Fig. 9
Pulse response of a higher characteristic impedance
line with two values of terminating impedances.

HELICAL WINDING EXPONENTIAL-LINE PULSE TRANSFORMERS
FOR MILLIMICROSECOND SERVICE *

J. Kukel and E. M. Williams
Carnegie Institute of Technology
Pittsburgh 13, Pennsylvania

Introduction

The transient properties of exponentially-tapered transmission lines have been described by Schatz and Williams[1] and the design of simple pulse transformers utilizing these transient properties has been discussed in a second paper by these authors.[2] It has been demonstrated that exponential-line sections offer a means of realizing the pulse transformation of milli-microsecond pulses which are beyond the range of application of conventional ferromagnetic-core transformers with present-day core materials and techniques. Where appreciable transformation ratios or pulse lengths of more than a few millimicroseconds are involved, the simple line sections[2] required have resulted in structures which are frequently unwieldy and difficult to fabricate. Attempts to increase the ratio of electrical length to physical length by using solid or liquid dielectric materials have met with some success, but the improvements attainable are limited in scope because of dielectric losses, frequency dependency of the dielectric constant, and structual difficulties. Furthermore the use of a dielectric to increase the ratio of electrical length to physical length results in a structure of undesirably low impedance level.

A study has been made of a slow-wave structure which uses single-layer shielded helices of varying turn density or varying pitch which is suitable for service as an exponential transmission line[3] for millimicrosecond pulse operation. The slow-wave properties of helical transmission lines of uniform pitch have long been recognized and have found use in the radio art for delay line structures[4,5,6] and in the traveling-wave tube.[7] A helical line of varying pitch has been suggested for use as a broadband transition structure from the conventional coaxial line to a helical line of uniform pitch.[8] The advantages of a tapered helical line for pulse transforming purposes are apparent when it is realized that relatively great electrical lengths are realized with short physical dimensions and that desirable impedance levels are readily obtainable. The helical winding exponential pulse transformer consists of a coaxial structure comprising a constant diameter

cylindrical outer shell which serves as the outer conductor and a concentric single-layer helical inner conductor. The inside conductor is of constant diameter and varying turn density in the case of the wire helix and varying pitch in the case of the tape helix.

Design Problems of the Tapered Helix

In utilizing tapered helices in exponential transmission line pulse transformers the major problem to be solved is the selection of the type and form of helix which will lead to an economical transformer design. This problem was approached through a study of the major types of helices which appear to be particularly appropriate for this application. This paper describes the properties of these helices and the results of some tests of experimental transformers.

The pulse behavior of exponentially-tapered transmission lines can be calculated when such parameters as inductance, capacitance, resistance, and conductance per unit length are known. In the transformers constructed to date resistance and conductance are negligible or nearly so. The effect of the other parameters is most conveniently expressed in terms of impedance level and propagation velocity along the line. The methods of this paper are concerned ultimately with these latter parameters.

A limited amount of information about the inductance and capacitance per unit length, the characteristic impedance, and velocity of propagation of coaxial cables with concentric single-layer uniform helicoidal inner conductors has been available, but the information is incomplete in that helical lines of large densities only, that is, many turns per unit line length, were considered. Recently, Kirschbaum[9] has made a rigorous analysis of uniform helical lines including cases of low turn density. Some of Kirschbaum's results are used in this paper.

Uniform Helical Lines

Inductance, Capacitance, and Impedance in the Uniform Wire Helical Line

A uniform wire helical line is one in which the inner conductor has a uniform turn density. For infinitely long lines in the axial direction, the unit inductance and unit

capacitance expressions (from Kirschbaum) are:

$$L = \frac{\mu_o R_1 \cos \psi}{2\tau} \left\{ 1 - \left(\frac{R_1}{R_2}\right)^2 + \tan^2 \psi \, \ln_e \left(\frac{R_2}{R_1}\right) - 4 \sum_{m=1}^{\infty} \left[1 - \frac{I_m'(mbR_1) K_m(mbR_2)}{I_m'(mbR_2) K_m'(mbR_2)} \right] \cdot K_m'(mbR_1) I_m'(mbR_1) \cos\left(\frac{m\tau r}{\cos \psi}\right) \right\} \quad (1)$$

$$\frac{1}{C} = \frac{R_1}{2\tau k \epsilon_o \cos \psi} \left\{ 2\ln_e\left(\frac{R_2}{R_1}\right) + 4\sum_{m=1}^{\infty} \left[1 - \frac{I_m(mbR_1) K_m(mbR_2)}{I_m(mbR_2) K_m(mbR_1)} \right] \cdot K_m(mbR_1) I_m(mbR_1) \cos\left(\frac{mbr}{\cos \psi}\right) \right\} \quad (2)$$

where

L = inductance per unit length along the helical conductor.

C = capacitance per unit length along the helical conductor.

ϵ_o = permittivity of free space.

μ_o = permeability of free space.

k = dielectric constant of the material filling up the region between the inner conductor and the sheath.

$1/\tau$ = number of turns of the helix per unit length in the axial direction.

R_1 = radius of the helix, the tranverse distance between the axis of the helix and the wire center of the helix.

R_2 = inner radius of the sheath.

r = radius of the wire of the helix.

ψ = helix angle = $\cot^{-1}\left(\frac{2\pi R_1}{\tau}\right)$

I_m = modified Bessel function of the first kind of order m.

I_m' = derivative with respect to the argument of I_m.

K_m = modified Bessel function of the second kind of order m.

K_m' = derivative with respect to the argument of K_m.

$b = \left(\frac{2\pi}{\tau}\right)$

The expressions for L and C were found after making the following assumptions:

1. Wire radius $r \ll R_1$.

2. The sheath and wire are of infinite conductivity.

3. The dielectric occupying the space between the wire helix and the sheath is homogenous, isotropic, and of zero conductivity and loss.

4. The inductance is calculated under steady current conditions (frequency $\rightarrow$ 0) and the capacitance under electrostatic conditions.

The characteristic impedance of the wire helix is given by

$$Z_o = \sqrt{L/C} \quad (3)$$

The complexity of the expressions for L and C results in an inconvenient expression for the characteristic impedance. Kirschbaum has made graphical plots of Z_o in terms of various dimensionless parameters contained in the expression for Z_o. Four sets of these plots are given in Figures 1-4. For the region to the right of the dotted dividing line of Figures 1-4, the inductance and capacitance can be expressed with sufficient accuracy with

$$L = \frac{\mu_o R_1 \cos \psi}{2\tau} \left(1 - \frac{R_1^2}{R_2^2}\right) \quad (4)$$

$$\frac{1}{C} = \frac{R_1}{\tau k \epsilon_o \cos \psi} \ln_e\left(\frac{R_2}{R_1}\right) \quad (5)$$

The characteristic impedance is then given by

$$Z_o = \sqrt{\frac{\mu_o}{\epsilon_o}} \frac{R_1}{\tau \sqrt{k}} \sqrt{\frac{1}{2}\left[1 - \left(\frac{R_1}{R_2}\right)^2\right] \ln_e\left(\frac{R_2}{R_1}\right)} \quad (6)$$

where

$\sqrt{\frac{\mu_o}{\epsilon_o}}$ = impedance of free space = 376.7 ohms.

$1/\sqrt{\mu_o \epsilon_o}$ = velocity of light = 3 X 10^8 meters per second. The expressions given in equations (4) - (6) are often employed with helical lines of the type under consideration without consideration as to their limits of application. Figures 1-4 indicate the region of application together with slight deviations due to wire size.

Velocity of Propagation in the Uniform Wire Helix

The velocity of propagation along the helical conductor is dependent upon the turn density of the helix. The velocity at a turn density of zero for which the line becomes an ordinary coaxial line with a straight inner conductor not concentric with the sheath, that is, having eccentricity, is

$$V_{\tau \rightarrow \infty} = \frac{1}{\sqrt{k \mu_o \epsilon_o}} \quad (7)$$

The velocity for infinite turn density is

$$V_{\tau \rightarrow 0} = \frac{1}{\sqrt{k \mu_o \epsilon_o}} \sqrt{\frac{\ln_e\left(\frac{R_2}{R_1}\right)^2}{1 - \left(\frac{R_1}{R_2}\right)^2}} \quad (8)$$

The velocity of propagation for values of turn density other than those of zero and infinity lies somewhere between the values given in equations (7) and (8). For those values of turn density for which equations (4), (5), and (6) are valid the velocity of propagation is

$$V = \frac{1}{\sqrt{k \mu_o \epsilon_o}} \frac{1}{\cos \psi} \sqrt{\frac{\ln_e\left(\frac{R_2}{R_1}\right)^2}{1 - \left(\frac{R_1}{R_2}\right)^2}} \quad (9)$$

Properties of the Uniform Tape Helix

The uniform tape helix differs from the uniform wire helical line in that a perfectly conducting tape is used instead of a wire conductor. The expression for capacitance is given by

$$\frac{1}{C} = \frac{R_1}{\tau k \epsilon_o \cos \psi} \ln_e\left(\frac{R_2}{R_1}\right) \quad (10)$$

The velocity of propagation on the helix in the tape direction is

82

$$v = \frac{1}{\sqrt{k\mu_o\epsilon_o}\,\sin\psi\,\sqrt{1+\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\frac{\cot^2\psi}{\ln_e\left(\frac{R_2}{R_1}\right)^2}}} \qquad (11)$$

where

   C = capacitance per unit length on the helix in the tape direction.

   v = velocity of propagation on the helix in the tape direction.

   R_1 = radius of the helix.

   R_2 = inner radius of the sheath.

   T = pitch of the tape, the tape width measured in the longitudinal direction.

   k = dielectric constant of the material between the tape and the sheath.

   ψ = helix angle = $\cot^{-1}\left(\frac{2\pi R_1}{T}\right)$

The assumptions made in calculating the expressions given in equations (10) and (11) are:

   1. Tape is of infinitesimal thickness.
   2. Coaxial line is infinite in extent.
   3. The sheath and tape are of infinite conductivity.
   4. The dielectric occupying the space between the tape and the sheath is homogeneous, isotropic, and of zero conductivity and loss.
   5. The tape helix is a butted-tape helix; one in which the spacing between successive turns is zero.
   6. Capacitance is calculated assuming a uniform charge density on the tape and the velocity of propagation (frequency $\longrightarrow 0$) assuming that the current is restricted to flow in the helix direction.

   Equation (10) is exactly the same as the capacitance expression for a wire helix of large turn density given in equation (5). A very close similarity is expected (in this instance exactly the same) since a wire helix of large turn density appears physically to be somewhat the same as a butted-tape helix. The velocity expressions, equations (9) and (11) also agree for large turn density for then

$$\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\frac{\cot^2\psi}{\ln_e\left(\frac{R_2}{R_1}\right)^2} \gg 1$$

in equation (11). The expression for the unit inductance in the tape direction is

$$L = \frac{\mu_o R_1 \ln_e\left(\frac{R_2}{R_1}\right)\sin^2\psi}{T\cos\psi}\left\{1+\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\frac{\cot^2\psi}{\ln_e\left(\frac{R_2}{R_1}\right)^2}\right\} \qquad (12)$$

The characteristic impedance is given by

$$Z_o = \sqrt{\frac{\mu_o}{\epsilon_o}}\frac{R_1\tan\psi}{T\sqrt{k}}\ln_e\left(\frac{R_2}{R_1}\right)\sqrt{1+\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\frac{\cot^2\psi}{\ln_e\left(\frac{R_2}{R_1}\right)^2}} \qquad (13)$$

Figure 5 gives a plot of the characteristic impedance, equation (13), for various ratios of (R_2/R_1). The equation for the dotted straight lines of Figure 5 is

$$Z_o = \sqrt{\frac{\mu_o}{\epsilon_o}}\frac{R_1}{T\sqrt{k}}\sqrt{\frac{1}{2}\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\ln_e\left(\frac{R_2}{R_1}\right)} \qquad (14)$$

This is the same as the special impedance expression for the wire helix, equation (6).

   Comparison of the impedance data of the wire helical line and the tape helical line shows that much lower impedance values can be obtained with the tape line than with the wire line.

Helical Exponential Lines

Impedance Level Distribution

   If the turn density of the wire helix or the pitch of the butted-tape helix is made to vary slowly, the characteristic impedance, or more appropriately, the impedance level at any particular point is given approximately by either the uniform wire helical line characteristic impedance expression or the uniform butted-tape helical line characteristic impedance expression evaluated at that point. Evaluation at the point in question implies the use of the turn density or the pitch at that particular point. Under the conditions of small variations of impedance level almost any impedance level distribution can be effected. The impedance level distribution of immediate interest is that of an exponential transmission line. Suppose that the values of the impedance level distribution desired lie in the range for which equation (6) is valid in one instance and equation (14) in the other instance. Equating the exponential transmission line impedance distribution to either equation (6) or (14),

$$Z_1 e^{\gamma x} = \sqrt{\frac{\mu_o}{\epsilon_o}}\frac{R_1}{T\sqrt{k}}\sqrt{\frac{1}{2}\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\ln_e\left(\frac{R_2}{R_1}\right)} \qquad (15)$$

where

   Z_1 = nominal impedance at the input of the exponential line.

   γ = flare coefficient (assumed positive).

   x = distance measured along the helix from the input of the line.

Solving for T, the turn spacing for the wire line and the axial tape width for the tape line,

$$T = \frac{\sqrt{\frac{\mu_o}{\epsilon_o}}}{Z_1\sqrt{k}}R_1 e^{-\gamma x}\sqrt{\frac{1}{2}\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\ln_e\left(\frac{R_2}{R_1}\right)} \qquad (16)$$

The distance X can be approximated by $2\pi R_1 n$ where n is the number of turns from the input end at which the position X occurs. The turn spacing or the axial tape width at the nth turn is given by,

$$T_n = \frac{\sqrt{\frac{\mu_o}{\epsilon_o}}}{Z_1\sqrt{k}}R_1 e^{-2\pi R_1 \gamma n}\sqrt{\frac{1}{2}\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\ln_e\left(\frac{R_2}{R_1}\right)} \qquad (17)$$

83

The total axial length of the helical line is found by summing T_m

$$\sum_{m=1}^{S} T_m = \frac{\sqrt{\frac{\mu_0}{\epsilon_0}}}{Z_1 \sqrt{k}} R_1 \sqrt{\frac{1}{2}\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\ln_e\left(\frac{R_2}{R_1}\right)} \sum_{m=1}^{S} e^{-2\pi R_1 \tau m} \quad (18)$$

where

  S = total number of turns of the line

  = $\dfrac{\text{length of the helical conductor}}{2\pi R_1}$

(to nearest integer).
The series given in equation (18) is a geometric progression and is summable.

$$\sum_{m=1}^{S} T_m = \frac{\sqrt{\frac{\mu_0}{\epsilon_0}}}{Z_1 \sqrt{k}} R_1 \sqrt{\frac{1}{2}\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\ln_e\left(\frac{R_2}{R_1}\right)} \left[\frac{1-e^{-2\pi R_1 \tau S}}{e^{2\pi R_1 \tau}-1}\right] \quad (19)$$

The ratio of axial length to the length of the helical conductor is given by,

$$\frac{\sum_{m=1}^{S} T_m}{2\pi R_1 S} = \frac{\sqrt{\frac{\mu_0}{\epsilon_0}}}{2\pi S Z_1 \sqrt{k}} \sqrt{\frac{1}{2}\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\ln_e\left(\frac{R_2}{R_1}\right)} \left[\frac{1-e^{-2\pi R_1 \tau S}}{e^{2\pi R_1 \tau}-1}\right] \quad (20)$$

The percent change in impedance level per turn based on the impedance level of the preceding turn is,

  $\%$ change per turn = $\left(1-e^{-2\pi R_1 \tau}\right)$ $\quad$ (21)

As implied earlier, expressions for impedance level, derived for uniform lines, lose some accuracy when applied even to gradually tapered lines. At present the relation between rate of taper and the accuracy with which uniform line expressions can be applied is not known. However experimental tests of the transformers constructed to date have disclosed no inaccuracies within the experimental error. The experimental error in studies of millimicrosecond pulse response are, however, quite high so that this problem is considered important for further study.

If impedance level variation per turn is small (approximately 5% or less), then equation (21) becomes
  $\%$ change per turn = $2\pi R_1 \gamma$ $\quad$ (22)
The value of the flare coefficient and the desired upper limit for the impedance level change per turn in conjunction with equation (22) gives the maximum allowable value for the helix radius. Equations (19) and (20) can be rewritten using the approximation of equation (22).

$$\sum_{m=1}^{S} T_m = \frac{\sqrt{\frac{\mu_0}{\epsilon_0}}}{2\pi \gamma Z_1 \sqrt{k}} \left(1-\frac{1}{a^2}\right) \sqrt{\frac{1}{2}\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\ln_e\left(\frac{R_2}{R_1}\right)} \quad (23)$$

$$\frac{\sum_{m=1}^{S} T_m}{2\pi R_1 S} = \frac{\sqrt{\frac{\mu_0}{\epsilon_0}}\left(1-\frac{1}{a^2}\right)}{(2\pi)^2 R_1 \gamma S Z_1 \sqrt{k}} \sqrt{\frac{1}{2}\left[1-\left(\frac{R_1}{R_2}\right)^2\right]\ln_e\left(\frac{R_2}{R_1}\right)} \quad (24)$$

where
  a^2 = nominal impedance level ratio of

transformation; the ratio of the output level to the input level of impedance.

For given values of impedance level ratio of transformation and helical line length, the axial line length, equation (23), becomes a minimum when

$$\ln_e \left(\frac{R_2}{R_1}\right)^2 = \left[\frac{\left(\frac{R_1}{R_2}\right)^2-1}{2\left(\frac{R_1}{R_2}\right)^2-1}\right]$$

or
$$\left(\frac{R_2}{R_1}\right) \cong 2.06 \quad (25)$$

This ratio then permits the evaluation of R_2 for minimum axial length since equation (22) gives the value of R_1 after the impedance level change per turn is specified.

In some instances the impedance level distribution desired may not lie in the region of application of equations (6) and (14). As a result a semi-graphical method of evaluation must be used. The impedance level for each turn is calculated from the input level of impedance and the percent change of impedance level per turn. The corresponding turn spacing or tape pitch is found from the curves of Figures 1, 2, 3, 4, and 5. By summing the turn spacing or tape pitch, the axial line length is found.

For those instances for which the desired impedance level distribution overlaps into the two regions discussed, the analysis can be broken into two parts. In the one part is used the semi-graphical analysis and in the other the analytical analysis.

The desirable low impedance level characteristic of the tape line and the ease of construction of a wire line can be combined to form a line containing both tape and wire. The impedance curves of both lines overlap for close wound helices, and hence for suitable impedance level distributions, one portion of the line can be tape and the other portion wire with the proper transition between the two (impedance levels the same at the junction).

Propagation Velocity

An exponential distribution of impedance level placed on a helical line does not necessarily imply that the line will operate as an exponential transmission line.* A basic requirement in the exponential transmission line is that the velocity of propagation be constant. For a uniform helical line the velocity of propagation along the helical conductor as well as on the sheath in the direction of the helix is not a constant with respect to the turn spacing or tape pitch, helix radius, and sheath radius but is dependent upon these quantities. The propagation velocity under

*Such a line would nevertheless operate as a transformer but prediction of its properties would be very difficult.

the assumption of small helix angle is given by equation (8) and shall be taken as the nominal velocity of propagation of the helical exponential line. For the special case $(R_1/R_2) = 2.06$ the propagation velocity is

$$\nu = \frac{1.37}{\sqrt{k \mu_4 \epsilon_0}} \qquad (26)$$

Since the velocity of propagation of a helical exponential is of the order of a constant, the line should operate to a first approximation as an exponential transmission line.

The ratio of the axial length to the length of the helix of a helical exponential transmission line is not the same as the ratio of the axial length of the helical line to the electrical length of a corresponding two wire or coaxial exponential line where the electrical length is the same as the length of the helix of the helical line. They differ because the nominal helical propagation velocity is not the same as the propagation velocity of the two wire and coaxial lines.

Frequency, Attenuation, and Breakdown Effects

The inductance of a uniform helical line was calculated assuming a constant current. If the current is given a monochromatic characteristic, the effective inductance becomes a function of frequency. When the frequency is increased to a level so high that turns having currents which are appreciably out of phase with each other and which are magnetically coupled with each other, then the effective inductance is adversely affected and is thus decreased. A plot of inductance versus wave length for unshielded coils but applicable for shielded coils is given by Blewett and Rubel.[4] In the use of the helical line as a pulse transformer the decrease in unit inductance for the most important frequency components of the pulse to be used must be small in order that the exponential transmission line theory be applicable and in order that unnecessary phase distortion be small.

No mention has been made of attenuation effects due to the finite conductivity of the conductors and the finite power factor of the dielectric. If the conductors are made of copper or something comparable the attenuation effects due to conductor energy loss can usually be considered negligible. For small real energy losses an analysis of the type given by Schatz and Williams[1] is applicable.

A butted tape line of the type that has been considered is not physically realizable. However, if a line is constructed with spacing between successive turns small compared with the tape width, then the butted tape analysis is reasonably correct. A factor which influences the spacing between successive turns of the tape line as well as the wire line is dielectric breakdown. The spacing must be sufficiently large such that the greatest potential difference that will exist between two adjacent turns is not sufficient to cause dielectric breakdown.

Experimental Results

Two helical exponential transmission lines were constructed, one of the wire type and one of the tape type. A complete picture of the disassembled wire line is given in Figure 6 and a partial picture of the disassembled tape line is given in Figure 7. The helices were wound on lucite tubing in order to insure low loss and dielectric constant not far from one. Incorporated in the low impedance level end of both lines for broadband matching is a 5C22 hydrogen thyratron which is suitable for short pulse work[10] and which serves in this instance as a switch between the helical line and a pulse forming line (not shown). Pictures of input and corresponding output pulses for the line terminated in its output level of impedance are given in Figures 8 and 9. The pulse results compare favorably with expected results. Table 1 contains the calculated parameters along with some experimental data. A close examination of the given pictures and the given data indicates that the operation of helical transmission lines with exponential impedance level distribution is comparable to that of exponential transmission lines and is suitable as a pulse transformer for millimicrosecond service.

References

1. E. R. Schatz and E. M. Williams, "Pulse transients in exponential transmission lines," Proc. I.R.E., vol. 38, pp. 1208-1212; October, 1950.
2. E. M. Williams and E. R. Schatz, "Design of exponential line pulse transformers," Proc. I.R.E., vol. 39, pp. 84-86; January, 1951.
3. E. Von Keutner, "Hochfrequenzkabel mit veränderlichem wellenwiderstand," Europaeischer Fernsprecherdienst, no. 62, pp. 3-9; March, 1943.
4. J. B. Blewett and J. H. Rubel, "Video delay lines," Proc. I.R.E., vol. 35, pp. 1580-1584; December, 1947.
5. K. H. Zimmerman, "Spiral delay lines," Electrical Communications, vol. 23, p. 327; September, 1946.
6. H. Von Kaden, "Properties of a R.F.cable whose inner conductor is a helix," Telegraphen Fernsprech Funk-und Fernseh-Technik, vol. 32, no. 9, p. 195; September, 1943.
7. J. R. Pierce, "Traveling-wave tubes," D. Van Nostrand Company, New York, N. Y.; 1950.
8. C. O. Lunt, "A broadband transition from coaxial line to helix," R. C. A. Review, vol. 11, pp. 133-142; March, 1950.

9. H. S. Kirschbaum, Ph.D. thesis at Carnegie Institute of Technology, Pittsburgh 13, Pa.; 1953.

10. J. B. Woodford, Jr. and E. M. Williams, "The initial conduction interval in high speed thyratrons," J. of Appl. Phys., vol. 23, pp.722-724; July, 1952.
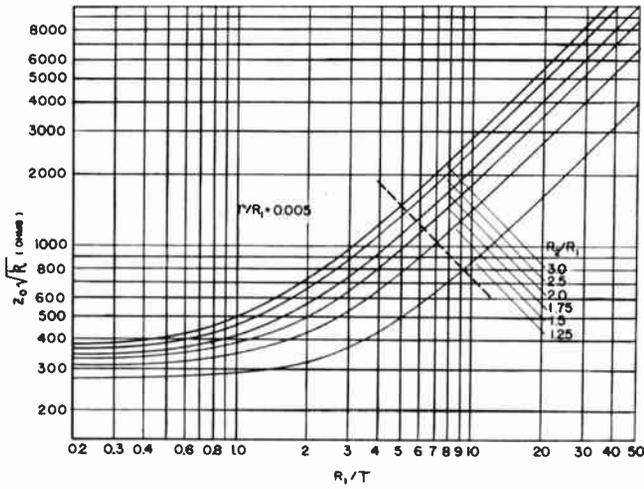
Fig. 1
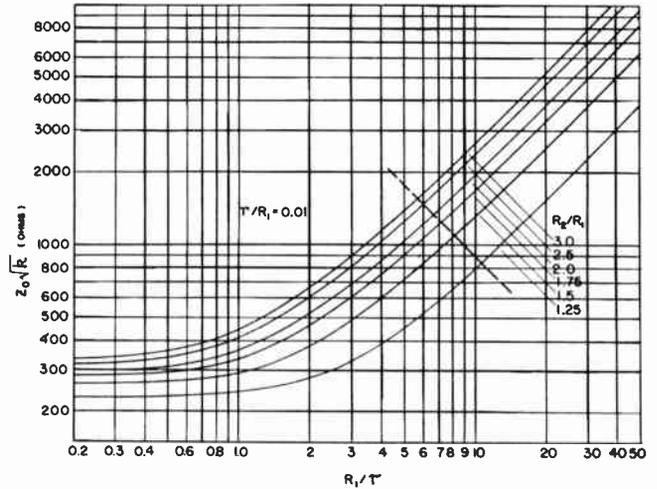
Fig. 2
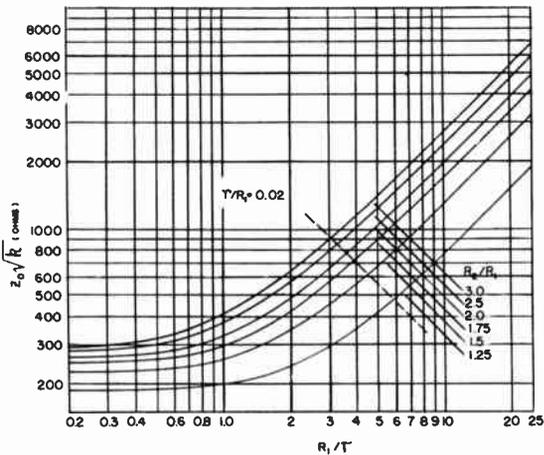
Fig. 3
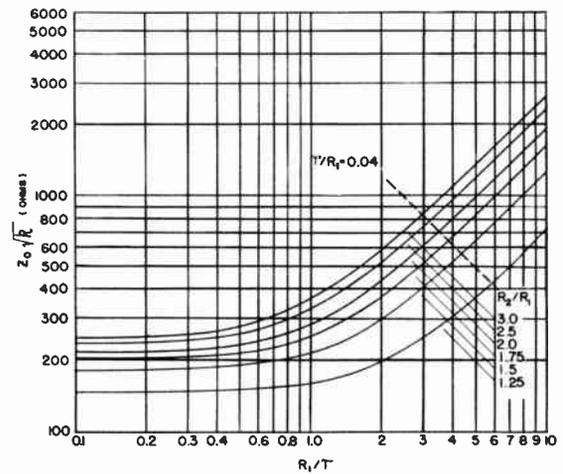
Fig. 4

Characteristic impedance of shielded wire helices.
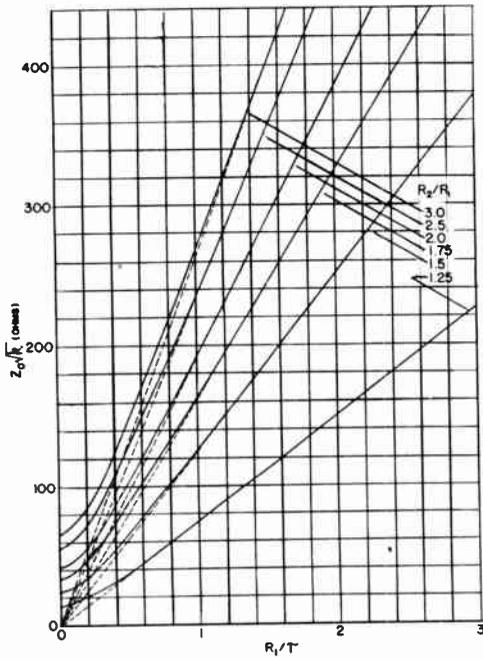
Fig. 5
Characteristic impedance of
shielded tape helices.
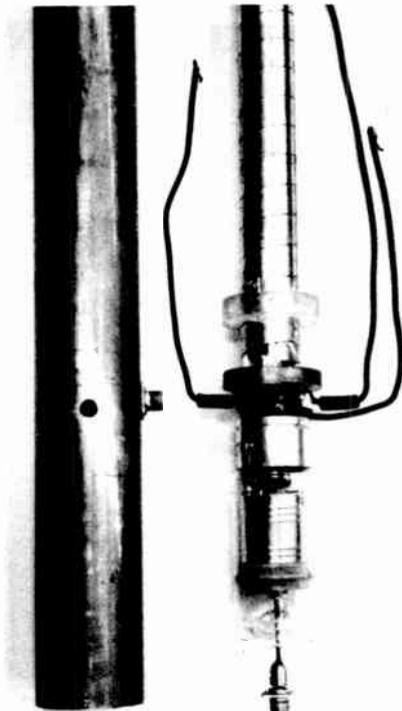
Fig. 7
Exponential-line section pulse
transformer consisting of tapered
tape and cylindrical outer
shell together with integral
hydrogen-thyratron.
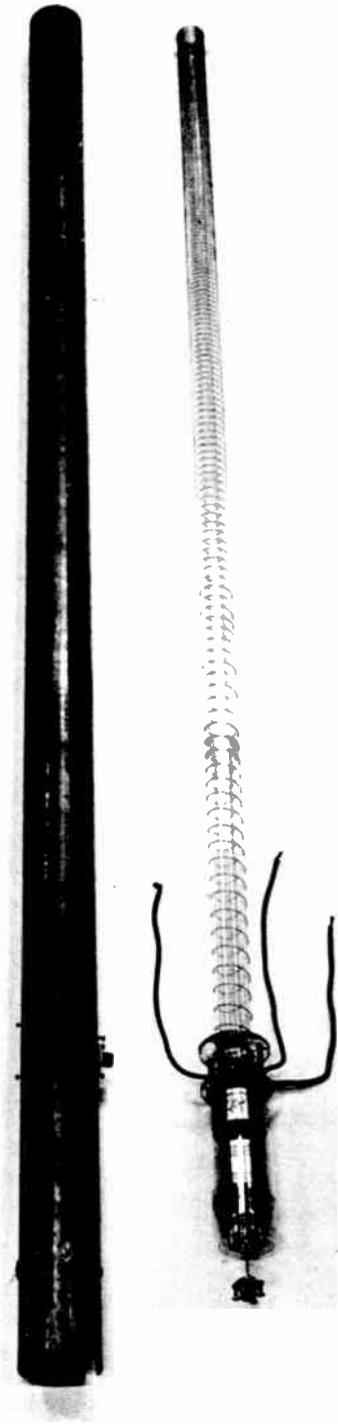
Fig. 6
Exponential-line section pulse
transformer consisting of tapered
wire helix and cylindrical outer
shell together with integral
hydrogen-thyratron.

87
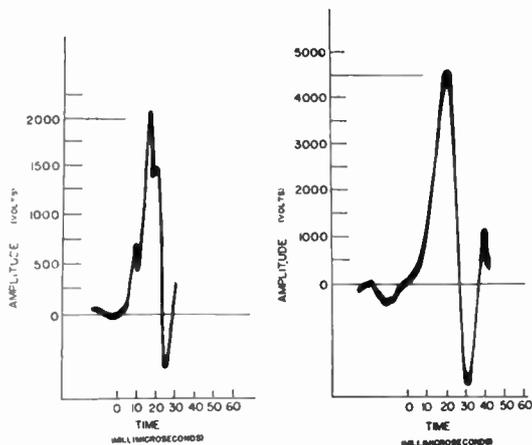
Fig. 8
Left, input pulse; right,
corresponding output pulse;
for test of experimental
helical-line wire pulse
transformer terminated in
its output level of
impedance.
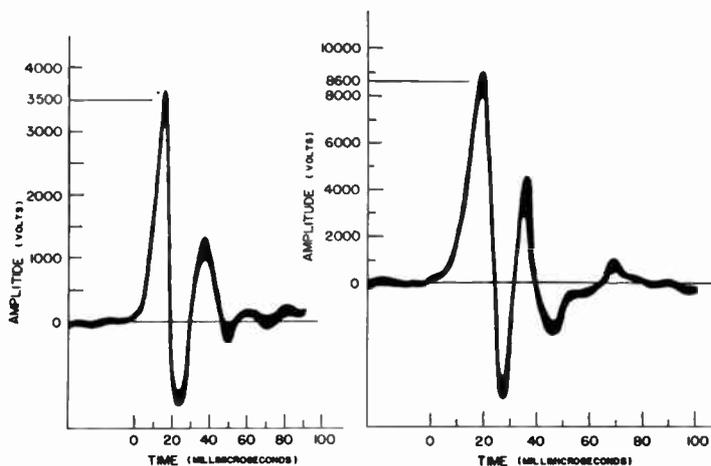
Fig. 9
Left, input pulse; right,
corresponding output pulse;
for test of experimental
helical-line tape pulse
transformer terminated in
its output level of
impedance.

Table I

LINE PARAMETERS	WIRE LINE	TAPE LINE
Helix radius (inch)	0.766	0.878
Sheath radius (inch)	1.47	1.47
Wire diameter (inch)	0.032	
Tape thickness (inch)		0.003
Tape spacing (inch)		0.01
Axial length (feet)	5.8	7.4
Axial / wire length tape	1/14.4	1/14.2
Number of turns	210	230
Velocity of propagation $\left(\frac{meter}{sec.}\right)$	4.0×10^8	3.8×10^8
Flare coefficient γ (per meter)	0.082	0.0643
% change in impedance level per turn	1.0	0.9
Calculated input level of impedance (ohm)	315	143
Experimental input level of impedance (ohm)	290	145
Calculated output level of impedance (ohm)	1600	1120
Experimental output level of impedance (ohm)	1590	1070

88

TIME DOMAIN APPROXIMATION BY USE OF PADE' APPROXIMANTS

R. D. TEASDALE
RCA VICTOR
CAMDEN, NEW JERSEY

Introduction

In engineering applications of networks or servomechanisms, it is often necessary to synthesize a complicated transfer function.

It is particularly convenient if the complicated transfer function can be approximated by a ratio of rational polynomials, which then can be physically realized as a network by Brune's method or otherwise. It is also desired that the approximation be effected in such a way that the error in the time domain is small and predictable.

It is possible to approximate a given transfer function to minimize the error in the complex frequency domain (the direct approximant), or to minimize the error in the time domain (the indirect approximant). The direct approximant was first developed by Pade 1 in 1892. However, the indirect approximant has been discussed only recently by Mathers[4].

Here, the basic theory is presented and is used to develop successive Pade approximants for several functions useful in network theory. The results are summarized in tables and the accuracy of approximation is illustrated by plots.

The Pade Method

Consider the approximation of a function $F(Z)$ (where Z may be complex) which is analytic about the origin. For every power series

$$F(Z) = a_0 + a_1 Z + \ldots \quad (a_0 \neq 0) \qquad (1)$$

and for every specified ordered pair of integers p,q, it is possible to find two rational polynomials $N(Z)$ and $D(Z)$ which satisfy two conditions:

1. The degree of N is p or lower and the degree of D is q or lower.
2. The coefficients of N and D are determined so that the power series of FD-N begins with the $(p + q + 1)$ st or a higher power of Z.--that is, so that

$$F(Z)D(Z) - N(Z) = \left\{ Z^{p + q + 1}_{+ \ldots} \right\} \qquad (2)$$

If in any way N and D can be found so that (2) is satisfied, then $G(Z)$, the corresponding Pade approximant to $F(Z)$, is the rational function

$$(p,q) = \frac{N(Z)}{D(Z)} = G(Z) \approx F(Z) \qquad (3)$$

To each $F(Z)$, then, there corresponds a sequence of Pade approximants for different choices of p and q. These successive approx-

imants can conveniently be entered in a double entry Pade Table of successive approximants. Each diagonal of this table provides an infinite sequence of distinct successive approximants for $F(Z)$.

$$p \longrightarrow$$

	(0,0)	(1,0)	(2,0)	...
q	(0,1)	(1,1)		
↓	(0,2)		(2,2)	
	(0,3)			
	...			

For example, the successive Pade approximants of $F(Z)$ are

$$(0,0) = a_0 \qquad (1,0) = a_0 + a_1 Z$$

$$(0,1) = \frac{a_0 b}{Z + b} \qquad (1,1) = \frac{a_0 d + cZ}{Z + d}$$

These are all normalized to give

$$F(0) = a_0 \text{ for } Z = 0$$

The basic problem of determining the coefficients of the polynomials N and D can be approached in different ways, depending on the form of $F(Z)$.

1. One can equate coefficients in equation (1) and solve simultaneously the resulting $p + q + 1$ linear equations. In actual problems this procedure is often lengthy.
2. Formal solutions for N and D can be written in the form of determinants which must then be expanded. See Appendix I.
3. It is sometimes possible to exploit the fact that an algebraic function is completely determined, up to a multiplying constant, when the locations of its poles and zeroes (and their multiplicity) are known. See Bellman and Straus[7] and Appendix II. The left-hand side must have a zero at $Z = 0$ if the right-hand side has one.
4. If by some artifice, one can obtain the continued fraction expansion of $F(Z)$, he can then find the required coefficients. See Perron[2] and Wall[9].

Advantages:

There are several definite advantages and useful features of the Pade method:

1. The approximant is given directly as a ratio of rational polynomials. This result is especially useful when dealing with transfer functions and when synthesis by a network is necessary.

2. Cne can specify in advance the number of zeroes p and the number of poles q i.e. the degree of the numerator and denominator of the rational function which is the approximant.

3. After building up a Padé Table for a given function one can get successive approximants by merely reading down any diagonal of the table. This procedure yields a rapidly converging sequence of approximants which have a small number of poles and zeroes.

4. If one approximant is known, others can often be obtained by use of recursion relations.

It is important to decide whether our approximations are to be made in the original domain (s) or in a transform domain (Z). We may approximate a transfer function F(s) to minimize the error in F(s) itself (a <u>direct</u> approximant in the s domain), or we may approximate F(s) to minimize the error in the corresponding f(t) (an <u>indirect</u> approximant using an auxiliary E plane).

The direct approximation is, of course, generally easier to obtain than is the indirect one. Both are discussed here.

Cne can define the "best" approximation in several different ways. The criterion of least mean squares is a useful one and is the one given by the Pade approximants.

The <u>Direct</u> Pade Approximant

If it is required to approximate a given function F(z) by G(z), a ratio of rational polynomials, so that the error is a minimum in the least squares sense, the procedure is straightforward.

With F(Z) = a_o + $a_1 Z$ + ... + $a_n Z^n$ + ...,

consider the integral I along the unit circle in the Z plane

$$ I = \int_{-\pi}^{+\pi} \left[F(Z) - Q_n(Z) \right]^2 \, d\theta \qquad (4) $$
$$ Z = 1e^{j\theta} $$

If all polynomials $Q_n(Z)$ of specified degree n are considered, it can be shown that to <u>minimize</u> I one must set $Q_n(Z) = q_n$, where q_n is the nth partial sum of the series for F(Z), i.e.

$$ q_n = \sum_{i=0}^{n} a_i Z^i $$

However, here G(z) is to be a rational function and <u>not</u> the polynomial $Q_n(Z)$. So one merely sets G(z) equal to the ratio of two rational polynomials N and D of specified degrees p and q, and determines the coefficients of these two polynomials in such a way that the power series expansion of $\left[F(z) - G(z) \right]$ about Z = 0 begins with Z^{p+q+1}, i.e., so that relation (2) is satisfied.

Then G(Z) = $\dfrac{N(Z)}{D(Z)}$ = (p,q) is called the direct Padé approximant of F(Z).

The <u>Indirect</u> Padé Approximant

The discussion is aided by reference to a diagram.
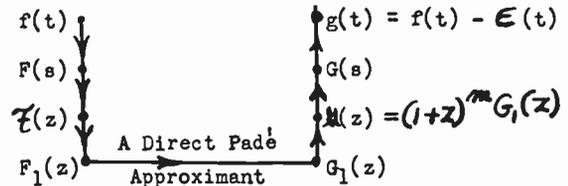
To a function f(t) there corresponds a Laplace transform F(s). The objective is to approximate F(s) by G(s), a ratio of rational polynomials, in such a way that $|\epsilon(t)|$, the error in the time domain, is a minimum in the least squares sense, where

$$ \left| \epsilon(t) \right| = \left| f(t) - g(t) \right| $$

The general procedure is as follows

1. From f(t), get F(s)

2. Use the conformal transformation S = $\dfrac{1 - Z}{1 + Z}$ and form $F\left(\dfrac{1 - Z}{1 + Z} \right) = \mathcal{F}(Z)$

3. Remove the m zeroes of $\mathcal{F}(Z)$ at Z = -1 by forming $F_1(Z) = \dfrac{\mathcal{F}(Z)}{(1 + Z)^m}$

4. Expand $F_1(Z)$ in a power series about Z = 0 and proceed as before to find $G_1(Z)$ by a <u>direct</u> Pade approximant.

5. Work back through $\mathcal{G}(Z)$ to G(S) = $K\mathcal{G}\left(\dfrac{1 - S}{1 + S} \right)$

The constant K has been introduced as a normalizing factor, to be chosen so that G(o) = F(o). This G(S) is the <u>indirect</u> Padé approximant of F(s) and will be, in general, different from the direct Padé approximant.

Examples and Plots

The time function $\dfrac{J_1(t)}{t}$ has been studied as the impulse response of a two-terminal network with maximum gain-bandwidth product. [4] The corresponding Laplace transform is

$$ F(S) = \sqrt{1 + s^2} - S $$
$$ = 1 - S + \frac{s^2}{2} - \frac{s^4}{8} + \frac{s^6}{16} - \dots \ (S \leq 1) \qquad (5) $$

It is instruc-

tive to find both the direct and the indirect
Pade approximants $(0,1)$ of this $F(S)$

The Direct Approximant of $F(s)$. Now
$(0,1) = \frac{a}{S + b}$, but a = b because $F(o) = 1$, and
so $(0,1) = \frac{b}{s + b} = \frac{N}{D}$. Also $p + q + 1 = 2$. Now
b must be determined so that $FD-N = \{S^2\}$ from (2).
Substitution yields b = 1, and so the direct Padé
approximant of $\sqrt{1 + S^2}$ - S is $(0,1) = \frac{1}{S + 1}$.

$$(0,1) = \frac{1}{S + 1} \qquad (\boldsymbol{5})$$

By continuing in this fashion it is possible
to form a Table of **direct** Padé approximants
for $\sqrt{1 + S^2}$ - S

$(0,0)$	$(1,0)$	$(2,0)$
1	$1-S$	$\frac{2-2S+S^2}{2}$
$(0,1)$ $\frac{1}{S+1}$	$(1,1)$ $\frac{-S+2}{S+2}$	
$(0,2)$ $\frac{2}{S^2+2S+2}$	$(1,2)$ $\frac{2}{S^2+2S+2}$	

The work of constructing such a table
can be reduced by using the relation for
the reciprocal of $F(S)$

$$(p,q) \text{ for } F(S) = \frac{1}{(q,p) \text{ for } \frac{1}{F(S)}}$$

Two other methods of obtaining a direct approx-
imant are given in Appendix I and II.

The Indirect Approximant of $F(S)$.

1. For $f(t) = \frac{J_1(t)}{t}$, $F(S) = \sqrt{1+S^2}$ - S

2. $\mathcal{F}(Z) = F\left(\frac{1-Z}{1+Z}\right) = \frac{\sqrt{2}\sqrt{1+Z^2} - (1-Z)}{1+Z}$

3. $\mathcal{F}(Z)$ has a zero at $Z = -1$, so form

$$F_1(Z) = \frac{\mathcal{F}(Z)}{1+Z} = \frac{\sqrt{2}\sqrt{1+Z^2} - (1-Z)}{(1+Z)^2}$$

In series $F_1(Z) = a+bZ+\ldots = 0.414+ 0.172Z$

4. As before set $(0,1) = \frac{d}{Z+c}$. But $d = ac$ because

$$F_1(0) = a, \text{ and so } (0,1) = \frac{ac}{Z + c} = \frac{N}{D} = G_1(Z)$$

and $p + q + 1 = 2$

Now c must be adjusted to satisfy relation (2).

Substituting the series yields $c = \frac{-a}{b} = -2.41$,

and so $G_1(Z) = \frac{-2.41a}{Z - 2.41}$

5. Then $\mathcal{G}(Z) = (1 + Z)^1 G_1(Z) = \frac{ac(1 + Z)}{Z + c}$

And so $G(S) = K \mathcal{G}\left(\frac{1 - S}{1 + S}\right) = \frac{2Kac}{(c-1)S + (c+1)} =$

$$\frac{.585K}{S + .414} = \frac{0.414}{S + 0.414}$$

Note that K has been set at 0.707 so that

$$G(0) = F(0) = 1$$

Thus for the function $\sqrt{1 + S^2}$ - S, the **indirect**
Padé approximant $(0,1)$ is $0.414/S + 0.414$. This
result may be compared with the **direct** approximant
$\frac{1}{S + 1}$ found earlier.

The results of this part of the discussion are
summarized in the curves. Figure 1 shows the
direct and indirect $(0,1)$ approximants to $F(S)$
in the S domain, while Fig.2 shows the correspond-
ing fits to $f(t)$ in the time domain. It is
evident that the indirect approximant gives the
better fit to $f(t)$, as it should. With $S = j\omega$,
the amplitude and phase curves for $F(j\omega)$ are as
given in Fig. 3.

These first approximants of low degree should not
be expected to yield close fits. However, the
fit improves considerably for successive approxi-
mants. For example, the second indirect Padé
approximant is

$$(1,2) = \frac{0.512 (S + 1)}{S^2 + 0.852S + 0.512}$$

for which the corresponding time function is

$$g(t) = 0.723e^{-0.426t} \sin(0.575t + 45°)$$

Figures 4 and 5 show how much the fit is improved
in both the S domain and the time domain by the
second successive approximant.

Mathers[4] has shown that the error in the time
domain for the third successive approximant $(2,3)$
has a maximum value of 1.1% at t = 7. For higher
order approximants the error was found to be too
small to determine with a slide rule.

Bibliography

1. Padé, H.; "Sur la representation approche'e
   d'une function par des fractions rationnelles"
   Ann. de l'Ecole Normale (3) Vol. 9,1892,pp.1-93

2. Perron, O.; "Die Lehre von den Kettenbruchen"
   2d edition Leipzig and Berlin 1929

3. Wall, H.S.; "General Theorems on the Conver-
   gence of Sequences of Pade Approximants" Trans.
   Amer. Math. Society, Vol.34, 1932, pp.409-416

4. Mathers, G.W.C.; "Synthesis of Lumped-Element

Circuits for Optimum Transient Response", ERL,
Stanford University, Stanford, California;
November 1, 1951

5. Forbenius, G.; "Uber Relation zwischen den
Naherungsbruchen von Potenzreihen," Journal
fur Mathematik, Vol. 90, 1881, pp. 1-17

6. Wall, H.S.; "On The Padé Approximants Associ-
ated with a Positive Definite Power Series",
Trans. Amer. Math. Society, Vol. 33, 1931,
pp. 511-532

7. Bellman, R. and Straus, G.; "Continued
Fractions, Algebraic Functions, and the Padé
Table", Proc. Nat. Acad. Science, U.S.A.,
Vol. 35, pp. 472-476, August 1949

8. Wiener, N.; "Extrapolation, Interpolation and
Smoothing of Stationary Time Series", Wiley,
New York, New York, 1949

9. Wall, H. S.; "Continued Fractions", Van Nos-
trand, New York, New York, 1948, Chapter 20

Appendix I

In the body of the paper one method of find-
ing a direct Padé approximant was given. It is
possible to formalize the process by the use of
determinants.

We wish to approximate a function
$F(x) = a_0 + a_1 x + a_2 x^2 + \ldots$, which is analytic
about the origin, by a ratio of rational poly-
nomials N and D

$$N = a_0 u_0 + t_1 x + \ldots + t_p x^p$$
$$D = u_0 + u_1 x + u_2 x^2 + \ldots + u_q x^q \qquad (7)$$

Here N and D are two arbitrary polynomials
whose $p + q + 1$ coefficients are to be deter-
mined so that the Taylor series of the function
$FD - N$ begins with x^{p+q+1} or a higher power of
x i.e. so that relation (2) is satisfied.

After substitution of (7) into relation (2),
it is possible[5] to show that both N and D can
be written in determinant form as

$$N = \begin{vmatrix} a_{p-q+1} & \cdots & a_p & \sum_{i=0}^{p-q} a_i x^{i+q} \\ a_{p-q+2} & & a_{p+1} & \sum_{i=0}^{p-q+1} a_i x^{i+q-1} \\ \vdots & & & \\ a_{p+1} & \cdots & a_{p+q} & \sum_{i=0}^{p} a_i x^i \end{vmatrix} \qquad (8)$$

and

$$D = \begin{vmatrix} a_{p-q+1} & a_{p-q+2} & \cdots & a_p & x^q \\ a_{p-q+2} & a_{p-q+3} & & a_{p+1} & x^{q-1} \\ \vdots & & & & \\ a_{p+1} & a_{p+2} & & a_{p+q} & 1 \end{vmatrix} \qquad (9)$$

As an example of this determinant method,
find the direct Padé approximant $(0,1)$ for the
function $\sqrt{1 + S^2} - S$

From (5) $a_0 = 1$, $a_1 = -1$, $a_2 = \frac{1}{2}$, $a_3 = 0$ (10)

From (8) $N = \begin{vmatrix} a_0 & 0 \\ a_1 & a_0 \end{vmatrix} = a_0^2 = 1$

From (9) $D = \begin{vmatrix} a_0 & S \\ a_1 & 1 \end{vmatrix} = -a_1 S + a_0 = S + 1$

So $(0,1) = \dfrac{N}{D} = \dfrac{1}{S + 1}$

As a further example, repeat the above process to
find the next direct Padé approximant $(1,2)$

From (8) and (10)

$$N = \begin{vmatrix} a_0 & a_1 & 0 \\ a_1 & a_2 & a_0 S \\ a_2 & a_3 & a_0 + a_1 S \end{vmatrix} = -\frac{1}{2}$$

From (9) and (10)

$$D = \begin{vmatrix} a_0 & a_1 & S^2 \\ a_1 & a_2 & S \\ a_2 & a_3 & 1 \end{vmatrix} = -\frac{1}{4}(S^2 + 2S + 2)$$

So $(1,2) = \dfrac{N}{D} = \dfrac{2}{S^2 + 2S + 2}$

Appendix II

It is sometimes possible to find Padé
approximants by using information concerning the
zeroes and poles of an algebraic function i.e.
by using the fact that an algebraic function is
completely determined, up to a multiplying con-
stant, by the nature and location of its zeroes
and poles.

Consider finding a direct Padé approximant
of $F(S) = \sqrt{1 + S^2} - S$ with the change of variable
$S = \sinh u$, this function becomes
$F(S) = \cosh u - \sinh u$.

Then substitution into (2) gives a term

$(\sinh u)^{2n}$ on the right hand side. Then one can exploit the fact that all the real roots of $\sinh u$ must lie at the origin. Further details are given by Mathers[4].
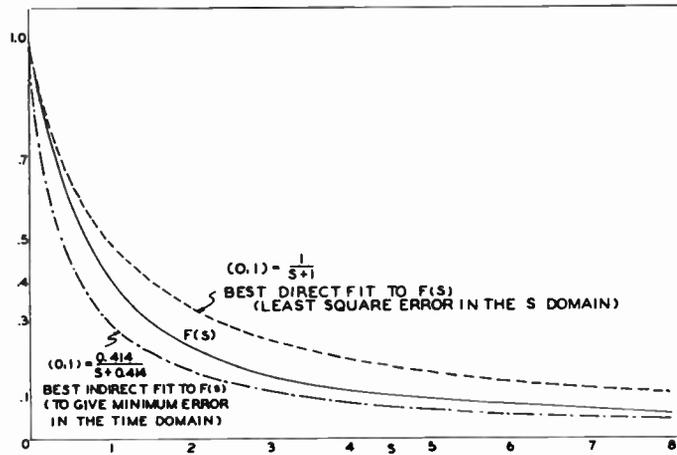
Fig. 1
Plot of $F(S) = \mathcal{L}\left[\dfrac{J_1(t)}{t}\right]$ and Padé approximants $(0,1) = \dfrac{\varepsilon}{S+a}$.
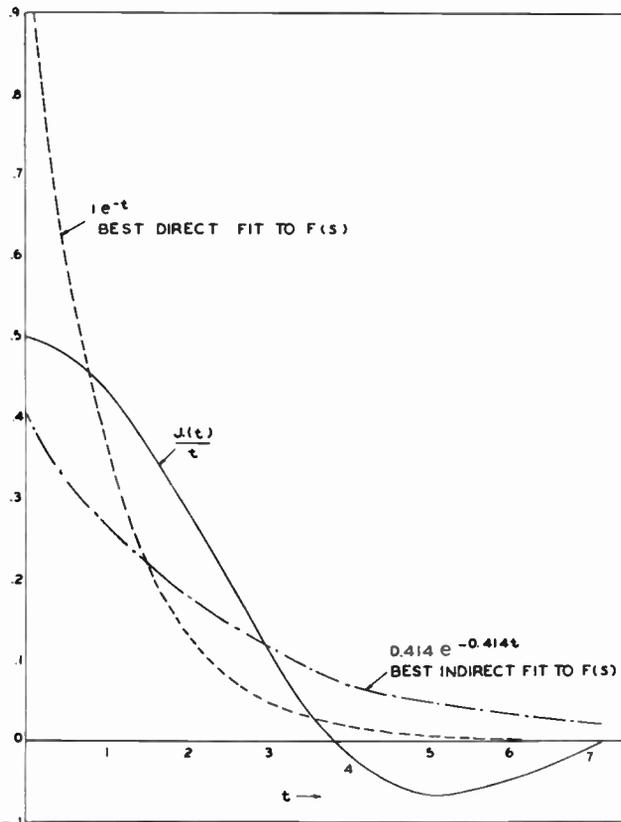
Fig. 2
Approximations in the time domain
corresponding to Fig. 1.
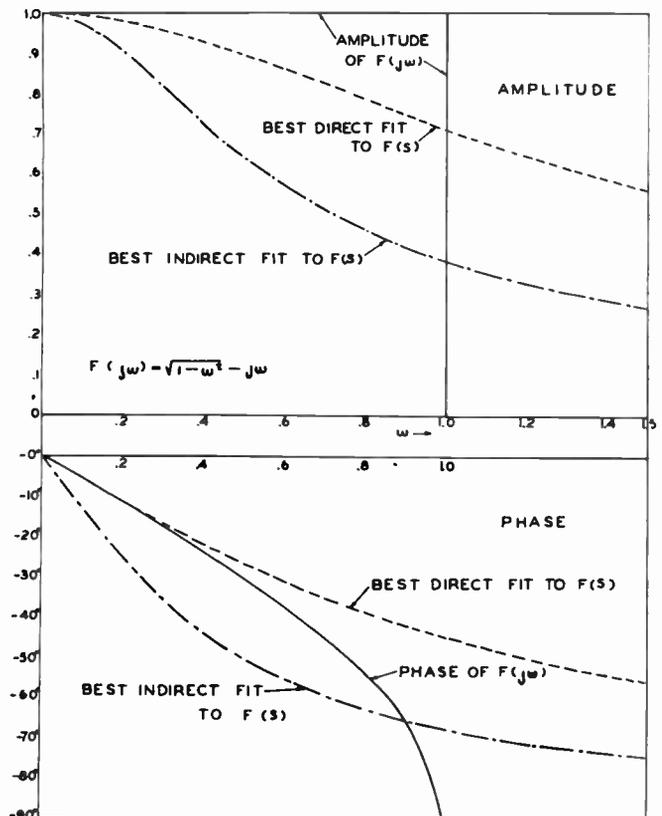
Fig. 3
Amplitude and phase characteristics
corresponding to Fig. 1.

93
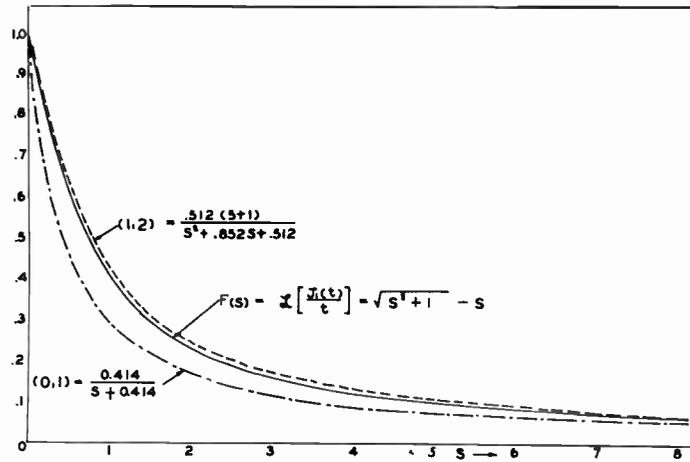
Fig. 4
Two successive indirect Padé approximants to F(S).
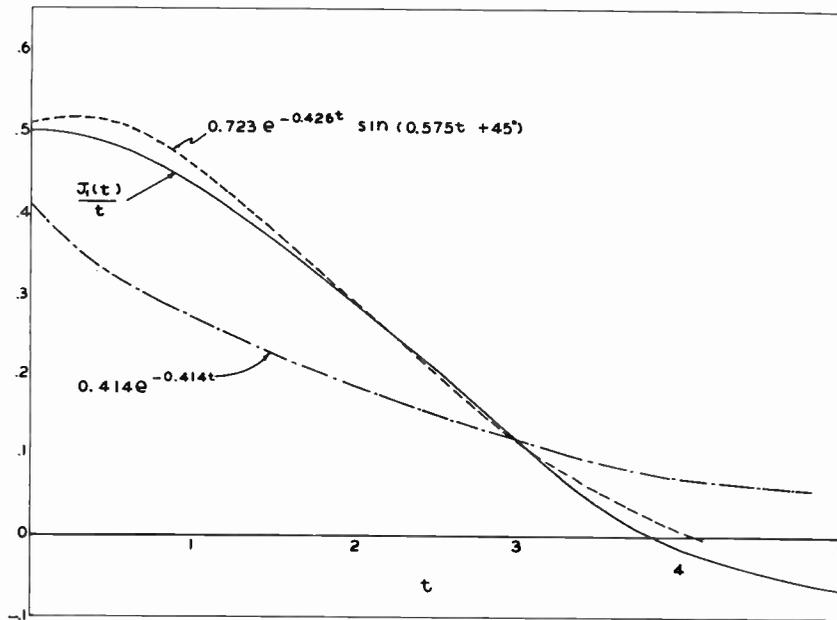
Fig. 5
Successive approximations in the time domain corresponding to Fig. 4.

FREQUENCY TRANSIENTS IN IDEALIZED LINEAR SYSTEMS

B. Gold

Electronics Department
Hughes Aircraft Company
Division Hughes Tool Company
Culver City, California

Of great value in the study of the transient behavior of systems is the information obtained about the inertia of the system. Thus, in pulsed communications, important use is made of the concepts of rise time and time delay. One might, perhaps, expect that the inertia of a given linear system would be similar regardless of the type of applied transient excitation. In reality, the information which is eventually to be extracted must be scrutinized before the truth of this statement is accepted or rejected. For example, in Slide 1, a single frequency ω_1 is passed through an i-f amplifier. If the amplitude of this sine wave is suddenly changed, the detected envelope will exhibit a transient-like behavior which would depend only on the time constant of the network. If, instead, the frequency is suddenly changed and the output of the network then passed through an appropriate frequency detecting device, the resulting transient behavior would depend in a marked way on the magnitude of the frequency change. A swing from ω_1 to ω_3 would result in a longer transient period, if we are interested in detecting instantaneous frequency.

Certainly this dependence of transient time on the magnitude of a change must be caused by a nonlinearity. In an FM detection system this nonlinearity is introduced by an amplitude limiter or any device which tends to eliminate the effects of amplitude modulation.

In 1942, a rather comprehensive study of the frequency transients problem was published[1]. Slide 2 shows a set of curves which give quantitative information concerning the inertia of a linear system to a sudden change in frequency. The linear system chosen is one which is often used in analysis (i.e., an idealized filter with a rectangular passband about some center frequency). As the ratio of the frequency shift to bandwidth increases, the percent overshoot increases and the total transient lasts longer. Of special significance is the fact that the time of initial rise to the new value is the same regardless of the magnitude of the transient. This result does not seem to be particularly reasonable physically and is one of the points of more detailed investigation in the present paper.

Another point of interest is the limiting case of the above result. When the ratio of deviation to bandwidth approaches zero, the result becomes identical with the corresponding result for the amplitude modulated case. Thus, for very small frequency shifts, the detecting system introduces no nonlinear distortion, and all the methods of linear system analysis may be used. A way of evaluating the degree of nonlinearity of a system and its subsequent effect on the desired information is by reference to the principle of superposition. If the response of a given system to a pulse is the same as the result obtained by superposing two step responses, it is quite safe to consider the system linear. As will be shown later, the region over which an FM system may be considered linear will depend on the duration of time as well as the magnitude of the disturbance.

Slide 3 shows some of the mathematics used to derive the results. In all cases when the carrier frequency is high, it is possible to write the output i (t) in the form given by Equation 1. The instantaneous frequency is then given by Equation 2. The block diagram illustrates the type of equipment needed to reproduce the instantaneous frequency.

Slide 4 shows the types of disturbances considered. In the first case, a frequency step is considered, wherein the frequency of the input carrier is suddenly changed. In the second case, the frequency is changed for a duration T and then suddenly brought back to its original value. In the general case of an arbitrary modulation, the chief difficulty lies in obtaining an analytic expression for i (t). This is a straightforward problem in Fourier Transform; however, the integrals are usually of too complicated a nature to treat. In our case the i (t) may be found for all networks for which the response to a suddenly applied sine wave is known, since the input voltage is merely a superposition of such waves. In Slide 5 the response of a physical single tuned circuit to a frequency step is shown for different values of frequency deviation. We note that the time to build up is smaller for larger deviations, although the overshoot and the length of the transient are larger. Comparing these results with those for the rectangular filter, it must be concluded that variations from constancy of amplitude response and linearity of phase response can change the response waveform quite radically. It is also interesting to note that the form of the response curves varies from overdamped to highly underdamped as the deviation increases. This is another interesting case of the effects of nonlinearity.

In Slides 6, 7, 8, and 9 is shown the instantaneous frequency when the input is a frequency pulse. The network considered is again

the rectangular filter with zero phase shift. Slide 6 illustrates the case when the disturbance lasts an appreciably shorter time than the natural lag of the network. The difference in wave shape between the actual response and that obtained by superposition is shown by comparing Curves 1 and 2. This difference became quite marked for $x > 4$ with respect to the damping. The more rapid decay of the true result and its greater smoothness may thus be attributed to nonlinear damping in the sense that the damping appears to depend on the pulse width.

In Slides 7 and 8, the width of the frequency pulse is increased. In both cases the ratio of deviation to bandwidth is one-half. It is noted that the shapes of the curves are not very different from the imaginary "linear" curves. In Slide 8, we note that the ripples on the true curve exceed those of the imaginary one. This leads to the fact that the degree of damping depends in some inverse manner on the pulse width.

Finally, Slide 9 treats the case $w = 0.091$, a relatively small deviation. As one would expect, the variation from the true to the linear result is very small.

All of the curves indicate that the resolution obtained is of the same order of magnitude for both an AM and an FM system through which a sequence of pulses is transmitted. One might conclude that the same bandwidth would be needed. On the basis of the curves shown, this conclusion seems to be acceptable provided that the deviation is not made larger than the bandwidth.

1. Salinger, H.: "Transients in FM", Proc. IRE, August, 1942.

Fig. 1

(1) $i(t) = A(t) \sin \omega_0 t + B(t) \cos \omega_0 t = C \sin(\omega_0 t + \phi)$

$C = \sqrt{A^2 + B^2}$ $\tan \phi = \dfrac{B}{A}$

(2) __Instantaneous Frequency__

$\omega = \dfrac{d\phi}{dt} = \omega_0 + \omega_i = \omega_0 + \dfrac{A\dfrac{dB}{dt} - B\dfrac{dA}{dt}}{A^2 + B^2}$

frequency transient → | linear system | → $i(t)$ → | ideal limiter | → | ideal discriminator | → ω_i

Fig. 3

RESPONSE of RECTANGULAR FILTER TO FREQUENCY STEP

a) w = 0   b) w = 1/2   c) w = 1/4

w = Deviation   D
W = Bandwidth   W

$\omega_i = \dfrac{d\phi}{dt}$ ϕ = Instantaneous Phase

Fig. 2

a) Voltage vs time plot of a frequency step.
b) Frequency vs time plot of a frequency step.

a) Voltage vs time plot of a frequency pulse.
b) Frequency vs time plot of a frequency pulse.

Fig. 4

RESPONSE OF SINGLY TUNED CIRCUIT
TO FREQUENCY STEP

a) w=0   b)w=1   c)w=2   d)w=4

x = a t

$\tilde{w} = \frac{1}{D} \frac{d\phi}{dt}$

D = Deviation

ϕ = instantaneous Phase

$w = \frac{D}{a}$

Slide 5

Fig. 5

Fig. C

Slide 6

F-M Square Pulse Transient Response

NOTE: Curve 1 is the F-M transient response of a radio frequency amplifier with rectangular amplitude characteristic. Curve 2 is the response obtained by superposition of the responses to two frequency steps of opposite polarity spaced by

$x = 0.01, \alpha = 0.1?$

$2f_c = $ Bandwidth $D = $ Deviation

$$\frac{2}{B} = \frac{1}{D} \frac{d\phi}{dt}$$

$X = 2\pi f_c T$

Fig. 7

Slide 8

F-M Square Pulse Transient Response

NOTE: Curve 1 is the F-M transient response of
a radio frequency amplifier with rectangular
amplitude characteristic.

Curve 2 is the response obtained by superposi-
tion of the responses of two frequency steps of
opposite polarity spaced by $x = 4.8$, $\omega = 0.5$

$2 f_c = $ Bandwidth $D = $ Deviation

$$\tilde{\omega} = \frac{1}{D} \cdot \frac{d\phi}{dt}$$
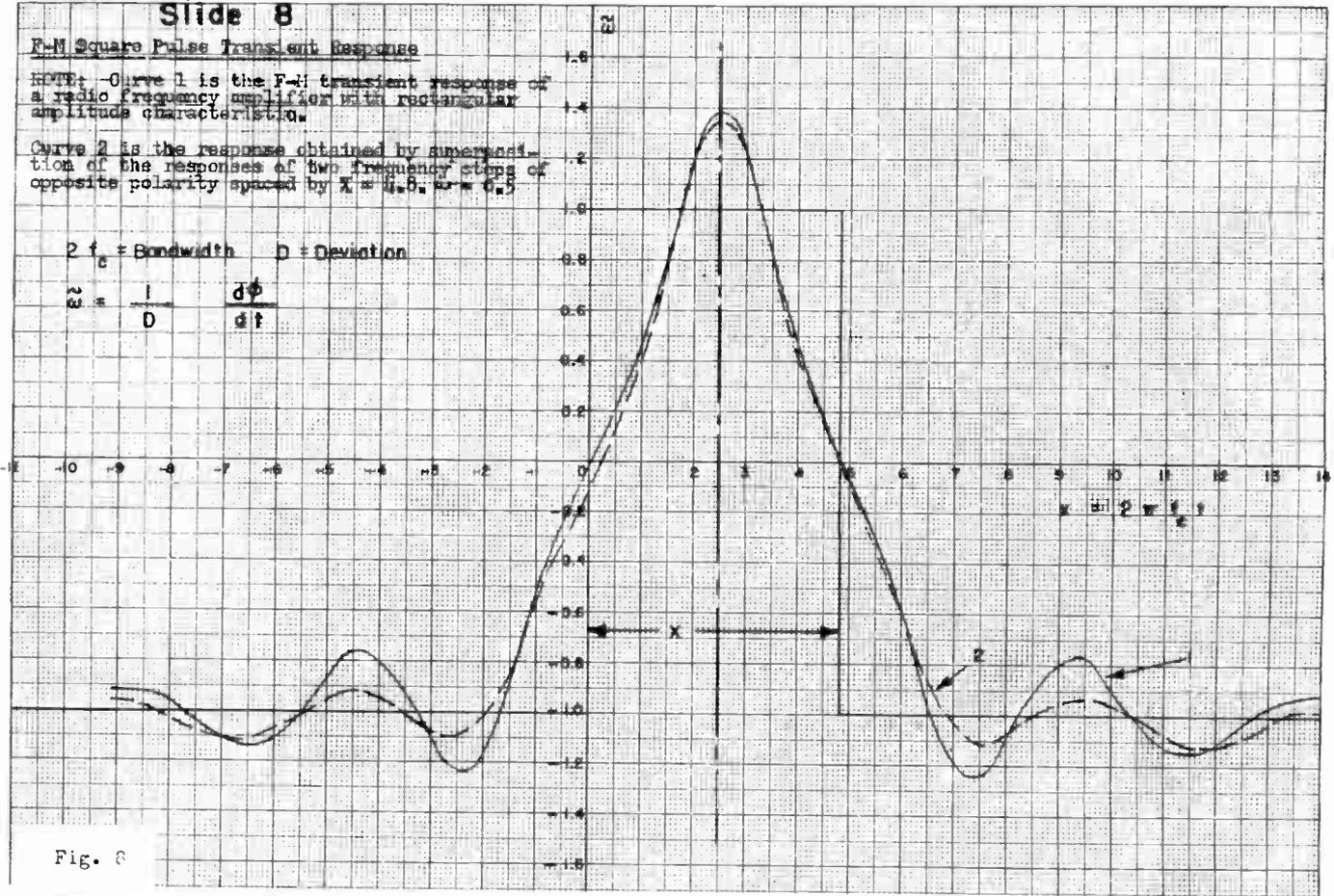
Fig. 8

Slide 9

F-M Square Pulse Transient Response

NOTE: Curve 1 is the F-M transient response of
a radio frequency amplifier with rectangular
amplitude characteristic.

Curve 2 is the response obtained by superposition
of the responses to two frequency steps of oppo-
site polarity spaced by $x = 4.8$, $\omega = 0.091$

$2 f_c = $ Bandwith $D = $ Deviation

$$\tilde{\omega} = \frac{1}{D} \cdot \frac{d\phi}{dt}$$

Fig. 9

101
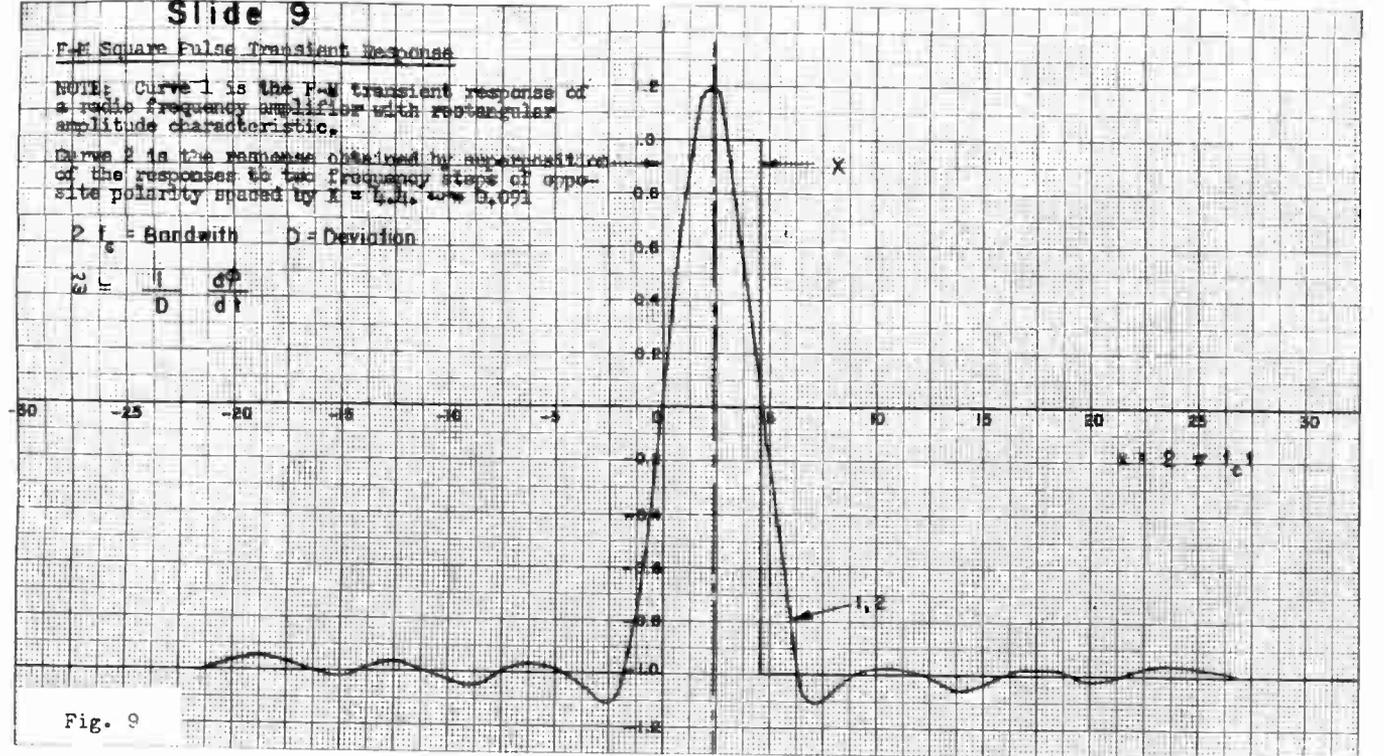
TRANSIENT ANALYSIS OF JUNCTION TRANSISTOR AMPLIFIERS

W. F. Chow
General Electric Company
Syracuse, New York

J. J. Suran
General Electric Company
Syracuse, New York

Introduction

The major sources of transient behavior in transistor amplifiers are considerably different from those in vacuum-tube amplifiers. Vacuum-tubes, operating in a frequency region below the U.H.F. band, are limited in their transfer characteristics by the existence of inter-electrode capacitances. These capacitances are functions of the electrode geometry as well as the operating point of the tube. As the operation of the tube approaches the ultra-high frequency region, transit-time effects will be encountered. Transit time considerations will materially change the frequency and transient response analysis of vacuum-tube circuits. However, transit time effects are negligible for most audio and radio frequency applications of vacuum tubes.

In a vacuum tube, electrons travel between the cathode and plate by virtue of an electric field gradient. Evacuation of air in the space between the cathode and plate minimizes obstruction of the electron path. Modulation of the electric field gradient causes a signal to appear in the plate circuit. In a transistor, however, carriers travel between the emitter and collector by virtue of a concentration, rather than electric field gradient. For n-p-n type transistor, the carriers are electrons and for p-n-p types the carriers are holes. The paths of the carriers are through the base region of the transistor, and hence the carrier paths are obstructed by the atomic structure of the semi-conductor and impurity materials which comprise the base layer. Modulation of the concentration gradient causes a signal to appear in the collector circuit.

Carriers move through the base of a transistor by a diffusion process, and transit time effects are of considerable importance even at low frequencies. In frequency and transient response analysis this fact constitutes the major difference between vacuum tubes and transistors. This paper will show how the transit-time phenomena in transistors may be taken into account in the equivalent circuit of the transistor, and how such an equivalent circuit may be simplified for practical transient response calculations. Although the analysis presented here has been restricted to junction transistors, it is applicable to point-contact types also insofar as the latter approximate the junction units in fundamental physical behavior. The circuit analysis is further restricted to small input signals, i.e., to transistors operating as linear (Class A) amplifiers.

Derivation of the Equivalent Circuit. If the effects of carrier transit in the base region of a transistor are completely neglected, the low-frequency a.c. equivalent circuit of the transistor may be represented by a T-network such as illustrated in Fig. 1a. (The equivalent π configuration may also be used). The circuit shown is for a grounded base amplifier where r_e, r_b and r_c are the a.c. resistances of the emitter, base and collector respectively. α is the current amplification factor and i_e and i_c are, respectively, the emitter and collector currents. Still neglecting the effects of carrier diffusion through the base region, the circuit of Fig.1a may be modified to include equivalent capacitances which result from the transistor geometry. Such an equivalent circuit is illustrated in Fig. 1b where C_{eb}, C_{ec}, C_{cb} are the emitter-base, emitter-collector and collector-base capacitances respectively. C_c is the collector capacity, and is due to the high-resistance barrier between the base and collector regions. In currently-available junction transistors, the effects of the interterminal capacitances C_{eb}, C_{ec} and C_{cb} are usually negligible compared to the effects of C_c and carrier diffusion. Consequently, the interterminal capacitances will be neglected in the ensuing analysis. The collector capacitance C_c, is a function of the collector voltage, but may be considered constant for a fixed operating point and for small-signal analysis [1,2,3].

In order that the equivalent circuit of Fig. 1b can be modified to include the transit time effects of carrier movement through the base, a study of the basic diffusion equation must be made. From the diffusion equation, it may be shown that the current amplification factor, α, is given by [3,4,5]:

$$\alpha(j\omega) = \frac{i_c}{i_e}(j\omega) = \operatorname{sech} \frac{W}{L_m}\sqrt{1+j\omega T_m} \quad (1)$$

where W is the width of the base region, L_m is the carrier diffusion length and T_m is the carrier lifetime. A mathematical analogue to Eq. 1 is found in the current transfer function of a distributed parameter RC transmission line. Thus, from the transmission line transfer characteristic:

$$\frac{i_o}{i_1}(j\omega) = \text{sech}\sqrt{Y(j\omega)\,Z(j\omega)} \qquad (2)$$

it is apparent that the diffusion process in the base may be represented by a transmission-line equivalent circuit such as illustrated in Fig. 2. When the transmission line analogue of the carrier diffusion process is included in the equivalent circuit for the transistor, the circuit of Fig. 3 is obtained. This is the basic equivalent circuit from which a simplified transient analysis technique will be derived.

It should be noted that Eq. 1 may be employed directly in the solution of transient response problems whenever the effects of circuit elements external to the transistor can be neglected. In particular, this applies to transistor amplifiers having small load resistances. Equation 1 has been solved for a step in emitter current, but the results are sufficiently involved to require omission from this paper[6].

The propagation of a pulse by a transmission line is characterized by two basic results: dispersion and delay. Dispersion refers to the attenuation and phase-shifting of certain frequency components in the transmitted signal by the selective character of the transmission line, and leads to a "spreading out" in time of the pulse wave front. Delay may be considered as due to the finite phase-velocity property of the transmission line and is manifested by a time lag between the input and output signals. Both dispersion and delay are intrinsic properties of the transient response of transistor amplifiers. These properties may be treated separately by a circuit approximation to the transmission line analogue as shown in Fig. 2. For example, only one section of the line may be used to represent the dispersion effect, and the remainder of the line may be replaced by an ideal delay line to represent the delay property. The ideal delay line has the transfer characteristic:

$$\frac{i_o}{i_1}(j\omega) = e^{-Kj\omega} \qquad (3)$$

where K is the delay time (seconds). Using the approximate representation, the transmission line of Fig. 2 may be illustrated as indicated in Fig. 4. The

current transfer function of this network is readily shown to be:

$$\frac{i_o}{i_e}(j\omega) = \frac{1}{1+GR} \cdot \frac{e^{-Kj\omega}}{1+\dfrac{RC}{1+GR}j\omega} \qquad (4)$$

At very low frequencies ($\omega \to 0$), Eq. 4 approaches the value $1/(1+GR)$, which will be denoted as α_o, the low-frequency current amplification factor. The frequency at which the current ratio given by Eq. 4 is 3 db below α_o, is the α-cutoff frequency, f_{α_o}. The time constant corresponding to f_{α_o} will be denoted by T_α. Equation 4 may, therefore, be written as:

$$\frac{i_o}{i_e}(j\omega) = \vec{\alpha} = \alpha_o \frac{e^{-Kj\omega}}{1+j\omega T_\alpha} \qquad (5)$$

where:

$$T_\alpha = \frac{1}{2\pi f_{\alpha_o}} \quad \text{and} \quad f_{\alpha_o} = \frac{1}{2\pi RC\,\alpha_o}$$

For most junction transistors currently available, α_o is very close to unity ($\alpha_o > 0.9$). Hence, the conductance G in Fig. 4 is usually very close to zero (infinite resistance) and can be neglected in the equivalent circuit.

Using the delay-line approximation, the a.c. transistor equivalent circuit illustrated in Fig. 3 is modified as shown in Fig. 5a. Denoting the current ratio given by Eq. 5 as $\vec{\alpha}$, the equivalent circuit of Fig. 5a is redrawn as shown in Fig. 5b. It is quite obvious that the circuit of Fig. 5 is a great deal simpler to deal with in circuit calculations than the transmission line equivalent of Fig. 3. As will be borne out later, calculations based upon the delay-line approximate circuit are sufficiently close to experimental results to warrant the circuit's consideration in practical design problems.

Frequency and Transient Response Calculations. To illustrate the use of the equivalent circuit of Fig. 5 in transient response calculations, consider the junction-transistor amplifier of Fig. 6a. The voltage generator of internal impedance R_g, generates a step voltage which is applied to the emitter of a grounded base transistor amplifier terminated by a load resistance, R_L. The voltage response, as measured across R_L will be determined from the equivalent circuit of Fig. 6b. In Fig. 6b, the current source $\vec{\alpha}\,i_e$ of Fig. 5b has been converted to a voltage source, $\vec{e}_c$ where:

103

$$\vec{e}_c = \vec{\alpha}\, i_e\, Z_c \qquad (6)$$

and Z_c is the parallel combination of r_c and C_c. Thus:

$$Z_c = \frac{r_c}{1 + r_c C_c\, j\omega} = \frac{r_c}{1 + T_c\, j\omega} \qquad (7)$$

The voltage transfer characteristic for the transistor amplifier circuit may be calculated from the Kirchoff equations for the network of Fig. 6b. In matrix form, these equations are:

$$\begin{bmatrix} e_i \\ e_c \end{bmatrix} = \begin{bmatrix} (R_g + r_e + r_b) & (-r_b) \\ (-r_b) & (Z_c + R_L + r_b) \end{bmatrix} \begin{bmatrix} i_e \\ i_c \end{bmatrix} \qquad (8)$$

When Eq. 8 is solved for i_c, it is found that the delay term $e^{-Kj\omega}$ appears in both the numerator and the denominator of the resulting expression. To simplify the transient calculation, it is desirable that the denominator of the transfer function be a polynomial in ($j\omega$) and contain no $e^{-Kj\omega}$ term. This can be achieved if the delay term is expanded as follows:

$$e^{-Kj\omega} = 1 - Kj\omega + \frac{(Kj\omega)^2}{2!} - \frac{(Kj\omega)^3}{3!} + \ldots \qquad (9)$$

Since K is usually very small (of the order of $0.1\mu s$ in junction transistors), all but the first three terms in Eq. 9 will be neglected. If this is done, the voltage transfer function of the approximate equivalent circuit illustrated in Fig. 6b is given by:

$$\frac{e_o}{e_i}(j\omega) = \frac{a_0 + a_1 j\omega + a_2(j\omega)^2 + C_o e^{-Kj\omega}}{b_0 + b_1\, j\omega + b_2(j\omega)^2} \qquad (10)$$

The coefficients a_1, b_1, C_o of Eq. 10 are in terms of the circuit and transistor constants:

$$a_0 = R_L\, r_b \qquad (11a)$$

$$a_1 = R_L\, r_b\, (T_\alpha + T_c) \qquad (11b)$$

$$a_2 = R_L\, r_b\, T_\alpha\, T_c \qquad (11c)$$

$$C_o = \alpha_0\, R_L\, r_c \qquad (11d)$$

$$b_0 = (R_g + r_e)(R_L + r_c + r_b) + r_b(R_L + r_c[1 - \alpha_0]) \qquad (11e)$$

$$b_1 = (T_\alpha + T_c)(r_b[R_g + r_e + R_L] + R_L[R_g + r_e]) + r_c\, T_\alpha\,(r_b + r_e + R_g) + \alpha_0\, K\, r_c\, r_b \qquad (11f)$$

$$b_2 = T_\alpha T_c\left[(R_g + r_e)(r_b + R_L) + R_L\, r_b\right] - \frac{1}{2}\cdot \alpha_0\, r_c\, r_b\, K^2 \qquad (11g)$$

An experimental transistor amplifier of the type illustrated in Fig. 6, was built to test Eq. 10. As measured within an estimated accuracy margin of $\pm 20\%$, the experimental circuit constants are as follows: $r_e = 30\,\Omega$, $r_b = 200\,\Omega$, $r_c = 600\,K\Omega$, $\alpha_0 = 0.9$, $T_\alpha = 10^{-6}$ sec., $T_c = 10^{-5}$ sec., $K = .5 \times 10^{-6}$ sec., $R_g = 50\,\Omega$, $R_L = 10\,K\Omega$. The voltage transfer function derived experimentally is compared to that derived theoretically from Eq. 10 in Fig. 7. Since the bandwidth of the experimental curve is somewhat greater than the theoretical bandwidth, it may be expected that the theoretical transient response calculation will lead to a conservative estimate of the physical response.

Transient response analysis proceeds directly from Eq. 10. Assuming that the input voltage, e_i, is a step function given by

$$e_i = \frac{1}{S}$$

where S is the Laplace operator (d/dt), Eq. 10 becomes, in terms of S:

$$e_o(s) = \frac{a_0 + a_1 S + a_2 S^2}{S\,(b_0 + b_1 S + b_2 S^2)} + \frac{C_o\, e^{-KS}}{S\,(b_0 + b_1 S + b_2 S^2)} \qquad (12)$$

The inverse transformation of the first term in Eq. 12 leads to a time function which is the response of the equivalent circuit to the step input when the transistor is considered as a passive circuit, i.e. when $\vec{e}_c$ is short-circuited. For good amplifiers, this term will be negligible. Inverse transformation of the second term in Eq. 12 results in:

$$e_o(t) = \frac{C_o}{b_2 A\Gamma}\left[1 + \frac{\Gamma e^{-A(t-K)} - A e^{-\Gamma(t-K)}}{A - \Gamma}\right]\mu(t-K) \qquad (13a)$$

where:

$$A,\ \Gamma = \frac{b_1}{2 b_2} \mp \sqrt{\left(\frac{b_1}{2 b_2}\right)^2 - \frac{b_0}{b_2}} \qquad (13b)$$

It is generally found that $\Gamma \gg A$; for this condition, Eq. 13a becomes:

$$e_o(t) = \frac{C_o}{b_2 A\Gamma}\left[1 - e^{-A(t-K)}\right]\left[\mu(t-K)\right] \qquad (14)$$

In Eqs. 13a and 14, $\mu(t-K)$ is a step function which is delayed in time, from the reference $t = 0$, by K seconds. Thus, the delay property of the transistor appears as a factor of the dispersion

104

function, the latter being a simple exponential in Eq. 14. Using Eq. 14, the theoretical transient response is compared to that obtained experimentally in Fig. 8. The same circuit and circuit constants are used here as were employed in the frequency-response calculations relating to Eq. 10. From Eq. 14, it is apparent that the total rise time, T_R, is:

$$T_R = \frac{1}{A} + K \qquad (15)$$

Defining the response rise time, t_R, as simply $1/A$, Fig. 9 illustrates the variation of t_R with collector resistance, collector capacitance, cut-off frequency and load resistance as calculated from Eq. 14. It should be cautioned that the curves illustrated in Fig. 9 are approximate relationships calculated for a specific case, and should therefore be referred to as indicative of trends rather than as references for universal calculations.

Conclusion. The equivalent circuit for a junction transistor may be modified by the inclusion of a transmission line analogue for the representation of diffusion phenomena within the base region. Such an equivalent circuit may then be used for transient response calculations. However, the resultant analysis would be so algebraically involved as to make the technique impractical for most design purposes. If the transmission line is replaced by a simple lumped-parameter R-C network in series with an ideal delay line, an equivalent circuit will be obtained which is much more amenable to simple design calculations than the transmission-line circuit. Calculations based upon the delay-line approximate circuit lead to conservative frequency-and-transient-response calculations. However, these

are close enough to experimental results to justify the design utility of the delay-line equivalent circuit.

Acknowledgement

References

1) R. L. Wallace and W. J. Pietenpol, "Some Circuit Properties and Applications of n-p-n Transistors", Proc. of I.R.E., Vol. 39, p 753, 1951.

2) J. S. Saby, "Fused Impurity p-n-p Junction Transistors", Proc. of I.R.E., Vo. 40, p. 1359, 1952.

3) R. F. Shea et al, "Principles of Transistor Circuits", John Wiley & Sons, Inc., N.Y., to be published, 1953.

4) W. Shockley, "Electrons and Holes in Semiconductors", D. Van Nostrand Co., Inc., N.Y., 1950.

5) W. Shockley, M. Sparks, G. K. Teal, "P-n Junction Transistors", Phys. Rev. 83, Section IX, p 161, 1951.

6) J. S. Schaffner and J. J. Suran, "Transient Response of the Grounded Base Transistor Amplifier With Small Load Impedance", to be published.
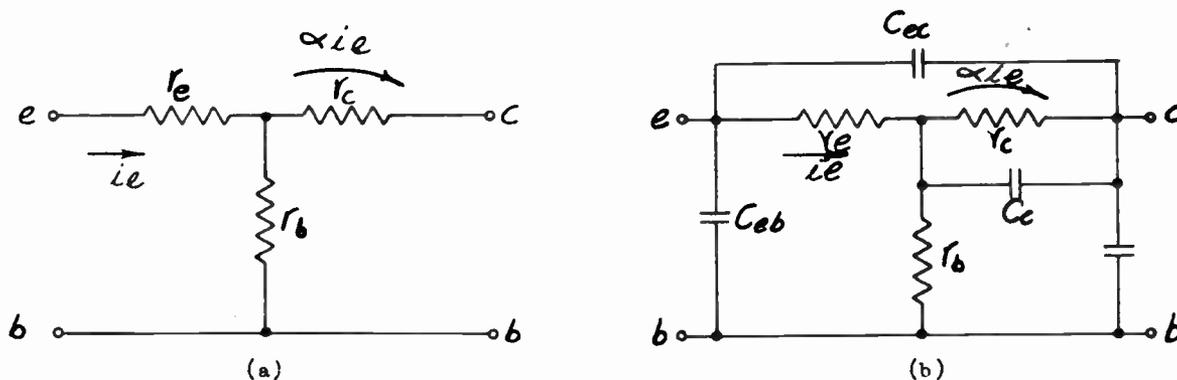
Fig. 1
AC equivalent circuit of a transistor neglecting diffusion
effects of the carriers.
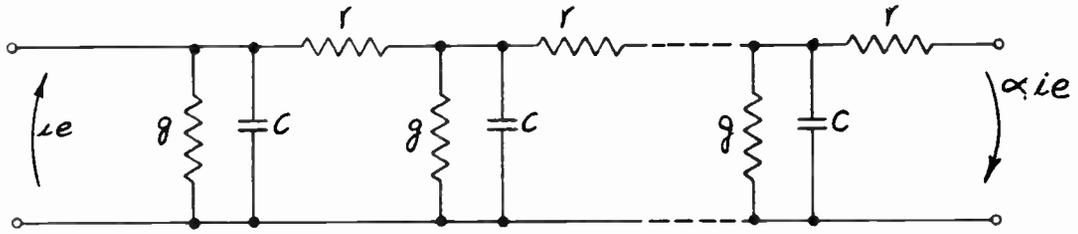
Fig. 2 – Transmission line analogue of equation 1.
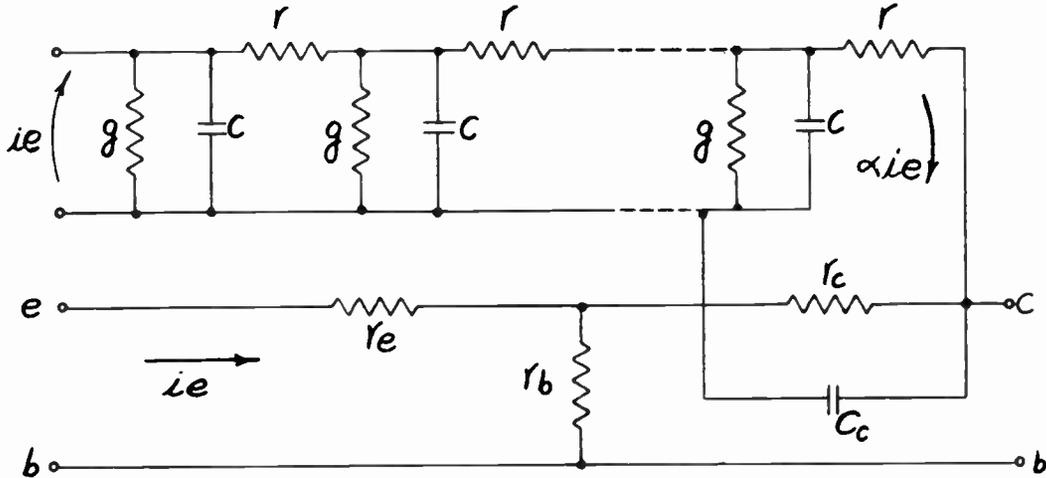
Fig. 3 – Basic ac transistor equivalent circuit.
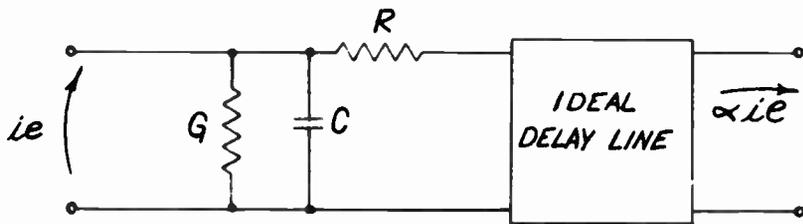
Fig. 4 – Delay-line approximation of the transmission line analogue shown in Fig. 2.

Fig. 5 – AC delay-line approximate equivalent circuit of the junction transistor.
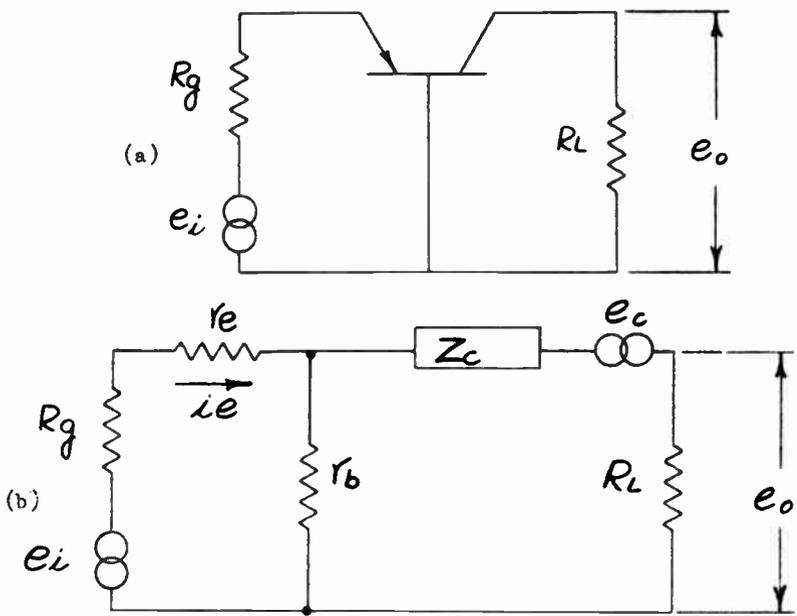
Fig. 6 - AC circuit for a grounded base transistor amplifier.
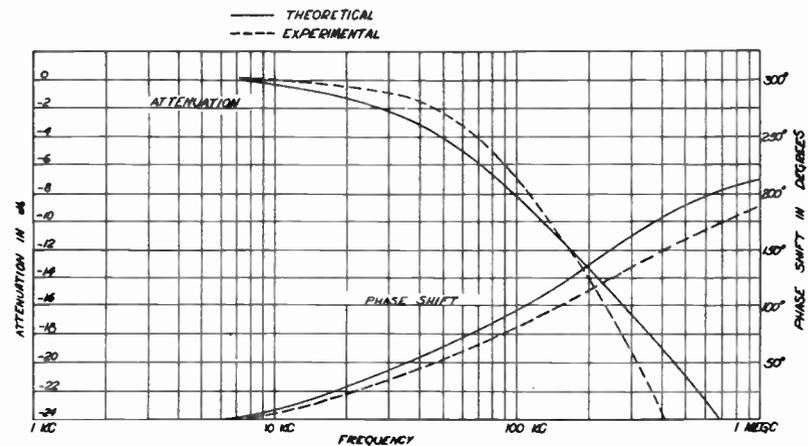
Fig. 7

Experimental and theoretical phase attenuation characteristics.
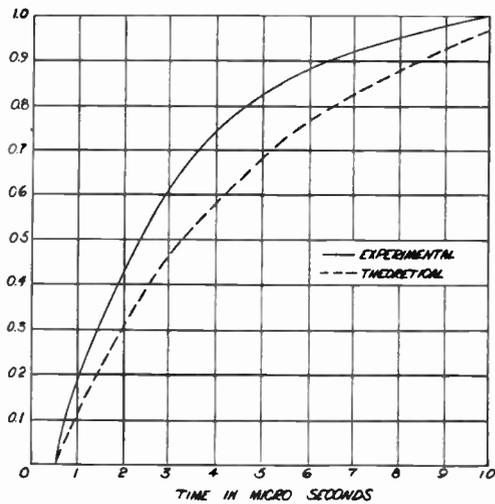
Fig. 8 - Theoretical and experimental transient response.
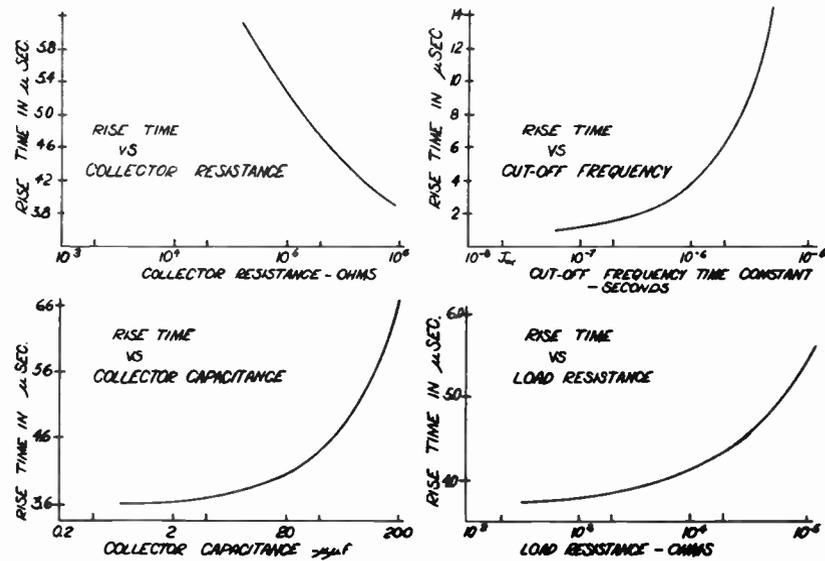
Fig. 9 - Variation of rise time with several transistor parameters.

THE GROUNDED-COLLECTOR TRANSISTOR AMPLIFIER
AT CARRIER FREQUENCIES

F. R. Stansel
Bell Telephone Laboratories, Inc.,
Murray Hill, N. J.

Summary

Expressions are derived for input resistance, output resistance and ratio of input to output voltage and current at low frequencies for transmission in both the base-to-emitter and the emitter-to-base directions. These expressions are extended to the carrier frequency range (up to approximately twice the alpha cutoff frequency) by considering the effect of the variation of alpha with frequency, of collector capacitance and of load capacitance. Experimental evidence is presented which verifies the equations obtained and indicates that the method of computing the effect of frequency may be applied to other transistor circuits.

Introduction

Of the three fundamental transistor circuits, grounded-base, grounded-emitter, and grounded-collector, only the grounded-collector circuit may have a high input impedance. This high input impedance, together with its approximately unity voltage amplification makes the grounded-collector transistor amplifier similar in both operation and use to the vacuum-tube cathode-follower circuit. Unlike the cathode-follower, the grounded-collector circuit is a bidirectional device, that is, it is capable of transmitting signals in either of two directions. There are also questions of stability in the use of grounded-collector transistor circuits which do not exist in vacuum-tube cathode-follower circuits.

The grounded-collector circuit has been discussed in a general way by previous writers.[1,2] It is the purpose of this article to extend these treatments and particularly to discuss the effects resulting from operating the grounded-collector amplifier at frequencies above the audio-frequency range. In addition to obtaining relations which may be useful for design purposes, the experimental data presented verifies a method of computing the effect of frequency variation which may be applied to other transistor circuits.

In this article the low frequency equations for the input and output resistance, and for the ratio of input to output voltage and current will be derived and their physical significance discussed. Following this there will be a discussion of the effect of frequency on transistor circuits and the basic principles cited will be applied to the low frequency equations previously derived. In the concluding portion of the article some experimental observations which verify the derived results will be given.

As previously mentioned, the grounded-collector is bidirectional, that is, the input may be either applied between the base and ground and the output taken from emitter to ground as shown in Figure 1, or the input may be applied between the emitter and ground and the output taken from base to ground as shown in Figure 2. The convention used in this article is to refer to the first connection (Figure 1) as transmission in the XY direction and the second connection (Figure 2) as transmission in the YX direction.

Input Resistance - XY Transmission (Output - Resistance - YX Transmission)

The quantity R_{xy} as defined in Figure 1 is both the input resistance of a grounded-collector amplifier when transmitting in the XY direction and the output resistance when transmitting in the YX direction. Wallace and Pietenpol[2] have given the value of this quantity. Using the notation of Figure 1.

$$R_{xy} = r_b + \frac{r_c}{1 + (r_c - r_m)/(R_y + r_e)} \qquad (1)$$

It is frequently desirable to express this relation in terms of the current amplification factor α which is defined as $(\partial i_c/\partial i_e)_{e_c}$ and is equal to

$$\alpha = \frac{r_m + r_b}{r_c + r_b} \qquad (2)$$

By solving (2) for r_m and substituting in (1) the following is obtained

$$R_{xy} = r_b + \frac{r_c}{1 + (r_c + r_b)(1 - \alpha)/(R_y + r_e)} \qquad (3)$$

R_{xy} consists of the base resistance r_b in series with an equivalent resistance R'_{xy} whose value is

$$R'_{xy} = \frac{r_c}{1 + (r_c + r_b)(1 - \alpha)/(R_y + r_e)} \qquad (4)$$

108

Generally R'_{xy} is so large that the initial r_b term of (1) or (3) may be neglected and R_{xy} may be considered approximately equal to R'_{xy}.

Equation (4) can be further simplified by introducing the approximation $r_b \ll r_c$ valid for most transistors of both point contact and junction types. Equation (4) then becomes

$$R'_{xy} \cong \frac{r_c}{1 + r_c(1 - a)/(R_y + r_e)} \qquad (5)$$

which may be rearranged in the form

$$\frac{1}{R'_{xy}} \cong \frac{1}{r_c} + \frac{1 - a}{R_y + r_e} \qquad (6)$$

or using conductance notation

$$G'_{xy} \cong g_c + (1 - a) G_{ye} \qquad (7)$$

in which $G_{ye} = 1/(R_y + r_e)$. These equations shows that the input resistance for XY transmission, neglecting the series resistance r_b, can be considered as composed of two resistors connected in parallel. One resistor is the collector resistance of the transistor r_c and the second is the resistance $(R_y + r_e)$ divided by $(1 - a)$. Since r_e is small, generally of order of 25 ohms, it frequently is negligible compared to R_y.

Consider the case of junction transistors. For this type r_c is quite large, often one or more megohms, so that the effect of this resistance is frequently of second order importance. a is generally only slightly less than unity so that $(1 - a)$ is a small quantity and $(R_y + r_e)/(1 - a)$ is large increasing in magnitude the nearer a is to unity. Therefore the input resistance of a grounded-collector amplifier using junction transistors is high and is determined largely by the load resistance and a, but can not exceed r_c.

Figure 3 shows the value of R'_{xy}/r_c plotted as a function of $(R_y + r_e)/r_c$ for values of a found in both junction and point-contact transistors. Junction type transistors have values of a less than unity and the curves applicable are the solid lines in the center portion of the figure. Point-contact transistors, on the other hand, generally have values of a greater than unity and the curves applicable to these transistors are the dotted lines in the upper and lower portion of Figure 3.

For a less than unity (junction transistors), R'_{xy} increases as a and $(R_y + r_e)/r_c$ are increased. As the load resistance R_y is increased the curves for various values of a converge and a has increasingly less influence on the value of R'_{xy}.*

For a greater than unity (point-contact transistors), the input resistance is more complicated and may under some conditions be negative. For these types of transistors collector resistance r_c is generally lower and therefore this term has more effect than in a junction transistor. The quantity $1 - a$ is negative so that R'_{xy} consists of a positive resistance r_c and a negative resistance $-(R_y + r_e)/(a - 1)$ in parallel.

For very small values of R_y, less than the value given by the following expression,

$$R_y = r_b(a - 1) - r_e \qquad (8)$$

R_{xy} is positive. As R_y is increased, R_{xy} decreases becoming zero at the value given in (8). For values greater than (8), R_{xy} is negative and increases in magnitude as R_y is increased until R_y is increased to the value given by

$$R_y = (r_c + r_b)(a - 1) - r_e \qquad (9)$$

For R_y equal to (9), R_{xy} is infinite. For greater values of R_y, R_{xy} is positive decreasing and approaching r_c asymptotically as R_y is increased. The corresponding variation of R'_{xy} is shown by the dotted curves of Figure 3.

Input Resistance - YX Transmission (Output Resistance - XY Transmission)

The quantity R_{yx} as defined in Figure 2 is both the input resistance of a grounded-collector amplifier when transmitting in the YX direction and the output resistance when transmitting in the XY direction. In the same manner as (3) its value is found to be

$$R_{yx} = r_e + \frac{(1 - a)(r_c + r_b)}{1 + r_c/(R_x + r_b)} \qquad (10)$$

As in the case at R_{xy}, R_{yx} consists of the emitter resistance r_e in series with an equivalent resistance R'_{yx} whose value is

$$R'_{yx} = \frac{(1 - a)(r_c + r_b)}{1 + r_c/(R_x + r_b)} \qquad (11)$$

In contrast with equation (5), in (11) the term $1 - a$ is in the numerator rather than in the denominator so that R'_{yx} has a low value. It is therefore not correct to consider R'_{yx} as an approximation for R_{yx} in the same manner that R'_{xy} was considered an approximation for R_{xy}.

Introducing the approximation $r_b \ll r_c$ (11) may be rearranged in the form

$$\frac{1}{R'_{yx}} \cong \frac{1}{(1-\alpha)} \left[\frac{1}{r_c} + \frac{1}{R_x + r_b} \right] \qquad (12)$$

or using conductance notation

$$G'_{yx} \cong \frac{g_c + G_{xb}}{1 - \alpha} \qquad (13)$$

in which $G_{xb} = 1/(R_x + r_b)$

Equations (12) and (13) indicate that R'_{yx} can be considered as the two resistors $(1 - \alpha) r_c$ and $(1 - \alpha)(R_x + r_b)$ connected in parallel.

For a junction transistor the collector resistance r_c is generally high enough so that the first branch has little effect on the value of R'_{yx} and the following approximation is usually valid for the total input resistance

$$R_{yx} \cong r_e + (1 - \alpha)(R_x + r_b) \qquad (14)$$

As $1 - \alpha$ is a small fraction and r_e has a value of 10 to 50 ohms, R_{yx} will in general be quite small.

The variation of the R'_{yx}/r_c as a function of $(R_x + r_b)/r_c$ is shown in Figure 4. As in the previous case, the curves applicable to junction transistors are shown in solid lines and the curves applicable to point-contact transistors are in dotted lines. Note that the curves applicable to junction transistors are plotted to a different scale of R'_{yx}/r_c than those applicable to point-contact transistors.

For α less than unity (junction transistors), R'_{yx} is always positive and increases as $(R_x + r_b)$ is increased with a limiting value of $(1 - \alpha) r_c$. As transistors having increasingly higher values of α are used, R'_{yx} is decreased. The limit is reached when α is unity. For this case R'_{yx} is zero for all values of $(R_x + r_b)$.

For α greater than unity (point-contact transistors), R'_{yx} is always negative. The total input resistance R_{yx} is positive for values of R_x less than

$$R_x = \frac{r_e - r_b (\alpha - 1)}{(\alpha - 1) - r_e/(r_c + r_b)} \qquad (15)$$

For R_x equal to the value given by (15), R_{yx} is zero. For larger values of R_x, R_{yx} is negative increasing in absolute magnitude as R_x increases and approaching the limiting value $-(\alpha - 1) r_c$. The corresponding variation of R'_{yx} is shown by the dotted lines in Figure 4.

Ratio of Input to Output Voltage and Current

These ratios may be found by solving the mesh equations for a grounded-collector transistor [1]. For xy transmission these ratios are

$$\left[\frac{e_x}{e_y} \right]_{xy} = \frac{r_c + r_b}{r_c} \left\{ 1 + \frac{r_b(1-\alpha) + r_e}{R_y} \right\} \qquad (16)$$

$$\left[\frac{i_x}{i_y} \right]_{xy} = (1-\alpha)(1 + r_b/r_c) + (R_y + r_e)/r_c \qquad (17)$$

For most transistors $r_b \ll r_c$ and $r_e \ll r_c$ so that the term outside the bracket in (16) and the $(1 + r_b/r_c)$ term in (17) are essentially unity. For large values of R_y the voltage ratio approaches unity while the current ratio is approximately

$$\left[\frac{i_x}{i_y} \right]_{xy} \cong (1-\alpha) + R_y/r_c \qquad (18)$$

For α less than unity (junction transistors), the output voltage and current are both in phase with the input voltage and current for all values of R_y. For α greater than unity (point-contact transistors) the output voltage and current are in phase with the input voltage and current if R_y has a value greater than (9). If R_y is greater than (8) and less than (9) the input and output voltages are in phase and the input and output currents are 180° out of phase. If R_y has a value between zero and (8) the output voltage and current are both 180° out of phase with the input voltage and current.

For YX transmission

$$\left[\frac{e_y}{e_x} \right]_{yx} = 1 + \frac{r_b}{R_x} + \frac{r_e}{(r_c + r_b)(1-\alpha)} \left[1 + \frac{r_c + r_b}{R_x} \right] \qquad (19)$$

$$\left[\frac{i_y}{i_x} \right]_{yx} = \frac{1}{1-\alpha} \left[1 + \frac{R_x}{r_c + r_b} \right] \qquad (20)$$

if $r_b \ll r_c$ and $R_x \ll r_c$, (19) is approximately

$$\left[\frac{e_y}{e_x} \right]_{yx} \cong 1 + \frac{r_b}{R_x} + \frac{r_e}{R_x(1-\alpha)} \qquad (21)$$

As in the case of XY transmission the input and output voltages are approximately equal.

For α less than unity (junction transistors) the output voltage and current are both in phase with the input voltage and current for all values of R_x. For α greater than unity (point-contact transistors) the output voltage and current are both 180° out of phase with the input voltage and current when R_x is less than the value given by (15). For values of R_x greater than (15) the

110

input and output voltages are in phase while the input and output currents are 180° out of phase.

Effect of Frequency

All of the preceding equations have assumed that the frequency is low. As the frequency is increased, four effects modify these equations. They are:

(1) Variation of α both in magnitude and in phase with frequency.

(2) Effect of capacitance shunted across r_c.

(3) Effect of capacitance shunted across the load and generator resistances, that is, across R_x and R_y.

(4) Effect of variation of r_e and r_b with frequency.

Ryder and Kircher [1] have discussed the frequency characteristic of α for point-contact transistors. Their observations were that the phase shift of α is "related to the amplitude in the same way as if the characteristic were that of a 'minimum phase' passive circuit." Based on these observations, D. E. Thomas [3] has suggested that α can be represented by the equation

$$\alpha = \frac{\alpha_0}{1 + j\Omega} \qquad (22)$$

in which α_0 is the value of α at low frequency, Ω is the ratio of the operating frequency to the cutoff frequency, f_c, and f_c the frequency at which the magnitude of α is $1/\sqrt{2}$ that of the low frequency value.*

A more exact expression for α

$$\alpha = \alpha_0 \ \text{sech} \ (j \ 2.43 \ \Omega)^{1/2} \qquad (23)$$

has been derived by Pritchard [4] from the theoretical treatment of NP junctions by Shockley, Sparks and Teal.[5] By introducing the cosh function and expanding into an infinite series it may be shown that (22) is a first approximation for (23). In an unpublished communication with the author Pritchard has pointed out that while this approximation is quite good in magnitude, the difference being less than .1 db at 1.8 f_c, the agreement in phase is not so good. At the cut-off frequency f_c the two expressions differ by 13°. Hence on the basis of both theoretical considerations and experimental data which follows the use of the simpler expression (22) is justified at lower frequencies. At higher frequencies a more complicated expression such as (23) may be required.

- -

*This cut-off frequency can also be defined as the frequency at which α has a phase angle of 45°, that is the real and imaginary parts are equal.

The author has made a series of measurements on NPN junction and experimental PNP alloy (or "diffused junction") transistors. The circuit used, which is shown in Figure 5, consisted essentially of a grounded-base amplifier excited from an oscillator through a terminated attenuator and a high-value resistor R2. Because of this resistor the input current is proportional to the attenuator setting. The load resistor R3 in the collector circuit is small and the output current is measured by the voltmeter shunted across this resistor. To measure cutoff frequency the oscillator is set at a low frequency and the attenuator is adjusted for a suitable output voltage. The attenuator setting is then reduced 3 db and the oscillator frequency increased until a frequency is reached at which the output voltmeter reads the original value.

Measurements were made on six NPN junctions and two PNP alloy transistors. At each frequency the ratio $|\alpha/\alpha_0|$ was computed using equation (22) and compared with the measured ratio. For frequencies less than twice the cutoff frequency the computed and the observed results were well within the limits of experimental error, often less than .1 db which was the smallest step on the attenuator. Typical data for three transistors are listed in Table I. At frequencies above twice the cutoff frequency the difference between the observed values and the values computed by (22) increases, but if (23) is used instead of (22) the agreement is still good. The variation in phase was checked by remeasuring these transistors in a grounded-emitter circuit similar to Figure 5. It can be shown that assuming α varies in accordance with (22), the cutoff frequency of a grounded-emitter stage f'_c (defined in the same manner as f_c except that the transistor is operated in a grounded-emitter circuit) is given by

$$f'_c \ \doteq \ (1-\alpha) \ f_c \qquad (24)$$

In all cases the computed and measured values of f'_c were within the limits of experimental error.

For operation of transistors in circuits at higher than audio frequencies, the cutoff frequency is one important parameter. Figure 6 shows the variation of cutoff frequency of a typical NPN junction transistor as a function of emitter current and for various collector voltages. Some variation in these curves exists from unit to unit, but all units checked show an increase in cutoff frequency with collector voltage and a maximum cutoff frequency at some value of emitter current between .5 and 2 milliamperes. Cutoff frequency may also be affected by ambient temperature but this variation is not uniform from unit to unit. In some units the cutoff frequency has been observed to increase with increasing temperature and in some to decrease

TABLE I

Measured and Computed Values of α for Three Transistors

NPN Junction Type				NPN Junction Type				PNP Alloy Type			
f_c = 1800 kc				f_c = 1000 kc				f_c = 160 kc			
α = .985				α = .971				α = .64			
Loss in α (db)				Loss in α (db)				Loss in α (db)			
KC	Cal.	Meas.	Dif.	KC	Cal.	Meas.	Dif.	KC	Cal.	Meas.	Dif.
200	.05	0	-.05	250	.21	.2	-.01	50	.40	.4	0
600	.46	.5	+.04	400	.64	.6	-.04	60	.57	.6	+.03
800	.78	.9	+.12	500	.97	.9	-.07	70	.76	.6	+.04
1000	1.17	1.3	+.13	650	1.53	1.5	-.03	80	.97	1.0	+.03
1200	1.60	1.8	+.20	800	2.15	2.2	+.05	90	1.19	1.3	+.11
1500	2.29	2.4	+.11	900	2.58	2.6	+.02	100	1.43	1.5	+.07
1800	3.01	3.0	-.01	1000	3.01	3.0	-.01	120	1.96	2.0	+.04
2200	3.96	3.6	-.36	1300	4.30	4.3	0	150	2.74	2.7	-.04
2500	4.67	4.4	-.27	1500	5.12	5.2	+.08	180	3.55	3.5	-.05
3000	5.77	5.5	-.27	1800	6.27	6.4	+.13	200	4.09	4.0	-.09
4000	7.74	7.7	-.04	2000	6.99	7.1	+.11	250	5.37	5.4	+.03
4500	8.60	8.5	-.10	2500	8.60	8.9	+.30	300	6.55	6.6	+.05
5000	9.40	9.5	+.10	3000	10.00	10.5	+.50	400	8.60	8.8	+.20
								500	10.32	10.8	+.48

with increasing temperature. This change is not large, generally not exceeding a total of 20% for the temperature range of 80°F to 120°F.

J. M. Early [6] has shown that because of secondary effects the complete expression for r_e and r_b includes a correction factor which varies with frequency. The corresponding correction factors for r_c and α are negligible. In this article these correction factors have not been introduced since in most cases these terms are small enough to be negligible. For more a exact treatment the effect of the variation of r_e and r_b with frequency will have to be considered.

Input Admittance - XY Transmission

The generalized input admittance is obtained from (5) by substituting for α equation (22), for $1/r_c$ the value $g_c + j\omega c_c$ and for $1/R_y$ the value $G_y + j\omega C_y$. Assuming that $r_e \ll R_y$ and after some purely mathematical steps, the following equivalent input conductance and capacitance are obtained:

$$G'_{xy} \cong g_c + \frac{(1 - \alpha_o + \Omega^2)G_y}{1 + \Omega^2} - \frac{2\pi f_c C_y \alpha_o \Omega^2}{1 + \Omega^2} \quad (25)$$

$$C'_{xy} \cong c_c + \frac{\alpha_o G_y}{2\pi f_c (1 + \Omega^2)} + \frac{(1 - \alpha_o + \Omega^2)C_y}{1 + \Omega^2} \quad (26)$$

in which $\Omega = f/f_c$.

It was previously shown in equations (6) and (7) that, at sufficiently low frequencies, R'_{xy} is equivalent to a circuit composed of two parallel resistances. Equations (25) and (26) are generalizations of (6) and (7). This generalized equivalent circuit, which is shown in Figure 7, consists of three parallel resistances representing the three terms of (25) and three parallel capacitances representing the three terms of (26). To make the equivalent circuit complete Figure 7 shows the impedance r_b although as previously discussed this term can generally be neglected.

The table in Figure 7 lists the values of these six parallel branches (giving the resistance branches in terms of either their resistance or their conductance) for three conditions. The first line marked "all frequencies" are the generalized terms obtained from (25) and (26). The last line marked "zero frequency" are the values obtained from (6) and (7) which are identical with those in the first line with Ω set equal to zero.

A useful approximate set of values can be obtained from (25) and (26) by assuming that the Ω^2 terms in $1 + \Omega^2$ and $1 - \alpha_o + \Omega^2$ can be neglected. This approximation is valid up to the order of .05 to .1 of the cutoff frequency depending on the value of α_o and is frequently useful in design of circuits in the frequency range up to 100 to 200 kilocycles. The values of the components in Figure 7 using this approximation are given in the second line of the table in this figure. This approximation shows that at low frequencies the input conductance and capacitance of a grounded-collector amplifier are nearly independent of frequency. The only term in this approximate equivalent circuit dependent on

frequency is R_3 (or G_3). As this term is also dependent on the load capacitance C_y, the effect of frequency can be minimized by making C_y small.

For higher frequencies equations (25) and (26) are too complicated to discuss in general terms. Rather some illustrative curves will be given for a grounded-collector amplifier using a typical NPN junction transistors whose parameters are

$a_o = .98$ $f_c = 2$ megacycles

$r_c = 1$ megohm $c_c = 5$ micromicrofarads

Initially the case when the load capacitance C_y equals zero will be considered.

Figure 6 shows the resistance component of the input impedance as a function of frequency for several values of load resistance. As the load resistance R_y is increased the input resistance increases but an upper limit on the input resistance is placed by branch R_1 (see Figure 7) which is numerically equal to the value of r_c of the transistor.

Figure 9 shows the variation of input resistance with frequency for transistors having other values of a_o than $a_o = .98$ with the load resistance R_y fixed at 50,000 ohms. Again the limiting value of input resistance is r_c.

Figure 10 shows the input capacitance for the case given in figure 8. For small values of load resistance R_y the input capacitance is largely determined by the component C_2. As the load resistance is increased the value of C_2 decreases and the input capacitance approaches the collector capacitance c_c as a limiting value. The variation of input capacitance with frequency, while not large even with small values of R_y, decreases as R_y is increased so that for large values of R_y the input capacitance is essentially independent of frequency.

Next consider the modifications of the input resistance (or conductance) when a capacitance C_y is added across the load. The effect of this capacitance is given by the third term in (25). Because this term has a minus sign the effect will be to decrease the input conductance, or to increase the input resistance. If the value of C_y is increased sufficiently the negative value of the third term in (25) can be made greater than the sum of the two positive terms thus making the input conductance negative. Such a negative input conductance has been measured experimentally.

When discussing the zero frequency case for point-contact transistors it was pointed out that since a is greater than unity one of the parallel resistances in the equivalent circuit has a negative sign. This condition is also true for the general case (25) in which R_2 (or G_2) is negative when a_o is greater than unity.

Input Admittance - YX Transmission

The input admittance for YX transmission can be obtained in the same manner as for XY transmission assuming that $r_b \ll R_x$. The results are

$$G'_{yx} \approx \frac{(1-a_o+\Omega^2)(g_c+G_x) + 2\pi f_c \Omega^2 a_o(c_c+C_x)}{(1-a_o)^2 + \Omega^2} \quad (27)$$

$$C'_{yx} \approx \frac{(1-a_o+\Omega^2)(c_c+C_x)}{(1-a_o)^2 + \Omega^2} - \frac{a_o(g_c+G_x)}{2\pi f_c[(1-a_o)^2+\Omega^2]} \quad (28)$$

Equations (27) and (28) define an equivalent circuit in the same manner as in the case of XY transmission. This equivalent circuit is shown in Figure 11 and the values of resistance or conductance and of capacitance for the several branches are given in the table of this figure. As in the previous case when Ω is equal to zero (27) and (28) reduces to the zero frequency case previously discussed in (12) and (13).

The approximation previously used in the case of XY transmission can not be used in this case as Ω^2 is associated with the term $(1 - a_o)^2$ which at low frequencies is of the same order as Ω^2.

The characteristics of the input resistance and capacitance for YX transmission are in many ways inverse of those for XY transmission. This may be seen by comparing Figures 8, 9, and 10 for XY transmission with Figures 12, 13 and 14 which are similar curves for the same transistor with YX transmission.

The input resistance for XY transmission (see Figures 8 and 9) is nearly constant at low frequencies and decreases as the frequency is increased. In contrast, for YX transmission the input resistance (see Figures 12 and 13), increases quite markedly with frequency at low frequencies. At frequencies higher than about half the cutoff frequency the input resistance becomes nearly constant approaching an asymptotic value. This asymptote is essentially R_x for low values of R_x. For higher values of R_x this asymptote is reduced in value by the parallel resistance components R_1 and R_3 (see Figure 11).

The input capacitance also varies quite markedly with frequency. Because of the negative sign of C_2 and C_3, the capacitance has a negative sign over a large portion of the frequency range. At frequencies higher than about half the cutoff frequency the magnitude and variation of the input capacitance diminishes and it approaches the value $c_c + C_x$ asymptotically,

For point-contact transistors, the denominator of R_1 and R_2 (or the numerator of G_1 and G_2) in Figure 11 becomes $\Omega^2 - (a - 1)$. For low

113

frequencies where Ω^2 is less than $(\alpha - 1)$ these terms will have a negative sign as in the zero frequency case. As frequency is raised a value of Ω will be reached above which these components will have a positive sign. The component R_3 (and G_3) of Figure 11 has a positive sign for both point-contact and junction transistors. Since this component depends on the load capacitance C_x, it is possible at carrier frequencies to adjust C_x so that the positive conductance of G_3 cancels the negative conductance of G_1 and G_2.

For point-contact transistors C_2 and C_3 retain their negative sign. C_1 may have a positive or negative sign in the same manner as R_1 and R_2.

Experimental Observations

A series of measurements were made to verify the foregoing mathematical treatment. The larger portion of these measurements was made using NPN junction transistors although some measurements were also made using point-contact and PNP alloy transistors. All observations checked computed results within the limits of experimental observations.

Figure 15 shows the measured ratio of input to output voltage of a grounded-collector amplifier as a function of the load resistor. These measurements were made at 30 kilocycles using an NPN junction transistor with a constant output voltage of .005 volts. The two curves show the theoretical ratios for transmission in the XY and in the YX directions, computed using (16) and (19) respectively. Except for low values of load resistance, the measured values, as indicated on Figure 15, check very closely the computed curve. For XY transmission there is a possibility that some of the error was because of neglecting the effect of stray inductance in series with the load which at low values of load resistance may be appreciable.

Figure 16 shows the ratio of input to output voltage as a function of frequency for three transistors, a NPN junction transistor, an experimental PNP alloy transistor and a point-contact transistor. These curves were all taken transmitting in the XY direction and with a load of 5620 ohms. In all cases, the voltage ratio is very close to unity, increasing very slightly as the frequency increases. From 10 kilocycles up to the cutoff frequency the variation in voltage is in the order of a few tenths of a decibel while even out to several times the cutoff frequency there is only a small change in voltage ratio.

Figures 17 and 18 show a series of measurement of input admittance of a grounded-collector amplifier using a NPN junction transistor and a 10,000 ohm load resistance. These measurements were made using a precision admittance bridge suitable for use at frequencies up to at least 500 kilocycles.

According to (25) and (26), both the conductance and the capacitance components of the admittance are linearly related to the load capacitance. Figure 17 shows observed values of input conductance and Figure 18 observed values of input capacitance both as functions of load capacitance and at several frequencies. It will be noted that the predicted linear relation was observed.

In equation (25) it will be noted that the term containing the load capacitance, C_y, has a negative sign. Hence, as load capacitance is increased the input conductance should decrease and if the load capacitance is made large enough, it should be possible to obtain a negative input conductance. Figure 17 shows that this decrease in conductance with increasing load capacitance was observed and for frequencies of 75 kc and above negative values of input conductance were measured.

More extensive measurements were made at 300 kilocycles of the input admittance of a grounded-collector amplifier using a NPN junction transistor. At any one frequency equations (25) and (26) may be written in the form.

$$G'_{xy} = g_c + k_1 G_y - k_2 C_y \qquad (29)$$

$$C'_{xy} = c_c + k_3 G_y + k_4 C_y \qquad (30)$$

in which k_1, k_2, k_3, k_4 are functions of α, the operating frequency and the cutoff frequency. Equation (29) indicates that a plot of G'_{xy} as a function of G_y should be a family of parallel straight lines having slopes equal to k_1 and intercepts proportional to the value of the parameter C_y. Similarly the plot of G'_{xy} vs C_y, of C'_{xy} vs G_y and of C'_{xy} vs C_y are all families of straight lines.

The experimental measurement when plotted shows such families of straight lines. Table II summarizes these data. The measured value of each of the four slopes, k_1, k_2, k_3 and k_4 is tabulated for several values of the parameter. The values of these slopes computed from measured values of α_o and cutoff frequency are included in this table. In general the computed values are somewhat higher than the measured values but the results are all within limitations of experimental error.

Acknowledgment

TABLE II

Experimental Verification of Equations (25 and (26)
by Input Admittance Measurements at 300 kc.

$$G'_{xy} = g_c + k_1 C_y - k_2 C_y$$

$$C'_{xy} = c_c + k_3 C_y + k_4 C_y$$

C_y	k_1	k_3		G_y	k_2	k_4
0	.1128	$.1016 \times 10^{-6}$		100 μmho	$.378 \times 10^6$	.1175
50.3μμf	.1160	$.1030 \times 10^{-6}$		316.5μmho	$.365 \times 10^6$	.1158
98.7μμf	.1160	$.1004 \times 10^{-6}$		561.8μmho	$.364 \times 10^6$	.1168
199.6μμf	.1203	$.0949 \times 10^{-6}$		826.4μmho	$.344 \times 10^6$	.1072
Computed value	.1340	$.1190 \times 10^{-6}$		1000 μmho	$.338 \times 10^6$	.1175
				Computed value	$.421 \times 10^6$	.1340

References

1. R. M. Ryder and R. J. Kircher - Some Circuit Aspects of the Transistor. B.S.T.J., Vol. 28, pp. 367-401, July, 1949.

2. R. L. Wallace, Jr. and W. J. Pietenpol - Some Circuit Properties and Applications of NPN Transistors. Proc. I.R.E., Vol. 39, pp. 753-767, July, 1951. Also B.S.T.J., Vol. 30, pp. 530-563, July, 1951

3. D. E. Thomas - Transistor Amplifier - Cutoff Frequency. Proc. I.R.E., Vol. 41, pp.1481-1483, November, 1952

4. R. L. Pritchard - Frequency Variations of Current-Amplification Factor for Junction Transistors. Proc. I.R.E., Vol. 41, pp. 1476-1481, November, 1952

5. W. Shockley, M. Sparks and G. K. Teal - P-N Junction Transistors. Phy. Rev., Vol. 83, pp. 151-162, July 1, 1951

6. J. M. Early - Effects of Space Change Layer Widening in Junction Transistors. Proc. I.R.E., Vol. 41, pp. 1401-1406, November, 1952
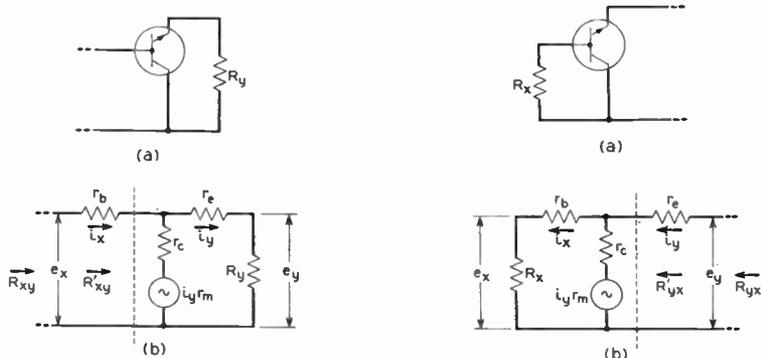
Fig. 1
Grounded-collector amplifier
transmitting in the XY direction.
(a) schematic;
(b) equivalent circuit.

Fig. 2
Grounded-collector amplifier
transmitting in the XY direction.
(a) schematic;
(b) equivalent circuit.
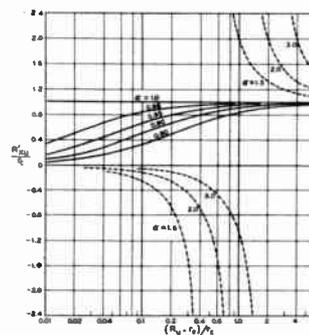
Fig. 3
Variation of input resistance
as a function of load resistance -
XY transmission.

115
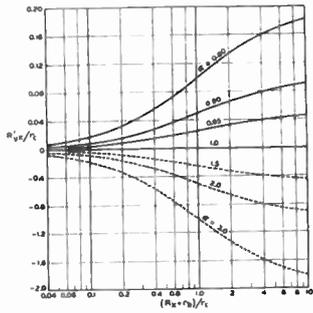
Fig. 4
Variation of input
resistance as a function
of load resistance -
YX transmission. Note
that scale for R'_{yx}/r_c
is different for
positive and negative
values.

Fig. 5
Circuit for measurement
of cutoff frequency.
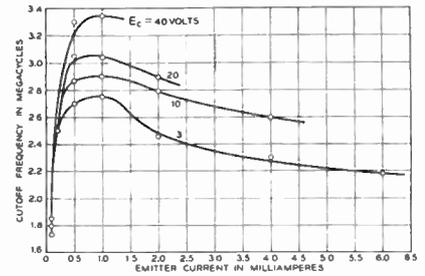
Fig. 6
Cutoff frequency of an
n-p-n transistor as a
function of emitter
current.

Fig. 7
Equivalent circuit for
input impedance of
grounded-collector
amplifier. XY
transmission.
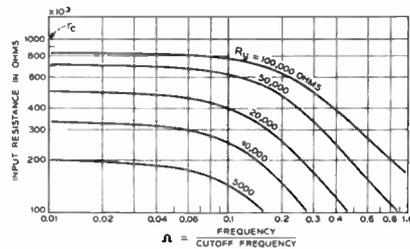
Fig. 8
Input resistance (XY transmission)
for transistor having
$\alpha_o = .98$, $r_c = 1$ megohm, $f_c = 2$ mc
$c_c = 5$ mmf.
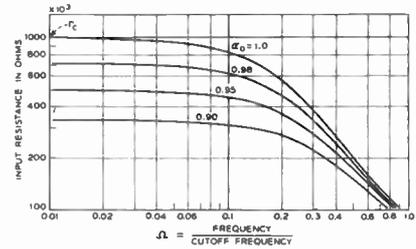
Fig. 9
Input resistance (XY transmission)
for transistor same as Fig. 8
except having other values of
α_o. $R_y = 50,000$ ohms.

Fig. 10
Input capacitance (XY transmission)
for same transistor as Fig. 8.
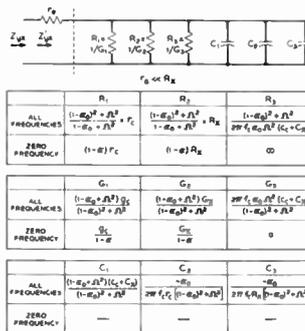
Fig. 11
Equivalent circuit for input
impedance of grounded-collector
amplifier. YX transmission.
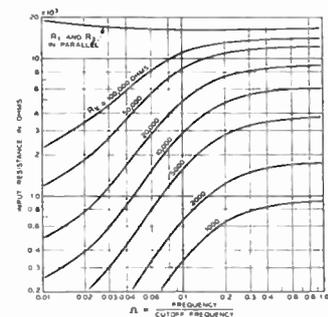
Fig. 12
Input resistance (YX transmission)
for same transistor as Fig. 8.

116
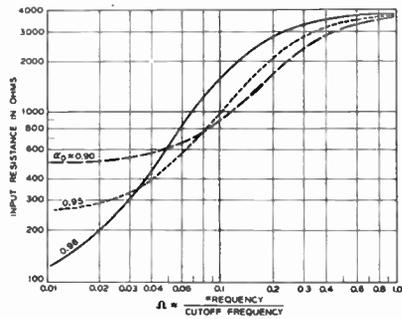
Fig. 13
Input resistance (YX transmission)
for transistor same as Fig. 8
except having other values
of α . R_x = 5000 ohms.
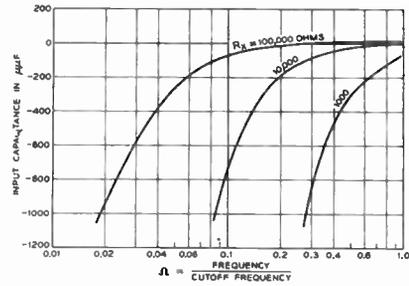
Fig. 14
Input capacitance (YX transmission)
for same transistor as Fig. 8.
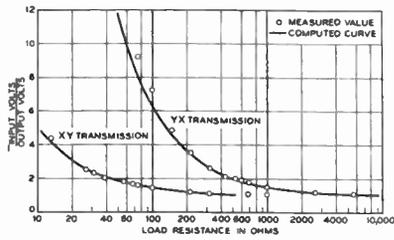
Fig. 15
Measured values of ratio of input
to output voltage as function
of load resistance. Output
voltage - .005 v;
frequency - 30 kc.
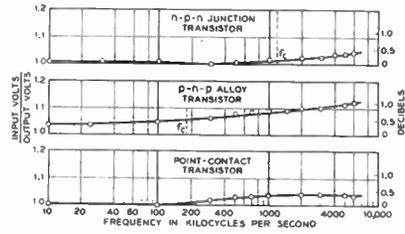
Fig. 16
Measured values of ratio of input
to output voltage as function
of frequency. XY transmission.
R_y = 5620 ohms.
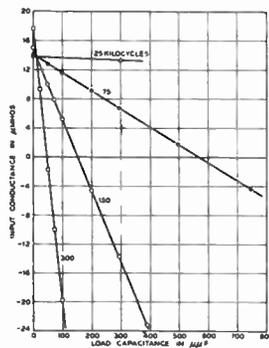
Fig. 17
Input conductance as a function of
load capacitance. XY transmission,
R_y = 10,000 ohms.
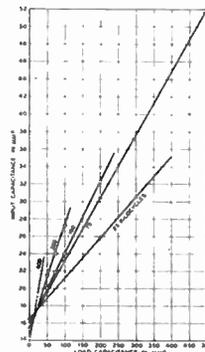
Fig. 18
Input capacitance as a function of
load capacitance for same
amplifier as Fig. 17.

117

SYMMETRICAL PROPERTIES
OF TRANSISTORS AND THEIR APPLICATION

George C. Sziklai

RCA Laboratories
Princeton, N. J.

ABSTRACT

There are certain transistor characteristics which are not present in vacuum tubes. Some of these characteristics may be best classified as symmetrical properties.

The first kind of symmetry may be found in the complementary characteristics of the NPN and PNP transistors. One is the symmetrical counterpart to the other, as a hypothetical positron triode would form a counterpart to the conventional electron tube.

This property may find a number of uses. As the base current is changed in the same direction in both units the emitter-collector current flow will increase in one and decrease in the other. A pair of these units fed from the same signal will provide a single ended push-pull output. Such a single ended push-pull circuit operates without a transformer or phase inverter.

Another application of the use of PNP and NPN transistors in combination is the use of two complementary units to form a direct coupled amplifier. The collector of a PNP transistor can be connected directly to the base of an NPN transistor and the power applied through the emitter. By this means a long chain dc amplifier can be built. This circuit provides a somewhat lower voltage or current gain than α cb per stage.

The complementary symmetry of transistors find an interesting application when it is applied both in cascading and for providing push-pull amplification. Such a two stage direct coupled Class B amplifier does not contain any parts other than the transistors themselves when operating from a high input resistive source directly into a 16 ohm loud-speaker voice coil.

Most transistors display another symmetrical property involving a single unit. This symmetry permits a current flow between the emitter and the collector in either direction and it is controllable by the base current. There is no comparable action in vacuum tubes since this would require an anode omitting electrons and a thermionic cathode accepting them. With transistors however, units with high degree of symmetry can be constructed. This is particularly true for the alloying process of transistor making.

The single unit symmetry has many interesting applications since it provides a fast bidirectional switch. A single symmetrical transistor can provide a sawtooth current with very high efficiency and with a minimum number of circuit components. A symmetrical transistor provides a very simple clamp circuit, balanced modulator, phase and FM detector.

A STUDY OF TRANSISTOR
CIRCUITS FOR TELEVISION RECEIVERS

George C. Sziklai,
Robert D. Lohman and Gerald B. Herzog

RCA Laboratories
Princeton, N. J.

ABSTRACT

After the discovery and a study of the symmetrical properties of junction transistors, several circuits were developed which are particularly useful in television systems.

Parallel with this development some experience was obtained with point contact transistors both in pulse and v-h-f circuits which appeared to be useful in television circuitry. In view of

118

these encouraging tests, it was decided to make a general study of transistors in television receivers. For this purpose, the development of a completely transistorized television receiver was undertaken. An experimental model using 37 developmental transistors and a 5-inch kinescope, housed in a plastic cabinet 13 x 12 x 7 inches, was constructed. This portable receiver operates on a single channel using a self-contained loop, and has a total battery-power consumption of 13 watts, more than 25 per cent of which is consumed by the kinescope heater.

The development of a complete experimental receiver, even with a number of compromises, provided an opportunity to deal with the problems found in every stage and circuit of the receiver. Although experimental point-contact transistors have recently been developed which will provide oscillations for the entire v-h-f television band, considerable difficulty was found in providing wide-band r-f gain using transistors at these frequencies. This problem was much less difficult at intermediate frequencies and at the intercarrier-sound frequency. The second-detector problem of obtaining high rectification efficiency with low load impedances was solved by using a transistor detector. The video amplifier problem was complicated by the requirement for a high input impedance; however, with a combination of junction and contact transistors a stable, high gain video amplifier with a relatively high input impedance was built. An audio system using complementary symmetrical junction transistors was designed to produce high output with good efficiency.

In the synchronization and deflection portion of the receiver circuits were devised for using transistors in ways which differ from the analogues of amplifier tubes. A single transistor was used as a d-c setter, sync separator and sync amplifier. A simple and reliable horizontal a-f-c system was developed by utilizing the symmetrical properties of transistors. Point-contact transistors were found to be particularly economical pulse and sawtooth oscillators. The complementary symmetry principle was used to provide vertical deflection with high linearity and efficiency. In the horizontal deflection circuits, the fast high-current switching ability of transistors was used advantageously. An efficient circuit using the symmetrical property of the transistor, which has no analogue in electron tubes, was devised for horizontal deflection.

Fig. 1

119
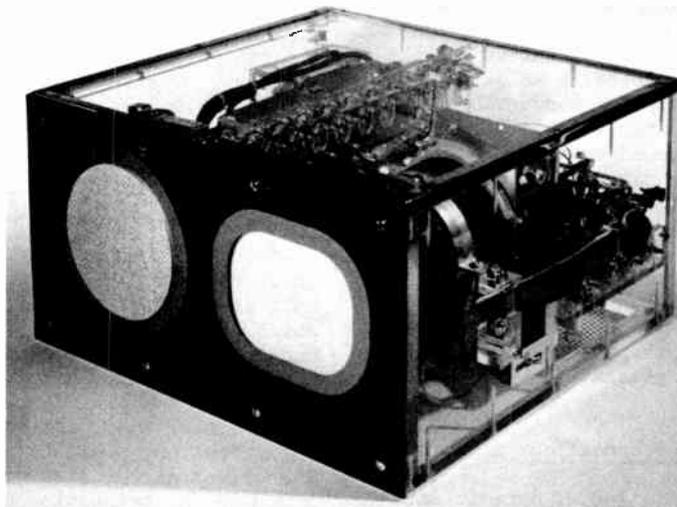
CONDUCTANCE CURVE DESIGN OF RELAXATION CIRCUITS

Keats A. Pullen
Ballistic Research Laboratories
Aberdeen Proving Ground, Md.

Summary. Design of non-linear repetitive circuits using electron tubes requires data not readily available on static tube characteristics curves. A technique for use of the recently developed conductance curves to this design problem has been developed.

The application of the technique to the design of multivibrators and blocking oscillators requires knowledge of the dynamic loop gain, the plate voltage swings, the dynamic tube conductances, and the static circuit characteristics. Determination of the switching time and the initiation bias are considered. The effect of the conductance of the positive grid and the effects of tube conductances are studied.

Several examples of multivibrator and blocking oscillator designs illustrating use of the method are presented. Experimental confirming data are included. The agreement with theory is examined.

The Relaxation Type Circuit

The relaxation circuit is an electronic configuration having clearly definable active periods for its active elements and clearly definable quiescent periods for the active elements. The active elements may be triode, tetrode, pentode, or multielement electron tubes, transistors, fieldistors, magnetic amplifiers, or any other form of element capable of controlling the transformation of energy.

Analysis and design of relaxation circuits may be divided into two phases. The first phase might be called the passive phase. The second phase is the active phase. The technique of calculation of the passive phase is well understood. The correction for plate conductance on the passive phase decay characteristic in multivibrators has been handicapped by the lack of adequate plate conductance data, however. Determination of the initiation point and the active transition characteristics has likewise been hampered by lack of adequate data on dynamic parameters of the active element. The determination of initiation conditions and the characteristics of the active transition is studied after a discussion of dynamic data presentation for electron tubes for use in relaxation type oscillators.

Dynamic Data Presentation

The problem of presentation of dynamic data for application to the design of comparatively linear tube circuits has already been described[1].

1. For Bibliography See Proceedings of National Electronics Conference, 1950, page 120

Where a wide degree of variation of the range of the parameters of the tube must be available, however, previously described forms of curves may not prove adequate.

The use of a logarithmic plate current scale on the standard plate or screen conductance type characteristics curves provides the required range of data on at least the static parameters. Choice of suitably scaled contour values for the conductance contours could then provide the dynamic data.

A set of logarithmic curves on the 6SN7 (6J5) triode has been prepared on one triode section of a 6SN7 tube (Fig. 1). As can be noted, the plate current range is from one microampere to ten milliamperes. The grid bias range is from zero to minus 26 volts. The transconductance and plate conductance contours plotted are those having 1, 2, 5, 10, 20, 50, etc micromhos up to the maximum value provided in the tube at zero bias.

Although the requirement of a curved load line would initially appear to lower the usefulness of the semi-logarithmic plot, the operation of plotting the curved load line does not appear to be difficult. The usefulness of the spreading of the data in the low transconductance area more than offsets the inconvenience in cases where a wide range of data must be provided.

During the initiation of transition in the multivibrator, dynamic characteristics on both active elements must be known. The initiation of the transition is controlled by the element which has been switched off and is slowly drifting back toward a conducting condition. The conducting element is providing a static amplification during the initial phases of triggering of the multivibrator. The triggering of the blocking oscillator, discussed later, develops in somewhat similar manner.

Electron Tube Multivibrator Initiation and Transition

The initiation of multivibrator transition in a free-running multivibrator occurs as a result of the slow drift of the bias of the tube which has been switched off back toward the conducting condition. Transition may be expected to start immediately upon establishment of an open circuit loop amplification of approximately unity. The drift of bias, neglecting the effect of loop amplification, would be comparatively uniform. The effect of the rise of loop amplification is to speed up the transition by introduction of a voltage in series with the discharging capacitor which acts in such a direction as to change the bias more rapidly. If, for example the amplification of triode one is minus A_1, and triode two is minus

A_2, the voltage change, neglecting the effects of capacitances, in the plate of triode two from a change in voltage E_o in the grid of triode one is $E_o A_1 A_2$. Consequently, the apparent voltage applied at the grid of tube one now becomes $E_o(1 + A_1 A_2)$. Expanding this equation with successive cycles of feedback, one gets a product series.

The voltage rise due to amplification of course does not occur instantaneously. The rate of rise depends on the stray circuit capacitances as well as the amplifications of the two amplifier tubes. Strictly, at least two build up functions will interact to control the time delay around the loop. The difference equation governing one cycle of amplification may be written

$$(dE_n/dt)/(dE_{n-1}/dt) = 1 + A_{1n} A_{2n}$$
$$\left(1 - e^{-\delta_1(t-tn)}\right)\left(1 - e^{-\delta_2(t-tn)}\right) \qquad (1)$$

From (1) the amplification at any instant may be written as

$$dE_n/dt = \prod_{j=0}^{n} \left[1 + A_{1j} A_{2j} \left(1 - e^{-\delta_{1j}(t-tj)}\right) \right.$$
$$\left. \left(1 - e^{-\delta_{2j}(t-tj)}\right)\right]\left(dE_o/dt\right) \qquad (2)$$

or the amplification may be written as

$$(dE_n/dt)/(dE_o/dt) = VA_n = \prod_{j=0}^{n} \left[1 + A_{1j} A_{2j}\left(1-e^{-\delta_{1j}(t-tj)}\right)\right.$$
$$\left. \left(1-e^{-\delta_{2j}(t-tj)}\right)\right] \qquad (3)$$

In these equations $\delta_{1j} = G_{1j}(t)/C_{1j}(t)$; $\delta_{2j} = G_{2j}(t)/C_{2j}(t)$, where the G's and the C's, being variable, must be calculated from the circuit parameters by the use of equations (6).

If the value of $(t - t_j)$ is chosen such that $\delta_{1j}(t-t_j)$ and $\delta_{2j}(t-t_j)$ are very small compared to unity, then (3) may be simplified. Taking $t - t_j = \Delta t_j$, then (3) becomes

$$VA_n = \prod_{j=0}^{n} \left[1 + A_{1j} A_{2j} \delta_{1j} \delta_{2j} \Delta t_j^2\right] \qquad (4)$$

Taking the logarithm both sides of (4) converts the product into a series.

$$Log\ VA_n = \sum_{j=0}^{n} Log\left[1 + A_{1j} A_{2j} \delta_{1j} \delta_{2j} \Delta t_j^2\right] \qquad (5)$$

The right hand side of (5) could be expanded into a power series, the value of Δt_j allowed to approach zero, and definite integral obtained if convergence may be assumed. Since in one transition, A_{1j} varies between zero and a finite value, and A_{2j} between a finite value and zero, the assumption of convergence appears to be reasonable.

The physical significance of (4) and (5) is rather interesting. The effect of amplification in the multivibrator is to cause a multiplication of voltage change on successive loops around the feedback circuit. Consequently, relatively small rates

of change of bias can be converted into extremely high rates of change at the steepest point in the transition. This steepness may be great enough to render the steepest part of the transition difficult to observe on even the highest speed oscilloscopes.

Practical use of (5) does not require transformation into a definite integral. The choice of a value of Δt_j to satisfy the relation $A_{1j} A_{2j} \delta_{1j} \delta_{2j} \Delta t_j^2 = c \leq 0.1$ on an iterative basis enables one to determine the time required per incremental ratio of amplification change. The relation in its present form does not readily indicate the rate of change of bias with time, but gives the rate of change of amplification. Conversion of the rate of change of amplification to rate of change of voltage requires division of Δt_j by $(1 + c)^j$ since the initiating voltage for each pass around the loop should be reduced to $1/(1 + c)^j$ of the final output voltage of the previous passage around the loop.

A table giving values of A_{1j}, A_{2j}, δ_{1j}, δ_{2j}, Δt_j, and $1/(1+c)^j$ may be formulated for a given multivibrator. The values of C_{1j}, C_{2j}, G_{1j}, and G_{2j} must be determined in order that the deltas may be used in the determination of both amplification values and the net bias values may be found. The equations for these functions are

$$G_{1j} = 1/R_{L1} + G_{P1} + 1/R_{C2}$$
$$G_{2j} = 1/R_{L2} + G_{P2} + 1/R_{C1}$$
$$C_{1j} = C_{pk1} + C_{pc1} + C_{ck2}$$
$$\qquad + (1 + A_{2j}) C_{cp2} + C_d \qquad (6)$$
$$C_{2j} = C_{pk2} + C_{pc2} + C_{ck1}$$
$$\qquad + (1 + A_{1j}) C_{cp1} + C_d$$

The correct value of j must be determined on the basis of the chosen value of c, and the amount of the total bias change. If the basic increment ΔE_o is taken as 10^{-3} volts, for example, $E_{cj} - E_{co} = (1 + c)^j \Delta E_o = (1 + c)^j \times 10^{-3}$ volts.

Proceeding in the indicated manner, one finds several interesting results. The first is that the rate of transition in a multivibrator may reach 1000 volts per microsecond. The second is that the second tube apparently is cut-off before the first tube reaches zero bias.

Examination, however, of the plate voltage change in the second tube required to carry the first tube to grid conduction shows that a nominal bias change on the second tube of the order of between one and five volts is sufficient to provide full transition. Integration of $A_{1j} A_{2j}$ as a function of input bias shows the bias change at tube one required for the active portion of the

121

transition. When the integral of $A_{1j}A_{2j}$ with respect to the bias on the first tube numerically equals the initial negative bias on tube one, transition should be nominally completed.

The input capacitance of the second tube prevents the nominal voltage changes read from the tube characteristics contours from being realized. If $\delta_{1j} = \delta_{2j}$, the loss of voltage change due to time constant would be the same, percentagewise, on each tube. Where δ_{1j} is not equal to δ_{2j}, however, approximately $\delta_{1j}/(\delta_{1j}+\delta_{2j})$ of the potential voltage change available in tube two will be realized for a time delay producing approximately $\delta_{2j}/(\delta_{1j} + \delta_{2j})$ of the potential voltage change available in tube one. Proper timing and switching results if this correction is made.

Effect of Plate Conductance on the Relaxation Time of a Multivibrator

The decay circuit which controls triggering in the conventional multivibrator is a resistance-capacitance combination which holds a decaying negative potential on a tube grid. Three resistors are included in the decay path for the capacitance. The first of these is the conducting tube load resistance, the second the conducting tube plate conductance, and the third the grid resistance in the grid of the non-conducting tube. If the coupling circuit is reduced to its simplest form, the plate load resistor and the plate conductance in parallel are effectively connected in series with the grid resistor. The rate of decay is determined by this combination. The plate conductance may be read from the plate characteristics curve as in Fig. 1 or Fig. 2. The decay action may be assumed to continue until the unity loop amplification condition is reached.

Experimental Tests- Multivibrator

The main purposes of the experimental tests on the multivibrator were two. The first of these was to examine the predictability of the initiation point. The second was to attempt to learn something about the triggering of multivibrators. The tests were made with the circuit shown in Fig. 3. One set of tests was performed with 27,000 ohm load resistors and a 250 volt supply, the second with 1000 ohm load resistors and an 85 volt supply. The reason for the difference was to establish one design requiring near cutoff bias at the initiation point, and the other design requiring very little bias at the initiation point. Table I gives initiation conditions for the four tests made.

Table I
Multivibrator Initiation

Load Impedance	27000 ohms	1000 ohms
Approx. Initiation bias		
tube one	-25.5 volts	-5.4 volts
Initiation Loop Gain	1.21	1.26
Corrected Loop Gain	0.88	1.003

In the initial tests giving loop amplifications of 1.21 and 1.26, the initiation point could be moved back and forth smoothly across the initiating pulse by shifting the applied bias. Fig. 4 shows a series of oscillograms of this test with different values of bias. The time constant involved proved to be that of the 0.01 microfarad capacitor and the 270,000 ohm resistor effecting the grid of tube one (Fig. 3). The circuit behaved as if the capacitor were returned to ground potential instead of appearing to have zero signal potential difference across the capacitor which would be expected at unity loop gain. Reduction of the coupling capacitor size to 0.001 microfarads proved the correctness of the analysis.

The requirement of slightly less than unity amplification for initiation with the 27,000 ohm load resistors at first appears curious. The possibility of existence of this condition with a large gain margin had been postulated on the basis that with loop amplifications very close to unity, the system might not reach stability before the unity amplification point had been reached. This condition apparently can occur. The corrected initiation points, corrected for the time constant, were calculated from the time constant of the pulse build-up and were also checked by experiment. Amplifications checked within about five percent.

The Blocking Oscillator

The blocking oscillator uses a tube and a transformer to initiate and form a pulse through uncontrolled regeneration (Fig. 5). Initiation, as in the multivibrator, results from slow decay of the charge stored in a capacitor. The capacitor in this case is in the grid return lead.

The presence of the transformer provides several interesting characteristics to the blocking oscillator. For very slow rates of change of bias, the transformer will behave as if it were without appreciable resistance, reactance, or magnetic coupling. When the grid voltage leaks off to a point that the effect of the grid voltage change on the transconductance produces sufficient change of plate current, the transformer behaves as a unity coupled transformer-phase inverter. Its load is then the grid leak resistor until zero bias is reached, then the grid conductance and the grid leak resistor. After cycle reversal, the loading exits in reverse order.

Answers to two questions were desired in the tests performed on the blocking oscillator. The first of these was: "Does the blocking oscillator initiation point vary with the slope of the applied triggering pulse?" A corollary question was "If so, does it vary in such a way that $LG_M(dE_c/dt)$ is approximately constant?" The second question was "What is the appearance of the blocking oscillator load line?"

The significance of the first question is

122

that $LG_M(dE_c/dt)$ will be constant if a loop gain, with the transformer behaving purely as an inductance, of a fixed value is required for initiation of the blocking oscillator. The measurements made indicate that the triggering of the blocking oscillator tested does appear to depend on the slope of the trigger pulse. In fact, to at least a rough approximation, the function $LC_M(dE_c/dt)$ does appear to be a constant at the initiation point. To date, the pressure of other work has prevented determination of the initiation gain required with the blocking oscillator.

The circuit of Fig. 5 was designed to provide access to both the cathode current and the plate voltage to permit oscillograph recordings to be made of the load line. Two typical load lines obtained are shown in Fig. 6. The first of these was made without the two clamper diodes (high vacuum) across the sawtooth generator circuit. As can be noted, the load line rises as

expected, but then drops vertically and follows the zero current axis. This condition apparently is a result of the load line drift with accumulation of rectification bias. With the clamping diode to discharge this bias, the load line takes a more conventional appearance.

Conclusions

The design of relaxation oscillators of at least two types appears to be aided by the use of a logarithmic plate current conductance type of plate or screen characteristics curve set. The presence of the dynamic conductance information enables one both to obtain a superficial feel of the functioning of the circuit and in at least some cases to obtain considerable detail information not otherwise calculable about a circuit design. The information obtainable, in fact, may be of sufficient use in many design problems to justify taking the conductance type contours as preliminary data for the determination of the design.
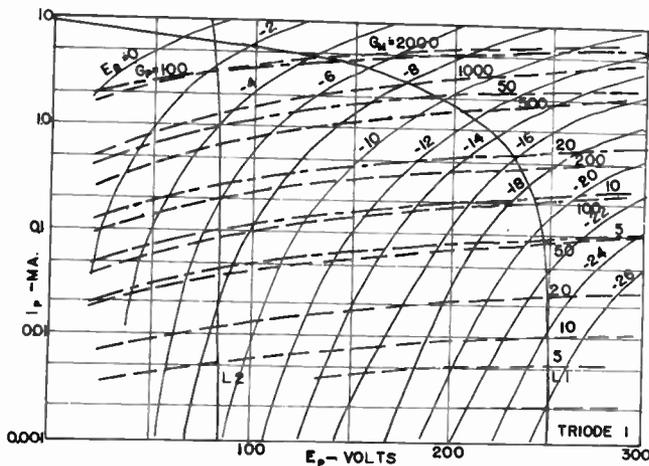
Fig. 1
Logarithmic current plate characteristics
curves - 6SN7 tube - triode one.
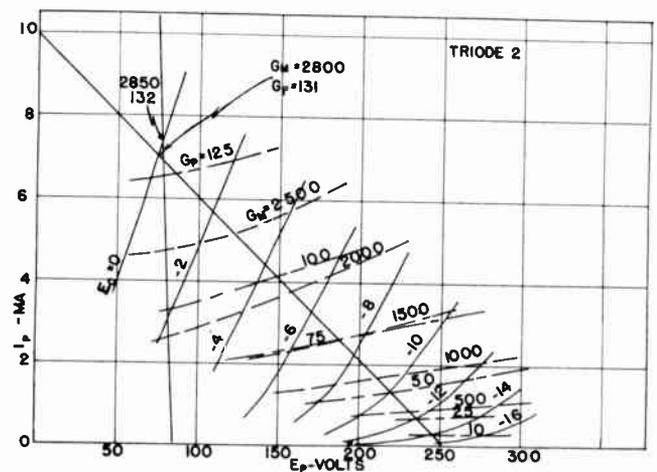
Fig. 2
Linear plate characteristics curves -
6SN7 tube - triode two.

123
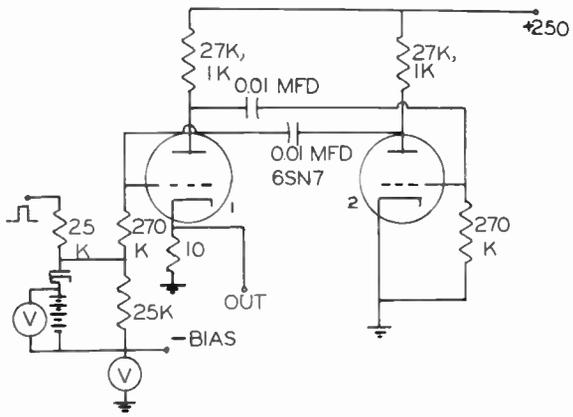
Fig. 3
Test multivibrator circuit.
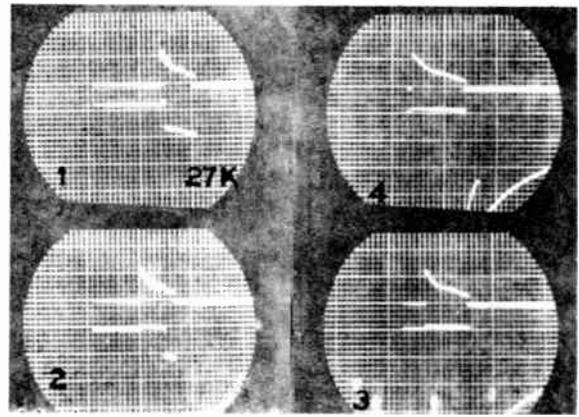
Fig. 4
Multivibrator waveform oscillograms.
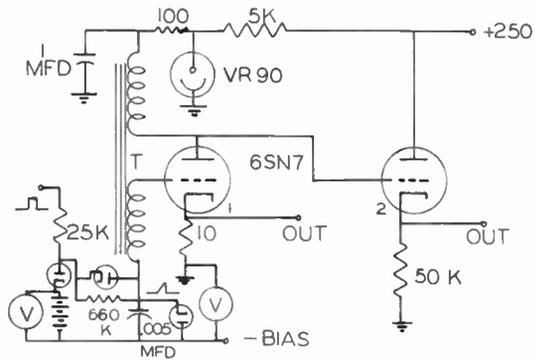
Fig. 5
Test blocking oscillator circuit.
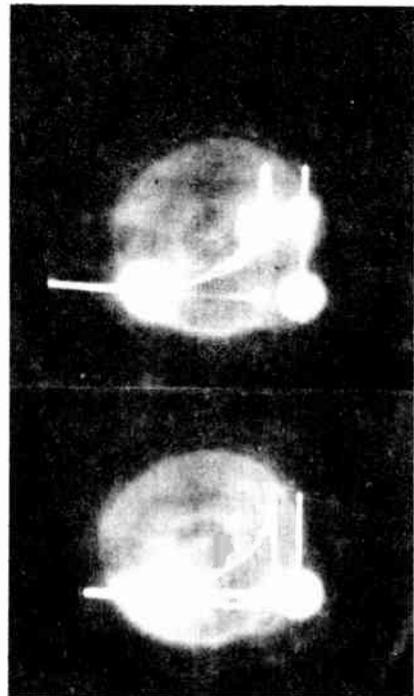
Fig. 6
Test blocking oscillator
load line oscillograms.

124

TRANSISTOR RELAXATION OSCILLATORS

Stanley I. Kramer

Fairchild Guided Missiles Division
Wyandanch, New York

Introduction

Several variations of the basic relaxation oscillator are discussed which provide certain advantages. These include lower peak currents and a closer realization to idealized rectangular waveforms.

Review of Conventional Circuit

The conventional circuit places a capacitor across the emitter input and utilizes the negative resistance for switching. This configuration is shown in Fig. 1(a) and its operation
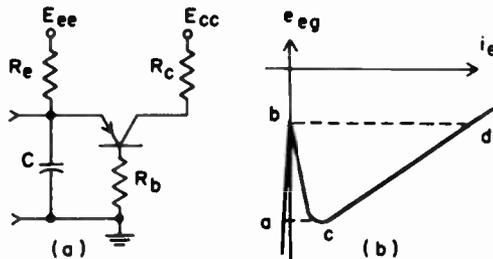
Fig. 1 - Basic Relaxation Oscillator
and Emitter N-Curve

can be analyzed in conjunction with the emitter N-curve. The emitter N-curve is the volt-amp characteristic of this circuit taken from emitter-to-ground and is shown in Fig. 1(b). This concept for analysis has been described elsewhere,[1] but is reviewed here briefly as a basis for the discussion to follow.

The curve described by the solid line abca is a plot of emitter current versus emitter voltage to ground and can be obtained point by point or dynamically. The segment ab represents the high input resistance at cut-off and is generally of the order of 1 megohm, bc is the negative resistance region which is largely a function of R_b and alpha, and cd represents saturation, the slope being a function principally of r_e and the external circuitry.[2]

Considering the operation of the circuit, capacitor C slowly charges from point a to the peak point b. At this instant the emitter resistance breaks down and the current suddenly increases. The voltage across C, however, cannot change instantly and corresponds in effect to a zero impedance load line which intersects the N-curve at d. The capacitor then proceeds

to discharge through the emitter, whose impedance to ground is defined by the slope of dc, to the valley point c. Here the current suddenly decreases and the capacitor again prevents any change of voltage which causes the current to return to the value corresponding to point a and the cycle is complete.

Two undesirable features of this circuit are the sloping top in the collector waveform (Fig. 2) caused by the changing emitter current in segment a-c (Fig. 1(b)) and the very high peak
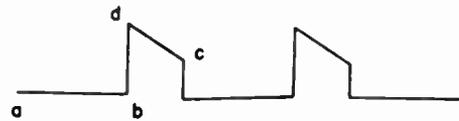
Fig. 2 - Basic Collector Waveform

currents, especially of the emitter, which exist at point d.

Modified Circuit

Fig. 3(a) shows a modification of the basic circuit having most of its desirable features and few of its shortcomings. The changes in the
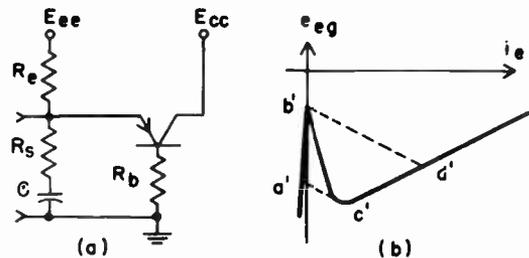
Fig. 3 - Modified Relaxation Oscillator
and Emitter N-Curve

basic circuit are the addition of a loading resistor, R_s in series with C and the elimination of R_c. Fig. 3(b) is the corresponding N-curve for this circuit and is referred to in the analysis which follows. Beginning at a', the voltage at the emitter rises as capacitor C charges from E_{ee}. At b' the current is suddenly increased by the switching action, but this time, although the capacitor voltage remains momentarily constant, the voltage across the emitter terminals is subject to change due to the drop in R_s. Thus b'd' has a slope corresponding to

the resistance R_s. From d' the voltage falls due to the discharge of C to c'. Here, again, the emitter resistance changes abruptly and c'd' is generated parallel to b'd' to complete the cycle.

Emitter Loaded Circuit

If $R_s \ll R_e$, this circuit degenerates to the simple case of a loading resistance in series with the emitter as shown in Fig. 4. The N-curve of this circuit taken at the points shown in the
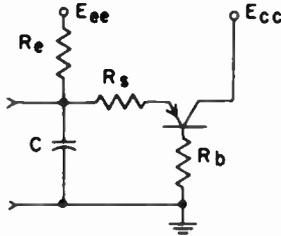
Fig. 4 - Emitter Loaded Relaxation Oscillator

diagram appears somewhat different than the previous one, but is actually equivalent. This is shown in Fig. 5. (a) and (b) are oscilloscope traces of the operating circuits of Figs. 3 and 4, respectively, and (c) is a superposition
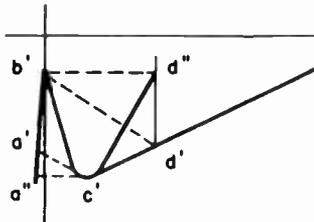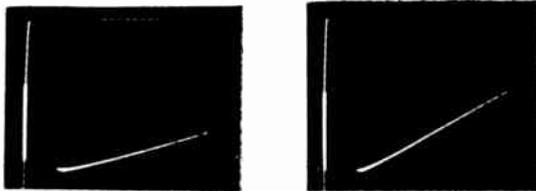
Fig. 5 - Emitter N-Curves of Oscillating Circuits; (a) Circuit of Fig. 3, (b) Circuit of Fig. 4, (c) Superposition of (a) and (b)

of the two curves. Curve a'b'd'c' is Fig. 3(b), whereas a"b'd"c' is obtained across C in Fig. 4. It will be noted that the current end points are substantially the same except for those corresponding to a' and a" which is generally second order and due to the approximation that R_s is very much less than the emitter resistance in ab. This shows, therefore, that resistance loading of the emitter is equivalent to altering the N-curve by a controllable amount in the saturation region. The slope of the curve in this case is

approximately equal to R_s if $R_c = 0$. This is shown by the following analysis which makes use of the equivalent circuit of Fig. 6.
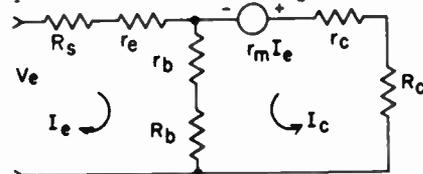
Fig. 6 - Equivalent Circuit of Emitter Loaded Relaxation Oscillator

The loop equations are

$$V_e = (R_s+r_e+r_b+R_b)I_e + (r_b+R_b)I_c$$

$$0 = (r_m+r_b+R_b)I_e + (r_b+R_b+r_c+R_c)I_c$$

Solving for V_e/I_e

$$R_{in} = \frac{V_e}{I_e} =$$

$$\frac{(R_s+r_e+r_b+R_b)(r_b+R_b+r_c+R_c)-(r_b+R_b)(r_m+r_b+R_b)}{r_b + R_b + r_c + R_c}$$

In the saturation region the internal parameters r_e, r_c, r_m, and r_b are very small, generally less than 100 ohms.[3] If these are small compared with the external circuitry the expression reduces to

$$R_{in} = R_s + \frac{R_b R_c}{R_b+R_c}$$

and if, in addition, R_c is made zero, the final expression reduces to

$$R_{in} = R_s$$

In a similar manner it can be shown that with $R_c = 0$, alpha approaches unity during saturation. Alpha by definition is

$$\frac{r_b + r_m}{r_b + r_c}.$$

In the subject circuit, however, this becomes

$$\frac{r_b + R_b + r_m}{r_b + R_b + r_c}$$

and if r_b, r_m and r_c are small compared with R_b, alpha is very nearly unity. The consequence of this fact is that during saturation $\frac{\Delta i_c}{\Delta i_e} = -1$, and the net current through the base is constant. This results in the flat topped waveform illustrated in Fig. 7. Thus, two effects have been achieved. First, the N-curve has been altered to permit large amplitude waveforms without excessive peak currents by controlling the saturation region and secondly, the waveform has been

Fig. 7 - Typical Waveform of Emitter Loaded
        Circuit

flattened by adjusting alpha to unity by removal
of R_c and by taking the output at the base.

Time Constants

The foregoing discussion using the N-curve
analysis makes no mention of the time constants
involved. This is more readily analyzed with
the aid of the simplified equivalent circuit of
Fig. 8 which assumes; 1. r_e is infinite during
cut-off and zero during saturation, 2. r_c is
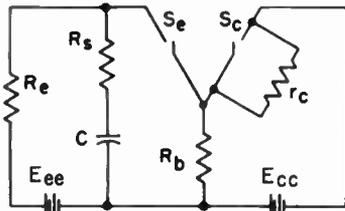
Fig. 8 - Simplified Equivalent Diagram

zero during saturation and 3. $r_b \ll R_b$. Then,
starting from cut-off (point a' in Fig. 3(b)), S_e
and S_c are open and capacitor C charges toward
E_{ee} with a time constant $(R_e + R_S)C$. When the
peak point is reached, the switches close and C
begins to discharge toward E_{cc} with a time
constant $R_S C$. The discharge continues until the
valley point is reached at which time the
switches re-open and the cycle is completed.
It should not be concluded from the foregoing
that the charge time is necessarily longer than
the discharge time, since the voltages involved
are different for the two cases. By making E_{ee}
and R_e small, it is possible to obtain a sym-
metrical waveform or even to go beyond. To
summarize, the cut-off time is a function of the
time constant $(R_S + R_e)C$ and voltage E_{ee}, whereas
the time spent in saturation depends upon time
constant $R_S C$ and voltage E_{cc}. This is not exact
since the transistor characteristic, naturally,
enters into the analysis.

Performance

The amplitude of the output voltage is
comparable with that obtainable from the basic
one and in practical circuits may approach the
collector supply voltage. However, the waveform
has a flat top unlike the sloping top of Fig. 2
and so the useable amplitude for most appli-
cations is considerably greater.

The top of the waveform for $R_S = 2K$ is
generally flat within 10% for W.E. 1698 and 1768
transistors. By trimming R_S it is possible to

make the top absolutely flat or even give it a
positive slope. A typical waveform is shown in
Fig. 7.

The pulse width is determined largely by the
product $R_S C$ and it is essential that the ratio
be controlled. The limitation being that large
values of C should not be used with very small
values of R since this may cause excessive dissi-
pation by maintaining high peak currents for too
long a time. The waveform in Fig. 7 has a period
of 1 millisecond and was obtained with R_S = 1K,
C = .1 ufd. The rise and fall times are compar-
able with those obtained in other transistor
relaxation oscillators and vary from .1 to .4
microsecond for the rise time with W.E. 1698 and
1768 transistors, and roughly twice this figure
for the fall time.

One of the principal advantages of this
circuit is the reduction of both the peak and
average currents compared with those obtained
in the classical circuit. In this case the
peak emitter current can be held at any desired
value beyond the valley point by proper choice
of R_S (see Fig. 3(b)).

While the discussion up to now has been
directed at free running oscillators, these
circuits are equally adaptable to mono-stable
or triggered operation by proper choice of
emitter bias. Synchronization pulses may be
injected either at the base or emitter depending
on the polarity of the pulse voltage. This is
often a distinct advantage over the circuit of
Fig. 1 which can only be triggered at the base,
the emitter being effectively bypassed by the
capacitor.

Finally, the reliability is increased by
controlling the slope of the saturation region
of the N-curve with stable elements external to
the transistor. This can be used in conjunction
with other, more direct, schemes of stabilization
to obtain even better uniformity.

A comparison between this circuit and the
basic one is shown in Table I as a function of
transistor characteristics. The particular
batch of transistors used in the test has
considerably poorer uniformity than is usually
encountered, but helps to demonstrate the
improvement in uniformity obtainable.

Modified Circuit with Split Load

A variation of this circuit is shown in
Fig. 9 where a finite R_c is inserted and the $R_S C$
network is positioned between emitter and
collector. Much of what has been said con-
cerning emitter loading also applies to this
circuit. The principal difference is that R_c is
now finite. Performance of this circuit is
similar to the previous one, but two separate
outputs may be taken; one from the base and the

other from the collector. The waveshapes are practically identical except for opposite polarity
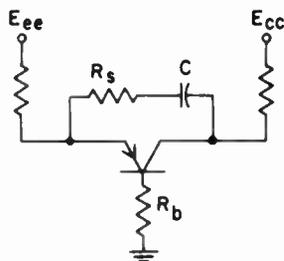
Fig. 9 - Relaxation Oscillator with Out-of-Phase Outputs Obtainable from Collector and Base

and their relative amplitudes may be adjusted by proper choice of R_c and R_b. It is not immediately obvious that the waveforms should be equal and opposite but if R_e is large compared with the other circuit parameters as is generally the case, nearly all of the emitter current must flow through R_c and of course through R_b.

The flat topped waveform is not theoretically realized in this circuit since alpha is not necessarily unity during saturation. This is a consequence of R_c being finite, but practically, the slope can be reduced to 10% or less of the total amplitude.

Narrow-Pulse Generator

The final circuit to be discussed is useful for generating short, high amplitude, rectangular pulses. The circuit diagram is shown in one form for Fig. 10(a). Quite arbitrarily, ground has been chosen to be the base. By a simple
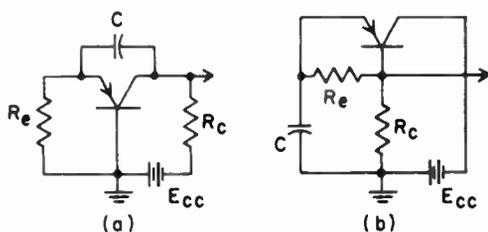
Fig. 10 - Narrow-Pulse Generator

transformation, which consists primarily of moving the signal ground to the collector, the circuit of Fig. 10(b) is obtained. This is similar to the circuit of Fig. 3 with two main differences; first R_s has disappeared and second R_e is returned to the base instead of to ground. This circuit now combines the features of the basic circuit (Fig. 1) in that R_s is omitted, with those of the modified circuit (Fig. 3) where R_c is omitted and output is taken from the base. Analysis of this circuit follows the ones described earlier and need not be repeated. The omission of R_s results in rather narrow pulses (order of a microsecond) but with high peak currents. The peak current is especially high with zero collector load since this forces the saturation segment of the N-curve to have a very small slope and pulls point a, Fig. 1(b) far out to the right. In order to prevent failure of the transistor due to excessive current, it has been painfully determined that the maximum safe capacity for the 1768 and 1698 transistors is no more than .01 ufd. The 1768 transistor while designed for low frequency work is nearly as fast as the 1698 and is less subject to damage from overloads. The maximum pulse amplitude that was obtainable using a 45 volt supply was 30 – 40 volts depending on the transistor. This is pushing to the limit and should not be attempted without an adequate supply of transistors. At lower voltages or with larger values of R_c the circuit is quite safe, however. Either version of Fig. 10 may be used according to the polarity desired. With the positive side of E_{cc} grounded, positive pulses are obtained in Fig. 10(a), negative in Fig. 10(b).

Returning R_e to the base in 10b, results in better stability and was first discussed by Anderson.[5] Bias is obtained from the collector current flowing through R_e. Most of the transistors used in this circuit operated as monostable oscillators with R_e greater than about 20K. If astable operation is desired an additional positive bias may be shunted from emitter to ground.

Triggering or synchronization is applied to the emitter or base according to which version is being used and the sensitivity of the mono-stable circuit is dependent on the bias.

Fig. 11 is a typical waveform showing a 30 volt pulse. The markers are spaced at 1 microsecond intervals.
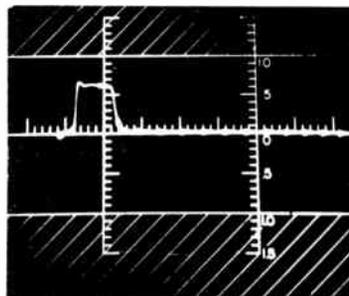
Fig. 11 - Pulse Waveform

128

1. A. E. Anderson, "Transistors in Switching Circuits", B.S.T.J. Vol. XXXI, No. 3 pp 411 - 422.

2. ibid

3. ibid

4. D. R. Brown, J. F. Jacobs, N. T. Jones, "Study of a Transistor Blocking Oscillator", Report E-435. Digital Computer Laboratory, M.I.T., pp 6, 7.

5. A. E. Anderson, "A Stabilized Transistor Delay and Switching Circuit", The Transistor, a collection of papers prepared by the Bell Telephone Laboratories. pp 429 - 436.

Table I

Trans- istor	$\propto$	R_{co} (K)	T (us)	Circuit of Fig. 1 Amplitude (relative)	T (us)	Circuit of Fig. 3 Amplitude (relative)
1*	2.0	8	750	2.0	--	-- **
5	3.8	12	875	2.0	1180	1.9
7	2.8	7	875	1.9	--	--
29	2.0	11	875	2.2	500	1.4
39	2.4	13	825	2.0	--	--
45	2.8	14	1150	2.0	1300	1.0
53	2.2	7	675	1.8	--	--
33	3.3	9	850	2.0	750	1.6
59	1.8	6	650	1.9	--	--
69	2.3	16	1335	2.3	1600	2.6
67	1.6	8	240	1.6	--	--
11	3.6	10	800	1.9	950	1.6
25	2.6	10	775	2.0	700	1.7
2	2.0	5	545	1.7	--	--
52	3.1	14	825	2.2	1100	2.1
30	4.3	20	1000	2.3	2100	2.6
40	4.1	20	890	2.0	1400	2.2
24	4.2	20	1125	2.2	2400	2.4
50	2.1	10	875	2.1	--	--
18	3.3	11	675	2.0	--	--
56	2.8	10	900	2.0	1200	1.7
12	3.1	8	750	2.0	--	--
38	4.3	20	1025	2.3	2800	2.6
60	2.0	6	725	2.0	--	--
32	3.4	8	860	2.0	1550	1.6
28	2.4	7	775	2.0	--	--
20	3.4	8	712	2.0	1300	1.7
42	2.8	10	800	2.1	1050	1.6
22	2.5	11	835	2.2	1200	1.8
46	4.5	17	1000	2.2	2300	2.2
4	2.6	9	560	1.9	--	--
34	3.6	10	.800	2.0	1100	1.6
44	3.1	14	912	2.2	1100	1.9
36	3.2	6	750	1.8	--	--

* odd numbered transistors are type 1698, even are type 1768

** these transistors could not be made to oscillate under the prescribed conditions.