

Convention Record



of the I·R·E

1954 NATIONAL CONVENTION

Part 2— Circuit Theory

LIBRARY
JUL 7 1954
R. B. PALMAY CENTER

SESSIONS ON . . .

Circuit Theory I — Symposium: Network Equalization

Circuit Theory II — Circuit Theory

Circuit Theory III — Network Synthesis

Circuit Theory IV — Transistor Circuits

SPONSORED BY

IRE PROFESSIONAL GROUP ON . . .

Circuit Theory

Presented at the IRE National Convention, New York, N.Y., March 22-25, 1954
Copyright 1954, by The Institute of Radio Engineers, Inc., 1 East 79 Street, New York 21, N.Y.

The Institute of Radio Engineers

Additional Copies

Additional copies of 1954 Convention Record Parts may be purchased from The Institute of Radio Engineers, 1 East 79 Street, New York 21, N. Y., at the prices listed below.

<i>Part</i>	<i>Title</i>	<i>Sponsoring Groups</i>	<i>Prices for Members (M), Libraries (L), & Nonmembers (NM)</i>		
			<i>M</i>	<i>L</i>	<i>NM</i>
1	Antennas & Propagation	Antennas & Propagation	1.25	3.00	3.75
2	Circuit Theory	Circuit Theory	1.25	3.00	3.75
3	Electron Devices & Component Parts	Electron Devices Component Parts	1.50	3.60	4.50
4	Electronic Computers & Information Theory	Electronic Computers Information Theory	1.50	3.60	4.50
5	Aeronautical Electronics & Telemetry	Aeronautical & Navigational Electronics Radio Telemetry & Remote Control	1.50	3.60	4.50
6	Audio & Ultrasonics	Audio Ultrasonics Engineering	1.50	3.60	4.50
7	Broadcasting & Television	Broadcast Transmission Systems Broadcast & Television Receivers	1.50	3.60	4.50
8	Communications & Microwave	Communication Systems Microwave Theory & Techniques Vehicular Communications	1.50	3.60	4.50
9	Medical & Nuclear Electronics	Medical Electronics Nuclear Science	1.50	3.60	4.50
10	Instrumentation & Industrial Electronics	Instrumentation Industrial Electronics	1.25	3.00	3.75
11	Engineering Management & Quality Control	Engineering Management Quality Control	1.00	2.40	3.00
			15.25	36.60	45.75

Responsibility for the contents of papers published in the Convention Record of the I. R. E. rests solely upon the authors, and not upon the IRE or its members.

CONVENTION RECORD OF THE I.R.E.

1954 NATIONAL CONVENTION

LIBRARY

PART 2 - CIRCUIT THEORY

JUL 7 1954

TABLE OF CONTENTS

U. S. PATENT OFFICE

Session 28: Circuit Theory I - Symposium: Network Equalization
(Sponsored by the Professional Group on Circuit Theory.)

Limitations on Amplitude Equalizers	Herbert J. Carlin	3
Synthesis of Resistively-Terminated RLC Ladder Networks	Er-Chun Ho and DeForest L. Trautman	14
Equalization of Video Cables	Philip W. Rounds	16
Application of a Minimum Phase Matrix to Adjustable Equalizer Design	W.R. Lundy	25
Equalization in the Time Domain	Murlan S. Corrington, T. Murakami, and Richard W. Sonnenfeldt	30

Session 35: Circuit Theory II - Circuit Theory
(Sponsored by the Professional Group on Circuit Theory.)

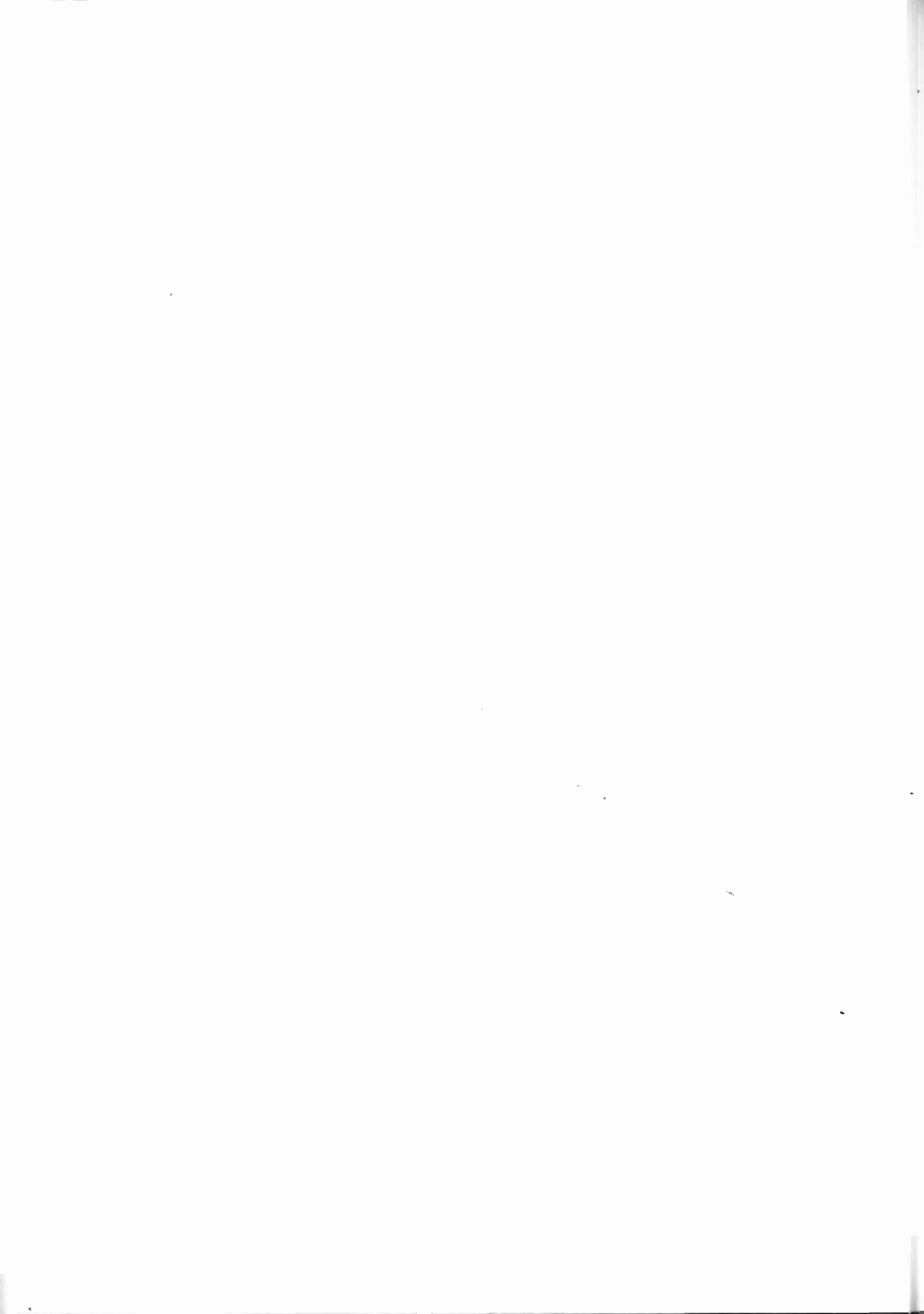
The Group Theoretical Aspect of Linear Four-Pole Theory	Wolfgang Gaertner	36
A Mathematical Technique for the Analysis of Linear Systems	John R. Papazzini and Arthur R. Bergen	44
Weighting Functions for Time Varying Feedback Systems	J.A. Aseltine and R.P. Favreau	52
Interconnection of Linear Transducers	Herbert Kurss	58
Dynamic Characteristics of Four-Terminal Networks	W.W. Happ	60

Session 39: Circuit Theory III - Network Synthesis
(Sponsored by the Professional Group on Circuit Theory.)

Some Techniques for Network Synthesis	George L. Matthaei	77
An Iterative Method for RC Ladder Network Synthesis	R.E. Scott and N. DeClaris	86
Networks Terminated in Resistance at Both Input and Output	Louis Weinberg	90
Approximating Band-Pass Attenuation and Phase Functions	V.H. Grinich	96
An Application of Modern Network Synthesis to the Design of Constant-Time-Delay Networks with Low-Q Elements	Leo Storch	105

Session 46: Circuit Theory IV - Transistor Circuits
(Sponsored by the Professional Group on Circuit Theory.)

A Transistor Analog (Abstract)	R.D. Lohman	118
Junction-Transistor Multivibrators and Flip-Flops	Eugene W. Sard	119
A Synthesis Procedure for Linear Transistor Circuits	J.P. Burnett	125
Network Partitioning Techniques Applied to the Synthesis of Transistor Amplifiers	H. Markarian	130
A New Equivalent Circuit for Junction Transistors	Ge Yao Chu	135



JUL 7 1954

Herbert J. Carlin
 Microwave Research Institute
 of
 Polytechnic Institute of Brooklyn

E. S. PAULI OFFICE

Abstract

If an equalizer amplitude response curve is specified, it will be shown that the minimum flat loss obtainable with physical networks is determined. This flat loss, or scale factor on the response curve, is a function of the equalizer output terminating impedance which is arbitrary but prescribed, and the specified tolerance on input mismatch.

If the output impedance is purely reactive, the limitations on maximum voltage transfer are obtained from a consideration of the open circuit impedance parameters of the system. If power or voltage transfer to a load with finite real part is to be optimized, the scattering parameters of the system are used to determine the limits of performance.

Examples will be given comparing the performance of matched and lossless equalizers. In many practical cases the latter do not have substantially higher gain than the matched equalizer.

I Definition of Equalizer Problem

The equalizer problem considered here concerns the transfer of voltage, current, or power from a prescribed generator with resistive internal impedance to a load whose impedance is a given function of frequency. It is presumed that a real frequency function is specified which defines the shape of transfer gain characteristic desired, and it is required to find a passive linear reciprocal equalizer network (a two terminal pair transducer, or two-port) which when placed between generator and load produces the specified gain shape and does so with maximum scale factor i.e. minimum flat loss. A gain characteristic which ideally is constant over a finite frequency band and zero elsewhere will be of major interest, and many of the results given may therefore be regarded as generalizations of the concept of maximum "gain-bandwidth product".⁽¹⁾

An additional specification is the tolerance on input mismatch. It is only because the equalizer networks investigated here are not limited to the lossless case that this specification can be set independently of the others.

In most of the examples given below the extremes of a matched input dissipative equalizer (zero mismatch) and a completely lossless equalizer (but not matched) will be compared.

Various aspects of the equalizer problem have been previously considered. Bode⁽¹⁾ discusses the limitations on "gain-bandwidth" product imposed by a load with shunt capacitance when a lossless equalizer is used, and also gives some consideration to matched equalizers for voltage transfer to a reactive load⁽²⁾. Fano⁽³⁾ has treated the problem of optimum match of an arbitrary load with a lossless network and La Rosa and Carlin⁽⁴⁾ (5) have examined this problem when the lossless restriction on the matching network is removed. Norde⁽⁷⁾ has treated matched minimum phase voltage equalizers for reactive loads. Other work on special aspects of "gain-bandwidth" product is too extensive to be given here. Wheeler⁽⁸⁾ and Hansen⁽⁹⁾ are typical references.

The present paper considers the general approach to any equalization problem and stems directly from the references cited above. The results presented on optimum voltage transfer to an arbitrary load (including the purely reactive load case) have not been given elsewhere.

II General Approach To Equalization of an Arbitrary Load

There are two basic restrictions which govern the design of an equalizer network. One of these is the general requirement of physical realizability on the overall network which includes both equalizer and prescribed load. The other is the total set of constraints specifically imposed by the load and this should be entirely independent of the equalizer network. If these constraints are satisfied, then when the overall network is synthesized, and the given load removed, the remaining circuit (the equalizer alone) is physically realizable. The form in which these restrictions are stated must be such as to explicitly (and preferably in a simple way) involve the transfer gain function whose scale factor is to be maximized. The process of finding the optimum equalization is then to adjust the gain function within these general restrictions until the limits of physical realizability are attained.

The constraints which apply to the equalization of a load containing dissipation (the purely reactive load case is considered

* This work was sponsored by Office of Naval Research under Contract Nonr-839(05), Proj. Designation NR-075-216

later) are most readily obtained by representing the prescribed load over the infinite frequency spectrum as a purely reactive 2-port with fixed elements terminated in a unit resistor. (Hereafter the generator resistance will be presumed normalized to unity) (10) (11). This representation is always possible and the single resistor is sufficient to account for all the power dissipated in the load. The "overall network" is now defined as the equalizer plus the reactance two port portion of the load.

In order that the "overall network" be physically realizable, it must have an array of scattering coefficients⁽¹²⁾ $S_{11}(p)$, $S_{22}(p)$, $S_{12}(p)$ which form the matrix of a positive definite or semi-definite Hermitian form for $\text{Re } p \geq 0$ (5) (6) (12) (13). The algebraic expression of this requirement gives the general set of realizability constraints previously referred to.

The specific load constraints are obtained from the fact that at certain real and complex frequencies no power can be transferred to the load no matter what equalizer network is used. These frequencies are the points on $j\omega$ and in the right hand half of the p plane at which the reactive 2-port portion of the load has zeros of transmission. At these frequencies the transmission factor of the overall network $S_{12}(p)$ must generally have a zero of transmission of order $2n$ if the load zero is of order n . Further the reflection factor looking in at the back end of the "overall network" i.e. $S_{22}(p)$ and generally its first $2n-1$ derivatives are completely determined by the reactance 2-port portion of the load. These properties follow from a consideration of the scattering equations for the cascade connection of a pair of two-ports⁽³⁾, and constitute the "load constraints" referred to earlier. It must be emphasized that load constraints are independent of the equalizer.

The load constraints amount to the statement that essentially the first $2n$ Taylor coefficients of the back end reflection factor of both load and overall network are equal in the series expansion about a load zero of transmission.

These requirements may be expressed in terms of the Cauchy formulas for the Taylor coefficients, and as a final result one obtains integral formulas for the logarithm of the amplitude of the back end reflection factor. If the overall network is specified so that it satisfies the general realizability requirements and in addition meets the limits on $\ln \left| \frac{1}{S_{22}(j\omega)} \right|$ imposed by the integral formulas, then the prescribed load can always be separated from the overall network leaving a physically realizable equalizer 2-port. A statement of these realizability conditions in the form of a theorem essentially as given by La Rosa⁽⁴⁾ (5)

is as follows:

Theorem 1

The necessary and sufficient conditions that a scattering matrix $[S(p)]$ ($p = \sigma + j\omega$) represent an overall network composed of an equalizer in tandem with a prescribed lossless 2-port (the reactance 2-port portion of the prescribed load) is:

i.e. (a) Matrix $[S]$ should be realizable
 $I - S(j\omega)^* S(j\omega)$ must be the matrix of a positive definite or semi-definite hermitian form, with $[S]$ symmetric and its elements rational functions of p with real coefficients, and no right half plane poles.⁽¹³⁾

(b) Right hand and boundary zeros of transmission of the load must appear in the transmission factor $S_{12}(p)$ of the overall network with at least the same multiplicity.

(c)* A set of integral restrictions on $\ln \left| \frac{1}{S_{22}(j\omega)} \right|$ of the form

$$\int_0^{\infty} f_1(\omega) \ln \left| \frac{1}{S_{22}(j\omega)} \right| d\omega = K_1 \quad (1)$$

must be simultaneously satisfied at all the zeros of transmission of the load. Each n^{th} order zero contributes N_1 integral equations with

$$N_1 = \begin{cases} n & \text{for a zero at zero or infinity} \\ 2n & \text{for a zero on } j\omega \\ 2n - n_0 & \text{for a right hand zero on the real axis} \\ 4n - 2n_0 & \text{for a conjugate pair of zeros in the right half plane.} \end{cases}$$

n_0 is the order of any right half plane zero of load transmission coincident with a zero of back end load reflection factor, i.e. of the reactance 2-port portion of the load.

This theorem can be applied in a direct and simple fashion to a variety of equalization problems involving a load containing dissipative elements. Special consideration will be required for problems involving a purely reactive load.

* The weighting functions $f_1(\omega)$ are tabulated by Fano⁽³⁾. The K_1 are related to the Taylor coefficients of the load at the zeros of transmission and are also tabulated in the same reference.

III Power Transfer Equalization

The application of the theorem given in the preceding section requires a determination of the relationships between the transfer function which is to be optimized and the reflection factor amplitude of the overall network $|S_{22}(j\omega)|$. The integral equations can then be used to determine "gain-bandwidth" type of restrictions on the equalizer. Part (a) of the theorem contains the necessary information for relating the reflection factor function $S_{22}(j\omega)$ to the equalization response of the overall network. In the case of power transfer from a generator with unit internal impedance to a load (represented in Darlington form), the insertion power gain at real frequencies of the overall network normalized to the available generator power is

$$\frac{P_L}{P_C} = |S_{12}(j\omega)|^2 \quad (2)$$

P_L is the power delivered to the load, P_C is the available generator power $(\frac{V_G^2}{4R_G})$, where V_G

is the generator open-circuit voltage) and $|S_{12}(j\omega)|$ is the amplitude of the voltage transmission coefficient of the overall network. Since this power transfer function is an element of the scattering matrix, it is directly related to $|S_{22}(j\omega)|$ by the general realizability constraints of part (a) of the theorem. La Rosa(5) (6) has shown that this portion of the theorem leads to the following necessary requirement on $|S_{12}(j\omega)|$ for an equalizer which maximizes the scale factor of power function when the shape is specified.

$$\frac{|S_{12}(j\omega)|^2}{|S_{22}(j\omega)|^2} \geq \frac{(1 - |S_{22}(j\omega)|^2)(1 + |S_{11}(j\omega)|^2)}{|S_{11}(j\omega)|^2} \quad (3)$$

In equation 3, $|S_{11}(j\omega)|$ is a specified input reflection factor amplitude function which sets the tolerance on input mismatch. In the special case that the equalizer is lossless, $|S_{11}(j\omega)| = |S_{22}(j\omega)|$ and equation 3 becomes

$$|S_{12}(j\omega)|^2 = 1 - |S_{22}(j\omega)|^2 \quad (4)$$

(lossless equalizer)

In another special case where the input mismatch is zero i.e. $S_{11}(p) \equiv 0$ equation 3 reduces to

$$|S_{12}(j\omega)|^2 = 1 - |S_{22}(j\omega)|^2 \quad (5)$$

(matched equalizer)

Equations 3,4,5 give the desired relations between $|S_{22}(j\omega)|$ and the power transfer function $|S_{12}(j\omega)|^2$. The integral equations, 1, may therefore be expressed in terms of the power transfer function and solved to obtain the maximum scale factor. The details of this procedure as well as examples are given

in references 4,5,6. The solution of the equations always gives a unique maximum-gain scale factor for a prescribed shape of power transfer function and this cannot be exceeded by any physical equalizer. In the case where the equalizer is to produce a flat pass band with zero gain outside this band, the solution for the scale factor is particularly simple and is found directly in terms of a minimum constant value of $|S_{22}(j\omega)| = |S_{22}|$ over the prescribed band with $|S_{22}(j\omega)| = 1$ elsewhere. In this case it is interesting to compare the optimum lossless and matched equalizers using equations 4 and 5

$$10 \log \frac{|S_{12}|^2 \text{ (lossless)}}{|S_{12}|^2 \text{ (matched)}} = 10 \log(1 + |S_{22}|^2) \quad (6)$$

< 3 db.

since $|S_{22}| \leq 1$. In any practical design of a flat power equalizer $|S_{22}|$ is considerably less than one so that over a specified band the gain of an optimum matched design is much closer to the gain of an optimum lossless equalizer than the outside limit of 3 db. given by equation 6.

IV Voltage Transfer Equalization of General Dissipative Load

a) Integral constraints for voltage transfer

The theorem given in section 2 may be applied to the problem of voltage equalization provided the voltage transfer function can be related to the scattering coefficients of the overall network. The voltage transfer function at real frequencies is taken to be

$$e = \frac{V_2}{V_1} \quad (7)$$

where $|V_2|$ is the amplitude of the voltage appearing across the load and $|V_1|$ is the open circuit (fixed) voltage amplitude of a normalized generator with unit internal impedance (pure resistance).

The generator produces a voltage V'_2 across the one ohm resistance in the Darlington representation of the load as a reactance 2-port terminated in unit resistance as shown in fig. 1. This voltage is related to V_1 by the voltage scattering function $S_{12}(p)$ of the overall network (fig.1). Thus

$$V'_2 = S_{12}(p) \frac{V_1}{2} \quad (8)$$

Since the power at the input to the load is the same as that delivered to the one ohm resistor in the Darlington representation

$$|V_2|^2 g(\omega) = |V'_2|^2 \quad (9)$$

where $g(\omega)$ is the input conductance of the load. Combining equations 7,8 and 9

$$e^2 = \left| \frac{V_2}{V_1} \right|^2 = \frac{|S_{12}(j\omega)|^2}{4 g(\omega)} \quad (10)$$

Since $|S_{12}(j\omega)|^2 \ll 1$ it is immediately clear that in any physical network

$$e^2 \ll \frac{1}{4g(\omega)} \quad (11)$$

In equation 10 $g(\omega)$ is specified by the load alone and e is directly proportional to $|S_{12}(j\omega)|$. Thus a necessary requirement for maximum voltage transfer is to maximize $|S_{12}(j\omega)|$ consistent with the general theorem on realizability given earlier. This is accomplished precisely as in the power transfer problem when $|S_{12}(j\omega)|$ and $|S_{11}(j\omega)|$ (prescribed) are related to $|S_{22}(j\omega)|$ by equations 3, 4, and 5. The general equation for optimum voltage transfer may then be written as

$$e^2 = \frac{(1 - |S_{22}(j\omega)|)(1 + |S_{11}(j\omega)|)}{4g(\omega)} \quad (12)$$

$$|S_{22}(j\omega)| \geq |S_{11}(j\omega)|$$

The lossless and matched cases are then given as

$$e^2 = \frac{1 - |S_{22}(j\omega)|^2}{4g(\omega)} \quad (\text{lossless}) \quad (13)$$

$$e^2 = \frac{1 - |S_{22}(j\omega)|}{4g(\omega)} \quad (\text{matched}) \quad (14)$$

The integral equations for the latter two cases using equation 1 are:

$$\frac{1}{2} \int_0^{\infty} f_1(\omega) \ln \left(\frac{1}{1 - 4e^2(\omega)g(\omega)} \right) \cdot d\omega = K_1 \quad (\text{lossless}) \quad (15)$$

$$\int_0^{\infty} f_1(\omega) \ln \left(\frac{1}{1 - 4e^2(\omega)g(\omega)} \right) \cdot d\omega = K_1 \quad (\text{matched}) \quad (16)$$

The weighting functions $f_1(\omega)$ and the parameters K_1 are those tabulated in reference 3.

The only difference in form for the integral constraints in the two special cases is the factor 1/2. However, since e appears under the integral sign the solution for maximum scale factor is generally formed from a transcendental equation so that there is no direct relation between the voltage gain of lossless and matched equalizers even in the flat transfer case. In this latter case

$$e^2(\omega) = \begin{cases} C^2 & \omega_1 \leq \omega \leq \omega_2 \\ 0 & 0 \leq \omega < \omega_1, \omega_1 < \omega \leq \infty \end{cases} \quad (17)$$

where C is the voltage gain constant to be maximized. The integrals 15 and 16 are then

$$\frac{1}{2} \int_{\omega_1}^{\omega_2} f_1(\omega) \ln \left(\frac{1}{1 - 4C^2g(\omega)} \right) \cdot d\omega = K_1 \quad (\text{lossless}) \quad (18)$$

$$\int_{\omega_1}^{\omega_2} f_1(\omega) \ln \left(\frac{1}{1 - 4C^2g(\omega)} \right) \cdot d\omega = K_1 \quad (\text{matched}) \quad (19)$$

In effect the problem of a flat voltage equalizer reduces to the solution of a power transfer problem where a non-flat power gain curve shape is specified.

b. Example - Flat Voltage Equalizer for R-L load

As an example of a voltage equalizer problem consider the case of a load consisting of the series combination of coil L and resistor R . The voltage transfer characteristic is to be a high pass one specified by:

$$e(x) = \begin{cases} C & x \leq x_c \\ 0 & 0 \leq x < x_c \end{cases} \quad (20)$$

where x is a normalized frequency variable and x_c is its cut off value:

$$x = \omega \frac{L}{R} \quad (21)$$

$$x_c = \omega_c \frac{L}{R} \quad (22)$$

The load conductance is

$$g(x) = \frac{1}{R} \frac{1}{1 + x^2} \quad (23)$$

Since the load has only a simple zero of transmission at infinity the weighting function $f(\omega)$ is unity and the integration constant K_1 is given by⁽³⁾:

$$K_1 \leq \frac{\pi R}{L} \quad (24)$$

The possibility of attaining the equal sign in equation 24 is dictated by equation 11. If equation 23 is substituted in that equation then an upper bound is set on $e^2(x)$ for any value of x :

$$e^2(x) \leq \frac{R(1 + x^2)}{4} \quad (25)$$

The permissible value of voltage gain increases with x , and thus for flat response

$$C^2 \leq \frac{R(1 + x_c^2)}{4} \quad (26)$$

since the lowest permissible gain occurs at cut-off. The upper limit for $K = \frac{\pi R}{L}$ can only be attained if the value of C required in the integral equations does not violate Eq. 26. The integral relations given by Eq. 18 and 19 become:

$$\frac{R}{L} \int_{x_c}^{\infty} \ln \frac{x^2 + 1}{x^2 + a^2} dx \leq \frac{\pi R}{L} \quad (27)$$

where

$$A = \begin{cases} 1 & \text{Matched case} \\ \frac{1}{2} & \text{Lossless case} \end{cases} \quad (28)$$

$$\text{and } a^2 = 1 - \frac{4C^2}{R} \quad (29)$$

For $a^2 \gg 0$ integration of Eq. 27 gives

$$\pi(1-a) - \chi_c \ln \frac{1 + \chi_c^2}{a^2 + \chi_c^2} - 2 \tan^{-1} \chi_c + 2a \tan^{-1} \frac{\chi_c}{a} = \pi/A, \quad a^2 \gg 0 \quad (30)$$

The equal sign is used in order to determine whether the value of a (hence C by Eq. 29) exceeds the limit of Eq. 26. When $A = 1$, the only real solution for a^2 in equation 30 occurs when $\chi_c = 0$. In that case:

$$a = a^2 = 0, \quad C = \frac{R}{2} \quad (31)$$

The limit of Eq. 26 for $\chi_c = 0$ is also $c = R$, so that for flat transfer over $0 \leq \chi_c < \infty$, the maximum value permitted by the integral constraint for a matched equalizer can be obtained and this flat gain is precisely the d.c. gain. A lossless equalizer would give no gain advantage, since the solution of Eq. 30 with $A = 1/2$ results in a value of C exceeding that permitted by Eq. 26.

For values of $\chi_c > 0$, negative values of a^2 are required to satisfy Eq. 30. Under these conditions the transcendental equation becomes:

$$\pi \left[\chi_c \cdot \ln \frac{\chi_c^2 + 1}{\chi_c^2 - b^2} + b \ln \frac{\chi_c - b}{\chi_c + b} + 2 \tan^{-1} \chi_c \right] = \frac{\pi}{A} \quad (32)$$

$$\text{Where } b^2 = -a^2 = \frac{4C^2}{R} - 1 \quad (33)$$

For values of $0 \leq \chi_c \leq 1.9$, the solution of Eq. 32 for C always exceeds that permitted by Eq. 26 for both values of A (lossless and matched cases). In this region, where the value of $\frac{\omega L}{R}$ is small at cut-off, the optimum flat gain is given by Eq. 26

$$C = \sqrt{\frac{R(1 + \chi_c^2)}{2}}, \quad 0 \leq \chi_c \leq 1.9 \quad (34)$$

and the lossless equalizer gives no advantage in gain over the matched equalizer. For the medium range of $1.9 < \chi_c \leq 4.8$, the flat gain of a matched equalizer as obtained from Eq. 32

is less than that given by Eq. 26, while the gain of the lossless equalizer is still limited by Eq. 26. Finally for the high range $4.8 < \chi_c \leq \infty$, both lossless and matched equalizers have gains limited by the solution of Eq. 31.

When χ_c is very large Eq. 32 is approximated very well by:

$$b^2 + 1 = 2 + \frac{\pi \chi_c}{A}, \quad \chi_c \gg 1 \quad (35)$$

Using Eq. 33, the optimum gain in this case is

$$C = \left. \begin{aligned} & \frac{\sqrt{R}}{2} \sqrt{2 + \frac{\pi \chi_c}{A}} \\ & \approx \frac{1}{2} \frac{\pi L}{A} \omega_c \end{aligned} \right\} \chi_c = \frac{\omega_c L}{R} \gg 1 \quad (36)$$

Thus as the load resistance becomes negligible compared to the load reactance at cutoff, the ratio of maximum flat gain of lossless ($A = 1/2$) and matched ($A = 1$) equalizers becomes:

$$\frac{C(\text{lossless})}{C(\text{matched})} = \sqrt{2} \quad (37)$$

The comparison of performance of lossless and matched equalizers for a flat high pass gain characteristic is summarized in graphical form in Fig. 2. The heavy line is the bounding curve defined by Eq. 26, and the dashed curves show the maximum flat gain of the lossless and matched equalizers as prescribed by the integral constraints.

V Voltage Transfer Equalization of Reactive Load

The case of optimum voltage transfer from a finite generator to a reactive load is an important practical problem. However, the basic realizability theorem quoted in section 2 is not readily applicable. Accordingly this equalization problem will be treated in a somewhat different way though the general point of view outlined in Section 1 will still be used. The optimum voltage equalization of an arbitrary lossless termination has not been considered in any complete fashion elsewhere, so that some details of the derivation of the realizability criteria (general constraints plus load constraints) will be given here. This will also serve to further illuminate the basis for the realizability theorem 1 of Section 2 since the two derivations parallel each other. It will be seen that whereas the scattering coefficients were a natural tool for handling the equalization of a dissipative load, the open circuit impedance elements are more directly applicable to the reactive load problem.

In Fig. 3a a finite generator of voltage V_1 and unit internal resistance is shown driving an equalizer terminated in an arbitrary reactance. The amplitude ratio of output load voltage to open circuit generator voltage is

$$\rho = \frac{|V_2|}{|V_1|} = |\bar{Z}_{12}(j\omega)| \quad (\text{general}) \quad (38a)$$

where $|\bar{Z}_{12}(j\omega)|$ is the amplitude of open circuit transfer impedance of an overall network shown in Fig. 3b, consisting of the equalizer shunted at the generator side by a 1 ohm resistor and at the output side by the reactive load.

If the equalizer is designed to produce an input match, then the voltage ratio may be written as:

$$\rho = \frac{|V_2|}{|V_1|} = \frac{1}{2} |Z_{12}(j\omega)| \quad (\text{matched equalizer}) \quad (38)$$

where $|Z_{12}(j\omega)|$ is the amplitude of the open circuit transfer impedance of the matching equalizer plus lossless load. (Generator resistance is not included.)

If the equalizer is lossless then in Fig. 3b with excitation at the load side of the overall network, the power delivered here is equal to that dissipated in the 1 ohm shunting resistor. Thus

$$\rho^2 = |\bar{Z}_{21}(j\omega)|^2 = |\bar{Z}_{12}(j\omega)|^2 = \bar{R}_{22}(j\omega) \quad (\text{lossless equalizer}) \quad (39)$$

where $\bar{R}_{22}(j\omega)$ is the real part of the open circuit output driving point impedance of the overall network on $j\omega$:

$$\bar{R}_{22}(j\omega) = \text{Re } \bar{Z}_{22}(j\omega) \quad (40)$$

Since the equalization functions to be optimized (Eq. 38, 39) involve the open circuit impedance parameters, the requirements on overall network physical realizability will be expressed in terms of these parameters. The necessary and sufficient conditions that an open circuit impedance matrix correspond to a physical 2-port is that the matrix be positive real. That is

$$a.) \quad R_{11}(j\omega) R_{22}(j\omega) - R_{12}^2(j\omega) \geq 0, \quad R_{11}(j\omega) \geq 0 \quad (41)$$

b.) The open circuit impedances have no poles in the right half plane.

c.) Impedance element poles on the boundary be simple and the residues of $Z_{11}(p)$ and $Z_{22}(p)$ at these poles satisfy

$$a_{11} a_{22} - a_{12}^2 \geq 0 \quad a_{11} \geq 0$$

Eq. 41 is the major physical constraint in the equalizer problem since if this is satisfied the remaining two requirements can in general be met by suitable design of the actual equalizer network.

The constraints of the load can be easily established by observing that the reactive load impedance $Z_L(p)$ is in parallel with the impedance seen looking in at the back of the equalizer. Thus

$$Z_{22} = \frac{z_{22} Z_L}{z_{22} + Z_L} \quad (42)$$

$$\bar{Z}_{22} = \frac{\bar{z}_{22} Z_L}{\bar{z}_{22} + Z_L} \quad (43)$$

where Z_{22} is the back end open circuit driving point impedance for the equalizer plus load with the 1 ohm resistor at the input removed, and \bar{Z}_{22} is a similar quantity for the overall network of Fig. 3b. (z_{22} is backend equalizer impedance)

Any zero of the reactive load must be simple and occur on $j\omega$. Further if Z_L is expanded in a power series at this zero, the first non vanishing Taylor coefficient is positive.* Inspection of Eq. 42 and 43 show that the back end impedance Z_{22} or \bar{Z}_{22} must vanish at this point and in the vicinity of the zero, $p_1 = j\omega_1$:

$$Z_{22}(p) = \bar{Z}_{22}(p) = Z_L(p) = a_1(p-p_1) \quad (44)$$

$p \rightarrow p_1$

Equation 27 is in fact entirely independent of the equalizer and its input termination.

The requirement that the back end impedance of the overall network be constrained by Eq. 27 can be expressed in integral form by using the Cauchy formulas for Taylor coefficients. These take the form in the present instance of:

$$\frac{1}{2\pi j} \oint g_1(p) Z_{22}(p) dp = a_1 \quad (45)$$

where the closed path of integration is along $j\omega$ (avoiding by small semicircular indentations any boundary poles) and is completed around the semicircle of infinite radius enclosing the right half p plane. $g_1(p)$ are the weighting functions to select the appropriate Taylor coefficient of the load, a_1 . If the contributions of the small indentations are accounted for, and $g_1(p)$ is modified slightly to be even in ω along $j\omega$, Eq. 45 may be written in terms of $R_{22}(j\omega)$ as follows:

$$\int_0^{\infty} \frac{1}{\omega^2} R_{22}(j\omega) d\omega = \frac{\pi a_0}{2} - \pi \sum_m \frac{b_m}{\omega_m^2} \quad (\text{zero at } p=0) \quad (46)$$

$$\int_0^{\infty} R_{22}(j\omega) d\omega = \frac{\pi a_{\infty}}{2} - \pi \sum_m b_m \quad (\text{zero at } p=\infty) \quad (47)$$

* This follows from Foster's reactance theorem. (15)

$$\int_0^b \frac{1}{(\omega^2 - \omega_1^2)^2} R_{22}(\omega) d\omega = \frac{\pi A_1}{4 \omega_1^2}$$

$$-\pi \sum_m \frac{b}{(\omega_m^2 - \omega_1^2)^2} \quad (\text{zero at } p = \pm j\omega_1)$$

$$\omega_m \neq \omega_1 \quad (48)$$

In every case b_m is the residue of any Z_{22} poles on $j\omega$ and is always positive. Such poles can only reduce the permissible integration constant. The only other possibility of reducing the right hand sides of Eq. 46, 47, 48 is for Z_{22} (or \bar{Z}_{22}) the equalizer to have a zero coincident with Z_L . Since these relations must all be satisfied simultaneously by Z_{22} (or \bar{Z}_{22}), it may be necessary to introduce elements in the equalizer which cause coincident zeros and/or additional poles on $j\omega$. Another point which should be mentioned here is that referring to equation 44, equations 45, 46, 47, 48 are valid both for the barred and unbarred quantities.

Finally since any zero of the load is a zero of voltage transfer, the following physical realizability theorem may be stated:

Theorem 2

The necessary and sufficient conditions that a combined network consisting of a two port and a prescribed reactive termination be physically realizable is

- The combined network must have a positive real open circuit impedance matrix.
- All zeros of the reactive load impedance must be contained in the open circuit transfer impedance of the combined network. (The zeros are all simple.)
- The integral constraints given by Eq. 46, 47, 48 and summarized by the form

$$\int_0^\infty f_1(\omega) R_{22}(\omega) d\omega = K_1 \quad (49)$$

must be simultaneously satisfied by the back end resistance of the combined network at each zero of the load.

The similarity between this theorem and theorem 1 is obvious.

VI Flat gain equalization of a reactive load

a) Form of integral constraints

As an example of the application of theorem 2, the case of flat voltage transfer for a pro-

totype low pass* equalizer defined by:

$$P = \left| \frac{V_2}{V_1} \right| = |\bar{Z}_{12}(\omega)| = \begin{cases} C & \text{for } 0 \leq \omega \leq 1 \\ 0 & \text{for } 1 < \omega \leq \infty \end{cases} \quad (50)$$

will be examined.

In the case of the matched equalizer condition a of theorem 2 leads to

$$R_{22}(\omega) - R_{12}^2(\omega) \geq 0 \quad (\text{Matched equalizer}) \quad (51)$$

This is obtained from equation 41 with $R_{11}(\omega) = 1$, $0 \leq \omega \leq \infty$.

$$R_{12}(\omega) = Z_{12}(\omega) \cos(\varphi(\omega) + \theta(\omega)) \quad (52)$$

Where φ is the minimum phase characteristic of the combined network (equalizer and load), and $\theta(\omega)$ any phase characteristic obtained with a tandem combination of unit voltage transfer networks, considered as part of the equalizer (i.e. the latter need not be a minimum phase network.)

The function $\varphi(\cdot)$ is directly found graphically or analytically from the shape of the $Z_{12}(\omega)$ amplitude function and is not affected by the scale factor on this function. $\theta(\omega)$ is independent of $\varphi(\omega)$ and is continuous with positive slope. ⁽¹⁾ If equation 52 is substituted into 51, it is clear that a necessary condition for maximum $Z_{12}(\omega)$ is that the equal sign be used. Thus

$$R_{22}(\omega) = |Z_{12}(\omega)|^2 \cos^2(\varphi + \theta) = R_{12}^2(\omega) \quad (53)$$

This means that the combined matched equalizer-load network will require no more than 1 resistor. ⁽¹⁴⁾ Suppose that $|Z_{12}(\omega)|$ is represented as

$$|Z_{12}(\omega)| = \begin{cases} C & 0 \leq \omega \leq 1 \\ \frac{C}{\omega^k} & 1 < \omega \leq \infty \end{cases} \quad (54)$$

(k a positive integer)

where

$$C = 2 \bar{C} \quad (55)$$

As $k \rightarrow \infty$, the transfer characteristic defined by equation 54 approaches the gain shape given by equation 50 (it differs by the factor 1/2).

The minimum phase characteristic defined by equation 54 is well approximated* by: ⁽¹⁾

* Identical forms for the high pass and band pass cases are obtained by a similar analysis.

^{**}At this point there is a certain degree of non-rigidity, but in the final evaluation of the integrals the exact dependence of phase on ω is unimportant provided it is proportional to k.

$$\varphi(m) = \begin{cases} -k \frac{2w}{\pi} & 0 \leq w \leq 1 \\ -k \frac{\pi}{2} & 0 < w \leq \infty \end{cases} \quad (56)$$

Thus the integral equation 49 of theorem 2 for the matched flat voltage equalizer is:

$$C^2 \int_0^1 f_1(m) \cos^2 \left(-\frac{2w}{\pi} k + \theta(m) \right) dm + C^2 \int_1^\infty \frac{f_1(m)}{m^k} \cos^2 \left(-\frac{k\pi}{2} + \theta(m) \right) dm = K_1$$

The second integral is easily seen to approach zero as $k \rightarrow \infty$. Expanding the cosine term, the first integral may be written as:

$$\frac{1}{2} \int_0^1 f_1(m) dm + \frac{1}{2} \int_0^1 (f_1(m) \cos 2\theta(m)) \cdot \cos 2k \frac{2w}{\pi} dw + \frac{1}{2} \int_0^1 (f_1(m) \sin 2\theta(m)) \cdot \sin 2k \frac{2w}{\pi} dw$$

The second and third integrals in the above expression are merely k th order Fourier coefficients for a periodic function equal to $(f_1(m) \cos 2\theta(m))$ or $(f_1(m) \sin 2\theta(m))$ over a finite interval, hence these integrals go to zero as k becomes infinite. The final representation of the integral constraint for a matched equalizer with the low pass flat voltage gain of equation 50 is therefore:

$$2 \bar{C}^2 \int_0^1 f_1(m) dm = K_1 \quad (\text{matched case}) \quad (57)$$

Letting

$$\int_0^1 f_1(m) dm = A_1 \quad (58)$$

equation 57 may then be written

$$\bar{C}^2 = \frac{K_1}{2 A_1} \quad (\text{Matched}) \quad (59)$$

This equation is independent of whether minimum or non-minimum phase networks are used.

For a lossless equalizer, equation 39 may be used directly in connection with equation 50 and placed into the expression of theorem 2. Thus the lossless equalizer with flat transfer characteristics given by equation 50 must satisfy:

$$\int_0^\infty f_1(m) \bar{R}_{22}(m) dm = \int_0^\infty f_1(m) |\bar{Z}_{12}(m)|^2 dm =$$

$$\bar{C}^2 \int_0^1 f_1(m) dm \quad (60)$$

$$\text{or } \bar{C}^2 = \frac{K_1}{A_1} \quad (\text{Lossless case}) \quad (61)$$

and the ratio of optimum matched and lossless equalizer gains for the flat transfer case and reactive load is:

$$\frac{\bar{C}(\text{lossless})}{\bar{C}(\text{matched})} = \sqrt{2} \quad (62)$$

b) Example of a C-LC load.

As an example of the application of the material given above suppose it is required to design a flat voltage equalizer over the band $0 \leq w \leq 1$ for a three element load. The load consists of a capacitor C' in shunt with a series circuit of coil L and condenser C .

This load impedance has zeros at $p = \infty$, and $p = \pm j\omega_1$ with $\omega_1 = \frac{1}{LC}$.

The Taylor coefficients at these zeros are:

$$(p = \infty) \quad a_\infty = \frac{1}{C'} \quad (63)$$

$$(p = \pm j\omega_1) \quad a_1 = \frac{2L}{C} \quad (64)$$

Referring to equations 47 and 48 and neglecting the pole residue terms:

$$K_1 = \frac{\pi a_\infty}{2} = \frac{\pi}{2 C'} \quad (65)$$

$$K_2 = \frac{\pi a_1}{4 \omega_1^2} = \frac{\pi}{2} L^2 \quad (66)$$

For the zero at infinity the weighting function is unity and

$$A_1 = \int_0^1 dm = 1 \quad (67)$$

For the zero at $\pm j\omega_1$, the weighting function is $f_2(m) = \frac{1}{(w^2 - \omega_1^2)^2}$ and

$$A_2 = \int_0^1 \frac{dw}{(w^2 - \omega_1^2)^2} = \frac{1}{2\omega_1^2(\omega_1^2 - 1)} + \frac{1}{4\omega_1^3} \cdot \ln \frac{\omega_1 + 1}{\omega_1 - 1} \quad (68)$$

* Presumed outside of pass-band.

Consider a specific numerical case with

$$C' = \frac{1}{2}$$

$$L = C = \frac{1}{\sqrt{2}} \quad (69)$$

For these conditions the various constants may be evaluated to give:

$$\frac{K_1}{A_1} = \pi \quad (70)$$

$$\frac{K_2}{A_2} = \frac{\pi/4}{0.35} = \frac{\pi}{1.40} \quad (71)$$

The only way to satisfy these constraints simultaneously is to reduce the value of K_1 by using parasitic elements in the equalizer. This can be done for example by use of a shunt condenser across the equalizer output and in parallel with the load of value $C'' = 0.2$. As a result the maximum gain of this equalizer is not limited by the capacitance C' , but by the series LC circuit and the maximum value of this gain for a low pass matched equalizer is

$$\bar{C} = \sqrt{\frac{\pi}{2.80}} \quad (\text{matched}) \quad (72)$$

and for a lossless equalizer the maximum gain is

$$\bar{C} = \sqrt{\frac{\pi}{1.40}} \quad (\text{lossless}) \quad (73)$$

It is up to the designer to decide whether a matched input to the equalizer system is worth sacrificing for a 40% increase in gain.

c. Design of a finite equalizer network

In the previous discussion ideal characteristics were assumed for the equalizer amplitude response. The voltage gain was presumed absolutely flat in the pass band, and the cut-off was taken as infinitely sharp. In order to apply theorem 2 to the design of a finite network, a gain characteristic may be assumed in analytic form (i.e. an even rational function of ω which is always positive). The analysis in such a case involves determination of the R_{22} characteristic and its substitution in the integral equations to determine the maximum gain scale factor. To illustrate this procedure a simple example will be given for the voltage equalization and match of a unit capacitive termination*. A Butterworth type of low pass response is assumed for a matched equalizer

$$|Z_{12}(j\omega)|^2 = \frac{C^2}{1 + \omega^4} \quad (74)$$

* This example is taken from a thesis by L. Norde for the degree M.E.E. at the Polytechnic Institute of Brooklyn, June 1953. (7)

The amplitude of $Z_{12}(j\omega)^2$ is down by $\frac{1}{2}$ at $\omega = 1$ which is considered the normalized cut-off frequency. C is a gain constant whose maximum value is to be determined by applying theorem 2.

Equation 74 may be factored to give the open circuit transfer impedance $Z_{12}(p)$ of the required combined matched equalizer-load network. This function must have no right half plane poles.

$$Z_{12}(p) = \frac{C}{p^2 + \sqrt{2}p + 1} \quad (75)$$

The impedance $Z_{12}(p)$ has the zero of the load at infinity as prescribed by theorem 2 because of the particular choice of amplitude function in equation 74.

The real component of $Z_{12}(p)$ along $p = j\omega$ is

$$R_{12}(j\omega) = C \frac{1 - \omega^2}{1 + \omega^4} \quad (76)$$

and applying equation 53 for the optimum matched equalizer

$$R_{22}(j\omega) = R_{12}^2(j\omega) = C^2 \frac{(1 - \omega^2)^2}{(1 + \omega^4)^2} \quad (77)$$

The complex impedance $Z_{22}(p)$ may be obtained by expanding $R_{22}(j\omega)$ in partial fractions as described on pages 204-205 of reference 1. The result* is

$$Z_{22}(p) = C^2 \frac{0.353 p^3 + p^2 + 1.065 p + 1}{p^4 + 2.82 p^3 + 4p^2 + 2.82 p + 1} \quad (78)$$

This function has the load zero at infinity as prescribed by theorem 2.

The integral constraint (equation 47) gives for unit capacitive load:

$$\int_0^{\infty} R_{22} d\omega = C^2 \int_0^{\infty} \frac{(1 - \omega^2)^2}{(1 + \omega^4)^2} \cdot d\omega = \frac{\pi}{2} \quad (79)$$

If the definite integral is evaluated, the gain factor C is found to be

$$C = 1.68 \quad (80)$$

or

$$\bar{C} = \frac{C}{2} = 0.84 \quad (\text{matched case}) \quad (81)$$

* In general if the resulting $Z_{22}(p)$ function does not contain the required load zeros, it is necessary to form new rational functions for $|Z_{12}(j\omega)|^2$ (possibly by the use of common numerator and denominator factors) until this requirement is satisfied.

The maximum gain for the completely ideal low pass characteristic (in the matched case) obtained from equation 59 is

$$\bar{c} \text{ (ideal)} = 0.89 \text{ (matched case)} \quad (82)$$

so that the simple characteristic of equation 74 is a reasonable compromise.

The final matched equalizer plus load is specified by equations 75, 73 and

$$Z_{11}(p) = 1$$

The resultant network is shown in Fig. 4 a. Observe that the load capacitance is removable leaving a physical equalizer, and only 1 resistor appears in the network.

The design of a lossless equalizer from a specified $|\bar{Z}_{12}(w)|^2$ characteristic which satisfies theorem 2 is straightforward. The equalizer network is merely synthesized from the back end as a reactance 2-port terminated in a 1 ohm resistor by the Darlington procedure for the special case of an infinite impedance generator.⁽¹⁰⁾ The 1 ohm resistor (input generator impedance) and load are then removed, and the remaining lossless network is the required equalizer.

For the example of equation 74 with

$$|\bar{Z}_{12}(w)|^2 = \frac{\bar{c}^2}{1+w^2} = R_{22}(w) \quad (83)$$

The integral requirement gives

$$\int_0^{\infty} \frac{\bar{c}^2}{1+w^2} dw = \frac{\pi}{2}$$

and*

$$\bar{c} = 2^{0.25} = 1.19 \text{ (lossless equalizer)} \quad (84)$$

The network is shown in Fig. 4b.

References

1. Bode, H.W. - "Network Analysis and Feedback Amplifier Design" (Book) D. Van Nostrand Co. N.Y., 1945.
2. Bode, H.W. - U.S. Patent 2,337,965, 1942.
3. Fano, R.M. - "Theoretical Limitations on the Broadband Matching of Arbitrary Impedances" - Jour. of E'kln Inst. v. 249 n 1, Jan. 1950; v.249 n 2, Feb. 1950.
4. La Rosa, R. - "Broadband Dissipative Matching Networks" D.E.E. Thesis Poly. Inst. of B'klyn, May 1953.
5. La Rosa, R. and Carlin, H.J. - "A General Theory of Wideband Matching" - Microwave Research Inst., Poly. Inst. of B'klyn, Report R-308-53 PIB-247 Feb. 1953 (Submitted for Publication to Jour. of Math & Phys.)
6. Carlin, H.J. and La Rosa, R. - "Broadband Reflectionless Matching" Proc. of Sym. on Modern Net. Syn. v 1, Poly. Inst. of B'klyn, April 1952.
7. Norde, L. - "Broadband Matching with Maximum Gain and Reactive Load" M.E.E. Thesis, Poly. Inst. of B'klyn, June 1953.
8. Wheeler, H.A. - "The Maximum Speed of Amplification in a Wideband Amplifier" Wheeler Monographs v 1, 1953. (Original publication 1949)
9. Hansen, W.W. - "On Maximum Gain Bandwidth Product in Amplifiers" Jour. App. Phys. v 16, Sept. 1945.
10. Darlington, S. - "Synthesis of Reactance 4-Poles" - Jour. of Math & Phys. v 18, pp 257-353 Sept. 1939.
11. Carlin, H.J. and La Rosa R. - "On the Synthesis of Reactance 4-Poles" Jour. of Applied Phys. v 24, n 10, Oct. 1953.
12. Laemmel, A.E. - "Scattering Matrix Formulation of Microwave Networks" Proc. of Sym. on Modern Net. Syn. v 1, Poly. Inst. of B'klyn, April 1952.
13. Belevitch, V. - "Synthese des Reseaux Electriques Passifs a n-Paires de Bornes de Matrice de Repartition Predeterminee", Annales des Telecommunications, v 6, n 11, p 302 (Nov. 1951).
14. Gewertz, C.M. - "Synthesis of a Finite, Four-Terminal Network from its Prescribed Driving-Point Functions and Transfer Function", Jour. Math & Phys., v 12 (1932-33) pp 1-257. also "Network Synthesis" (Book) Williams and Wilkins, Baltimore, 1933.
15. Foster, R.M. - "A Reactance Theorem", Bell System Technical Jour., April 1924.

* This constraint could have been directly obtained from the network synthesis by choosing \bar{c} to give the prescribed load capacitance.

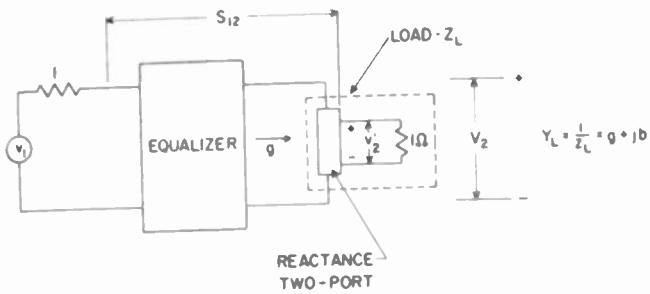
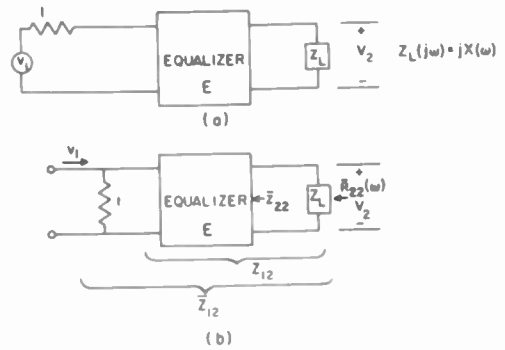


Fig. 1 - Voltage transfer - general load.



$$\frac{V_1}{V_2} = \bar{Z}_{12} = \frac{1}{2} Z_{12} \text{ (MATCHED CASE)}$$

$$|\bar{Z}_{12}|^2 = R_{22} \text{ IF } E \text{ LOSSLESS}$$

Fig. 3 - Voltage transfer - reactive load.

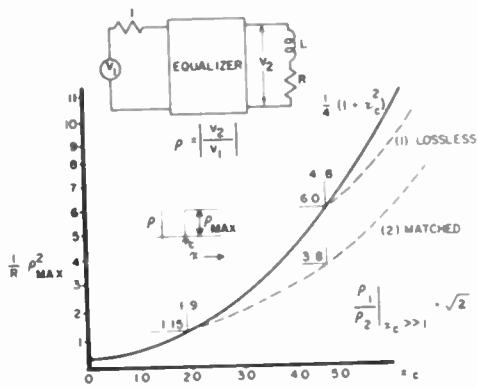
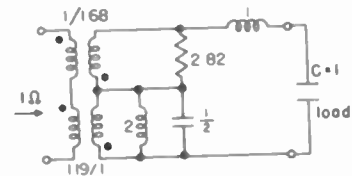
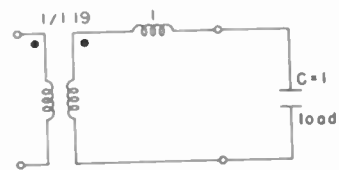


Fig. 2 - Voltage equalizer R-L load.



(a) MATCHED EQUALIZER A = 0.84



(b) LOSSLESS EQUALIZER A = 1.19

Fig. 4
Equalizers for capacitive load and
voltage gain = $\frac{A}{\sqrt{1+W^4}}$

SYNTHESIS OF RESISTIVELY - TERMINATED
RLC LADDER NETWORKS*

Er-Chun Ho and DeForest L. Trautman
Department of Engineering
University of California
Los Angeles, California

Many equalizer problems reduce to the design of a network with specified transmission characteristics between given resistive terminations. In this note, we shall present a new technique by which one can synthesize a ladder network employing no ideal transformers to meet exactly the requirements of (a) a prescribed general minimum-phase transmission function $\frac{E_2}{E_1}$ (except within a constant multiplier) and (b) arbitrarily given, finite, purely-resistive terminations R_1 and R_2 .

The general characteristics of the synthesis problem under consideration may be represented by Figure 1.



Figure 1

Any rational function of λ ($\lambda = \sigma + j\omega$), regular at $\lambda = \infty$, having its poles restricted to the left half-plane and its zeros restricted to the left half-plane and on the j -axis is acceptable as $\frac{E_2}{E_1}$. (These poles and zeros need not be simple). The constant multiplier of $\frac{E_2}{E_1}$ is not always arbitrary, but subject to the physical restrictions on maximum power transfer through a passive network when the terminations are finite resistances, i.e., $\left| \frac{E_2}{E_1} \right|_{\lambda = j\omega}$ has a definite upper bound.

The first step for achieving this realization is to extract a suitable A-matrix (general circuit parameter matrix) from the given transmission function and terminations. Then, a network having the characteristic specified by the extracted A-matrix is synthesized by a matrix factorization procedure. This approach has been employed by other investigators¹⁻⁴ for the synthesis of lossless networks. The more general RLC ladder development will be treated in this note. The ladder development involving no ideal transformers requires an entirely different technique in the decomposition of the A-matrix.

If we specify the RLC ladder network by the general circuit parameter matrix $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, then by network analysis,

$$\frac{E_0}{E_2} = A_{11} + \frac{1}{R_2} A_{12} + R_1 A_{21} + \frac{R_1}{R_2} A_{22} \quad (1)$$

and $A_{11} A_{22} - A_{12} A_{21} = 1 \quad (2)$

Premultiplying a constant k to $\frac{E_2}{E_1}$, suitable A-functions constructed from (1) and (2) are

$$A_{11} = \frac{1}{2} a k \frac{E_0}{E_2} \quad (3)$$

$$A_{12} = \frac{R_2}{2} (\sqrt{2a} - a) k \frac{E_0}{E_2} + \sqrt{R_1 R_2} \quad (4)$$

$$A_{21} = \frac{1}{2 R_1} (\sqrt{2a} - a) k \frac{E_0}{E_2} - \frac{1}{\sqrt{R_1 R_2}} \quad (5)$$

$$A_{22} = \frac{R_2}{2 R_1} \frac{(\sqrt{2a} - a)^2 k \frac{E_0}{E_2}}{a} \quad (6)$$

where k and a , also a constant, are to be determined later to insure the physical realizability of the synthesized network.

In developing this A-matrix into a ladder network, preliminary factorization of A_{11} is generally necessary. Because A_{11} has the properties of $\frac{E_2}{E_1}$, it may always be factored into a product of a finite number of non-minimum resistive positive-real functions, eq. (7), by appropriately choosing a set of positive real numbers $C_1 C_2 C_3 \dots C_n$.

$$\begin{aligned} A_{11} &= \frac{1}{2} a k \frac{E_0}{E_2} = \frac{1}{2} a k \frac{1}{H} \frac{\lambda^m + \alpha_1 \lambda^{m-1} + \dots + \alpha_m}{\lambda^2 + \beta_1 \lambda + \beta_2} \quad (m \geq 0) \\ &= \frac{1}{2} a k \frac{1}{H} \frac{(\lambda^2 + \beta_1 \lambda + \beta_2) (A + \beta_3) (A + \beta_4)}{(\lambda^2 + \beta_1 \lambda + \beta_2) (\lambda^2 + \beta_5)} \\ &= \left[a k \frac{1}{2 k_2 k_3 k_4} \frac{(\lambda^2 + \beta_1 \lambda + \beta_2)}{(\lambda + C_1)(\lambda + C_2)} \right] \left[k_2 \frac{(A + C_3)(A + C_4)}{\lambda^2 + \beta_1 \lambda + \beta_2} \right] \\ &\times \left[k_3 \frac{A + C_5}{\lambda + C_3} \right] \left[k_4 \frac{A + C_6}{\lambda + C_4} \right] \left[k_5 \frac{\lambda^2 + \beta_5}{\lambda^2 + \beta_5} \right] \left[k_6 \frac{(A + C_7)(A + C_8)}{\lambda^2 + \beta_5} \right] \\ &\times \left[k_7 \frac{A + \beta_3}{\lambda + C_5} \right] \left[k_8 \frac{A + \beta_4}{\lambda + C_6} \right] \\ &= (a k A_1) (k_2 A_2) (k_3 A_3) (k_4 A_4) \quad (k_5 A_5) \quad (7) \end{aligned}$$

where $A_1 A_2 A_3 \dots A_n$ represent the non-minimum resistive positive real functions and $a, k, k_2, k_3 \dots k_n$ are constant multipliers to be chosen later. In eq. (7), typical factorization is shown for $m = p$. For the case $m > p$, factors like $k_9 (\lambda + C_9)$ are also included. It should be noted that surplus factors are introduced only when needed.

The ladder representation is now obtained by matrix factorization as follows:

$$\begin{aligned} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ A_{21} & 1 \end{bmatrix} \begin{bmatrix} a k A_1 & 0 \\ 0 & \frac{1}{a k A_1} \end{bmatrix} \begin{bmatrix} k_2 A_2 & 0 \\ 0 & \frac{1}{k_2 A_2} \end{bmatrix} \\ &\times \begin{bmatrix} k_3 A_3 & 0 \\ 0 & \frac{1}{k_3 A_3} \end{bmatrix} \begin{bmatrix} k_4 A_4 & 0 \\ 0 & \frac{1}{k_4 A_4} \end{bmatrix} \begin{bmatrix} 1 & A_5 \\ \frac{1}{A_5} & 1 \end{bmatrix} \quad (8) \end{aligned}$$

* Work supported in part by Office of Naval Research.

EQUALIZATION OF VIDEO CABLES

Philip W. Rounds
Bell Telephone Laboratories
Murray Hill, N. J.

Summary

This paper describes the equalization of local intracity video cables. An analysis is made of cable performance at video frequencies, upon which the equalization plan is based. A network function is derived for approximating a straight-line loss slope on a logarithmic frequency scale. Using this result, a design procedure for fixed and adjustable cable equalizers is described.

Introduction

The local wire transmission of television signals between studios, control rooms, transmitters, and long-distance offices is handled by the Bell System on a video basis over special 16-gauge, shielded, polyethylene-insulated pairs which may be included within the sheath of normal telephone cables. Figure 1 illustrates the physical construction of these pairs. The lengths of the video links and the number of links that may be connected in tandem are such that a total cable distortion of several hundred db may need to be equalized to an accuracy of better than .1 db. The equalization of cables to this precision requires a clear understanding of the nature of the cable performance, both as regards its nominal characteristics and the types of variation that may be encountered in service. It also calls for an equalization plan flexible enough to take into account all the factors of variation. Furthermore the video band is logarithmically very wide, extending for more than 5 decades from 30 cps to 4.5 mc. An equalizer design technique is needed which is suited to the wide-band approximation problem and which will lead to equalizer structures capable of yielding the desired performance in manufacture.

Cable Characteristics

Cable Terminations

The loss of a cable section with arbitrary terminating impedances is equal to the sum of the cable attenuation and the reflection and interaction losses which are functions of the ratio of the cable image impedance to its terminations. Figure 2 shows the image impedance of the 16-gauge video pairs of Figure 1.

Termination of these pairs in a pure resistance at each end leads to low-frequency reflection and interaction losses which are not proportional to cable length and which ripple with frequency. These losses would be difficult or impossible to equalize successfully. Consequently, in the system under consideration here, the terminals and repeaters have been designed to match the image impedance of the cable, thus eliminating the reflection and interaction losses. Under these conditions, the total cable loss is equal to the cable attenuation and is therefore strictly proportional to cable length.

Nature of Cable Losses

The attenuation of a typical 1000-ft length of video cable is shown in Figure 3. In this figure, frequency is plotted horizontally and attenuation vertically, both to a logarithmic scale. At low frequencies the attenuation increases along a straight line corresponding to a square-root-of-frequency increase. At higher frequencies the attenuation also increases as the square root of frequency, but at a different level. Between the two regions there is a smooth, or fairly smooth, transition. The attenuation behavior in these regions may be associated with the resistance losses in the 16-gauge conductors. At the low end of the scale, the distributed capacitance and resistance in the cable are the contributing factors; at higher frequencies, the loss change results from an increase in the conductor resistance through skin effect. In addition to the conductor losses, there is another increment of cable attenuation which increases approximately as the first power of frequency and leads to the rise in the curve at the extreme right of Figure 3. The magnitude of the linear component in the figure has been grossly exaggerated for the sake of clarity. For 16 PSV cable, the actual value of the linear component is about one-tenth that indicated in the figure.

Cable Loss Function

With this picture of the cable performance as a guide, an analytical expression has been found which is capable of representing the actual cable attenuation to good accuracy. This expression appears

as equation (1) below.

$$A = K \left\{ \frac{f+f_2}{f+f_1} \sqrt{f} + k_1 f^{1-\Delta} \right\} \quad (1)$$

where A = attenuation constant
 f = frequency
 K, k_1 , f_1 , f_2 , and Δ are constants

The residual error using equation (1) in place of the actual cable data is shown on Figure 4.

Cable Phase Function

An accurate determination of cable phase as well as loss is important in video design, since the faithful transmission of television signals requires that the phase be held linear to a precision comparable with the required loss precision, counting one radian of phase non-linearity as equivalent to one neper of loss distortion. Owing to the fact that the non-linear component of the cable phase is less than 1% of the total phase, direct measurement of cable phase yields little significant information. The existence of an analytic expression for the cable attenuation, however, allows the phase to be determined to an accuracy equivalent to the loss determination.

The cable phase is taken as the minimum phase² associated with the loss expression of equation (1), modified by the minimum phase corresponding to the loss-correction factor of Figure 4. Substituting the loss, A, of equation (1) in the minimum-phase integral

$$B_c = \frac{2\omega_c}{\pi} \int_0^{\infty} \frac{A-A_c}{\omega^2 - \omega_c^2} d\omega \quad (2)$$

the corresponding phase is found to be (consistent units such as nepers and radians must be used)

$$B = K \left\{ \left[1 - \frac{2\sqrt{f} (f_2 - f_1)}{(f+f_2)(\sqrt{f} + \sqrt{f_1})} \right] \left[\frac{f+f_2}{f+f_1} \sqrt{f} \right] + k_1 \left[\tan \left((1-\Delta) \frac{\pi}{2} \right) \left[f^{1-\Delta} \right] \right\} \quad (3)$$

Although the phase given by the second term in equation (3) becomes infinite as Δ approaches zero, the non-linear component remains finite. The character of the non-linearity may be brought out by subtracting a linear phase and writing the limit. Thus

$$\lim_{\Delta \rightarrow 0} \left\{ k_1 \left[\tan \left((1-\Delta) \frac{\pi}{2} \right) \left[f^{1-\Delta} \right] \right] \right\} = -k_1 \frac{2}{\pi} f \log f$$

The minimum phase associated with the loss of Figure 4 may be obtained readily to sufficient accuracy by graphical integration or by matrix multiplication.

Equalization Plan

The loss shown in Figure 3 applies to a representative sample of 16 PSV cable. Other types of cable, or other samples of the same cable, might be expected to exhibit a similar loss pattern. This has been found to be the case for the cables that have been examined. In each instance, equation (1), with appropriate values of the constants, has been found to be capable of describing the loss behavior to an accuracy comparable with that shown in Figure 4.

From these facts a pattern of equalization emerges. By designing fixed blocks of equalization based on nominal cable characteristics, and further by providing adjustable equalizers at the receiving terminals of each video link to compensate for the effects of variations in the constants of equation (1), it is expected that any type of cable likely to be encountered in service can be successfully equalized. The fixed equalizers can be designed in suitable units to handle various lengths of cable. Through the use of these fixed equalizers, any circuit length may be equalized to within one-half the size of the smallest unit. The adjustable equalizers at the receiving terminal can be used to provide the final incremental adjustment for length.

Equalizer Configuration

Following this plan of equalization, the fixed equalizers must be designed to high precision (in the order of .005 db) so that the requisite number of equalizers may be connected in tandem without exceeding the overall distortion tolerances. The desired loss characteristic for a 20-db equalizer is shown in Figure 5. For equalizers having monotonically decreasing loss characteristics over such wide bands, the bridged-T configuration shown in Figure 6 has been found to yield good results. The bridging arm consists of a resistance paralleled by resistance-capacitance branches which reduce the bridge-arm impedance, and consequently the equalizer loss, in a step fashion at successive

points along the working frequency band. The final resonant-frequency branch is added to bring the loss to zero just above the top useful frequency in order to conserve system gain by holding the total equalizer loss to as low a value as possible.

Having selected the general configuration of the equalizer, the remaining problem is to determine how many R-C branches are needed to obtain the required precision of match and how the elements are to be proportioned to secure the desired characteristics.

Infinite Slope Approximation

Selection of Loss Function

As a guide in answering these questions and by way of laying a foundation for the design procedure, we may consider the idealized problem of approximating a slope of $k \times 20$ db per decade extending from zero to infinity, as shown on Figure 7. A characteristic of this sort may be approximated by an infinite network with a repetitive pattern of real zeros and poles of loss (corresponding to an R-C structure). The loss expression takes the form of an infinite product.

$$e^{\theta} = \prod_{n=-\infty}^{\infty} \frac{p - p_{2n}}{p - p_{2n-1}} \quad (4)$$

where $\theta = A + jB = \text{loss and phase}$
 $p = j2\pi f$
 p_{2n} and p_{2n-1} are negative real numbers

The corresponding loss is found by taking the logarithm of the absolute value of equation (4). The resulting expression becomes

$$\begin{aligned} \text{loss in db} &= 10 \log_{10} \prod_{n=-\infty}^{\infty} \frac{(f_{2n})^2}{(f_{2n-1})^2} \\ &+ 10 \log_{10} \prod_{n=-\infty}^{\infty} \frac{1 + (f/f_{2n})^2}{1 + (f/f_{2n-1})^2} \quad (5) \end{aligned}$$

where $f_{2n} = |p_{2n}/2\pi|$

To approximate the desired characteristic, the frequencies f_{2n} and f_{2n-1} , corresponding to the zeros and poles, are selected as shown in Figure 7 to make the asymptotic representation of the loss zig-zag about the desired infinite slope with a uniform period on a logarithmic frequency scale. The asymptotic representation consists of lines having slopes of 0 db per decade and 20 db per decade alternately, with the

points of transition being made at the frequencies f_{2n} and f_{2n-1} .

Determination of Loss Error

The actual loss as computed from equation (5) will round the corners of the asymptotes. It will thus ripple about the desired curve with the same period as the asymptotic representation but with a much lower amplitude of ripple. The amplitude of the ripple will depend on the number of zero-pole pairs per decade of frequency, being smaller the greater the number of such pairs. The problem, then, is to determine the amplitude of the loss ripple in terms of the number, h , of zero-pole pairs per decade.

To avoid mathematical difficulties, this problem is approached indirectly by way of considering the minimum-phase ripple associated with the loss ripple of equation (5). The justification for this procedure is that a sinusoidal loss ripple on a logarithmic frequency scale

$$\Delta A = a[\sin(2\pi h \log_{10} f)] \text{ nepers} \quad (6)$$

is accompanied by a minimum-phase ripple

$$\begin{aligned} \Delta B &= [\tanh(.4343\pi^2 h)] a[\cos(2\pi h \log_{10} f)] \\ &\approx a[\cos(2\pi h \log_{10} f)] \text{ radians} \quad (7) \end{aligned}$$

which is of approximately equal amplitude. The periodic loss ripple may be expressed as a sine series if the end points of the period are properly chosen. For a slope of 10 db per decade ($k = \pm 1/2$), the loss ripple has even symmetry about the quarter-period points, so that the sine series contains only odd order terms. In this case the shape of the phase ripple will be identical with the shape of the loss ripple, and equation (7) gives the exact relation between their amplitudes. Since the loss and phase ripples are in quadrature, it is sufficient to determine the phase error at the points where the loss error is zero and equate the maximum loss error to the phase error so obtained. For other values of slope, the relation is approximate.

The minimum phase associated with the loss of equation (5) may be obtained directly from equation (4). Thus

$$B = \sum_{n=-\infty}^{\infty} \left[\tan^{-1}(f/f_{2n}) - \tan^{-1}(f/f_{2n-1}) \right] \quad (8)$$

From the symmetry of the curves in Figure 7, it can be seen that the loss error will be zero at the points where the asymptotes

intersect the line of desired infinite slope. There are two such sets of points, one at $f^2 = (f_{2n})(f_{2n-1})$ and the other at $f^2 = (f_{2n})(f_{2n+1})$, illustrated by the points A and B in Figure 7. The frequencies f_{2n} and f_{2n-1} may be defined mathematically as follows. As before, let $k \times 20$ db per decade be the slope of the line to be approximated, and let h represent the number of zero-pole pairs per decade. (In Figure 7, $k = -1/4$ and $h = 1$). For convenience we can, without loss of generality, choose the values of frequency for $n = 0$ such that $(f_0)(f_{-1}) = 1$. Then

$$(f_{2n})(f_{2n-1}) = 10^{2n/h}$$

and

$$f_{2n}/f_{2n-1} = 10^{-k/h}$$

from which

$$f_{2n} = 10^{(2n-k)/2h}$$

and

$$f_{2n-1} = 10^{(2n+k)/2h}$$

For the first frequency, A, in Figure 7, $f = 1$ and

$$B = \sum_{n=-\infty}^{\infty} \left[\tan^{-1} 10^{-(2n-k)/2h} - \tan^{-1} 10^{-(2n+k)/2h} \right] \quad (9)$$

For the second frequency, B, in Figure 7, $f = 10^{-1/2h}$ and

$$B = \sum_{n=-\infty}^{\infty} \left[\tan^{-1} 10^{-(2n+1-k)/2h} - \tan^{-1} 10^{-(2n+1+k)/2h} \right] \quad (10)$$

By noting that $\tan^{-1} X = 90^\circ - \tan^{-1} 1/X$, these formulas may be reduced to

$$\frac{B}{2} = \tan^{-1} 10^{k/2h} - 45^\circ + \sum_{n=1}^{\infty} \left[\tan^{-1} 10^{-(2n-k)/2h} - \tan^{-1} 10^{-(2n+k)/2h} \right] \quad (11)$$

and

$$\frac{B}{2} = \sum_{n=0}^{\infty} \left[\tan^{-1} 10^{-(2n+1-k)/2h} - \tan^{-1} 10^{-(2n+1+k)/2h} \right] \quad (12)$$

The numerical solution of equations (11) and (12) can be simplified by writing them in a form which will permit the values to be read from a table of log tangents.

Thus

$$\frac{B}{2} = (\log \tan)^{-1}[k/2h] - 45^\circ + \sum_{n=1}^{\infty} \left\{ (\log \tan)^{-1}[-(2n-k)/2h] - (\log \tan)^{-1}[-(2n+k)/2h] \right\} \quad (13)$$

and

$$\frac{B}{2} = \sum_{n=0}^{\infty} \left\{ (\log \tan)^{-1}[-(2n+1-k)/2h] - (\log \tan)^{-1}[-(2n+1+k)/2h] \right\} \quad (14)$$

These series converge rapidly, so that their numerical solution is quite simple.

Equations (13) and (14) give the phase at the points A and B respectively. The straight line being approximated is accompanied by a constant phase² of $k \times 90^\circ$. Using this result, it is possible to compute the phase error for various assumed values of h . Finally, the phase errors can be converted into equivalent loss errors as explained previously.

Curve of Loss Error

This has been done for an assumed slope of 10 db per decade ($k = \pm 1/2$) and the results plotted in Figure 8. For the special value of $k = \pm 1/2$, equations (13) and (14) give identical results. For other values of k , equations (13) and (14) give ripples of unequal amplitude, indicating that the phase ripples are alternately large and small along the frequency scale. Considering the larger amplitude only, it has been found by computation that this amplitude varies as $\sin(kx180^\circ)$. Using this result along with the curve of Figure 8, it is possible to determine the complexity of network required to approximate an infinite slope to any desired precision of match.

Design of Fixed Equalizers

Degree of Equalizer Loss Function

The solution of this idealized problem does not give a direct answer to the problem in hand, since the desired equalizer characteristic as shown in Figure 5 is not a constant infinite slope, but rather a slope which is changing steadily throughout the band. The results can be used as a guide, however, in determining the proper density of zeros and poles, if the equalizer characteristic is viewed over a limited frequency range as a segment of an infinite slope and the density of critical frequencies determined accordingly. This establishes the number of elements required to meet a desired precision of match. To match the curve on Figure 5 to the accuracy required, a configuration of the complexity of Figure 9 was found to be needed. The frequencies corresponding to the geometric mean of the zero-pole frequencies are noted in Figure 9.

Design Procedure

The next step in the design is the determination of the element values which will give the precision of match which the preceding theory indicates is possible. This problem might be handled in a variety of ways, but the availability of modern computing machinery makes the method of successive approximations appear the most effective and expeditious. Experience has shown that, starting with an orderly array of zeros and poles, rapid convergence on the desired result can be obtained provided that the array remains orderly during each step of the successive approximation. This concept of orderly array may perhaps be made more definite by tracing the actual steps in the equalizer design.

First Approximation

Considering the bridge arm of the equalizer shown in Figure 9, the shunt resistance is determined by the desired d-c loss of the equalizer. Next, assuming that L in Figure 9 is zero, C is selected somewhat arbitrarily to make the equalizer loss correct at the top of the band, in this case at 4.5 mc. After this is done, the lowest-frequency R-C branch is considered. Its reactance is made equal to its resistance at the branch frequency and its impedance-multiplying factor is adjusted so that the combination of the R-C branch with the initial two branches will give the desired loss at the branch frequency. This procedure is repeated successively adding one R-C branch at a time until each of the branches is

included. Finally the inductance L is added to produce a resonance at the point where the required equalizer loss as shown in Figure 5 passes through zero. This establishes the first approximation in the design. The orderliness of the array can be illustrated by plotting a series of points representing the reciprocal of the resistance in each R-C branch against the critical frequency in that branch, as shown in Figure 10. These points fall on a smooth curve.

Succeeding Approximations

The next step in the design process consists of computing the loss of the equalizer with the assumed element values, and also computing the loss change with changes in each element. With these data in hand, it is possible to arrive at a second approximation. If care is taken in adjusting the parameters to retain a uniform pattern of values as shown in Figure 10, the process can be made to converge very rapidly on a uniform-ripple type of characteristic. The result obtained by following this process through five successive approximations is shown in Figure 11.

Phase Equalization

The problem of phase equalization may merit brief consideration. If the attenuation equalizers were to equalize the cable loss from zero to infinite frequency there would be no phase problem, since both equalizers and cable are minimum-phase structures (ignoring a linear phase component in the cable) and the tandem combination would exhibit no phase distortion. Whatever phase distortion exists, therefore, is a result of the increasing loss of the equalized circuit above the working band. The magnitude of this distortion can be determined by adding the in-band cable phase to the computed equalizer phase. For each of the fixed equalizers, one all-pass phase section was found to be adequate to bring the phase distortion within tolerable limits.

Design of Variable Equalizer

Structure Used

As a further example of the application of the design technique described in the paper, we may consider the design of the variable equalizer used to provide the final incremental adjustment for cable length. The desired loss characteristic for this equalizer is the same as that for the fixed equalizers. Proportional variation of the characteristic is obtained by use of a constant-resistance form of one

of the variable-equalizer structures described by Bode⁴. The configuration of the equalizer is shown in Figure 12. The two boxes in the figure are in themselves four-terminal constant-resistance structures. If φ denotes the transfer constant of these component networks, the overall performance of the equalizer can be described by the approximate relation^{5,6}

$$\theta - A_0 = \rho K e^{-2\varphi} \quad (15)$$

In this expression, θ is the overall equalizer loss and phase, A_0 and K are constants and ρ is a numerical factor varying from -1 to +1 as the control resistances are changed over their complete range from zero to infinity. When $\rho = 0$, the equalizer loss is constant with frequency at the value A_0 . Other values of ρ will give proportionate changes in loss above and below this value as shown in Figure 13. Equation (15) permits the transfer constants φ , of the component four-terminal networks to be determined directly from the desired overall loss and phase, θ , of the complete structure.

Design Procedure

For the problem in hand, θ was determined from equations (1) and (3), modified to give reasonable out-of-band performance. Rather than going into this procedure in detail, it may serve the purposes of this paper sufficiently well to note the type of transfer constant, φ , required to give an overall loss varying as the square root of frequency.

In this case (16)

$$\theta - A_0 = H\sqrt{f} + jH\sqrt{f}$$

Substituting in equation (15),

$$e^{-2\varphi} = \frac{H}{\rho K} (\sqrt{f} + j\sqrt{f})$$

from which

$$\varphi = -\frac{1}{2} \log_e \left(\frac{H}{\rho K} \right) - \frac{1}{4} \log_e (2f) - j \frac{\pi}{8}$$

or in units of db and degrees

$$\varphi = \text{constant} - 5 \log_{10} f - j22.5^\circ \quad (17)$$

Equation (17) describes a network having an infinite slope of 5 db per decade of the type previously considered. An approximation technique for this type of network has already been described in which the amplitude of the approximation error may be determined in advance. The only further step necessary is to relate the

error in approximating φ to the error in the overall equalizer characteristic $\theta - A_0$. This may be obtained by differentiation of equation (15) from which

$$\frac{d(\theta - A_0)}{(\theta - A_0)} = -2 d\varphi \quad (18)$$

Thus an error of .1 db (.0115 nepers) in φ will produce an error of 2.3% in the overall loss of the equalizer. Knowing the maximum loss swing of the equalizer and the desired tolerance in loss, requirements may be placed on the transfer constant, φ , of the component four-terminal networks in the equalizer. The configuration of the four-terminal networks will be generally similar to the configuration of the fixed equalizers.

Problems of Construction

Much of the attractiveness of the design method described here lies in the fact that the equalizer configuration can be selected in advance. This permits an early evaluation of the problems of construction.

Physical inductors and capacitors are not ideal devices, but have associated with them dissipation and parasitic capacitance and inductance. It is important to choose an equalizer configuration which will allow the parasitic elements to be absorbed or at least will allow their effects to be minimized. This is particularly true in the case of video equalizers where the large low-frequency elements must be made to function compatibly with the high-frequency elements throughout the entire video band. The configuration of Figure 6 was selected with these factors in mind.

In the bridge arm of the equalizer, the R-C branches are in parallel, a condition which results in the minimum total capacitance. Even so some of the low-frequency branches require large paper capacitors. The effects of dissipation and series inductance in these paper capacitors is minimized, however, because they appear in series with large resistances. For the higher-frequency branches, small mica capacitors are used and the lead lengths kept short.

In the shunt arm of the equalizer a series resistor is shown. Part of its resistance may be combined with each of the R-L meshes by a transformation which results in a resistor appearing in series with each inductor. This resistance value may then be identified with the winding

resistance of the inductor. The series resistance available is not large enough to eliminate the need for magnetic-core inductors in the low-frequency branches. The distributed capacitance and high-frequency dissipation in the magnetic-core inductors is tolerable, however, since they are shunted by a resistor of low value. The higher-frequency inductors are single-layer solenoids and present no special problem.

By taking advantage of the favorable configuration of Figure 6, and by using care in the mechanical layout of parts and wiring, the precision of the manufactured product can be made consistent with the overall requirements for video system equalization.

Acknowledgment

The author wishes to acknowledge the help of J. A. Yungman, Miss G. L. Lakin, and C. L. Semmelman who were largely responsible for carrying out the development of the methods described in this paper.

References

1. H. A. Wheeler, "The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes." Proc. I.R.E., vol. 27, pp 359-385; June 1939.
2. H. W. Bode, "Network Analysis and Feedback Amplifier Design." D. Van Nostrand Co., Inc., New York, N. Y., chapter 14.
3. W. R. Lundry, "Application of a Minimum Phase Matrix to Adjustable Equalizer Design," presented concurrently with this paper.
4. H. W. Bode, "Variable Equalizers," Bell System Tech. Journal, vol. 18, pp 229-244; April 1938.
5. W. R. Lundry, "Attenuation and Delay Equalizers for Coaxial Lines," Trans A.I.E.E., vol. 68, pp 1174-1179; 1949.
6. R. W. Ketchledge and T. R. Finch, "The L3 Coaxial System - Equalization and Regulation," Bell System Tech. Journal vol. 32, pp 833-878; July 1953.

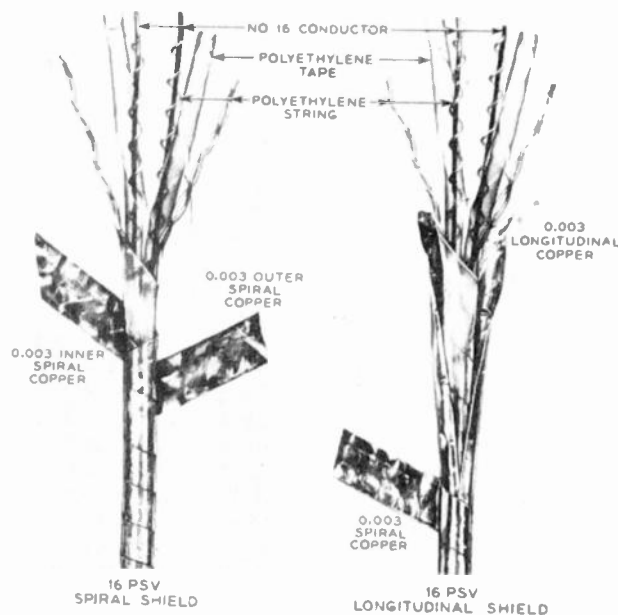


Fig. 1
Construction of video pairs.

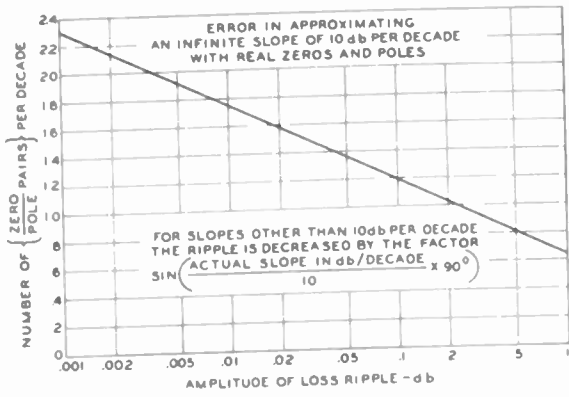


Fig. 8
Error in approximating infinite slope.

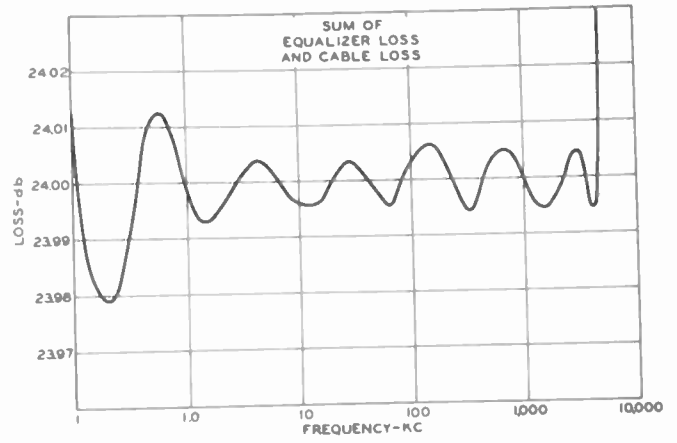


Fig. 11
Equalization error.

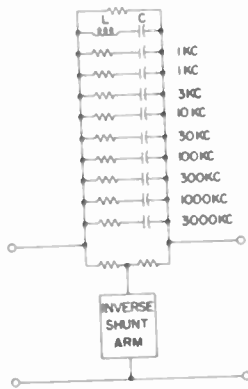


Fig. 9
Configuration of 20 db fixed equalizer.

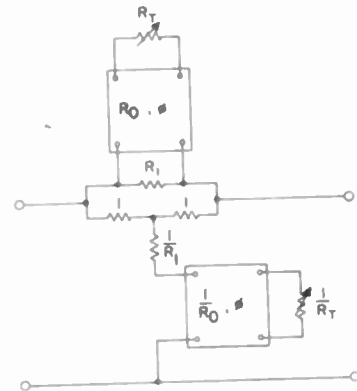


Fig. 12
Constant-resistance variable equalizer.

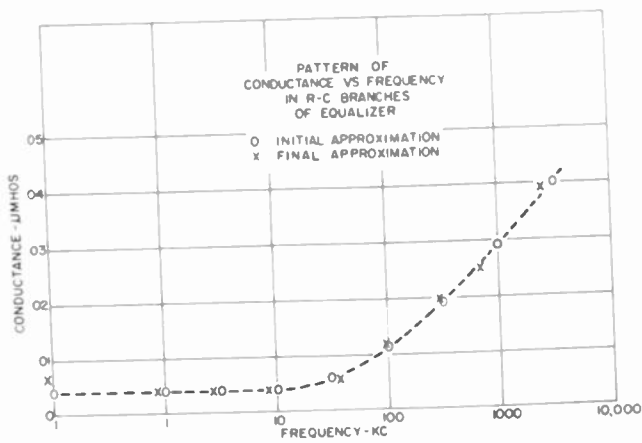


Fig. 10
Relation between parameters in R-C branches of equalizer.

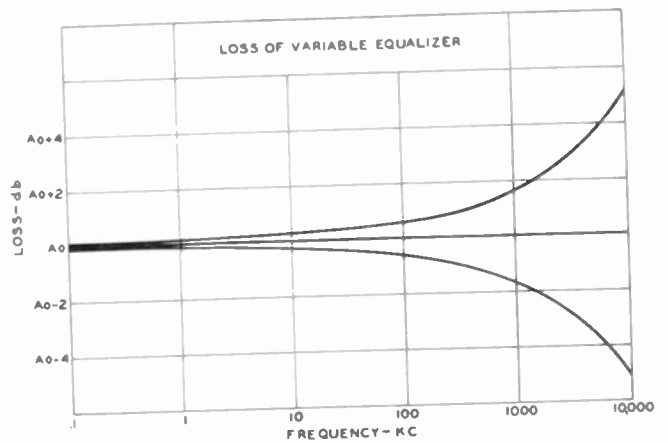


Fig. 13
Loss characteristics of variable equalizer.

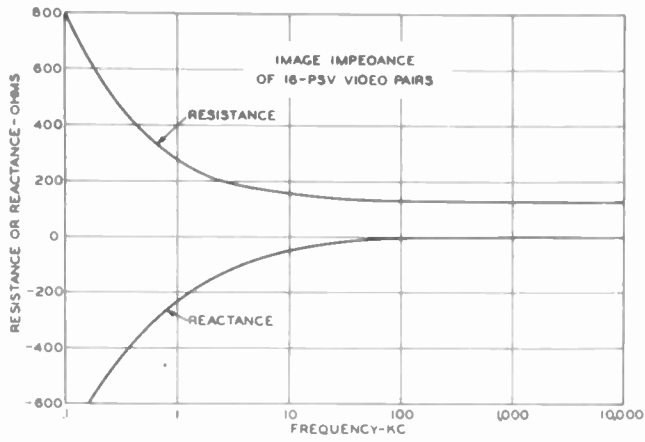


Fig. 2
Image impedance of video pairs.

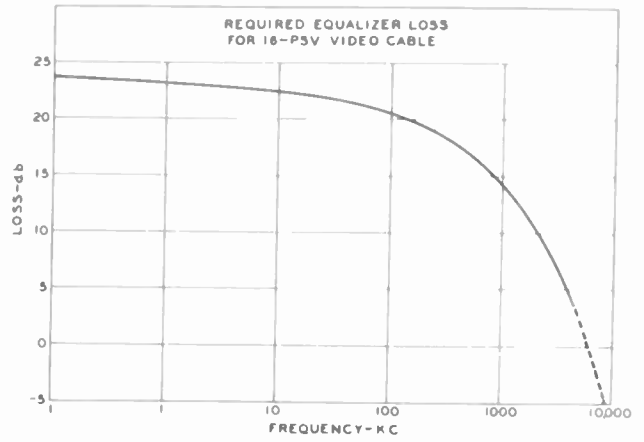


Fig. 5
Equalizer loss objective.

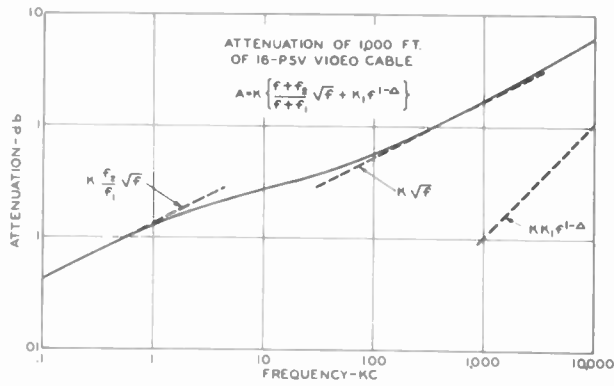


Fig. 3
Attenuation of video pairs.

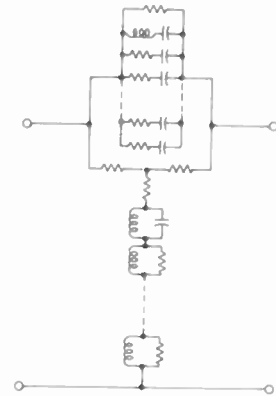


Fig. 6
Equalizer configuration.

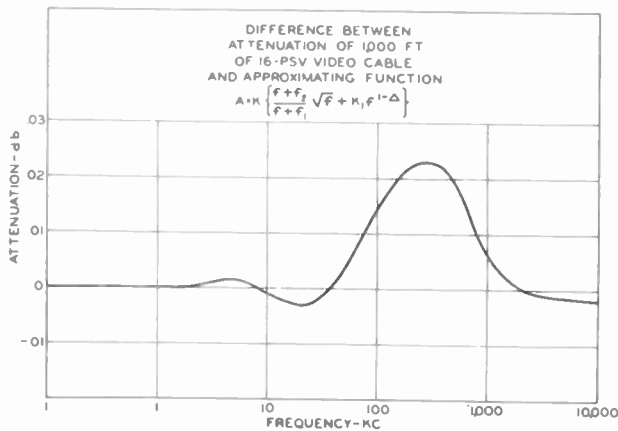


Fig. 4
Error in cable loss function.

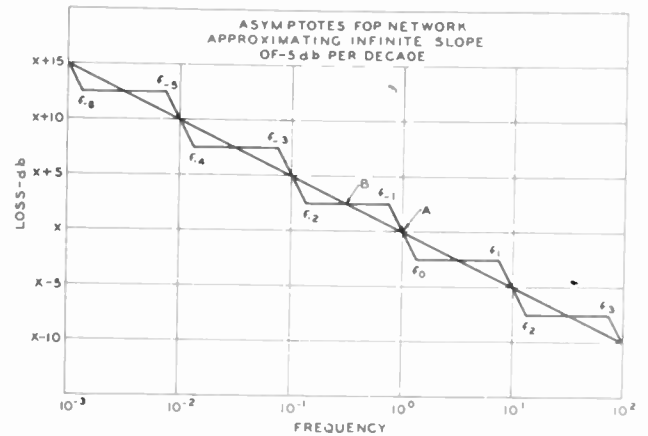


Fig. 7
Asymptotes for network approximating infinite slope.

APPLICATION OF A MINIMUM PHASE MATRIX TO ADJUSTABLE EQUALIZER DESIGN

W. R. Lundry
Bell Telephone Laboratories
Murray Hill, N. J.

SUMMARY

A method for calculating the "minimum phase" associated with a given loss characteristic by numerical methods is described. The process requires a matrix by vector multiplication which can be performed on a high-speed computer.

By successive applications of this process a relatively involved equalizer design problem is reduced to two simple problems.

INTRODUCTION

Recognition of the unique relation between the real and imaginary parts of the self- or transfer-impedance of "minimum phase" networks has led to a number of improvements in planning the equalization for transmission systems. The minimum phase matrix used herein was originally developed as a special tool for use in the analysis of a particular wide-band carrier system. Although a broad sketch of the method of calculating a minimum phase matrix is given, the primary objective of this paper is to show how minimum phase calculations can be used to simplify a design problem. The example is based on a network structure which occurs frequently in equalization systems but has received very little attention in the literature. The procedure depends on the free use of numerical methods and requires computing equipment comparable to the 604 Calculator.

I. Minimum Phase Relationships and their Calculation

Minimum phase network functions, as a matter of definition, are those having all their singularities in the left half of the p -plane. For such functions, an integral relating the real and the imaginary parts can be calculated.¹ In one form this relation is

$$\beta(\omega_c) = \frac{2\omega_c}{\pi} \int_0^{\infty} \frac{\alpha(\omega)}{\omega^2 - \omega_c^2} d\omega \quad (1)$$

where $\beta(\omega_c)$ is the phase shift at a frequency $\omega_c/2\pi$ and $\alpha(\omega)$ is the attenuation at the frequency $\omega/2\pi$.

In practical applications of this integral two difficulties arise. The first has to do with the limits of integration. The behavior of a circuit or a system in the neighborhood of infinity is always a matter for speculation and judgement. In many cases the behavior in the neighborhood of zero is equally uncertain. Fortunately, the denominator of the integrand acts as a weighting function which makes the behavior of $\alpha(\omega)$ near the limits of integration relatively unimportant. Assumptions based on good engineering judgement can be used in these regions without seriously affecting the accuracy of the calculation.

The second difficulty arises from the need to know $\alpha(\omega)$ in a form which can be integrated. Most design problems start with the statement of a desired loss characteristic in the form of curves or tabular data. Here difficulties can be avoided by the use of numerical methods which permit the evaluation of the integral without first solving an approximation problem.

In the main, the present problem can be solved with a high order of accuracy by the use of Gregory's Method of numerical integration.⁽³⁾ This method assumes, in effect, that the integrand can be matched by a polynomial of fixed degree and furnishes the coefficients required to calculate the integral as the sum of a series of products. If the polynomial is of first degree, Gregory's method gives the familiar trapezoidal rule. Over a limited range near ω_c this method is inadequate because of the pole in the integrand. This difficulty is avoided by assuming a polynomial match to $\alpha(\omega)$ and performing an exact integration over this interval to determine the Cauchy principal value.

Enough has been said to indicate the general nature of the numerical methods used to calculate this integral. In its

final form the phase is evaluated by using the formula

$$\beta_i = \sum_{j=1}^N C_{ij} \alpha_j \quad i = 1, 2, \dots, M \quad (2)$$

where β_i is the phase at the i th frequency and α_j is the loss at the j th frequency. The constants C_{ij} are independent of the particular loss characteristic involved. It should be noted that a different set of constants is required for each frequency at which phase is to be determined. This is because both the denominator in the integrand of (1) and the Gregory coefficients determine the C_{ij} and the former depends on ω_c .

Equation (2) describes β_i as equal to a matrix by vector product where the C_{ij} are the elements of an $M \times N$ matrix and the α_j are the elements of an $N \times 1$ matrix or vector. For brevity, the $M \times N$ matrix will herein be referred to as the "minimum phase matrix".

II. Machine Computation of Minimum Phase

The elements of a minimum-phase matrix have been computed for a normalized arithmetic frequency scale with j taking on all values between 1 and 63 and i ranging between 5 and 55. This gives a total of more than 3000 entries. Calculation of the minimum phase for problems of this magnitude becomes practical when there is available some large scale computing equipment such as the IBM 604 Calculator. With such equipment it is possible to have the matrix values stored on punched cards, to copy the loss data on the corresponding cards and then to perform the required multiplications at the rate of 100 a minute.

It turns out that the capacity of the equipment is such as to permit the solution of 3 simultaneous problems at this same rate. Multiplications are performed in the machine and punched on the cards containing the original entries. The cards are then transferred to a tabulator which lists and totals them giving, at the end of each 63 cards, printed sums equal to the phase at one frequency. The machine time required for the complete calculation of three problems is about one hour.

III. Adjustable Loss Equalizer Design

As an example of the application of the minimum phase matrix to design

problems, consider an adjustable equalizer requirement as shown by Figure 1. This illustrates a type of loss characteristic which is frequently required in a long transmission system which may be carrying both telephone and television circuits. The solid curve indicates the desired loss characteristic at one extreme setting of the equalizer and the dotted curve is its mirror image about the reference loss, here indicated as zero. An equalizer which can be adjusted continuously and linearly between these two curves is desired. It will be noted that at one frequency there is a pivot point where the relative loss is zero for all adjustments.

One of the best types of adjustable equalizer which can be used for this job is shown in Figure 2. This consists of a 2-terminal impedance composed of a resistance in parallel with the input terminals of a 4-terminal network. The 4-terminal network is required to have constant resistance image impedances at both pairs of terminals and is terminated at its output end by an adjustable resistance. By properly choosing the constants of this circuit the insertion loss can be expressed as

$$\alpha - \alpha_0 + j\beta = K\rho e^{-2(A+jB)} \quad (3)$$

Here α and β represent the insertion loss and phase of the 2-terminal network with α_0 the reference value of α ; K is a constant dependent on α_0 ; ρ is the reflection coefficient at the output of the 4-terminal network whose image transfer loss and phase are designated A and B respectively. This equation is a linearized approximation to the exact expression but is quite adequate for the usual design problem. (4), (5), (6)

Since only the loss characteristic, $\alpha - \alpha_0$, was specified, the relation between the two terminal network loss and the image transfer constant of the 4-terminal network is expressed by the second equation of this figure. It will be noted that the loss we are interested in depends on both the loss A and the phase B of the 4-terminal structure. For example, the pivot point of the loss characteristic can be obtained either by making A very large or by making B equal to 45° . The problem of choosing a structure to perform this job is substantially more difficult than the usual problem of fixed equalizer design because of the dependence of the final result on both A and B . Skill in the design art is acquired very slowly.

It is apparent from the structural form that the loss $\alpha - \alpha_0$ must be minimum phase (the insertion loss of any two terminal structure is minimum phase). This suggests that if we determine a value of β which is consistent with the required loss characteristic, it will be possible to solve for A and B.

IV. Minimum Phase of 2-Terminal Structure

Three minimum phase calculations were run for this particular problem. Two of these furnish information on the phase of the network for different out-band loss behavior corresponding to reasonable engineering assumptions. The other was used to obtain the low-frequency phase behavior in finer detail than provided by the other calculations. Collectively, they serve to emphasize the flexibility of the matrix through the use of different frequency normalizing constants for each case.

For this particular design problem, loss information up to 20 mc was available. Although the equalizer need not follow the system loss characteristic above 8.5 mc, departures must be considered in the light of their effect on the phase below 8.5 mc. The system carries television signals and failure of the equalizer to compensate corresponding system phase distortion would have to be corrected by an associated adjustable phase equalizer. In the hope that this could be avoided, the minimum phase associated with the solid curve of Fig. 3 was calculated.

A convenient choice of frequency interval for this calculation is 0.4 mc. That is, the α_1 correspond to loss values read from the curve at 0.4, 0.8, 1.2, ... 20 mc. This furnishes the first 50 values. A reasonable choice for the remaining 13 is the constant value, -6.6 db, corresponding to a constant loss of -6.6 db from 20 mc to infinity. These points serve to define the loss characteristic quite well over the higher part of the useful band (2 to 8.5 mc). In the region below 1 mc the coarse granularity forces one to accept a compromise match to the curve such as is indicated by the dotted line designated II. This will cause substantial errors in the minimum phase calculation at the lower frequencies.

An accurate determination of the minimum phase at low frequencies can be obtained by using α_1 read from the curve at 0.1 mc intervals as indicated by the

curves designated I. The last usable value, α_{63} , will be the loss at 6.3 mc. For reasons which can be found in reference (2) the values used are dropped below the actual loss in the range 5 to 6.3 mc.

Finally, a case which is more likely to correspond to a practical equalizer design is computed. This follows the solid curve up to 10.2 mc and dotted curve III from 10.2 to 12 mc. This can be realized with a frequency interval of 0.2 mc.

A composite presentation of results of the three computations is given in Fig. 4. Curves Ia and II together define the phase of the "best" equalizer, i.e., one which would surely compensate both the variable loss and phase of this system. Curves Ib and III furnish a composite look at the phase of a more practical equalizer. Curve Ib is derived by adding sufficient linear phase to Ia to produce an intersection with III at 3.2 mc (again justified by reasoning given in reference (2)).

V. The 4-Terminal Transfer Constant

The frequency band used for television transmission over this system runs from approximately 4 to 8.5 mc. Over this band the difference in phase between II and III of Fig. 4 is essentially linear. Hence no loss of transmission quality can be charged to an equalizer having the phase shown by the upper curve of Fig. 4. Therefore it was chosen as the basis for further design work.

From the values of loss given and the values of phase shown on this curve, the transfer constant $A + jB$ of the 4-terminal network was calculated by means of the formula

$$A + jB = (1/2)[\ln K_p - \ln(\alpha - \alpha_0 + jB)] \quad (4)$$

which follows directly from (3). The results are shown in Figs. 5 and 6.

The design of the 4-terminal network must realize* both A and B. This can be reduced to (a) the design of a minimum phase loss equalizer and (b) the design of an all-pass network giving the phase not

*An allowance of 1.7 db must be made for the term $-\frac{1}{2} \ln K_p$ to insure the required adjustment range. Additive constants do not affect the minimum phase computation.

realized in (a). Two additional minimum phase calculations were made to determine an appropriate split-up. Both, of course, follow the solid curve A of Fig. 5 but differ in the assumed out-band performance as indicated by the dotted lines IV and V. Only one normalized frequency interval, 0.2 mc, was used for these cases. The results are shown by dotted curves IV and V of Fig. 6.

The difference between these curves and the required phase, B, is shown in Fig. 7. The dots show the calculated phase of an all-pass section having a stiffness ratio $b = 1.3$ and a critical frequency $f_c = 19.0$ mc. The agreement with curve IV is good and indicates that the design of a minimum phase structure following the loss curves A and IV of Fig. 5 will solve the problem. This is a conventional, fixed, loss equalizer design which appears to offer no unusual problems.

VI. Further Possibilities

Although few experienced designers would be likely to need further reduction of the problem, it should be pointed out that the minimum phase matrix can be used in the design of the fixed equalizer.

Assume that this is to be realized by a single, constant-resistance section of the bridged-T type. The loss and phase are related to a 2-terminal impedance, $R+jX$, by the formula

$$\epsilon^{A+jB} = 1 + R + jX \quad (5)$$

Since A and B are known, both R and X can be calculated immediately. Values of R

multiplied by the minimum phase matrix give the "minimum reactance" associated with R and serve to resolve X into two parts. One will be realized automatically when R is synthesized and the other must be realized with a purely reactive network.

ACKNOWLEDGMENT

The basic program for minimum-phase calculation by numerical methods was originally proposed by Mr. R. W. Hamming. Calculation of the minimum-phase matrix was done under the direction of Mr. B. A. Kingsbury.

References

- (1) "Network Analysis and Feedback Amplifier Design", H. W. Bode. (Van Nostrand) Chapt. 11 and 14.
- (2) *ibid*, Chapt. 14
- (3) "Calculus of Observations", Whittaker and Robinson (Blackie and Son, Ltd.) Chapt. 7.
- (4) "Variable Equalizers", H. W. Bode, Bell Sys. Tech. Jl., Vol. XVII, April, 1938.
- (5) "Attenuation and Delay Equalizers for Coaxial Lines", W. R. Lundry, Trans. A.I.E.E., Vol. 68, pp. 1174-1179.
- (6) "The L3 Coaxial System - Equalization and Regulation", R. W. Ketchledge and T. R. Finch, Bell Sys. Tech. Jl., Vol. 32, July, 1953.

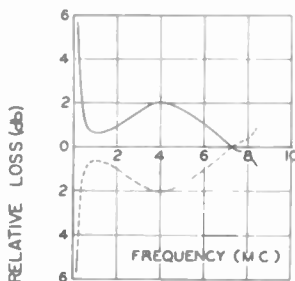


Fig. 1
Required loss of adjustable equalizer.

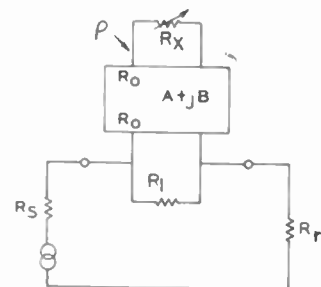


Fig. 2
Configuration of adjustable equalizer.

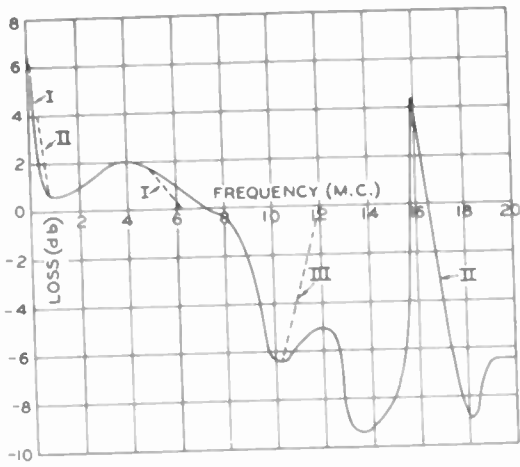


Fig. 3
Loss characteristics used for
minimum phase determination.

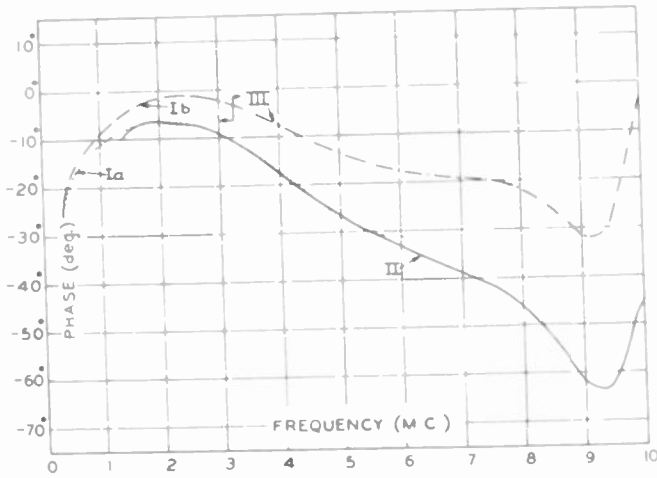


Fig. 4
Minimum phase of two-terminal network.

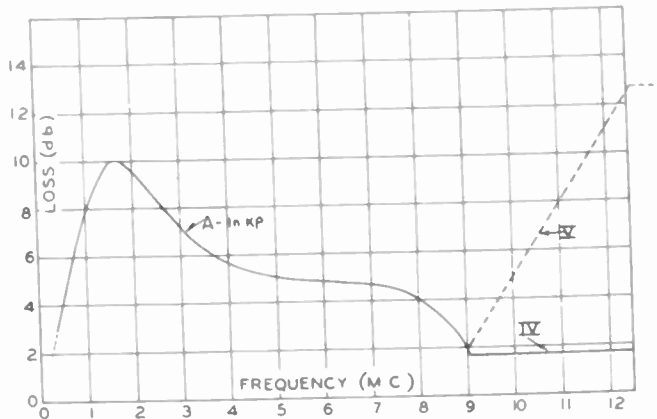


Fig. 5
Loss required in four-terminal network (A).

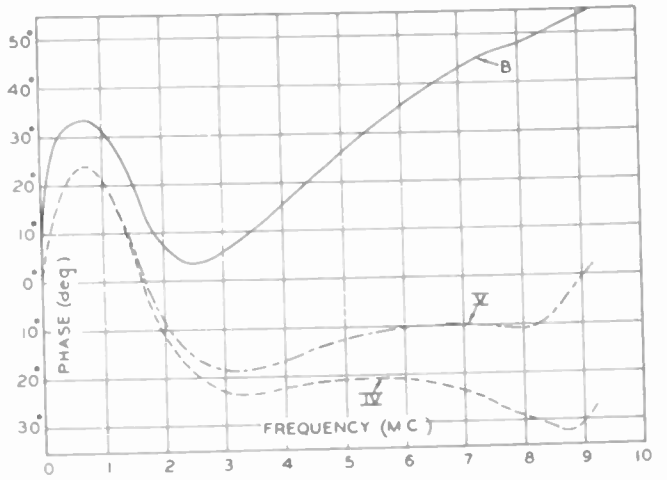


Fig. 6
Phase required in four-terminal network (B);
minimum phase of A (IV & V).

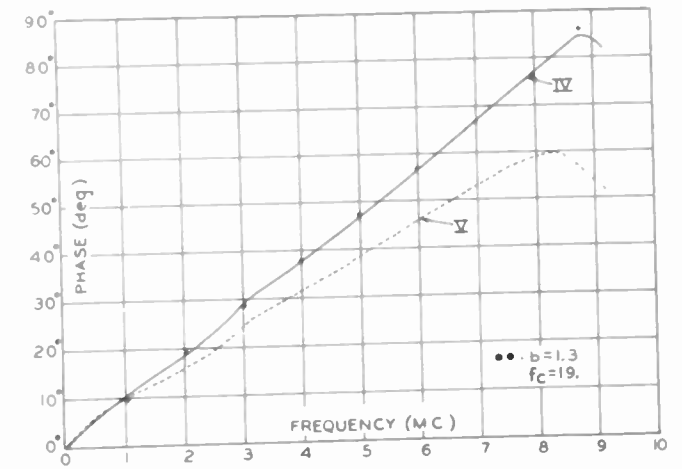


Fig. 7
Residual phase realized with
all-pass structure.

EQUALIZATION IN THE TIME DOMAIN

by

Murlan S. Corrington
T. Murakami
Richard W. Sonnenfeldt

Consumer Products Division
Radio Corporation of America
Camden, N. J.

Summary

Methods are described for compensating a network to restore the original transient driving force, except for a possible time delay, and for correcting parts of a transient where desired. Passive networks can be used to reshape a transient, to add or subtract any desired waveform anywhere along the transient, or to trap out undesired overshoots. The compensating network corrects the phase and amplitude characteristics of the network simultaneously.

Introduction

When a transient waveform is transmitted through a system it often happens that the transient output does not have the shape desired and it becomes necessary to equalize the system to change the transient to the desired shape. Two kinds of equalization will be discussed. The first case will be an attempt to restore the original signal, with a time delay permitted. The second case will be a method for changing a given transient to a desired one, which may be different from the original waveform.

Restoring a Unit-Step Function

If a tuned circuit has a very low Q , there will be a slight rise in the selectivity curve when the circuit is resonant to the driving force. In

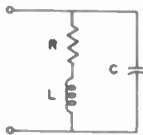


Fig. 1 Shunt-Peaked Circuit

Fig. 1 let the peaking coil L be added in series with the load resistor R . The normalized impedance of this circuit is

$$\frac{Z}{R} = \frac{1 + pL/R}{1 + pRC + p^2LC} = \frac{1 + pQ}{1 + p + p^2Q} \quad (1)$$

where $\omega_0 RC = 1$, $p = i\omega/\omega_0$, and $Q = \omega_0 L/R$.

If equation (1) is used to find the response of the network to a unit-step of current¹, the voltage output is as shown by Fig. 2. All curves

start out with unit slope, and the amount of overshoot is determined by the Q of the circuit. The highest Q can be without producing an overshoot is $Q = .25$. The curve labeled $A_1(\alpha)$ is for the semi-infinite constant-slope filter.²

If the reciprocal of the operational form of equation (1) is taken, the response to a unit-step function will be given by

$$A(\alpha) \supset \frac{1}{s} \frac{1 + s + s^2Q}{1 + sQ} = 1 + \frac{1}{s} - \frac{Q}{1 + sQ}$$

$$\supset \delta(\alpha) + U(\alpha) - e^{-\alpha/Q} \quad (2)$$

where $\delta(\alpha)$ is the unit-impulse function, $U(\alpha)$ is the unit-step function, and $\alpha = \omega_0 t$.

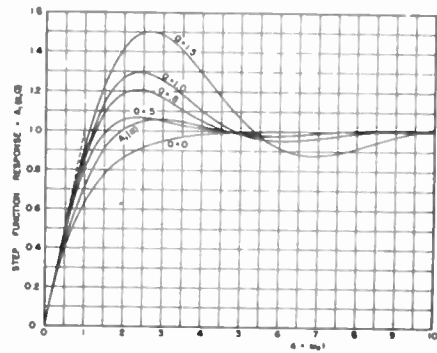


Fig. 2 Unit-Step Response of Shunt-Peaked Filter

The network having this response is shown by Fig. 3,

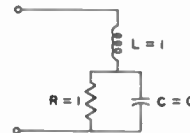


Fig. 3 Compensating Network

and the voltage response to a unit-step of current is shown by Fig. 4, for various values of Q .

There is a unit impulse at the origin followed by the exponential rise to the final value of unity.

This means that any network which has a transient response to a unit-step function as shown by Fig. 4 can be cascaded with a system having the transient response of Fig. 2 to restore the original signal, which was a unit-step function at the origin. Since the operational forms of the two networks are reciprocals, it is evident that cascading them produces an all-pass system with zero phase shift at all frequencies. Since the network has been compensated to restore one transient driving force, it will restore any other driving force.

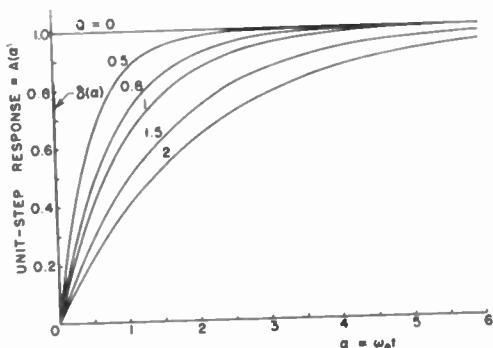


Fig. 4 Unit-Step Response of Compensating Network

If, in addition, a network is cascaded with the above two which has a linear phase shift at all frequencies and no selectivity, the result will be an undistorted unit-step response but delayed in time by an amount equal to the phase shift in the delay system, at a given frequency, divided by that frequency. In many applications such a delay is permissible.

The network of Fig. 3 is quite simple and is easy to construct. In a later section a general synthesis procedure will be given which will make it possible to synthesize any compensating transient.

Equalization of a Network to Obtain a Desired Transient

If a system has the unit-step response shown by Fig. 2, it can be compensated to change the response to a ramp function with arbitrary initial slope. This compensating network is illustrated by Fig. 5.

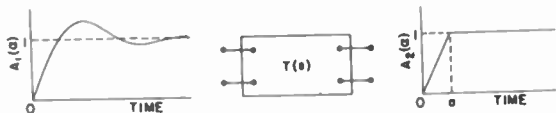


Fig. 5 Compensating Network

By equation (1) the operational form of the input signal is

$$A_1(a) = \frac{1}{s} \frac{1 + sQ}{1 + s + s^2Q} \quad (3)$$

where $a = \omega_0 t$ is the normalized time. The operational form of the ramp function is obtained by taking the operational form of a linearly-increasing driving force with slope $1/a$ and subtracting another with slope $-1/a$ but delayed in time by $a = a$. The operational form of this output response is

$$A_2(a) = \frac{1}{as^2} [1 - e^{-as}] \quad (4)$$

The operational form of the compensating network is the ratio of the output operational form to the input operational form, or

$$\begin{aligned} T(s) &= \frac{1 + s + Qs^2}{as(1 + Qs)} [1 - e^{-as}] \\ &= \frac{1}{a} \left[1 + \frac{1}{s} - \frac{Q}{1 + Qs} \right] [1 - e^{-as}] \quad (5) \end{aligned}$$

For an example of this type of compensation, the curve labeled $Q = 1$ in Fig. 2 will be changed to a ramp function of varying slope, as shown by Fig. 6. The first case, $a = 1.0$, will cause the curve to rise faster than

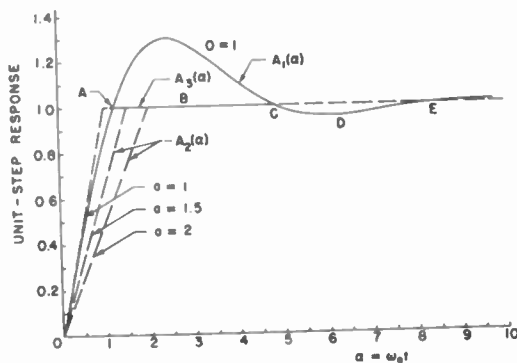


Fig. 6 Compensation to Give Ramp Function

the original signal, but the other two, $a = 1.5$ and 2.0 will decrease the initial rate of rise.

The inverse Laplace transform of equation (5) will give the impulse response of the compensating network. This is shown, for the three values of a , by Fig. 7. There is a positive impulse of strength $1/a$ at the origin, an exponential rise, and a negative impulse of strength $1/a$ at $a = a$, followed by an exponential decay to the final value of zero.

The integrals of the curves of Fig. 7 give the unit-step response of the compensating network. The three curves, for $a = 1, 1.5, 2.0$, are shown by Fig. 8; all start out at $1/a$ and rise exponentially until $\omega_0 t = a$. At this point there is a sudden drop of $1/a$ and an exponential rise to the

final value of unity. These curves mean that if a network has the unit-step response shown by

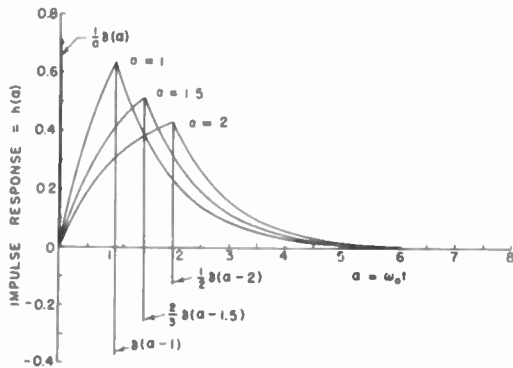


Fig. 7 Impulse Response of Compensating Network

Fig. 8, such a network will have just the right amplitude and phase characteristics to compensate the transient labeled Q = 1 on Fig. 6 and change it to the ramp function shown.

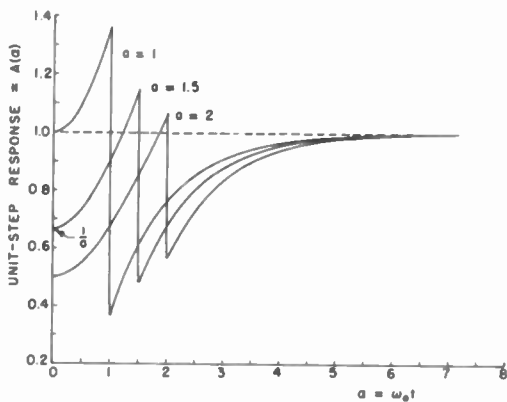


Fig. 8 Unit-Step Response of Compensating Network

To illustrate a more general solution, the curve labeled Q = 1 of Fig. 2 will be compensated by removing the first overshoot, leaving the rest of the curve without change. The new curve will be the curve QABCDE of Fig. 6.

By means of the convolution integral, an integral expression for the solution can be obtained. Thus, if $A_1(a)$ is the original transient, $A_3(a)$ is the unit-step response QABCDE of Fig. 6, and if $h(a)$ is the impulse response of the compensating network, we have by the convolution integral

$$A_3(a) = \int_0^a A_1(\lambda) h(a - \lambda) d\lambda \quad (6)$$

where the unknown function $h(a - \lambda)$ is in the integrand.

Solution of the Integral Equation for the Impulse Response

The interval up to point A of Fig. 6 was divided into a number of equal subintervals (say 10) and the response $A_1(a)$ was tabulated throughout the entire range at this interval. These points were then tabulated on a strip of paper as shown by Fig. 9. The unknown values of $h(a - \lambda)$ will be put on the second strip later as shown. When the two strips are moved along until they line up at a certain point a as shown (say $a = 4$) and the corresponding points are multiplied across, the resulting products are the successive points of the integrand of equation (6). The Newton-Cotes numerical integration formulas can be used to

			7
			6
			5
			4
			3
			2
			1
			0
a	$A_1(a)$	$h(a)$	a
0			
1			
2			
3			
4			
5			
6			
7			

Fig. 9 Sliding-Strip Method

evaluate this integral.³⁻⁴

To start the solution it should be noted that the curve of Fig. 6 is not to be changed for $0 \leq a \leq 1.2092$. This means that in the solution

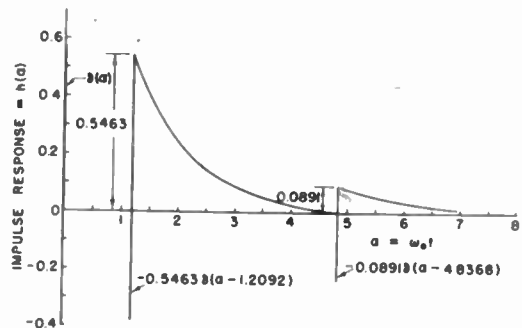


Fig. 10 Impulse Response of Compensating Network

for $h(a)$ there is a positive unit-impulse at the origin, as shown by Fig. 10, and nothing further until $a = 1.2092$. At point A there is a negative impulse of strength 0.5463 since there is a sudden change in the slope of the curve of this

amount at that point. Following the negative impulse the curve decreases smoothly from the value 0.5463, which is the second derivative of the curve for $Q = 1$ at $\alpha = 1.2042$, to zero at the point C. At point C there is a sudden change in slope of amount -0.0891 , and the second derivative is $+0.0891$, so there is an impulse of this amount followed by a smooth decrease from $+0.0891$ toward zero. Fig. 10 shows the final curve.

The initial part of the curve was obtained by assuming a power series expansion at $\alpha = 1.2092$. The Newton-Cotes numerical integration formulas were used to integrate equation (6), with the unknown ordinates of $h(\alpha - \lambda)$ of Fig. 10 appearing in the formula. By setting up the integration formulas for one tabular interval, and again for one half the tabular interval, simultaneous equations were obtained for the unknown ordinates of Fig. 10. Once the solution was started properly, the ordinates were entered on the sliding strip of Fig. 9. By using successively higher point

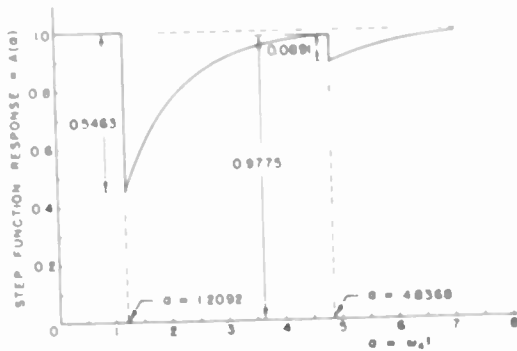


Fig. 11 Unit-Step Response of Compensating Network

formulas it was then easy to extend the solution point by point since there was only one unknown ordinate in each formula. This sequence of steps gave the complete curve of Fig. 10.

The integral of this curve is the unit-step response, shown by Fig. 11. Any network that has this unit-step response will compensate the transient of Fig. 6 to give the corrected curve QABCDE, if the original transient ($Q = 1$) is passed through it.

Delay Line Synthesis of Transients

If a distortionless transmission line, as shown by Fig. 12, is terminated at both ends with its characteristic impedance, any pulse that starts down the line will be propagated without distortion and will not be reflected from the end. The resistors R_0, R_1, R_2 , etc., are large compared to the characteristic impedance of the line and do not load the line. When a unit-step of voltage is applied to the input, the output from resistor R_0 will be a step of voltage of height depending on the value of R_0 . It can be applied to the upper bus, labeled -, or the lower one, labeled +,

by the single-pole double-throw switch shown.

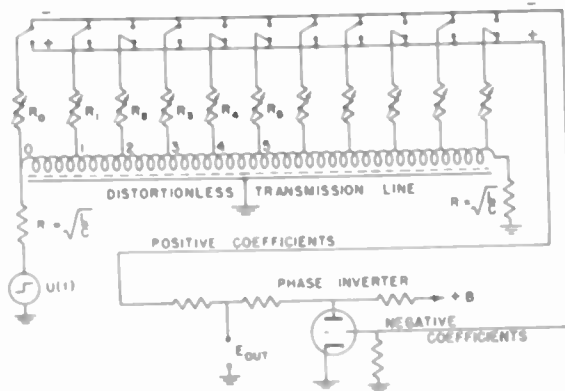


Fig. 12 Terminated Delay Line

At the next point on the line the response will also be a step of voltage, delayed one unit of time, and the height will be determined by the setting of R_1 .

By setting the successive R 's and the polarity switches properly, it is possible to construct any given transient as a series of steps of voltage. The approximation can be made as good as desired by taking enough steps or taps on the transmission line.

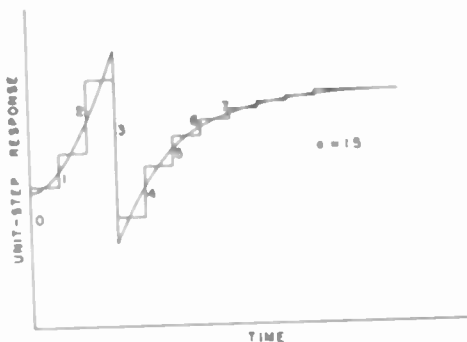


Fig. 13 Step Approximation

As an example, suppose the response of Fig. 8 is to be synthesized. As shown by Fig. 13 the response can be approximated by a series of steps 0, 1, 2, 3, etc. If the taps on the delay line of Fig. 12 are properly chosen, and the resistors and polarity switches are set properly, the response of Fig. 13 can be approximated with as many steps as desired. Because of the limited total bandwidth of the line the discrete steps will tend to smooth out and a surprisingly good approximation can often be obtained.

The oscillograms of Fig. 14 show several waveforms that were synthesized to show how the small steps disappear. Figs. 14(a) and 14(b) show how straight-line segments can be approximated. Careful examination will show that there is a slight

irregularity along the sloping lines. Fig. 14(c) shows how a pulse can be added in the middle of a square wave, without producing appreciable ringing anywhere else. Fig. 14(d) shows how to put several irregularities in the center of the square wave. Figs. 14(e) and 14(f) show how the polarity of an added pulse can be changed merely by changing the polarity switches of Fig. 12.

If the incoming signal, curve $A_1(\alpha)$ of Fig. 6, is sent through a tapped delay line that is adjusted

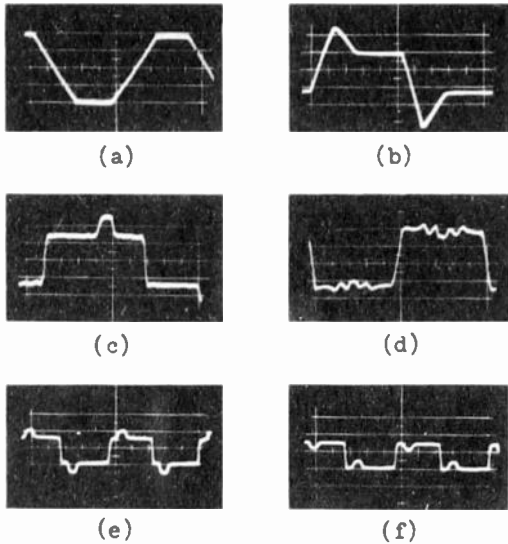


Fig. 14 Oscillograms of Waveforms

in accord with Fig. 13, the resulting transient output will be the ramp function of Fig. 6, $A_2(\alpha)$. In this method it is not necessary that the original step function, or the network that produced the response $A_1(\alpha)$, be available at the point where compensation is applied. The delay line simultaneously corrects the amplitude and phase of the network to produce the desired transient.

Parallel Method of Compensation

If the original signal and network are available,

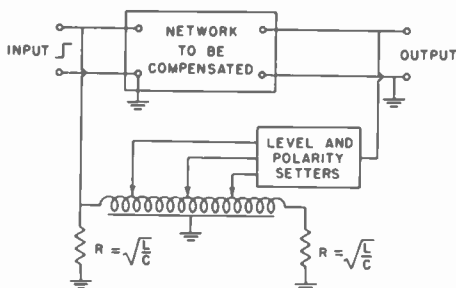


Fig. 15 Parallel Method of Compensation

a tapped delay line can be used to compensate a transient, as shown by Fig. 15. The error between

the response of the network to the unit-step function and the desired response is computed. The delay line is then adjusted with the circuits shown by Fig. 12, to produce a unit-step response corresponding to the compensation required. The level and polarity setters are shown in the block diagram of Fig. 15.

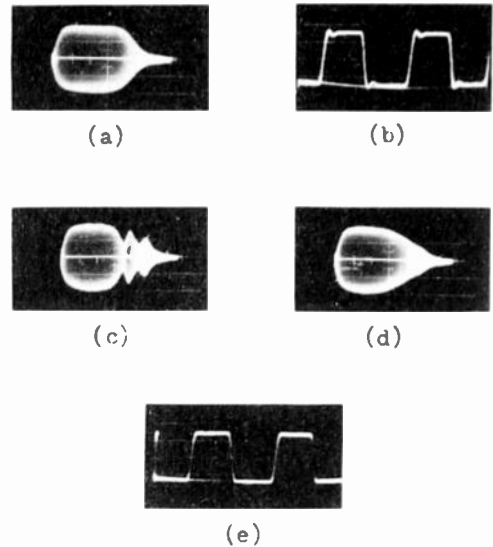


Fig. 16 Oscillograms of Trap Compensation

Trap Method of Compensation

If a transient overshoot, such as that shown by the portion of $A_1(\alpha)$ between points A and C of Fig. 6, is to be removed, a trap can be used to ring at just the right time, with just the right amount, to remove the overshoot.

Fig. 16(a) shows an oscillogram of the sweep response of a low-pass filter, which is fairly flat to cutoff but which has a transient overshoot, as shown by the square-wave response of Fig. 16(b). If two parallel traps are placed in one or two of the cathodes of the tubes in the amplifier stages, or series traps are connected from the plates to ground, the transient overshoot can be removed. Fig. 16(c) shows the traps in place, but with the Q's too high. Fig. 16(d) shows the traps properly adjusted, and Fig. 16(e) shows the resulting transient response. The corner on the square wave can be made quite sharp if the two traps are adjusted properly.

The first trap is tuned to a frequency corresponding to the duration of the first overshoot. The L/C ratio is adjusted to match the impedance of the filter so it cancels just the right amount of overshoot. Since the cancellation will not be perfect the filter will eventually start to ring again. A second trap is used to cancel this subsequent ringing. Additional traps can be used if more precise compensation is required.

Conclusions

It is often possible to restore the original driving force in a system if a compensating network is used which corrects the amplitude and phase of the system. The system may be a network, or propagation paths, and may include transducers. The resulting over-all response must have constant gain at all frequencies and a phase shift proportional to frequency at all frequencies.

When a given transient response is to be corrected to give a different transient, a compensating network can be used. If the transient is passed through such a network it will modify the transient to give the desired output. The original driving force and network need not be accessible.

If the original driving force is available, it can be passed through a network in parallel with the uncompensated network, to produce a correcting transient response. The two outputs are then added together.

When a low-pass filter has a transient overshoot it can often be removed by placing rejection traps in the filter at just the proper frequency to cancel out the transient. Two such traps will usually give very good suppression of the overshoot.

All of these methods use passive networks only. No diode clippers or other active elements are

necessary to shape the pulse in the network. Ripples can be added or subtracted at any part of the transient. A very simple method of synthesis enables the simultaneous correction of the amplitude and phase of the network by turning knobs and throwing polarity switches on a delay line.

References

1. Murray F. Gardner and John L. Barnes, Transients in Linear Systems, Appendix A, pp. 332-356, vol. I; John Wiley & Sons, Inc., New York, N. Y., 1942.
2. Murlan S. Corrington, Transient response of filters, RCA Review, vol. 10, no. 3, pp. 397-429; Sept. 1949.
3. W. Woolsey Johnson, On Cotesian numbers; their history, computation, and values to $n=20$, Quart. Jour. Pure and Applied Math., vol. 46, pp. 52-65; 1915.
4. National Bureau of Standards, Tables of Lagrangian Interpolation Coefficients, Columbia University Press, New York, 1944, pp. xxxii-xxxiii.

THE GROUP THEORETICAL ASPECT OF LINEAR FOUR-POLE THEORY

Wolfgang Gaertner
Signal Corps Engineering Laboratories
Fort Monmouth, New Jersey

Summary - A few basic relations are derived which show how group theoretical concepts may be introduced into four-pole theory. The possible perspectives of the new approach are pointed out.

A. Introduction¹

This paper is essentially a suggestive one. Its purpose is to show how group theoretical considerations may be applied to four-pole theory which to the author's knowledge has not previously been done to any considerable extent². The new approach to the well established field justifies itself not only by the results of its derivations but also by adding to the understanding and insight and thus providing ideas and suggestions on how to handle certain problems. Furthermore it will become apparent that group theory may introduce an additional classifying principle into the extensive field of four-poles. In some respects the problem resembles the group theoretical approach to the symmetry question in crystallography³. If we can show that certain elements - in our case the four-pole matrices - satisfy the group postulates, all the methods and theorems of group theory may be applied to these elements.

The task of systematically treating four-pole theory on the basis of the group concept is very big and in this short discussion we only want to make the first step by actually deriving a few basic theorems which will serve as a starting point for the whole development. We also restrict ourselves here to passive four-poles.

B. Resistanceless Four-Poles in Cascade Arrangement

We start by considering the cascade arrangement of linear resistanceless four-poles at a given frequency.

THEOREM I.

The linear passive resistanceless four-poles form a group G with respect to cascade arrangement.

We represent the four-pole by the matrix

$$F_i = \begin{pmatrix} A_i & B_i \\ C_i & D_i \end{pmatrix}, \quad (1)$$

where for simplicity we assume B_i and $C_i \neq 0$. Equation (1) represents the following relations:

$$\begin{aligned} V_1 &= A_i V_2 - B_i I_2 \\ I_1 &= C_i V_2 - D_i I_2 \end{aligned} \quad (2)$$

The direction of the currents and voltages is as indicated in Fig. 1.

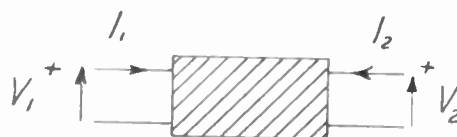


Fig. 1

If we define the vector

$$\begin{pmatrix} V_i \\ I_i \end{pmatrix} = S_i \quad (3)$$

and introduce the matrix

$$M = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (4)$$

The cascade arrangement of the four poles F_1, F_2, \dots, F_n as indicated in Fig. 2

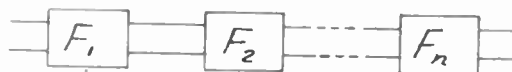


Fig. 2

is described by the equation

$$S_1 = F_1 \cdot F_2 \cdots F_n \cdot M S_2 \quad (5)$$

where the F_i stand for the matrices (1). Since the F_i are to represent linear passive resistanceless four-poles we have the requirement that

$$Re \left[V_1 I_1^* + V_2 I_2^* \right] = 0 \quad (6)$$

for any relation (2). Let us see of what form $A, B, C,$ and D must be to satisfy equation (6). We know from an article by TELLEGÉN and KLAUSS⁴ that (6) is equivalent to

$$V_1 = Z_{11} I_1 + Z_{12} I_2$$

$$V_2 = Z_{21} I_1 + Z_{22} I_2$$

(7)

with Z_{11}, Z_{22} imaginary and

$$Z_{21} = -Z_{12}^*$$

(8)

The asterisk (*) denotes the complex conjugate. We may transform (7) into (2) and find

$$V_1 = \frac{Z_{11}}{Z_{21}} V_2 - \frac{|Z|}{Z_{21}} I_2$$

(9)

$$I_1 = \frac{1}{Z_{21}} V_2 - \frac{Z_{22}}{Z_{21}} I_2$$

where (Z) stands for the determinant

$$|Z| = \begin{vmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{vmatrix}, \text{ real} \quad (10)$$

and we assume that $Z_{21} \neq 0$.

We see therefore that F must have the general form

$$F = C \begin{pmatrix} \frac{A}{C} & \frac{B}{C} \\ 1 & \frac{D}{C} \end{pmatrix} \quad (11)$$

with $A/C, D/C$ imaginary, B/C real, C complex. (12)These relations look a little different from the usual form since the reciprocity relation is not assumed. To prove theorem I we have to show that the set of the matrices F_i form a group with respect to multiplication, which means that they have to fulfill the following postulates:

- Every product of any two matrices and the square of every matrix is a member of the set.
- The associative law holds: $F_1(F_k F_l) = (F_1 F_k)F_l$
- The set contains a unit matrix F_u for which $F_u F_i = F_i F_u = F_i$ for every matrix of the set.
- Every matrix of the set has an inverse, $F_x = F_i^{-1}$, so that $F_i F_x = F_x F_i^{-1} = F_u$

We now prove these four postulates.

Proof of property a) If F_1 and F_k are elements of the group then also $F_l = F_1 F_k$ is also an element of the group:

$$F_i = \begin{pmatrix} A_i & B_i \\ C_i & D_i \end{pmatrix} \quad (13)$$

$$F_k = \begin{pmatrix} A_k & B_k \\ C_k & D_k \end{pmatrix}$$

$$F_l = \begin{pmatrix} A_l & B_l \\ C_l & D_l \end{pmatrix}$$

 F_i and F_k have the same form as equation (11)

$$F_i = C_i \begin{pmatrix} \frac{A_i}{C_i} & \frac{B_i}{C_i} \\ 1 & \frac{D_i}{C_i} \end{pmatrix} \quad (14)$$

$$F_k = C_k \begin{pmatrix} \frac{A_k}{C_k} & \frac{B_k}{C_k} \\ 1 & \frac{D_k}{C_k} \end{pmatrix} \quad (15)$$

where $\frac{A_i}{C_i}, \frac{D_i}{C_i}, \frac{A_k}{C_k}, \frac{D_k}{C_k}$ are purelyimaginary, $\frac{B_i}{C_i}$ and $\frac{B_k}{C_k}$ are real, and C_i, C_k

are complex. By multiplying equations (14) and (15) we get

$$F_l = F_i F_k \quad (16)$$

or

$$F_l = C_i C_k \begin{pmatrix} \frac{A_i}{C_i} \frac{A_k}{C_k} + \frac{B_i}{C_i} \frac{B_k}{C_k} & \frac{A_i}{C_i} \frac{B_k}{C_k} + \frac{B_i}{C_i} \frac{D_k}{C_k} \\ \frac{A_k}{C_k} + \frac{D_i}{C_i} & \frac{B_k}{C_k} + \frac{D_i}{C_i} \frac{D_k}{C_k} \end{pmatrix} \quad (17)$$

where now

$$\frac{A_i}{C_i} \frac{A_k}{C_k} + \frac{B_i}{C_i} \frac{B_k}{C_k} \quad \text{is real,}$$

$$\frac{A_i}{C_i} \frac{B_k}{C_k} + \frac{B_i}{C_i} \frac{D_k}{C_k} \quad \text{is imaginary,} \quad (18)$$

$$\frac{A_k}{C_k} + \frac{D_i}{C_i} \quad \text{is imaginary, and}$$

$$\frac{B_k}{C_k} + \frac{D_i}{C_i} \frac{D_k}{C_k} \quad \text{is real}$$

If we now reduce equation (17) to the form of equation (11), it follows that

$$F_l = C_i C_k \left(\frac{A_k}{C_k} + \frac{D_i}{C_i} \right) \begin{pmatrix} \frac{A_l}{C_l} & \frac{B_l}{C_l} \\ 1 & \frac{D_l}{C_l} \end{pmatrix} \quad (19)$$

and we see that in F_1

$$\begin{aligned} \frac{A_\ell}{C_\ell} &= \frac{A_i A_\kappa + B_i}{C_i C_\kappa + C_i} \left/ \frac{A_\kappa + D_i}{C_\kappa + C_i} \right. \text{ is imaginary,} \\ \frac{B_\ell}{C_\ell} &= \frac{A_i B_\kappa + B_i D_\kappa}{C_i C_\kappa + C_i} \left/ \frac{A_\kappa + D_i}{C_\kappa + C_i} \right. \text{ is real,} \\ C_\ell &= C_i C_\kappa \left(\frac{A_\kappa + D_i}{C_\kappa + C_i} \right) \text{ is complex,} \\ \frac{D_\ell}{C_\ell} &= \frac{B_\kappa + D_i D_\kappa}{C_\kappa + C_i} \left/ \frac{A_\kappa + D_i}{C_\kappa + C_i} \right. \text{ is imaginary,} \end{aligned} \quad (20)$$

Therefore F_1 is an element of the group defined by the invariance of (6).

Proof of property b) The associative law holds for matrix multiplication in general and need not be demonstrated separately in this special case.

Proof of property c) A unit element exists

$$F_u = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (21)$$

such that

$$F_u F_i = F_i F_u = F_i, \quad (22)$$

and this is readily verified. Since (21) cannot be put into form (11) the invariance of (6) has to be proven separately in this case. (21) is the representation of

$$S_1 = F_u M S_2, \quad \begin{pmatrix} V_1 \\ I_1 \end{pmatrix} = \begin{pmatrix} V_2 \\ -I_2 \end{pmatrix} \quad (23)$$

the direct connection of the two pairs of terminals.



Fig. 3

Proof of property d) For every element F_1 there exists an inverse element F_1^{-1} such that

$$F_i F_i^{-1} = F_i^{-1} F_i = F_u \quad (24)$$

The inverse of a matrix

$$F_i = \begin{pmatrix} A_i & B_i \\ C_i & D_i \end{pmatrix} \quad (25)$$

is given by

$$F_i^{-1} = \frac{1}{A_i D_i - B_i C_i} \begin{pmatrix} D_i & -C_i \\ -B_i & A_i \end{pmatrix} \quad (26)$$

or

$$F_i^{-1} = \frac{B_i}{A_i D_i - B_i C_i} \begin{pmatrix} -D_i & C_i \\ B_i & B_i \\ 1 & -A_i \\ -B_i & \end{pmatrix} \quad (27)$$

in which

$$\begin{aligned} \frac{D_i}{B_i} &= \frac{D_i}{C_i} \left/ \frac{B_i}{C_i} \right. \text{ is imaginary,} \\ \frac{C_i}{B_i} &= 1 \left/ \frac{B_i}{C_i} \right. \text{ is real,} \\ \frac{A_i}{B_i} &= \frac{A_i}{C_i} \left/ \frac{B_i}{C_i} \right. \text{ is imaginary,} \end{aligned} \quad (28)$$

Therefore F_1^{-1} belongs to the group.

The physical significance of the inverse element is clear from Fig. 4.



Fig. 4

With this the proof of theorem I is completed.

We notice that the group G is infinite and non-commutative.

C. Passive Linear Four-Poles Containing Resistors

THEOREM II:

The passive linear four-poles containing resistors do not form a group with respect to cascade arrangement.

We prove theorem II by a special example:

The general conditions for the Z matrix of a passive four-pole are

$$Z = \begin{pmatrix} m_1 + jm_1', & n + jn' - l - jl' \\ n + jn' + l + jl', & m_2 + jm_2' \end{pmatrix} \quad (29)$$

where $m_1 \geq 0, m_2 \geq 0, m_1 m_2 - n^2 - l'^2 \geq 0$ (30)

We consider the four-pole

$$Z = \begin{pmatrix} m_1 & n \\ n & m_2 \end{pmatrix} \quad (31)$$

with $m_1 > 0, m_2 > 0, m_1 m_2 - n^2 > 0$ (32)

The corresponding F matrix is

$$F = \begin{pmatrix} \frac{m_1}{n}, & \frac{m_1 m_2 - n^2}{n} \\ \frac{1}{n}, & \frac{m_2}{n} \end{pmatrix} \quad (33)$$

and its inverse

$$F^{-1} = \begin{pmatrix} \frac{m_2}{n}, & -\frac{1}{n} \\ \frac{m_1 m_2 - n^2}{n}, & \frac{m_1}{n} \end{pmatrix} \quad (34)$$

Transforming back to the Z matrix we obtain

$$Z^{-1} = \begin{pmatrix} -\frac{m_2}{m_1 m_2 - n^2}, & -\frac{n}{m_1 m_2 - n^2} \\ -\frac{n}{m_1 m_2 - n^2}, & -\frac{m_1}{m_1 m_2 - n^2} \end{pmatrix} \quad (35)$$

or

$$Z^{-1} = \begin{pmatrix} M_1 & N \\ N & M_2 \end{pmatrix} \quad (36)$$

But here

$$M_1 < 0, M_2 < 0 \quad (37)$$

in contradiction with the condition (30) for the passive four-pole. Therefore theorem II is proven since we showed that for a special case no inverse element exists. Physically this is a very plausible result. It is impossible to restore the power lost in a four-pole containing resistors by means of a passive four-pole. The group theoretical

treatment of the passive four-poles containing resistors therefore has to be taken up together with the active four-poles, after the resistanceless four-poles are sufficiently investigated. For the present purpose we consider the passive four-poles

$$Z = \begin{pmatrix} m_1 + jm_1', & n + jn' - l - jl' \\ n + jn' + l + jl', & m_2 + jm_2' \end{pmatrix} \quad (38)$$

as the sum of the matrices $Z_0 + Z_{RES}$

$$Z = \begin{pmatrix} jm_1, & jn' - l \\ jn' + l, & jm_2 \end{pmatrix} + \begin{pmatrix} m_1, & n - jl' \\ n + jl', & m_2 \end{pmatrix} \quad (39)$$

the second of which is the zero matrix if the network does not contain resistors. Then we divide the passive four-poles into an infinite number of classes containing infinitely many elements, each class having one and only one matrix Z_0 and all the matrices Z_{RES} which satisfy the conditions for passivity. This classification can still be worked out in more detail. We chose the Z_0 as the representatives of these classes and discuss in the following the group theoretical properties of these representatives. In theorem I we have proven that they form a group. Let us now consider its most important sub-group.

D. Passive Linear Four-Poles Satisfying the Reciprocity Relation

THEOREM III:

The passive linear resistanceless four-poles which satisfy the reciprocity relation form a sub-group G_r of the Group G .

If we calculate $|F_1|$ from (9) we get

$$|F_1| = \frac{Z_{12}}{Z_{21}} \quad (40)$$

In the general case of passivity without resistors we have

$$\begin{aligned} Z_{12} &= -(l - jn') = -\alpha \\ Z_{21} &= l + jn' = \alpha^* \end{aligned} \quad (41)$$

and (40) becomes

$$|F_1| = -\frac{\alpha}{\alpha^*} \quad (42)$$

where α^* is the complex conjugate of α . The reciprocity relation for the Z-matrix requires that

$$Z_{12} = Z_{21} \quad (43)$$

or

$$\alpha = -\alpha^* \quad (44)$$

α is imaginary and

$$|F_i| = 1 \quad (45)$$

All that we have to show in addition to what has been shown in equations (13) through (28) is that the combination of two elements

$$F_i \text{ and } F_\kappa \text{ for which} \\ |F_i| = 1, |F_\kappa| = 1 \quad (46)$$

gives also the value 1 for the determinant

$$|F_\ell| = |F_i \cdot F_\kappa| \quad (47)$$

But it is a well known property of matrices that if

$$F_\ell = F_i \cdot F_\kappa \quad (48)$$

then

$$|F_\ell| = |F_i| \cdot |F_\kappa| \quad (49)$$

which with (46) proves (47).

The unit element

$$F_u = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (50)$$

also satisfies (45) and the inverse element with the determinant

$$|F_i^{-1}| = \frac{1}{|F_i|} \quad (51)$$

exists. With this, Theorem III is proven.

So far we found that the group theoretical treatment of passive four-poles containing resistors

has to be taken up together with the active four-poles, and that the passive resistanceless four-poles form a group of which the four-poles that satisfy the reciprocity relation constitute a subgroup. We may also point out that there are many more relations among the four-pole matrices which became apparent during the preparation of this paper but could not be included due to limitations in time and space.

E. The Ideal Gyator

Now let us consider matrices of the general form F_i , the determinant of which has the value

$$|F_i| = -\frac{\alpha}{\alpha^*} \quad (52)$$

This problem is of considerable interest since in connection with net-work synthesis the question has arisen, if the set of known four-pole elements is complete and its investigation resulted in the discovery of the ideal gyator⁵. We are going to describe how the problem is attacked by group theory.

We know already that four-poles which satisfy the reciprocity relation form a group with the determinant

$$|F_i| = 1 \quad (45)$$

It therefore becomes obvious that there must be at least one more element in addition to the classical four-poles which gives rise to a matrix of form (52). To cut the discussion short we at once consider the simplest case

$$\alpha = \alpha^* \quad (53)$$

$$\text{or } |F_i| = -1 \quad (54)$$

$$\text{and } F_i = \begin{pmatrix} 0 & B \\ C & 0 \end{pmatrix} \quad (55)$$

From (54)

$$-BC = -1 \quad (56)$$

or

$$B = \frac{1}{C} \quad (57)$$

and the conditions for the passive Z-matrix

$$Z_{12} = -B, Z_{21} = \frac{1}{C} \quad (58)$$

$$Z_{12} = -(l - jn')$$

(59) a sub-group G_g of G .

$$Z_{21} = l + jn'$$

(60)

$$Z_{12} = -Z_{21}^*$$

(61)

we get

$$B = \frac{1}{C}$$

From (62) and (57)

$$B = \frac{1}{C}$$

it follows that B and C are real.

To thus obtain an element

$$F_i = \begin{pmatrix} 0 & S_i \\ \frac{1}{S_i} & 0 \end{pmatrix}$$

(63)

or its more familiar T-matrix

$$Z = \begin{pmatrix} 0 & -S \\ S & 0 \end{pmatrix}$$

(64)

the ideal gyrator.⁵

It can now be shown that for the realization of the most general passive resistanceless four-pole no additional element is required.

This is an interesting example which shows how by purely group theoretical reasoning we are led to a new four-pole element. One may expect that the same procedure will lead to similar discoveries in the fields of active four-poles and multi-poles in general.

F. Illustrations of the Basic Principle

In the preceding argument the basis for the group theoretical approach to four-pole theory was sketched and we now want to add a few examples of how the whole structure of the set of four-pole matrices may be investigated and classified by group theoretical reasoning.

THEOREM IV:

The ideal gyrators and the ideal transformers form

The combination of two ideal gyrators F_1 and F_k with the gyration resistances s_1 and s_k

$$F_i = \begin{pmatrix} 0 & S_i \\ \frac{1}{S_i} & 0 \end{pmatrix}, F_k = \begin{pmatrix} 0 & S_k \\ \frac{1}{S_k} & 0 \end{pmatrix} \quad (65)$$

(62) gives

$$F_l = \begin{pmatrix} \frac{S_i}{S_k} & 0 \\ 0 & \frac{S_k}{S_i} \end{pmatrix} \quad (66)$$

(57)

an ideal transformer of transformation ratio s_i/s_k . The combination of an ideal transformer F_l and an ideal gyrator F_m

$$F_m = \begin{pmatrix} 0 & S_m \\ \frac{1}{S_m} & 0 \end{pmatrix} \quad (67)$$

gives

$$F_l F_m = \begin{pmatrix} 0 & \frac{S_i S_m}{S_k} \\ \frac{S_k}{S_i S_m} & 0 \end{pmatrix} \quad (68)$$

an ideal gyrator. The cascade arrangement of an even number of ideal gyrators gives an ideal transformer of transformation ratio

$$\frac{S_1 \cdot S_3 \cdots}{S_2 \cdot S_4 \cdots} \quad (69)$$

whereas an uneven number gives an ideal gyrator of the gyration resistance

$$\frac{S_1 \cdot S_3 \cdot S_5 \cdots}{S_2 \cdot S_4 \cdots} \quad (70)$$

The inverse element for

$$F_i = \begin{pmatrix} 0 & S_i \\ \frac{1}{S_i} & 0 \end{pmatrix} \quad (71)$$

is

$$F_i^{-1} = \begin{pmatrix} 0 & S_i \\ \frac{1}{S_i} & 0 \end{pmatrix} \quad (72)$$

The unit element

$$F_u = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (73)$$

belongs to the group. It may be shown that there exists a sub-group of rational gyration resistance. Further we see that the ideal gyrator transformer sub-group is infinite and is the direct product group of the cyclic sub-groups of order 2 generated by each individual ideal gyrator with the gyration resistance

$$F_i = \begin{pmatrix} 0 & S_i \\ \frac{1}{S_i} & 0 \end{pmatrix} = F_i^{-1} \quad (74)$$

$$F_i^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = F_u \quad (75)$$

There are several more types of sub-groups of G_g which were omitted in the present discussion. We only want to mention one more:

THE IDEAL TRANSFORMER SUB-GROUP

Among the matrices with

$$|F_i| = 1 \quad (76)$$

we especially consider the matrices of the form

$$F_i = \begin{pmatrix} a & 0 \\ 0 & \frac{1}{a} \end{pmatrix}, \quad a \text{ real,} \quad (77)$$

the ideal transformers the passivity of which can be proven directly. We thus have

THEOREM V:

The ideal transformers form a group G_t which is a sub-group of the gyrator - transformer sub-group with the determinant $|F_i| = 1$ and thus satisfying the reciprocity relation.

Proof:

$$F_i = \begin{pmatrix} a_i & 0 \\ 0 & \frac{1}{a_i} \end{pmatrix} \quad (78)$$

$$F_\kappa = \begin{pmatrix} a_\kappa & 0 \\ 0 & \frac{1}{a_\kappa} \end{pmatrix} \quad (79)$$

$$F_i F_\kappa = \begin{pmatrix} a_i a_\kappa & 0 \\ 0 & \frac{1}{a_i a_\kappa} \end{pmatrix} \quad (80)$$

We see that G_t is commutative.

The inverse element for

$$F_i = \begin{pmatrix} a_i & 0 \\ 0 & \frac{1}{a_i} \end{pmatrix} \quad (81)$$

is

$$F_i^{-1} = \begin{pmatrix} \frac{1}{a_i} & 0 \\ 0 & a_i \end{pmatrix} \quad (82)$$

The unit element belongs to the group. It can be shown that there exists a sub-group of rational transformation ratio.

G. Concluding Remarks:

The preceding paper may serve as a starting point for a program of investigating circuit theory from the basis of group theory. After discussing the passive four-poles and classifying them according to the different sub-groups analogous considerations will be applied to active four-poles and 2n-poles in general, as it is felt that the group theoretical approach may prove useful in systematizing network theory. At the same time connections with the many existing theorems of four-pole theory will have to be established. Finally the frequency dependence of four-poles as reflected in group theoretical language will have to be investigated and evaluated with respect to their contribution to the general problem of linear circuit synthesis.

Acknowledgements:

I would like to thank Dr. H. Fuchs of Siemens & Halske Ges. m. b. H., Vienna, in discussions with whom I received many valuable ideas which finally led to the preparation of this paper.

References

1. General references for group-theory: A. Speiser, "Theorie der Gruppen von endlicher Ordnung", J. Springer, Berlin 1927. W. Burnside, "The Theory of Groups", Cambridge Univ. Press, 1927. H. Margenau and G. M. Murphy, "The Mathematics of Physics and Chemistry", D. van Nostrand, New York, 1943.
2. W. Cauer, "Theorie der linearen Wechselstrom-schaltungen", AVG, Leipzig 1941; P. le Corbeiller, "Matrix Analysis of Electric Networks", Harvard Univ. Press 1950.
3. J. J. Burckhardt, Die Bewegungsgruppen der Kristallographie.
4. Philips Res. Rep. 5, 81, 1950
5. B. D. H. Tellegen, Philips Res. Rep. 5, 81; 1948

A MATHEMATICAL TECHNIQUE FOR THE ANALYSIS OF LINEAR SYSTEMS

John R. Ragazzini and Arthur R. Bergen
Department of Electrical Engineering
Electronics Research Laboratories
Columbia University
New York 27, New York

Introduction

The specification of a linear system or network often takes the form of its response in the time domain to a test function. A common test function is the unit step or ramp function although there are many cases where the input function is an arbitrary function of time. In any case, the determination of the response of the system is a relatively straightforward matter involving the solution of a set of linear differential equations. However, while straightforward, obtaining numerical values for the response as a function of time is an involved and tedious task not readily amenable to the use of desk calculators and similar computational aids. The alternative has been to apply numerical methods leading to numerical results which closely approximate the actual solution.

The application of numerical methods to the solution of linear differential equations is not new¹. It has also been recognized that the program for obtaining a solution can be described in terms of operational procedures^{2,3,7,8}. What has been lacking in previous work, however, has been an interpretation of the methods being employed in terms of the physical problem as well as a simplified and orderly technique for setting up the problem by the average engineer. By applying techniques which were developed primarily for the analysis and synthesis of sampled-data feedback control systems, this order and relationship to the physical problem can readily be established. The technique of the z-transformation^{4,5} or pulse transfer function⁶ has been employed to solve problems in feedback control systems in which data is being sampled at one or more points. This established and direct approach is the one which has been applied in this paper to the numerical solution of linear systems.

The z-Transformation

In order to describe the z-transformation, reference is made to Figure 1 where a time function $r(t)$ is sampled by means of a sampling switch S at equal intervals of time separated by a time T . The output of the switch $r^*(t)$ is a sequence of pulses which may be represented for mathematical purposes as a series of impulses or delta functions whose areas are equal to the amp-

litude of $r(t)$ at the respective sampling instants. Thus

$$r^*(t) = \sum_{n=-\infty}^{+\infty} r(nT)\delta(t-nT) \quad (1)$$

The Laplace transform of the pulse sequence is given as

$$\mathcal{L} [r^*(t)] = \sum_{n=-\infty}^{+\infty} r(nT) e^{-nTs} \quad (2)$$

It is noted that the Laplace transform of such pulsed functions always is in terms of e^{-nTs} and it has become customary to replace this exponential by a new variable $z = e^{Ts}$ thus accounting for the name "z-transform." A very useful characteristic of these transforms is that the infinite series given in equation 2 may usually be expressed in closed form so that tables of z-transforms can be constructed^{5,6} in much the same manner as tables of ordinary Laplace transforms.

One of the most useful properties of the z-transform is that it is possible to obtain the pulsed output of a linear system by expressing a pulse transfer function $G^*(z)$ which relates the input and output z-transforms. Schematically this is shown in Figure 2 where the output $c(t)$ is sampled synchronously with the input to give a pulse sequence $c^*(t)$. In terms of z-transforms, this relation is expressed as

$$C^*(z) = W^*(z) R^*(z) \quad (3)$$

It is readily shown that $W^*(z)$ must be the Laplace transform of the sampled output $c^*(t)$ when the input $r(t)$ is a unit impulse. The pulse transfer functions $W^*(z)$ can be expressed in closed form and tables of such transforms are available in standard works on sampled data systems^{5,6}. An abridged table is given in Table I. Thus, the process of obtaining the transform of the pulsed output of a linear system subjected to a pulsed input is exactly analogous to that for continuous systems employing the ordinary Laplace transform. To obtain the output pulse sequence in the time domain, the z-transform $C^*(z)$ can

be inverted by contour integration, by reference to available tables or far more simply for the purpose at hand, by a process of simple long division. This will be demonstrated later in an illustrative example.

Feedback systems can be treated by this technique by deriving relationships between the pulsed input and pulsed output for various configurations⁵. One such configuration which is of particular value to the application being discussed in this paper is shown in Figure 3b. The relation between the z-transform of the input and output is given by:

$$C^*(z) = \frac{RG^*(z)}{1 + HBG^*(z)} \quad (4)$$

where $RG^*(z)$ signifies the z-transform of $R(s)G(s)$ and $HBG^*(z)$ the pulse transfer function corresponding to $H(s)B(s)G(s)$. It is noted in passing that the z-transforms corresponding to $R(s)G(s)$ or $H(s)B(s)G(s)$ are not equal to the products $R^*(z)G^*(z)$ and $H^*(z)B^*(z)G^*(z)$ respectively. The whole transform $R(s)G(s)$ or $H(s)B(s)G(s)$ must be looked up in the tables. Inversion of equation 4 yields the sampled output time function in response to an input $r(t)$.

Application of Method To Continuous Systems

The application of the z-transform technique to the transient solution of a linear continuous system will be illustrated by applying it to feedback systems since this is where the most saving in complexity occurs. The technique used is to replace the continuous system shown in Figure 3a with the sampled model of Figure 3b whose response adequately approximates the actual response. The system is assumed to be low-pass so that the output $c(t)$ or $c_m(t)$ contains but few high frequency components. The output is sampled by S at a frequency which is high enough to cause the error introduced by the model to be small. This requires a choice of sampling frequency at least twice the highest frequency component expected at the output.

It is assumed that the frequency at which the system's frequency response is down approximately 40 db marks the highest expected output frequency. At this frequency, only the forward gain need be considered. Therefore, as a general rule the sampling frequency should be chosen such that it is at least twice the frequency for which the forward gain is down approximately 40 db.

The output $c(t)$ in Figure 3b which is fed back to the input through the switch S

is a pulse sequence at the point x and bears little resemblance to the output which is fed back in the actual system. To reconstruct more closely the actual function fed back, an element B is introduced. B can be chosen to reconstruct the original fed back function to as high a degree of accuracy as desired. However, to prevent undue and generally unwarranted complexity a simple element is introduced which reconstructs a polygonal approximation to the actual function by converting each pulse into a triangular generating function as shown in Figure 4. This technique^{2,3} can be represented by a block described by a transfer function $B(s)$. The requirement on $B(s)$ is that its response to a short pulse or, in the limit, an impulse of the same area, be a triangle as shown in Figure 5a. Such a response is not physically realizable since it is initiated at a negative time but this is of no concern since the model is only mathematical in nature. The combination of the sampling switch operating synchronously with a period T and the triangle function generator B replace the direct connection which existed in the actual system. Any errors in computation resulting from use of the model are caused by the failure of this combination to faithfully reproduce the output $c(t)$ in the feedback line.

The transfer function of the triangle function generator is obtained by taking the Laplace transform of the triangle time function in Figure 5a. It is observed by referring to Figure 5b that this impulsive response is the sum of three ramps, one advanced by the sampling time T and having a positive slope of $1/T$, the second, a ramp at the origin having a negative slope of $2/T$ and third, a ramp delayed by a sampling time T having a positive slope of $1/T$. The Laplace transform of these three components is

$$B(s) = \frac{1}{Ts^2} e^{Ts} - \frac{2}{Ts^2} + \frac{1}{Ts^2} e^{-Ts}, \quad (5)$$

which becomes

$$B(s) = \frac{e^{-Ts}}{Ts^2} (1 - e^{-Ts})^2 \quad (6)$$

The Laplace transform of the loop gain of a feedback system or the cascaded gain of an open loop system is the product of the transforms of the various components. Applying the z-transformation to these components as indicated in equation 4 is an operation which yields the pulse transfer functions for which tables are available. For transforms which contain only powers of s and exponential terms

such as those in equation 6 the z-transform is quite simple as shown by reference to Table I.

Thus, if it is desired to obtain the response of a feedback system such as that of Figure 3a to a test function whose Laplace transform is $R(s)$, the first step is to replace it with a model as shown in Figure 3b. A choice of sampling frequency T is made using the guides previously described. The pulse transfer functions $RG^*(z)$ and $HGB^*(z)$ are then computed. These will generally be in the form of a ratio of polynomials in z^{-1} . To obtain the output of the model system at sampling instants, substitution is made in equation 4 and an inversion of the output z-transform $C^*(z)$ is carried out. This may be done using standard inversion theorems but for this purpose, it is far more useful to first expand $C^*(z)$ into a power series in z by long division. $C^*(z)$ is then in the form

$$C^*(z) = c_0 + c_1 z^{-1} + c_2 z^{-2} + c_3 z^{-3} + c_4 z^{-4} + c_5 z^{-5} + \dots + c_n z^{-n} + \dots \quad (7)$$

The magnitude of the output of the model system at sampling instants is simply the coefficient of each of the terms in the expansion at a delay corresponding to the instant in question. As an example, the output at the fifth sample time is c_5 in equation 7.

Since the z-transform of the output pulse sequence $C^*(z)$ is the ratio of two polynomials in z , the process of expansion of $C^*(z)$ into a power series in z is carried out by long division. This may be done either by hand or by means of a desk calculator. The process may be terminated whenever the desired number of points is obtained. This procedure is clarified by means of the illustrative example which follows later. An important advantage which results from this method of inversion is that it is unnecessary to obtain the roots of the denominator of $C^*(z)$ or $C(s)$. This is generally the most time consuming operation in the inversion procedure for higher order feedback systems. On the other hand, unless extensive tables of z-transforms are available, the roots of the internal transfer functions $G(s)$ and $H(s)$ must be found. In feedback systems, this is only a minor problem since the linear system generally consists of decoupled components having relatively simple individual transfer functions. In obtaining the response of linear networks whose transfer functions can be expressed only in terms of higher order polynomials, however, the problem of obtaining roots still remains.

Illustrative Example

To illustrate the application of the technique, the response to a step function of the system shown in Figure 6a will be computed. Following the rule of selecting a sampling frequency at a value where the forward gain is about -40 db the convenient value of 1 sample per second is chosen. The forward gain is -30 db at this frequency. The sampling switch and triangle generator next replace the direct feedback path of the actual system as shown in the block diagram of the model in Figure 6b. The Laplace transform of the sampled output, $C^*(z)$ is related to the input by the z-transform relationship given in equation 4. The numerator of this expression, $RG^*(z)$ is the z-transform corresponding to the Laplace transform, $R(s)G(s)$:

$$RG^*(z) = \mathcal{Z} \left[\frac{1}{s^2(s+1)} \right] = \mathcal{Z} \left[\frac{1}{s^2} - \frac{1}{s(s+1)} \right] \quad (8)$$

From Table I this becomes (for $T = 1$)

$$RG^*(z) = \frac{z^{-1}}{1-z^{-1}} - \frac{z^{-1}(1-\epsilon^{-1})}{(1-z^{-1})(1-\epsilon^{-1}z^{-1})} \quad (9)$$

which upon substitution of values and simplification becomes

$$RG^*(z) = \frac{z^{-1}}{1-z^{-1}} \left[\frac{0.3679 + 0.2642z^{-1}}{(1-z^{-1})(1-0.3679z^{-1})} \right] \quad (10)$$

Similarly, the term $HGB^*(z)$ is the z-transform corresponding to the Laplace transform $H(s)G(s)B(s)$:

$$HGB^*(z) = \mathcal{Z} \left[\frac{\epsilon^8(1-\epsilon^{-8})^2}{s^3(s+1)} \right] \quad (11)$$

which from Table I simplifies to

$$HGB^*(z) = z(1-z^{-1})^2 \mathcal{Z} \left[\frac{1}{s^3} - \frac{1}{s^2(s+1)} \right] \quad (12)$$

which with further application of Table I and with the substitution of numerical values becomes

$$HGB^*(z) = \frac{0.1321 + 0.4198z^{-1} + 0.0802z^{-2}}{(1-z^{-1})(1-0.3679z^{-1})} \quad (13)$$

Substituting the individual components from

equation 10 and equation 13 into equation 4 and simplifying, the z-transform of the output sequence becomes

$$C^*(z) = \frac{0.3679z^{-1} + 0.2642z^{-2}}{1.1321 - 2.0802z^{-1} + 1.3962z^{-2} - 0.4481z^{-3}} \quad (14)$$

It is recalled that this expression is the Laplace transform of the output C taken at sampling instants only. Inversion of equation 14 yields the values of the output at each sampling instant. This is most conveniently done by expanding equation 14 into a power series in z^{-1} by means of simple long division. By carrying out the long division with the aid of a desk calculator, the z-transform of the output pulse sequence is obtained:

$$C^*(z) = 0.3250z^{-1} + 0.8305z^{-2} + 1.125z^{-3} + 1.1721z^{-4} + 1.094z^{-5} + 1.0111z^{-6} + 0.9718z^{-7} + \dots \quad (15)$$

The inversion of this expression into the time domain is simple since the heights of the output ordinates of the time function at each sampling instant is the coefficient of the corresponding term in z^{-n} . These ordinates are plotted in Figure 7. For purposes of comparison, the curve for the actual response obtained from the exact solution of the differential equation is also plotted. Correspondence of the two results in no case exceeds two per cent at sampling instants where both functions are accurately known.

It is to be noted that the example worked out above is a simple one which might be solved just as well without recourse to numerical methods. However, even this simple case demonstrates that at no time was it necessary to solve the characteristic equation of the overall system for its roots and that the operations involved are all simply handled on a calculating machine. The use of z-transform tables furthermore makes procedures used in solving the problem analogous to those normally employed in solving similar time domain problems.

Arbitrary Inputs

The previous discussions centered about those cases where the Laplace transform of the input function, $R(s)$, is known explicitly. Often, the input function is not so available but is given merely as an experimental curve or number series representing a curve which is not readily Laplace transformable. In that case, the

procedures outlined previously can be applied directly by using the number itself or sampled sequence in place of $R^*(z)$. For instance, if an arbitrary function designated by $r(t)$ were applied to the system, then the transform of the sampled function, $R^*(z)$ would be merely

$$R^*(z) = r(0) + r(T)z^{-1} + r(2T)z^{-2} + \dots + r(nT)z^{-n} + \dots \quad (16)$$

This cannot ordinarily be expressed in closed form but would be carried as a number series with as many terms as desired in the expressions for the output sequence transform.

For use in such cases, the computation models which more conveniently apply are shown in Figure 8 for the open and closed circuit cases respectively. The triangle function generator is $B(s)$ as before. For the feedback case, the location of the sampler and triangle generator shown here is not as desirable as that used with systematic inputs. The reason is that it is desirable, from the viewpoint of accuracy, to place the sampler in a position where the high frequency content of the time function is minimal. In the case of low-pass feedback systems, this generally occurs at the output of the forward transmission path. Nevertheless, the error introduced by the high frequency components of the sampler can be minimized by proper choice of the sampling frequency. For the cases shown in Figure 8, the z-transforms of the output $c^*(t)$ are given by

$$C^*(z) = BG^*(z) R^*(z) \quad (17)$$

and

$$C^*(z) = \frac{BG^*(z)}{1+HGB^*(z)} R^*(z) \quad (18)$$

respectively. It is noted that $R^*(z)$ appears here as a separate component and that it can be inserted independently as a power series in z^{-1} of the form given in equation 16.

Computation Errors

One of the problems common to numerical methods is that of determining the error introduced by approximating a differential equation by a corresponding difference equation. In cases under discussion, this resolves itself into the problem of ascertaining the error in the results introduced by a specified sampling frequency or, conversely, the determination

of the sampling frequency required to obtain a specified error. Explicit solutions for sampling frequency to obtain a specified error are not, in general, known. However, an approximate error can be easily obtained for various specific system configurations, using the techniques previously described. As an example an expression for the error in the system of the illustrative example will be derived.

The definition of error can be best understood by referring to Figure 9. Essentially, the difference between the output of the actual system and model system is \mathcal{E} , the computation error. The Laplace transform of this error is given by

$$\mathcal{E}(s) = C(s) - C_m(s) \quad (19)$$

The Laplace transform of the sampled output of the model, $C_m^*(s)$ can be expanded into an infinite series⁵

$$C_m^*(s) = \frac{1}{T} \sum_{n=-\infty}^{+\infty} C_n(s+nj\omega_0) \approx \frac{1}{T} C_m(s) \quad (20)$$

This approximation holds for the system considered here because the contributions of other than the central term of the series, $C(s)$, are small, since the sampling frequency is chosen to be high compared to the pass band of the system.

Substituting the central term only in equation 19, there results for the system error

$$\mathcal{E}(s) \approx \frac{G(s)R(s)}{1+G(s)} - \frac{R(s)G(s)}{1+\frac{1}{T}B(s)G(s)} \quad (21)$$

Simplifying this expression, and making use of the fact that $G(s)$ and $B(s)G(s)/T$ are approximately equal,

$$\mathcal{E}(s) \approx \frac{[\frac{1}{T}B(s) - 1] [G(s)]^2 R(s)}{[1+G(s)]^2} \quad (22)$$

The triangle function generator transfer function $B(s)$ is given in equation 6 in terms of exponential operators. If these exponentials are expanded into a power series, equation 22 becomes

$$\mathcal{E}(s) \approx \frac{T^2 s^2}{12} W(s)C(s) + \frac{T^4 s^4}{360} W(s)C(s) + \dots \quad (23)$$

where $W(s)$ represents the overall transmission function,

$$W(s) = \frac{G(s)}{1+G(s)} \quad (24)$$

Consistent with the approximation made in equation 20 only the first term of equation 23 need be considered. Thus a model for obtaining the approximate computation error in a low-pass feedback system is given in Figure 10. The output of the triangle function generator, $C_B(s)$ is a polygonal approximation of the actual output $C(s)$ and, for purpose of error computation, differs from it only negligibly. Thus, it may be stated that

$$C(s) \approx C^*(s)B(s) \quad (25)$$

The Laplace transform of the error function then becomes

$$\mathcal{E}(s) \approx \frac{T^2 s^2}{12} B(s)W(s)C^*(s) \quad (26)$$

$$\mathcal{E}^*(z) \approx \frac{T}{12} (z-2-z^{-1}) C^*(z)W^*(z) \quad (27)$$

where $W^*(z)$ is the pulse transmission function of the system.

Now if the original problem being computed was the response of the system to a unit step input, $W^*(z)$ can be found with acceptable accuracy by obtaining the first back difference of the approximate output already computed. While a delay of about one-half a sampling interval is introduced this is readily eliminated by shifting forward the first difference by one-half a sampling interval. Thus,

$$W^*(z) \approx z^{\frac{1}{2}} \left[\frac{1}{T} (1-z^{-1}) C_m^*(z) \right] \quad (28)$$

If this technique is applied to the illustrative problem given previously, the value of the error at sampling instants is shown in Figure 7. The error can be compared to an exact computation of error and it is noted that the results agree closely. At least insofar as the illustrative example is concerned, the general rule of sampling a low-pass system at a frequency at which the forward transmission is down approximately 40 db is a good one leading to errors in the order of a few per cent.

Conclusions

The physical interpretation of numerical techniques for obtaining solutions described by linear differential equation

is a fruitful one. It leads to an understanding of the effect of the sampling procedure on the solution. In addition, the expression of numerical techniques in terms of the z-transform leads to a solution in terms of a sampled-data system analogous to the original continuous system. These sampled-data systems have been studied extensively and are in a state of development which makes their use by engineers a straightforward procedure. Thus, the numerical solution of the output of linear systems in response to either systematic or arbitrary input functions is a practical process comparable to conventional transform methods.

Acknowledgement

This research was supported in part by the United States Air Force under Contract No. AF 18(600)-677 monitored by the Office of Scientific Research, Air Research and Development Command.

References

1. Salvadori and Baron, "Numerical Methods in Engineering," (Book) Prentice-Hall Inc., 1952.
2. Tustin, A., "A Method of Analyzing the

Behavior of Linear Systems in Terms of Time Series," Jour. IEE, Vol. 94, Part II-A, No. 1, May 1947.

3. Madwed, A., "Number Series Method of Solving Linear and Nonlinear Differential Equations," Report Number 6445-T-26, Instrumentation Laboratory, M.I.T., April, 1950.
4. Hurewicz, W., "Theory of Servomechanisms, James, Nichols and Phillips," Vol. 25, Chap. 5, Radiation Lab. Series, McGraw Hill, 1947.
5. Ragazzini and Zadeh, "The Analysis of Sampled-Data Systems," AIEE Trans., Vol. 71, Part II, 1952, pp. 225-232.
6. Barker, R. H., "The Pulse Transfer Function and its Application to Sampling Servo-Systems," Monograph No. 43, Measurement Section, Institution of Electrical Engineers, (London) July, 1952.
7. Piety, R. G., "A Linear Operational Calculus of Empirical Functions, Phillips Petroleum Co., Research Division, Report 106-12-51R, 1951.
8. Bubb, F. W., "A New Linear-Operational Calculus, AF Tech. Report 6581, WADC, May, 1951.

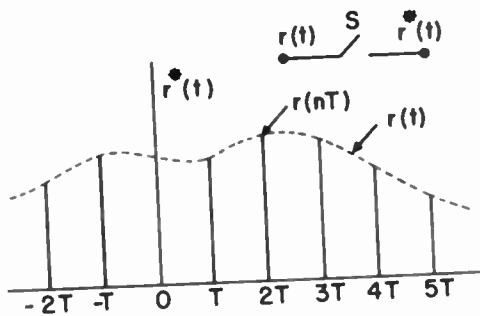


Fig. 1 - Sampled time function.



Fig. 2 - Pulsed linear system.

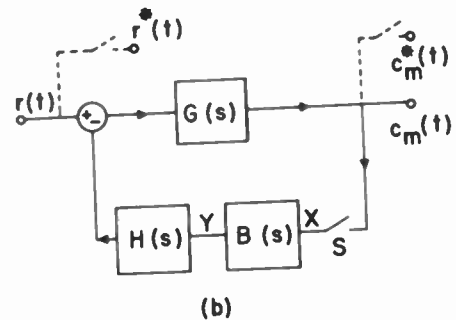
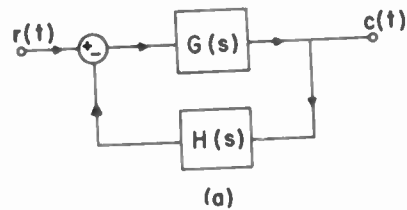


Fig. 3
(a) Continuous feedback system;
(b) sampled model for computation.

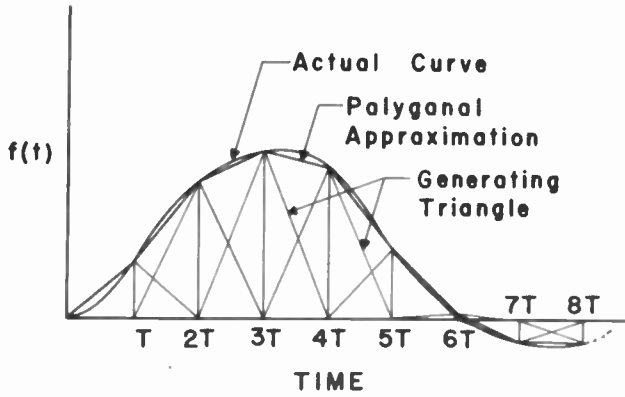


Fig. 4
Polygonal approximation of time function.

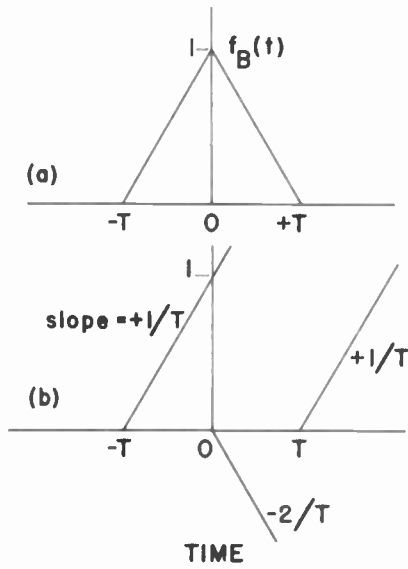


Fig. 5
Impulsive response of triangle generator.

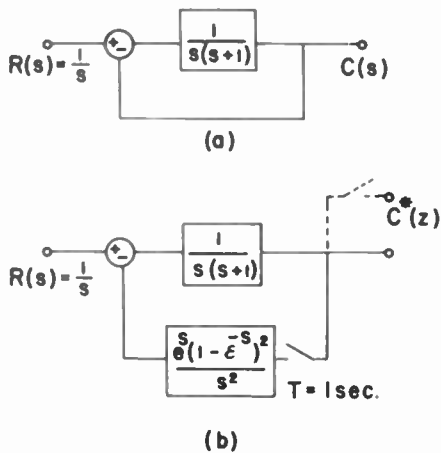


Fig. 6 - System used in illustrative example.

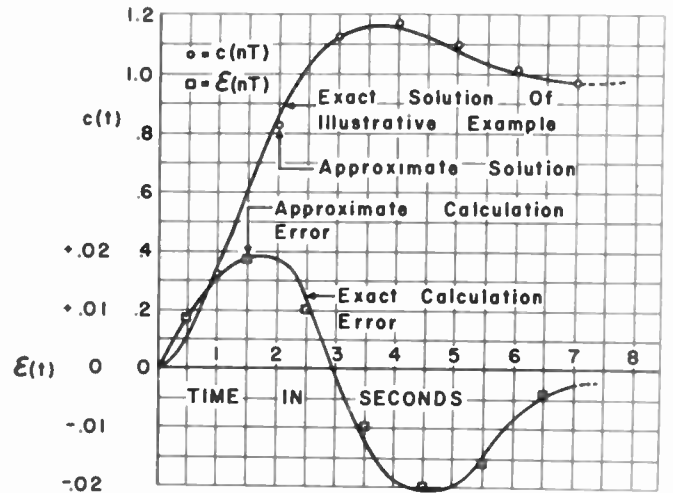


Fig. 7
Computed response and error curves for illustrative example.

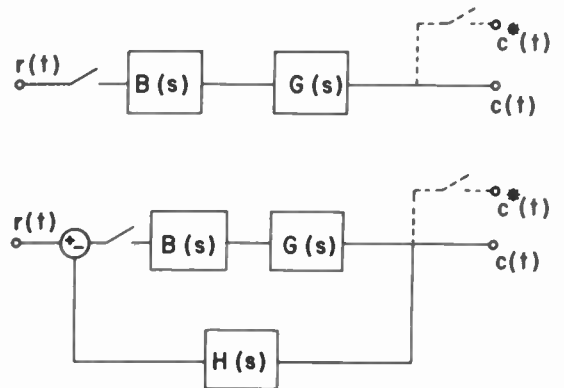


Fig. 8 - Models used for arbitrary inputs.

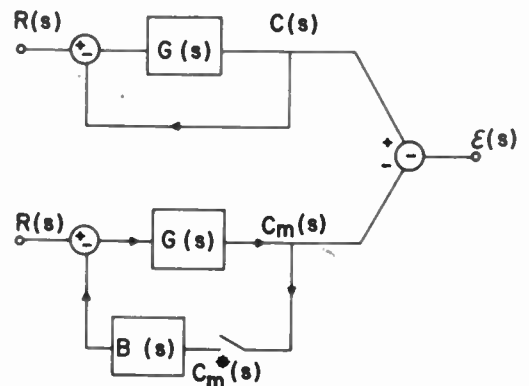


Fig. 9 - System for computation of error.

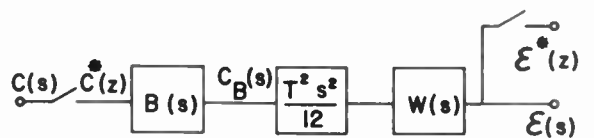


Fig. 10
Model for computing error in feedback system.

Table I
Table of corresponding Laplace and z-transforms.

Laplace Transform $F(s)$	z-Transform $F^*(z)$
$U(\epsilon^{Ts})V(s)$	$U^*(z)V^*(z)$
e^{-nTs}	z^{-n}
$\frac{1}{s}$	$\frac{1}{1-z^{-1}}$
$\frac{1}{s^2}$	$\frac{Tz^{-1}}{(1-z^{-1})^2}$
$\frac{1}{s^3}$	$\frac{T^2 z^{-1}(1+z^{-1})}{2(1-z^{-1})^3}$
$\frac{1}{s+a}$	$\frac{1}{1-e^{-aT}z^{-1}}$
$\frac{a}{s(s+a)}$	$\frac{z^{-1}(1-e^{-aT})}{(1-z^{-1})(1-e^{-aT}z^{-1})}$
$\frac{a}{s^2+a^2}$	$\frac{z^{-1} \sin aT}{1-2z^{-1} \cos aT+z^{-2}}$
$\frac{s}{s^2+a^2}$	$\frac{1-z^{-1} \cos aT}{1-2z^{-1} \cos aT+z^{-2}}$

WEIGHTING FUNCTIONS FOR TIME VARYING FEEDBACK SYSTEMS

By

J. A. Aseltine and R. R. Favreau

Research and Development
Guided Missile Division
Hughes Aircraft Company
Culver City, California

Introduction

The analysis of linear time varying systems is inherently more difficult than the analysis of constant coefficient systems. There are two reasons for this. In the first place the response of a linear time varying system is a function of two variables, namely the time at which a disturbance is applied, and the time at which the response is measured. This is contrasted to a constant coefficient system when the response is a function of only one variable, the time difference between the application of a disturbance and the measurement of the response. In the second place the response of a linear time varying system must usually be described by more complex functions than the exponential functions which describe the response of constant coefficient systems.

The pencil and paper analysis of linear time varying systems is usually ruled out because the complex functions which are required to describe their response are either not known or not tabulated. For this reason and because of time considerations the analysis of linear time varying systems is usually performed upon an analog computer. While the computer eliminates the difficulties associated with complicated functions required to describe the response, it does not overcome the added complexity involved because the response is a function of two variables. In this paper we will describe a technique which may be used to instrument a linear time varying system upon an analog computer which to a certain extent reduces the complexity associated with describing a system response which is a function of two variables.

The Analysis of Linear Systems

There are two basic methods of analyzing linear constant coefficient systems. The first method uses the steady state response to a sine wave of various frequencies to describe the system. The second method uses the transient response to an impulse or step function to describe the system. Once either of these characteristics is known the response of the system to any arbitrary input may be determined.

Either of these techniques may be used to analyze linear time varying systems. However, the

transient response method is by far the simplest when an analog computer is used in the analysis. There are two reasons for this. In the first place the frequency response of a linear time varying system is a function of time. If the response of the system is to be analyzed at more than one instant a variety of frequency response curves is necessary. In the second place an analog computer operates in the time domain and the obtaining of frequency response characteristics, while possible, represents a long and tedious procedure. In addition frequency response curves cannot be operated upon by the computer for further system analysis.

The transient response method for linear time varying systems lends itself quite readily to an analog computer because a time varying system usually appears to be in the transient state. In addition the impulse or step function responses which are used to characterize the system are functions of time and can therefore be operated upon by the computer for further analysis.

The analysis of a system usually reduces to determining the response of that system to arbitrary driving functions. The response of the system to initial conditions can be obtained by solving for appropriate equivalent driving functions. If the response of the system to an impulse is known, we may determine the response to an arbitrary input by using the convolution integral. Let $h_{-1}(t, t_1)$ be the response of the system at time t to an impulse applied at time t_1 . The response θ_0 to an arbitrary driving function of $f(t)$ is obtained from¹

$$\theta_0(t) = \int_0^{\infty} f(t_1) h_{-1}(t, t_1) dt_1 \quad (1)$$

Here $f(t_1)$ has been assumed to be a nonstatistical function. In the case where we have white noise input with spectral density W_0 , the standard deviation of the output (σ) may also be obtained from the weighting function.²

$$\sigma^2(t) = 1/2 \int_0^{\infty} W_0(t_1) [h_{-1}(t, t_1)]^2 dt_1 \quad (2)$$

In practice, we generally use normally distributed noise since in this case the probability

distribution of the input is not changed by the system and the response may be described by two simple parameters, the mean and the standard deviation.

If we attempt to use an analog computer to perform the above mathematical operations we run into some difficulty. The reason for this is that the integrating procedures described above are with respect to t_1 , the time at which the impulse is applied. If our system is simulated upon the computer and an impulse introduced, the resulting weighting function is a function of t , (the time of measurement of the response) for a particular value of t_1 . The computer could be operated a number of times always introducing the impulse at a different time t_1 and measuring the response at a particular value of t . The weighting function, $h_1(t, t_1)$ could then be determined point by point for a particular value of t as a function of t_1 . This procedure is long and does not allow the computer to perform the integrations described in Equations (1) and (2), except on a greatly expanded time scale.

The analog of a linear time varying system, therefore, does not provide us with the desired weighting function for easy system analysis. More desirable would be a system which provides us with the impulse response as a function of the time at which the impulse is introduced. This new system is called the adjoint system, and its validity and instrumentation have been described by Laning and Battin.³

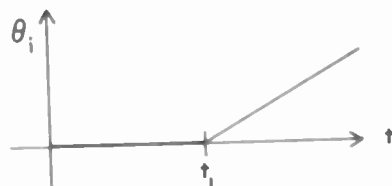
We will show how the adjoint system can be synthesized from the original system from the block diagram, using relations between various weighting functions. Long mathematical argument is avoided, so that the physical nature of the problem becomes more apparent. A short appendix contains a discussion of the mathematical aspects of the problem, but this material is not necessary to the development.

Some Relations Between Weighting Functions

We will need the relations developed below in the sections to follow.

In the transient analysis of constant coefficient systems one is usually concerned with the response of the system to impulses, steps, ramps, etc. Because superposition applies, it may easily be shown that the response to a step is the time derivative of the response to a ramp. In addition the response to an impulse is the time derivative of the response to a step. Similar characteristics may be shown to apply to linear time varying systems.

We call the response of a system to an impulse or any of its derivatives or integrals, "weighting functions". We first consider the system shown in Figure 1. ($u(t)$ is the unit step function).



$$\theta_i = (t - t_1) u(t - t_1)$$



Figure 1
Ramp Response

Combining two ramps of magnitude $1/\Delta t_1$, and separated by Δt_1 , we have the result shown in Figure 2.

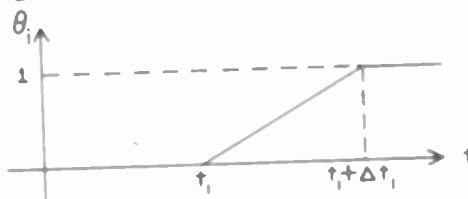


Figure 2
Step Response

Now we go to the limit as $\Delta t_1 \rightarrow 0$. The input in Figure 2 then becomes a unit step, and if we call the response to this $h_0(t, t_1)$, we have

$$h_0(t, t_1) = -\frac{\partial h_1(t, t_1)}{\partial t_1} \quad (3)$$

Similarly, (4)

$$h_{-1}(t, t_1) = -\frac{\partial h_0(t, t_1)}{\partial t_1} = \frac{\partial^2 h_1(t, t_1)}{\partial t_1^2}$$

It is interesting to note these derivative relations between the weighting functions are all with respect to t_1 , the time at which a disturbance is applied. These relationships apply to all linear time varying systems. It would now seem reasonable to look for some relationship between the weighting functions which is particular to the system to be analyzed. If this can be done we may then formulate a differential equation with the desired impulse response as the dependent variable and the time at which a disturbance is

applied, t_1 , as the independent variable.

Also of interest is the response of a time varying gain element to impulse and doublet functions. We have summarized the results in Figure 3. The result in Figure 3b follows from Figure 3a, if we differentiate input and output with respect to t_1 , a procedure justified by the superposition property of linear systems.

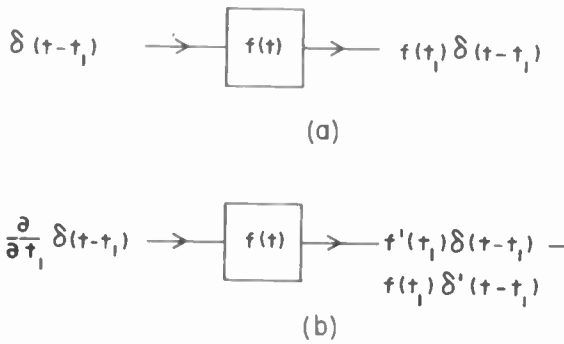


Figure 3
Impulse and Doublet Response

Time Varying Feedback System

Consider the feedback system shown in Figure 4.

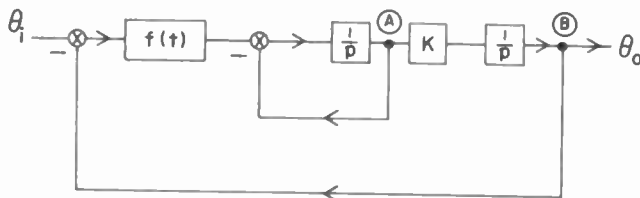


Figure 4
Time Varying Feedback System

Here $1/p$ is the transfer function representing integration. There are two integrations in the above loop and the system in general would have two initial conditions. Let us determine the driving functions which are equivalent to a unit initial condition at point (B) at time t_1 .

Suppose we have a δ function at time t_1 at the input:

$$\theta_i(t) = \delta(t-t_1) \quad (5)$$

It is easy to see that this particular input will

give an initial condition of magnitude $f(t_1)$ at point (A) and nothing at point (B)

Next consider a doublet at the input:

$$\theta_i(t) = \delta'(t-t_1) \quad (6)$$

The output of the $f(t)$ block is

$$f(t_1)\delta'(t-t_1) - f'(t_1)\delta(t-t_1)$$

The input (6) produces an impulse of magnitude $f(t_1)$ and an initial condition of magnitude

$-f'(t_1)$ at point (A). The impulse continues through the circuit to produce an initial condition of magnitude $Kf(t_1)$ at point (B) and magnitude $-f(t_1)$ at point (A). The input required to produce a unit initial condition at point (B) alone can be made up of a combination of (5) and (6) in the following manner.

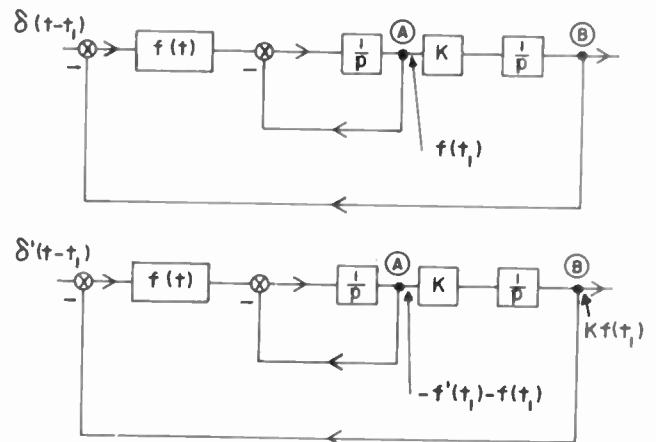


Figure 5
Effects of Impulse and Doublet

To obtain the desired unit initial condition at (B) we need, referring to Figure 5, an input of $\frac{1}{Kf(t_1)} \delta'(t-t_1)$. But this places at point (A) the initial condition:

$$-\frac{f'(t_1)}{Kf(t_1)} - \frac{1}{K} = -\frac{1}{K} \left(1 + \frac{f'(t_1)}{f(t_1)}\right)$$

Now this can be cancelled if we introduce a δ function at the input of magnitude

$$\frac{1}{f(t_1)} \left[\frac{1}{K} \left(1 + \frac{f'(t_1)}{f(t_1)}\right) \right]$$

Therefore, the total input required to place a unit initial condition at (B) is

$$\theta_i = \frac{1}{Kf(t_1)} \delta'(t-t_1) + \frac{1}{Kf(t_1)} \left(1 + \frac{f'(t_1)}{f(t_1)}\right) \delta(t-t_1) \quad (7)$$

The response due to this input is

$$\theta_0 = \frac{h_{-2}(t, t_1)}{Kf(t_1)} + \frac{h_{-1}(t, t_1)}{Kf(t_1)} + \frac{f'(t_1)}{Kf^2(t_1)} h_{-1}(t, t_1) \quad (8)$$

A second and different way to put a unit initial condition at point (B) is to introduce a unit step function as shown in Step 1 of Figure 6.

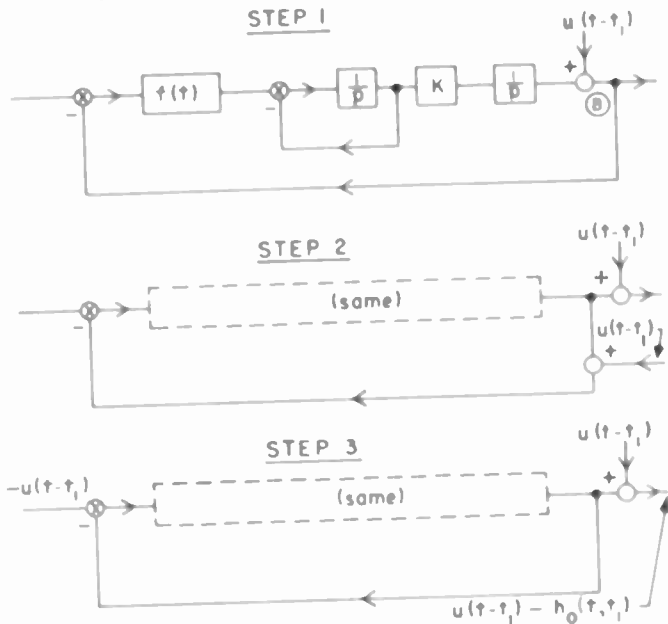


Figure 6
Second Input to Put Unit at (B)

This input can be moved around the loop as in Steps 2 and 3 of Figure 6 without changing the output. The net response, as shown in Step 3 is

$$\theta_0 = u(t-t_1) - h_0(t, t_1) \quad (9)$$

Now, since this response (9) is due to a unit initial condition at (B), as was (8), the two can be equated. We call this the method of equivalent inputs. The resulting equation is

$$\frac{h_{-2}(t, t_1)}{Kf(t_1)} + \frac{h_{-1}(t, t_1)}{Kf(t_1)} + \frac{f'(t_1)}{Kf^2(t_1)} h_{-1}(t, t_1) \quad (10)$$

$$= u(t-t_1) - h_0(t, t_1)$$

or, collecting terms

$$-\frac{1}{K} \frac{\partial}{\partial t_1} \frac{h_{-1}(t, t_1)}{f(t_1)} + \frac{h_{-1}(t, t_1)}{Kf(t_1)} + h_0(t, t_1) \quad (11)$$

$$= u(t-t_1)$$

We note in passing that (10) gives a relation between weighting functions, which can be used to find a third when two are already known. Now, differentiating with respect to t_1 and letting t_2 be the time of observation, (11) becomes

$$-\frac{1}{K} \frac{\partial^2}{\partial t_1^2} \left[\frac{h_{-1}(t_2, t_1)}{f(t_1)} \right] + \frac{1}{K} \frac{\partial}{\partial t_1} \left[\frac{h_{-1}(t_2, t_1)}{f(t_1)} \right] \quad (12)$$

$$- h_{-1}(t_2, t_1) = -\delta(t_2 - t_1)$$

This is a differential equation with the impulse response function as the dependent variable and t_1 , the time at which an impulse is introduced as the independent variable. The time of observation t_2 is a parameter of the equation. In order to put (12) into a form suitable for computer instrumentation, we make the change of variable,

$$t = t_2 - t_1$$

where t is real computer time. Then (12) becomes

$$\frac{1}{K} \frac{\partial^2}{\partial t^2} \left[\frac{h_{-1}(t_2, t_2 - t)}{f(t_2 - t)} \right] + \frac{1}{K} \frac{\partial}{\partial t} \left[\frac{h_{-1}(t_2, t_2 - t)}{f(t_2 - t)} \right]$$

$$+ h_{-1}(t_2, t_2 - t) = \delta(t)$$

A block diagram of this equation is shown in Figure 7.

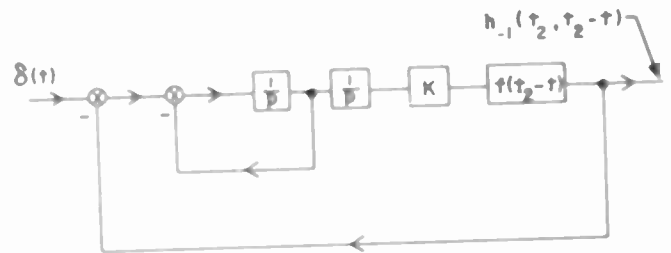


Figure 7
Block Diagram of Adjoint

And, rearranging, we have finally the block diagram in Figure 8.

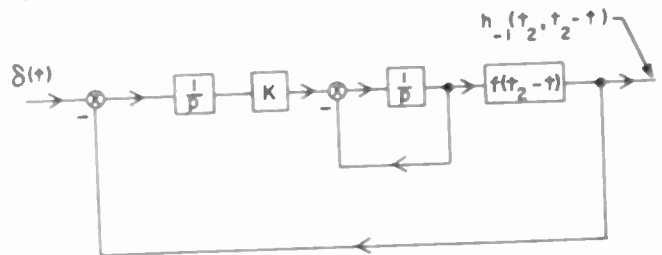


Figure 8
Adjoint System

Comparing Figure 8 with Figure 4, we note the following:

1. The order of elements around the loops is reversed.
2. The time varying element in Figure 8 starts with the value $f(t_2)$ which was its final value in Figure 4. The argument, (t_2-t) , decreases as time goes on.
3. The positions of system inputs and outputs are reversed in the two figures. The inputs to both systems are impulse functions.
4. There is available in the system of Figure 8, the function $h_{-1}(t_2, t_2-t)$. This is one of the weighting functions, but measured at $t = t_2$ for varying time of application of the impulse. This is a function which can be used in the way described in the introduction to find the response to a general input, or the mean square response to a statistical input.

The properties listed above are those of the adjoint system described by Laning and Battin.³ The adjoint, therefore, has then been arrived at through a physical argument. A mathematical discussion of the adjoint and an example of the method of drawing the block diagram for an adjoint system will be found in the appendix.

It is interesting to plot representative weighting functions produced by the two systems, the original, Figure 4, and the adjoint, Figure 8. These might be as shown in Figure 9.

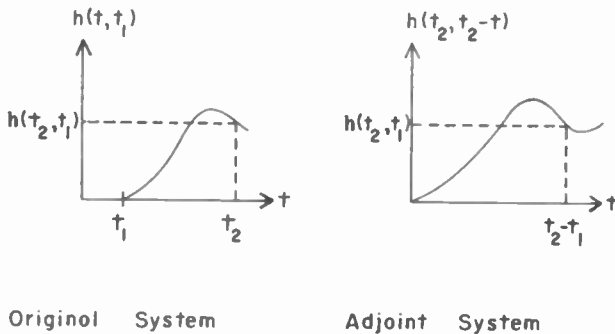


Figure 9
Typical Responses of System and Adjoint

Conclusions

The method discussed has led to a development of the adjoint system by a physical argument, avoiding the discussion of Green's functions and the like. It may, therefore, provide more physical intuition in understanding the adjoint system. The Equation (10) can be used to find other weighting functions graphically from ones already computed.

The system used for illustration of the method can be changed to fit specific problems. However, the adjoint system can be drawn immediately, and the counterpart of Equation (10) found from it.

The authors wish to thank R. K. Roney of Hughes Aircraft Company for his part in discussions relating to this work, and Barbara Pudewa for her help in preparation of the manuscript.

Appendix

It has been shown³ that the adjoint system may be used to obtain the weighting function for a linear time varying system as a function of the time of application of the impulse. In this appendix we will show for a simple system how the adjoint is obtained from the block diagram; how the equation for the adjoint system is related to that for the original system; and how the solutions of the two system equations are related.

Let us consider the servo with varying gain shown in Figure 10. The input is an impulse function at time t_1 .

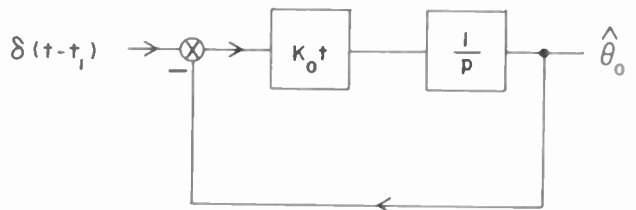


Figure 10
Time Varying Servo

The equation describing this system is

$$\frac{1}{K_0 t} \frac{d\hat{\theta}_0}{dt} + \hat{\theta}_0 = \delta(t-t_1) \quad (13)$$

The system could be simulated as shown in Figure 11.

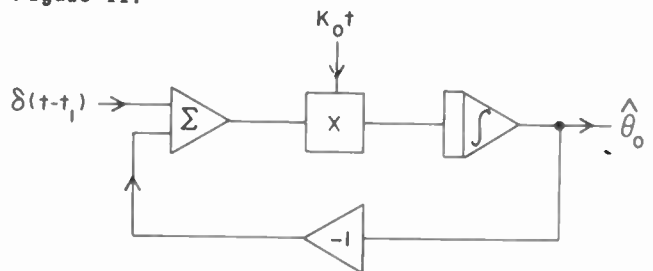


Figure 11
Computer Block Diagram for Figure 10

In Figure 11, we define the impulse response as

$$\hat{\theta}_0(t) = h_{-1}(t, t_1)$$

Now the adjoint system is found from the original by:

1. Turning each element in the loop around, and reversing the direction of signal flow.
2. Letting the variation of time varying elements start from some time t_2 , and run backwards relative to their action in the original system.
3. Interchanging the input and output of the system. The new input is $\delta(t)$. We will show that the output is $h_{-1}(t_2, t_2 - t)$.

Following these rules, we have in Figure 12 the adjoint system for Figure 11. We will call the impulse response of this new system $\theta_0(t)$.

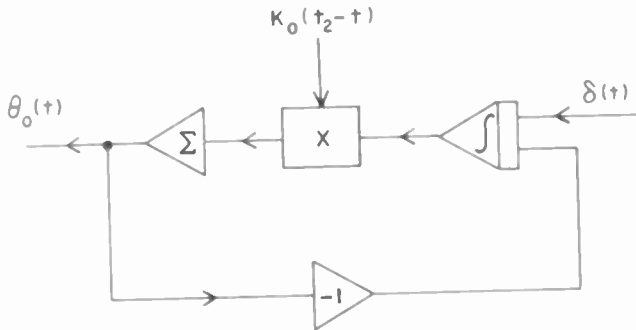


Figure 12
Adjoint System for Figure 11

The equation for the system in Figure 12 is

$$\delta(t) - \theta_0(t) = \frac{d}{dt} \left[\frac{1}{K_0(t_2 - t)} \theta_0(t) \right] \quad (14)$$

We change variable, $x = t_2 - t$, $\theta_0(t_2 - x) = \bar{\theta}_0(x, t_2)$, and (14) becomes

$$-\frac{d}{dx} \left[\frac{1}{K_0 x} \bar{\theta}_0(x, t_2) \right] + \bar{\theta}_0(x, t_2) = \delta(x - t_2) \quad (15)$$

This equation is the adjoint* of

$$\frac{1}{K_0 x} \frac{d}{dx} h_{-1}(x, t_1) + h_{-1}(x, t_1) = \delta(x - t_1) \quad (16)$$

* The adjoint equation is obtained by replacing each operator

$$f_n(x) \frac{d^n}{dx^n} \text{ by } (-) \frac{d^n}{dx^n} f_n(x)$$

where we have used h_{-1} for the solution since the equation is the same as (13). Now a property of the solutions of a pair of adjoint equations like (15) and (16)* is (under certain conditions on initial values which are satisfied here)

$$\bar{\theta}_0(\xi_1, \xi_2) = h_{-1}(\xi_2, \xi_1)$$

so that

$$\bar{\theta}_0(x, t_2) = h_{-1}(t_2, x)$$

or

$$\theta_0(t_2 - x) = h_{-1}(t_2, x)$$

Now, letting $x = t_2 - t$, we have

$$\theta_0(t) = h_{-1}(t_2, t_2 - t)$$

which is the desired result.**

References

- 1) Ince, "Ordinary Differential Equations," Dover, 1944.
- 2) R. R. Bennett, "Analog Computing Applied to Noise Studies," Proc. of I.R.E., Vol. 41, p. 1509, October 1953.
- 3) J. H. Laning and R. H. Battin, "An Application of Analog Computers to Problems of Statistical Analysis," Cyclone Symposium II, 1952.

* See, for example, Reference 1, p. 256

** The system considered is simple enough to yield an analytic solution. It can be verified that

$$h_{-1}(t, t_1) = K_0 t_1 \exp\left(-\frac{K_0}{2}(t^2 - t_1^2)\right)$$

$$\theta_0(t) = K(t_2 - t) \exp\left(\frac{K_0}{2}(t^2 - 2t_2 t)\right)$$

$$\bar{\theta}_0(x, t_2) = Kx \exp\left(\frac{K_0}{2}(x^2 - t_2^2)\right)$$

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

Summary

The analysis of a network whose internal structure is that of a linear graph is well known. In particular; the elimination of hidden meshes in a mesh analysis or hidden nodes in a nodal analysis has been amply discussed. The point of this paper is to extract and extend the algebra involved and thereby solve more general problems such as the interconnection of multiport linear transducers where the linear graph characterization is either invalid or irrelevant.

As a preliminary the framework of transducer theory is outlined. (Figure 1). The novel feature beyond the customary treatment¹ is the partitioning of the analytical variables into two sets suggestively denoted as concealed variables and accessible variables. Whereas the transducer constraints are imposed upon the entire assemblage of analytical variables it is only the accessible variables which directly participate in determining the observables². The natural occurrence of concealed variables and indeed the motivation for the accessible-concealed terminology is illustrated by the interconnection of transducers. (Figure 2).

It is essential, at this point, to develop methods by which one can modify the constraints and variables of a transducer in an unobservable manner. In particular those methods (Figure 3) which exploit the aforementioned partitioning are accorded special attention. The discussion of these latter methods was previously initiated³ and is supplemented here.

A simple but basic type of constraint is the vanishing of a single dependent variable. Sets of such constraints, hereafter referred to as concealed constraints, are removed by the Campbell (elimination of concealed circuits) formula, the

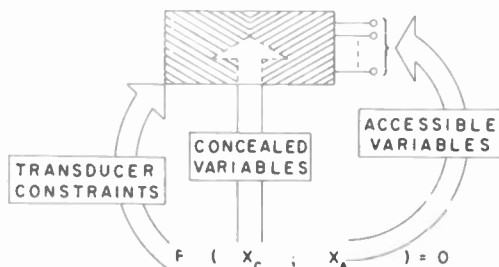
Kron (partitioned matrix) reduction formula, or a variant of Cramer's rule. (Figure 4) (For details see reference 3). It is then shown how problems with arbitrary but linear constraints are reducible to problems with concealed constraints. Indeed the general case can be so reduced by an imbedding procedure which regards the simultaneous occurrence of dependent and independent variables as a single vector in a higher dimensional space (Figure 5) (A variant of this scheme in which the imbedding is done symmetrically was employed by N.E. Reed⁴ in his nodal analysis of a linear graph when the branch impedances possess mutual inductance.) However, the general imbedding, though foolproof, may yet involve needless computational complications and hence it is wisest to use matrix algebra as an adjunct to or replacement of the imbedding procedure. For example the impedance-transducer connection of impedance-transducers is reduced to concealed type constraints when one writes the connecting equations in a judiciously imbedded form and merely adds this to the unconnected performance equations. (Figure 6) The same device suffices for the interconnection of scattering matrices (Figure 7). Finally the use of matrix multiplication, typified by the Kron connection "tensor", is discussed. (Figure 8)

References

1. Kerns, D.M., Basis of the Application of Network Equations to Waveguide Problems, Journ. Res. N.E.S. 42 (1949) 515-546.
2. Wiener, N., Harmonic Analysis and the Quantum Theory, Journ. Franklin Inst. 207 (1929) 525.
3. Kurss, H., Network Analysis with the Aid of the Matrix Generating Polynomial, Conv. Rec. I.R.E. 1953, Pt.V 39-43.
4. Reed, N.E., Node Equations, Proc.I.R.E. 32 (1944) 355-359.

* This work was performed for AFRC under Contract AF-19(604)-890.

I. ANALYTICAL VARIABLES



II. OBSERVABLES

$$x_o = f[x_a]$$

Fig. 1 - Transducer behavior.

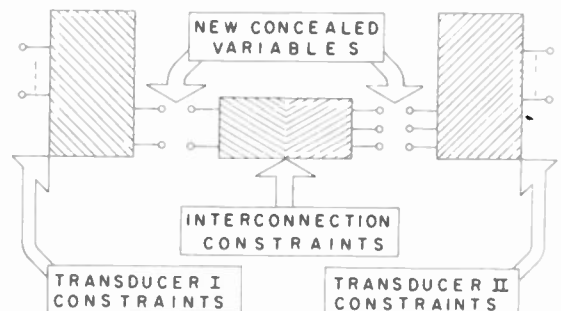


Fig. 2 - Transducer interconnections.

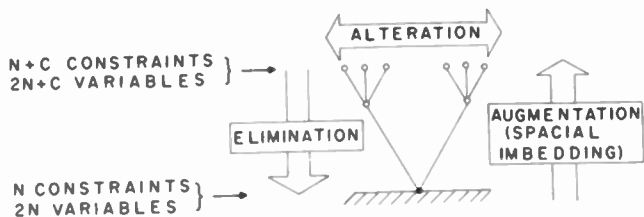


Fig. 3

Unobservable "concealed variable" operations.

IF

$$\begin{bmatrix} y_1 \\ y_2 \\ 0 \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \Rightarrow \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = A' \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

THEN

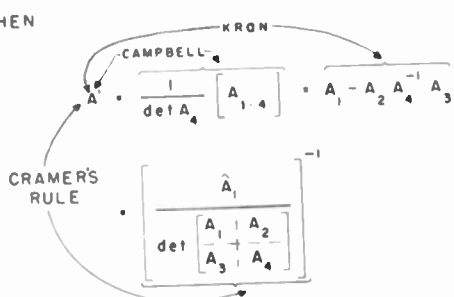


Fig. 4 - Elimination in a special case.

1. GIVEN

$$L \begin{bmatrix} Y_A \\ Y_C \end{bmatrix} = A \begin{bmatrix} X_A \\ X_C \end{bmatrix}$$

2. REWRITE AS

$$\begin{bmatrix} Y_A \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ L & -A \end{bmatrix} \begin{bmatrix} Y'_A \\ Y'_C \\ X'_A \\ X'_C \end{bmatrix}$$

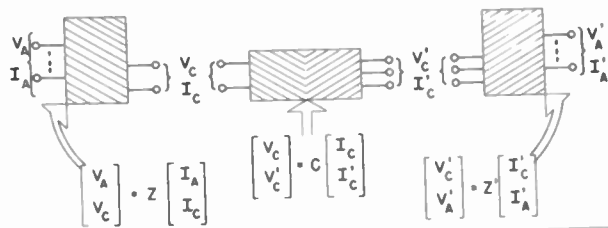
3. ELIMINATE

$Y'_A, Y'_C,$ AND X'_C

4. \Rightarrow

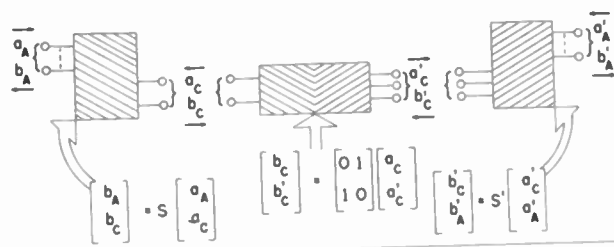
$$Y_A = A' X_A$$

Fig. 5 - General reduction to the special case.



$$\Rightarrow \begin{bmatrix} V_A \\ 0 \\ 0 \\ V'_A \end{bmatrix} \cdot \left\{ \begin{bmatrix} Z & 0 \\ 0 & Z \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right\} \begin{bmatrix} I_A \\ I_C \\ I'_C \\ I'_A \end{bmatrix}$$

Fig. 6 - Z-transducer connection of Z-transducers.



$$\Rightarrow \begin{bmatrix} b_A \\ 0 \\ 0 \\ b'_A \end{bmatrix} \cdot \begin{bmatrix} S & 0 & 0 \\ 0 & -I & 0 \\ 0 & 0 & S' \end{bmatrix} \begin{bmatrix} a_A \\ a_C \\ a'_C \\ a'_A \end{bmatrix}$$

Fig. 7 - Interconnection of scattering matrices.

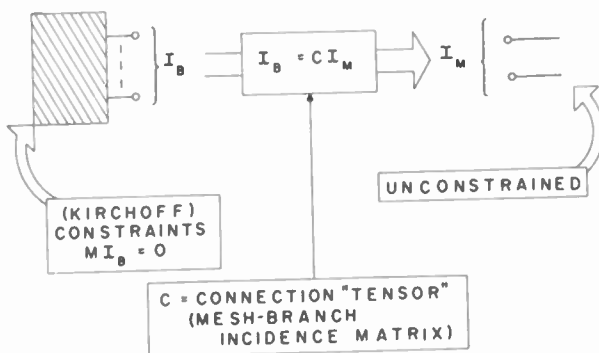


Fig. 8 - Elimination by matrix multiplication.

DYNAMIC CHARACTERISTICS OF FOUR-TERMINAL NETWORKS

W. W. Happ
Sylvania Electric Products, Inc.
Electronics Division
Ipswich, Massachusetts

Abstract

A set of six anti-commuting symbols is used to establish systematically relationships between dynamic characteristics of four-terminal networks. Numerous examples are given, several from transistor circuits. The method is particularly useful in conjunction with signal flow graphs.

1. Dynamic Characteristics: Notation

A set of dynamic characteristics of the form $(\partial x / \partial y)_z$ can generally describe the performance of a non-linear network over a suitably small region of operation. A four-terminal network, as shown in Fig. 1, imposes an interdependence between the four basic variables, E_1, E_2, I_1, I_2 , which may be interpreted as input and output voltage and the corresponding currents. Thus x, y , and z in the expression $(\partial x / \partial y)_z$ can be any of the four basic variables, yielding 4! or 24 dynamic characteristics. Of these, 12 are listed in Table 1, the remaining 12 are obtained by interchanging subscripts 1 and 2.

Relationships between dynamic characteristics can be obtained rapidly and systematically

(i) by defining an anti-commuting symbol, called the deviation of x with respect to y ,

$$(x, y) = -(y, x) \quad (1.1)$$

(ii) by expressing dynamic characteristics as ratios of two deviations:

$$\left(\frac{\partial x}{\partial y} \right)_z = \frac{(x, z)}{(y, z)} \quad (1.2)$$

As an example, consider the product of two dynamic characteristics. From (1.1) and (1.2) one obtains in the terminology of Table 1:

$$z_{12}^{oo} y_{11}^{oo} = \frac{(E_1, I_1)(I_1, I_2)}{(I_2, I_1)(E_1, I_2)} \frac{(I_1, E_1)}{(E_1, I_2)} = -A_{11}^{so}$$

Thus, the 24 dynamic characteristics of a four-terminal network can be expressed as ratios of six deviations:

$$(E_1, E_2)(E_1, I_1)(E_1, I_2)(E_2, I_1)(E_2, I_2)(I_1, I_2)$$

Table 2 lists $z, y, g, h, a,$ and b circuit parameters by stating the network equations in terms of deviations as defined by (1.1) and (1.2). The corresponding flow graph shows

- (i) The partial derivative, e.g. $z_{11} = (\partial E_1 / \partial I_1)_{I_2}$, marked by an arrow.
- (ii) The independent variable or source, marked (e), e.g. I_1 .

- (iii) The dependent variable or sink, marked (o), e.g. E_1 .
- (iv) The network equation, e.g. $E_1 = z_{11} I_1 + z_{12} I_2$, an equation being equivalent to all the arrows terminating at one point.

In addition Table 2 shows that the notation defined by (1.1) and (1.2) leads to a concise formulation for the following:

$$\text{Passive network } (E_1, I_1) = (I_2, E_2) \quad (1.3)$$

$$\text{Symmetrical network } (E_1, I_2) = (I_1, E_2) \quad (1.4)$$

$$\text{Uniqueness } (E_1, I_1)(E_2, I_2) + (I_1, E_2)(E_1, I_2) + (E_2, I_1)(I_1, I_2) = 0 \quad (1.5)$$

The uniqueness condition (1.5) states that the z, y, g, h, a, b parameters represent the same physical network, hence (1.5) is used when converting from one set of parameters to another and was used in obtaining the circuit determinant in Table 2.

The purpose of the investigation is:

- (i) to derive rigorously equations (1.1) to (1.5).
- (ii) to examine those properties of the (x, y) symbols which are relevant to the analysis of four-parameter networks.
- (iii) to demonstrate the usefulness of the method of deviations by examples, particularly in conjunction with the flow graph technique.

2. Derivation of Basic Relationships Between Dynamic Characteristics

The deviation of x with respect to y is here defined as

$$(x, y) = \left(\frac{\partial x}{\partial t} \right)_y \quad (2.1)$$

It follows that

$$\frac{(y, z)}{(x, z)} = \left(\frac{\partial y}{\partial x} \right)_z \quad (1.2)$$

which is independent of t . The parameter t may remain unspecified as far as calculations of dynamic characteristics are concerned. Also from (2.1)

$$(x, x) = 0 \quad (2.2)$$

$$(xy, z) = x(y, z) + y(x, z) \quad (2.3)$$

If a functional relationship between x, y and z exists,

$$f(x, y, z) = 0 \quad (2.4)$$

Denoting partial differentiation of f with respect

to x as f_x , we have from (2.1) and (2.4).

$$(x,s) f_x + (y,s) f_y + (z,s) f_z = 0 \quad (2.5)$$

Let s assume the values x, y, z consecutively, then the three equations (2.5) are not independent, and the determinant D_3 vanishes, where

$$D_3 = \begin{vmatrix} (x,x) & (y,x) & (z,x) \\ (x,y) & (y,y) & (z,y) \\ (x,z) & (y,z) & (z,z) \end{vmatrix} \quad (2.6)$$

Taking (2.2) into account, we have

$$D_3 = 0 \text{ if } (x,y) = -(y,z) \quad (1.1)$$

To show that (1.1) is not only a sufficient but also a necessary condition for (2.4), we may compute from (1.1) and (2.3)

$$(z,xy) = y(z,x) + x(z,y) \quad (2.7)$$

Setting $z = xy$, we have from (2.2) and (2.3)

$$(xy, xy) = xy [(x,y) + (y,x)] \quad (2.8)$$

which is satisfied if, and only if, (1.1) holds.

Several interpretations of the parameter t and of the concept of deviation are possible:

(a) The deviation may be regarded as the response (or fluctuation) of x due to a stimulus (or drive) t with the variable y held constant. Thus (x,y) is equal in magnitude but opposite in sign to the fluctuation of y with x kept constant under the same drive t , namely (y,x) .

(b) A geometrical interpretation of (2.1) is possible by considering a surface in the x, y, t space. The direction and the scale of that coordinate "axis" must be adjusted to satisfy at all points the relation

$$\left(\frac{\partial x}{\partial y}\right)_t = 1 \quad (2.9)$$

since (2.1) gives $(t,x) = 1$ and $(t,y) = 1$ (2.10)

The dimension of each quantity in (2.9) and (2.10) should be carefully noted. For the purpose of computation it is rarely necessary or desirable to specify t . Indeed the arbitrary nature of t in the deviation (x,y) is similar to the arbitrary nature of s in $(\partial x / \partial y)_s$ when a set of mutually dependent partial derivatives are all in the s plane, such that the above expression becomes a function of x and y only.

(c) Alternatively, the deviation of x with respect to y may be defined as the determinant

$$(x,y) = \begin{vmatrix} \left(\frac{\partial x}{\partial r}\right)_s & \left(\frac{\partial y}{\partial r}\right)_s \\ \left(\frac{\partial x}{\partial s}\right)_r & \left(\frac{\partial y}{\partial s}\right)_r \end{vmatrix} \quad (2.11)$$

where r and s are arbitrary functions which are similar to t in (2.1) and which may remain unspecified when evaluating dynamic characteristics. This definition already implies (1.1) and (2.12)

Most of the recent work^{1,2,3,4,5,6} on dynamic characteristics or 'small signal' circuit parameters of active four-terminal networks has been done in connection with equivalent circuits for transistors, hence several applications of this technique to transistor circuits will be given.

Anti-commuting symbols similar to the deviation (x,y) have been used for "small parameter" systems in thermodynamics^{7,8}, celestial dynamics^{9,10} and quantum dynamics^{11,12}, where the symbol (x,y) is frequently referred to as Jacobian, Poisson bracket or Perturbation. The term deviation appears more appropriate since a fluctuation of x at constant y is measured by the mean deviation or by the standard deviation of x from its equilibrium value. In the case of dynamic characteristics of circuit elements, it is also sometimes convenient to think of the deviation (x,y) as the RMS deviation of x from its equilibrium value as a result of a small test signal applied to the circuit.

3. The Four-Terminal Network: Uniqueness

Any four-terminal network is subject to a requirement of uniqueness; namely, given any two of the basic variables, there must be one, and only one, value for each of the other two basic variables. This requirement corresponds to the 'equation of motion' in mechanics and to the 'equation of state' in thermodynamics.

Thus if we assume that there are four basic variables as shown in Fig. 1 these are not independent but related by an equation

$$F(E_1, E_2, I_1, I_2) = 0 \quad (3.1)$$

Differentiating F with respect to t at constant s , we obtain

$$\frac{\partial F}{\partial E_1}(E_1, s) + \frac{\partial F}{\partial E_2}(E_2, s) + \frac{\partial F}{\partial I_1}(I_1, s) + \frac{\partial F}{\partial I_2}(I_2, s) = 0 \quad (3.2)$$

Let s equal E_1, E_2, I_1, I_2 consecutively. The four equations (3.2) are independent if the determinant

$$D^2 \neq 0 \quad (3.3)$$

where

$$D^2 = \begin{vmatrix} (E_1, E_1)(E_1, E_2)(E_1, I_1)(E_1, I_2) \\ (E_2, E_1)(E_2, E_2)(E_2, I_1)(E_2, I_2) \\ (I_1, E_1)(I_1, E_2)(I_1, I_1)(I_1, I_2) \\ (I_2, E_1)(I_2, E_2)(I_2, I_1)(I_2, I_2) \end{vmatrix} \quad (3.4)$$

Evaluating the determinant with the help of (1.1), one obtains $\pm D = (E_1, E_2)(I_1, I_2) + (E_2, I_1)x(E_1, I_2) + (I_1, E_1)(I_1, I_2)$ (3.5)

Since the four basic variables E_1, E_2, I_1, I_2 are

connected by two relations, or constraints, D^2 must have a double zero and hence

$$D = 0 \quad (1.5)$$

This is the uniqueness condition.

It is instructive to examine the alternative definition of the deviation as given by (2.11). Substituting (1.2) into (2.11) yields

$$(x,y) = \begin{vmatrix} \frac{(x,s)}{(r,s)} & \frac{(y,s)}{(r,s)} \\ \frac{(x,r)}{(s,r)} & \frac{(y,r)}{(s,r)} \end{vmatrix} \quad (3.6)$$

Clearly (1.5) and (3.6) are equivalent under the normalization

$$(r,s) = 1 \quad (2.12)$$

4. The Passive Network: Reciprocity

Table 2 shows that the condition for a passive network can be stated concisely in terms of deviations as

$$(E_1, I_1) = (I_2, E_2) \quad (4.1)$$

This reciprocity condition does not hold in circuits with active elements, such as vacuum tubes or transistors. In thermodynamics a similar reciprocity requirement exists, namely the Second Law of Thermodynamics, and in Mechanics, a reciprocity requirement occurs in Newton's Law of Action and Reaction. Analogous quantities and laws are compared in Table 3. Although the reciprocity condition has been experimentally well confirmed, its theoretical foundation has been a source of error and confusion. A summary and critical discussion of this problem was given by deGroot¹³, who contrasts satisfactory and unsatisfactory derivations of the reciprocity relation and remarks that "it seems rather amazing that one can arrive in so many incorrect ways at the correct answer" and shows "how one can get correct answers with incorrect methods". Since numerous derivations of the reciprocity relations given in the literature are based upon methods criticized by deGroot, a proof is here presented which is based upon the Principle of Least Dissipation¹⁴ and which follows closely the work of deGroot¹³ and Onsager¹⁵.

The power dissipated in a four-terminal network is

$$P = E_1 I_1 + E_2 I_2 \quad (4.2)$$

also, for a passive network, P is a minimum or

$$\left(\frac{\partial P}{\partial E_2} \right)_{E_1} = 0 \quad \text{when } I_1 = 0 \text{ or } I_2 = 0 \quad (4.3)$$

Noting that P is a function of E_1 and E_2 , we can evaluate (4.3) with the aid of Table 2 and we obtain, after differentiation and substitution from Table 2,

$$y_{12} = y_{21} \quad (4.4)$$

Alternatively using deviations we have from (4.3) and Table 2

$$(P, E_1) = E_1 (E_1, I_1) + (E_2, I_2) \quad (4.5)$$

hence, if P is a minimum, (4.1) follows at once. Similarly, if P is a function of E_1 and I_1

$$\left(\frac{\partial P}{\partial E_1} \right)_{I_1} = 0 \quad \text{when } I_2 = 0 \text{ or } E_2 = 0 \quad (4.6)$$

$$\text{hence } |a| = 1 \quad (4.7)$$

Using deviations, (4.6) yields

$$(P, I_1) = I_1 (E_1, I_1) + (E_2, I_2) \quad (4.8)$$

similarly, if P is minimum, (4.1) follows at once.

When dealing with dynamic characteristics of passive networks, it is frequently advantageous to recognize symmetry properties of the system. Equ. (1.1), (1.3), and (1.5) remain invariant, if each of the four basic variables E_1, E_2, I_1, I_2 take the place of another in a certain allowed sequence. This allowed sequence is given by the rotation of the square, Fig. 2, about any of its axes of symmetry, provided that the voltages E_1 and E_2 change sign if the stationary dotted line is crossed by one of the voltages.

5. Equivalent Dynamic Characteristics

It is often possible and desirable to replace certain dynamic characteristics by others. Frequently it is necessary to investigate the validity of such substitution. As an example consider

$$(y_{11}^{sc} - y_{11}^{oc}) y_{22}^{sc} = (y_{21}^{sc})^2 \quad (5.1)$$

Writing this in the notation of Table 1 with the aid of (1.2) gives

$$\frac{(I_1, E_2)}{(E_1, E_2)} - \frac{(I_1, I_2)}{(E_1, I_2)} = \frac{(I_2, I_1)}{(E_2, E_1)} \cdot \frac{(I_2, E_2)^2}{(E_1, E_2)^2}$$

and after cross multiplication

$$(I_1, I_2)(E_2, E_1) + (I_1, E_2)(E_1, I_2) + (I_2, E_2)(E_2, I_2) = 0$$

which is equivalent to (1.3) and (1.5). Hence (5.1) is true for passive (or compensated) networks, but does not hold for active networks.

Using the analogue thermodynamic quantities of Table 3, (5.1) gives Rankin's equations relating the specific heats, while the mechanical analogues of Table 3 give the relation between group and phase velocities, if substituted into (5.1).

More insight into the relationships between dynamic characteristics can be gained by the

effective use of signal flow graphs. This technique^{16,17} is extensively used in the analysis of servomechanisms and has recently been applied to circuit analysis. The basic rule is: a point represents a dynamic variable, which is determined by the arrows terminating at that point. As an example consider

$$\text{in Fig. 3a } I_1 = Y_{11}^{so} E_1 + Y_{12}^{so} E_2$$

$$\text{in Fig. 3b } I_1 = Y_{11}^{oo} E_1 + A_{11}^{so} I_2$$

Source or generator (G) and sink or load (L) are frequently omitted in drawing the flow diagram. If Fig. 3a and Fig. 3b represent equivalent circuits of the same set of currents and voltages, the resulting transmission between points must be the same, thus:

$$\left(\frac{\partial I_1}{\partial E_2} \right)_{E_1} = Y_{12}^{so} = A_{11}^{so} Y_{11}^{so} \quad (5.3)$$

$$\left(\frac{\partial I_1}{\partial E_1} \right)_{E_2} = Y_{11}^{so} = Y_{11}^{oo} + Y_{21}^{so} A_{11}^{so} \quad (5.4)$$

Eliminating A_{11}^{so} yields

$$(Y_{11}^{so} - Y_{11}^{oo}) Y_{22}^{so} = Y_{21}^{so} Y_{12}^{so} \quad (5.5)$$

Clearly (5.5) reduces to (5.1) if the circuit satisfies the reciprocity relation (1.3). Similar relationships can now be derived effortlessly by rotation of the square in Fig. 2. Consider, for instance the substitution.

$$\begin{matrix} E_1 & I_1 & E_2 & I_2 \\ \downarrow & \downarrow & \downarrow & \downarrow \end{matrix} \quad (5.5)$$

$$\begin{matrix} E_2 & I_2 & -I_1 & E_1 \\ \downarrow & \downarrow & \downarrow & \downarrow \end{matrix} \quad (5.6)$$

$$(Y_{22}^{oo} - Y_{22}^{so}) Z_{11}^{so} = A_{12}^{so} A_{21}^{so} \quad (5.6)$$

Since there are innumerable similar relations between dynamic characteristics, no attempt will be made to summarize or tabulate these. Rather, the aim of this investigation is to present a few typical examples which exhibit the power and directness of the method of deviations in conjunction with flow graphs.

6. Dynamic Characteristics of Devices: Examples

To investigate dynamic characteristics of devices, it is convenient to distinguish three types of four-terminal circuit elements. (For each type of device an illustrative example is given in Table 4.)

(i) active elements, such as the transistor. Generally all 24 dynamic characteristics can be computed. From an energy point of view, an active element is equivalent to a passive element plus an internal source of energy.

(ii) unilateral elements, such as the vacuum tube. These are active elements in which the signal flow occurs only in one direction. Feed-

back is absent and not all deviation can be computed. It is usually possible, however, to convert a unilateral element into an active network by the appropriate use of additional passive elements to provide feedback.

(iii) passive elements, such as the transformer. Energy is dissipated inside the system in accordance with the Principle of Least Dissipation, as no internal sources of energy are present. Feedback and feedforward compensate each other, as will be seen presently and networks are therefore often referred to as compensated networks.

The flow graphs shown in Table 4 are intended as an aid in visualizing the functional relationships between voltages, currents and dynamic characteristics of the device. It is always understood that these relationships are known at the outset of the analysis and that they are meaningful. The task is then to use the appropriate method for extracting from the flow graph the desired solution, which in our case is the evaluation of dynamic characteristics.

The problem of constructing a flow graph from known dynamic characteristics leads frequently to several alternative solutions for the same device. The desired form of the flow graph depends on how we perceive the pertinent functional relationships between the dependent and independent variables of the device. The following examples will illustrate this.

(i) The Transistor. Table 5 shows the conventional wiring (circuit) diagram and the corresponding flow graph. While comparatively little information is obtainable from the former, the flow graph permits straightforward evaluation of the dynamic characteristics. For instance, when the emitter is grounded and the load is connected to the collector, as shown in the second column of Table 5, three paths lead from I_o to V_o . Hence

$$\left(\frac{\partial V_o}{\partial I_o} \right)_{I_b} = \frac{(V_2, I_1)}{(I_2, I_1)} = r_o + r_e - r_m$$

Since all parameters in the flow graph are resistances, it is convenient to set $(I_1, I_2) = 1$ and to enter the value of (I_1, V_2) in Table 5. Similarly the other three deviations can be evaluated by following the flow path from each source to each sink. Finally (V_1, V_2) must be computed from (1.5). Thus Table 5 contains in concise form all 24 dynamic characteristics for each of the six possible circuit connections of the transistor.

Table 4 includes the flow graph of a junction transistor as described by Ryder¹⁸. In this particular case it was found desirable to describe the performance of the device in terms of six parameters. We can readily evaluate the flow path from each source to each sink, for instance

$$(E_1, E_2) = r_e + r_b (1-\alpha) \text{ provided } (I_1, E_1) = 1$$

An interesting alternative description of the junction transistor due to Chu¹⁹ makes use of the similarity of minority carrier diffusion in a transistor and signal propagation in a transmission line. Table 5 gives the flow diagrams: The generalized base width θ of the transistor corresponds to the generalized length of the transmission line. Indeed the flow graph will describe a transmission line, if $Z_0 = Z_\xi$, which is the characteristic impedance of a (passive) transmission line. For the transistor, $Z_0 \neq Z_\xi$, these equivalent characteristic impedances can be expressed in terms of design parameters (base width, diffusion constant, frequency, etc.) of the transistor.

The references listed at the head of each column in Table 4 contain a detailed explanation of each parameter shown in the corresponding flow graph. A physical interpretation of these parameters is therefore not attempted here.

(ii) The Vacuum Tube. Table 4 includes two examples of the vacuum tube as a circuit element. Flow graphs and deviations are given in each case. Some deviations cannot be computed when the vacuum tube is connected as a unilateral active element, i.e. when a signal in the output circuit will not affect the input circuit. Numerous other examples of flow diagrams for vacuum tube circuits are given by Mason¹⁶.

(iii) Energy Converters. The transformer is given as an example of an energy converter in Table 4 with the corresponding deviation. Additional examples are presented in Table 6. Energy converters contain no internal source of energy and hence must obey the reciprocity relation (1.3). In some applications it is preferable to define the direction of current flow as the direction of energy flow, this means replacing I_2 by $(-I_2)$ in Fig. 1, and changing (1.3) into $(E_1, I_1) = (E_2, I_2)$ (1.3*)
 (1.4) into $(E_1, I_2) = (E_2, I_1)$ (1.4*)
 This convention was used in Table 6. The uniqueness condition (1.5) remains invariant.

The functional relationship between variables of the device are often perceived as a closed chain of dependency, consider for instance, the (d.c. shunt) generator in Table 6. The torque (T) in the armature depends upon the speed (W), the current delivered (I) is determined by the torque (T) and in turn affects the back - emf (E). The chain of dependency is closed by the -- often linear -- relationship between E and W. The functional relationship $E \rightarrow W$ may be considered as feedback while the other link between the mechanical and electrical system, $T \rightarrow I$, is referred to as feedforward, the reference direction is that of energy flow in the device. By writing out the equations that correspond to a feedback flow graph, and by solving these equations in terms of dependent and independent variables, it is always possible to convert a closed chain into an open chain and viceversa. It is, therefore, incorrect to speak of a device as "containing" feedback; as pointed out by Mason¹⁶, "feedback is only present if we perceive

a close chain of dependence." The relationship between dynamic characteristics and feedback parameters is discussed in the following section. With the aid of the method of deviations, techniques will be developed to express four-terminal networks in terms of feedback parameters or, viceversa, to eliminate feedback parameters from a network.

7. Techniques for the Analysis of Four-Terminal Flow Graphs

Some of the results of this section are tabulated as follows:

Table 7: Elementary Operations - Two Terminal Flow Graphs

Table 8: Equivalent Junctions - Three Terminal Flow Graphs

Table 9 and 10: Equivalent Four-Terminal Flow Graphs

A number of techniques have proven helpful in the analysis and synthesis of flow graphs, these techniques are mainly based upon the elementary operations defined in Table 7.

A. Path Inversion

Recalling that each point or junction represents an equation, it is useful to distinguish between two types of junctions shown in Table 8.

(i) contributive junctions, such as

$$a I_1 + b I_2 = I_3 \quad (7.1)$$

(ii) distributive junctions, such as

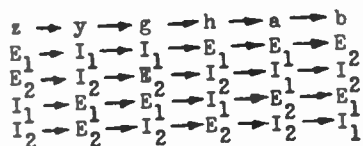
$$a I_1 = b I_2 = I_3 \quad (7.2)$$

Junctions connecting more than three variables can always be replaced by a series of contributive and distributive junctions. (7.1) and (7.2) must remain invariant under permissible transformations: An example of a permissible transformation is the inversion of path between I_1 and I_2 in Table 8, (i), (ii), (iii), and (iv). Since any path may be considered a sequence of distributive and contributive junctions, we can interchange a dependent variable and an independent variable by inverting a complete path connecting these two variables. A conversion of this type is shown in Table 9, column 2. This flow graph is obtained from the "g" parameters in Table 2 by inverting the path from E_1 to I_1 . Noting that $1/z_{11} = g_{11}$, it follows from the path inversions shown in Table 8, that there is a change of sign at the contributive junction, $E_{12} \rightarrow -E_{12}$, but no change of sign at the distributive junction, $E_{21} \rightarrow E_{21}$. Other equivalent flow diagrams with mixed or hybrid parameters can be obtained by repeated path inversion and are listed in Table 9.

B. Symmetry Inversion

Frequently it is possible to invert a flow pattern on the basis of its symmetry properties.

An examination of Table 5 will reveal that the right side and the left side of the table are mirror images of each other, provided that subscripts 1 and 2 are interchanged upon reflection. Similarly Table 2 follows from its first row, if we replace consistently the following symbols.



By applying these symmetry considerations to Table 9 and Table 10, we can increase sixfold the number of tabulated equivalent flow graphs.

C. Cascade Conversion

It is understood that flow is possible only in the direction of the arrow. A path will now be defined as one assigned route from a source (independent variable) to a sink (dependent variable) and it is numerically equal to the product of all its branches (dynamic characteristics). The transmission from given source to a given sink is defined as the sum of all paths between these given points. Thus, two distinct types of transmission exist.

(i) Cascade Transmission consists of a finite number of paths and is defined as the sum of all paths from source to sink provided that no individual path traverses any junction more than once. In Table 8, (v) and (vii), the cascade transmission from I_1 to I_2 is $a + bc$. In Table 8, (ix) and (xi), the cascade transmission from I_1 to I_2 is bc .

(ii) Feedback Transmission consists of an infinite number of paths, and is the sum of all paths connecting a given source to a given sink. In Table 8, (ix) and (xi), there are N paths from I_1 to I_2 each contributing $ob(abc)^N$. Thus the feedback or total transmission is the sum of all paths

$$\sum_{N=0}^{\infty} ob(abc)^N = \frac{ob}{1-abc} = \frac{ob}{R} \quad (7.3)$$

where R is the return difference²⁰ of a given path which is defined as:

$$\text{return difference} = \frac{\text{cascade transmission}}{\text{total transmission}} \quad (7.4)$$

Cascade conversion consists of replacing paths in parallel by a single path. Conversion of Fig. 3(a) to Fig. 3(b) is an example. By this method the deviations listed in Table 9 can be computed rapidly from the corresponding flow diagrams. For instance for column 2 in Table 9, we have

$$\epsilon_{22} = \epsilon_{22} - \epsilon_{21} \epsilon_{11} \epsilon_{12}$$

D. Feedback Conversion

Any part of a given path which can be traversed by the flow more than once is referred

to as a "loop". The total transmission once around the loop is defined as the loop transmission. For a given loop we have

$$\text{Loop transmission} + \text{return difference} = 1 \quad (7.5)$$

The return difference in (7.5) refers to a specific loop m and it is denoted by R_m . If a branch or a junction of a path is traversed by the feedback flow of the loop, then the return difference R of the path is sum of the return differences of all loops touched by that path.

$$R = \sum_m R_m \quad (7.6)$$

If, however, a loop links another second loop, the second loop must be replaced before evaluating the first.

In Table 10, the return difference of the loop in column 1 is given by

$$R=1 - \frac{(E_2, I_2)(E_1, I_1)}{(E_1, I_2)(E_2, I_1)} = \frac{(E_1, E_2)}{(I_1, E_2)} \frac{(I_1, I_2)}{(E_1, I_2)}$$

which was obtained with the aid of (1.5). From column 1 in Table 10, we obtain

$$\text{cascade transmission from } I_1 \text{ to } E_1 \text{ is } \frac{(E_1, E_2)}{(I_1, E_2)}$$

$$\text{total transmission from } I_1 \text{ to } E_1 \text{ is } \frac{(E_1, E_2)1}{(I_1, E_2)R}$$

$$\text{which is equivalent to } \frac{(E_1, I_2)}{(I_1, I_2)}$$

Rapid evaluation of flow graphs with feedback, such as those given in Table 6, is thus readily achieved. The most common forms of feedback flow graphs with the corresponding deviations are listed in Table 10.

8. Logarithmic Techniques

Dynamic characteristics have been expressed as the ratio of two deviations in accordance with the definition (1.2). On the other hand, it is possible to express impedances and admittances of passive networks as a ratio of two polynomials in ω , the frequency impressed upon the network. It is, therefore, permissible to equate each of the six deviations of a network to a polynomial in ω . Indeed it can be shown that the power of this polynomial in ω depends upon the elements that couple the two variables correlated by the deviation, while the magnitude of the coefficients of this polynomial depend upon the magnitude of the coupling elements. This association between the deviation of two variables on the one hand and a frequency characteristic on the other hand, has a very practical application. It permits us to describe the complete frequency response of all dynamic characteristics of a network by a set of at most six 'log-magnitude' versus 'log-frequency' curves.

As an example consider the transformer as analyzed by Mishkin²¹, for which the deviations are given in Table 4. No generality is lost by

taking the windings in the primary (N_1) and in the secondary (N_2) as unity. Using the notation of Table 4 and denoting the magnetic reluctance by $1/L_m$, we obtain,

$$(E_1, I_1) = (I_2, E_2) = Z_m = L_m \omega \quad (8.1)$$

$$(I_1, E_2) = Z_m + Z_2 = R_2 (1 + \tau_2 \omega) \quad (8.2)$$

$$(E_1, I_2) = Z_m + Z_1 = R_1 (1 + \tau_1 \omega) \quad (8.3)$$

$$(E_1, E_2) = R_1 R_2 (1 + \tau_3 \omega) (1 + \tau_4 \omega) \quad (8.4)$$

where $\tau_2 = 1/\omega_2 = (L_m + L_2)/R_2$

$$\tau_1 = 1/\omega_1 = (L_m + L_1)/R_1$$

Let $\tau_m^2 = L_m^2/R_1 R_2$ and $\tau_3 = 1/\omega_3$, $\tau_4 = 1/\omega_4$ substitution of (8.1) to (8.4) into (1.5) yields

$$\tau_1 + \tau_2 = \tau_3 + \tau_4$$

$$\text{and } \tau_1 \tau_2 - \tau_m^2 = \tau_3 \tau_4$$

This simply means that ω_1 and ω_2 are always 'sandwiched' between ω_3 and ω_4 . That is, if $\tau_1 > \tau_2$

then $\tau_3 > \tau_1 > \tau_2 > \tau_4$ or $\omega_3 < \omega_1 < \omega_2 < \omega_4$.

Note that $(E_1, I_1) = 1$ if $\omega = 1/L_m$ on account of (8.1). From the break frequencies, the asymptotes to the frequency response have been plotted in Fig. 4. By division of any two deviations shown, any desired dynamic characteristic is obtained as a function of frequency.

Another useful logarithmic technique in conjunction with deviations is to replace the basic variables in Fig. 1 by their logarithm. The uniqueness relation (1.5) remains invariant under this substitution and a set of 'logarithmic' dynamic characteristics can be computed. Similarly (1.5) remains invariant, if a set of variables is replaced by their respective exponentials, and an alternative set of dynamic characteristics results. Indeed, the "sensitivity" of a variable x at constant y can be expressed as $(x, y)/x = (\ln x, y)$ or equally well in terms of $(\ln x, \ln y)$.

The methods developed in the previous sections to analyze dynamic characteristics of networks can, therefore, be applied with equal validity to obtain relationships between the sensitivities of the circuit parameters of a network.

9. Synthesis of Four-Terminal Networks

Several ways of adding of two or more four-terminal networks are possible, for instance following combination of networks are frequently used:

- (i) in series
- (ii) in parallel
- (iii) as feedback elements
- (iv) in cascade

The method of deviations may be used to advantage to compute the dynamic characteristics

of a combined (or synthesized) network from the dynamic characteristics of its components.

As an example consider network A which was obtained by combining in cascade networks B and C. The dynamic characteristics of each network are:

$$\begin{aligned} A: & (E_1, I_2); (E_1, E_3); (E_1, I_3); (I_1, E_3); (I_1, I_3); (E_3, I_3) \\ B: & (E_1, I_1); (E_1, E_2); (E_1, I_2); (I_1, E_2); (I_1, I_2); (E_2, I_2) \\ C: & (E_2, I_2); (E_2, E_3); (E_2, I_3); (I_2, E_3); (I_2, I_3); (E_3, I_3) \end{aligned}$$

On cascading B and C to form A, the output from B had to equal the input to C, thus E_2 and I_2 were eliminated in forming A. The problem now facing us is to express the deviations of A in terms of the deviations of B. The solution can be written down by inspection, for instance

$$(E_1, E_3) = (E_1, E_2)(E_3, I_2) - (E_1, I_2)(E_3, E_2) \quad (E_2, I_2)$$

which follows directly from the uniqueness conditions (1.5) and demonstrates the advantage of this method over matrix and other methods in computing dynamic characteristics of combined networks.

10. Generalization: The Multi-Terminal Network

Consider a network with n independent voltages and n independent currents. To establish a more powerful notation, the case $n=2$ is first briefly reviewed.

$n=2$ Let the voltages and currents E_1, E_2, I_1, I_2 of a four-terminal network have a one-to-one correspondence to the variables k, j, m, n taken in any order. If $(\partial x / \partial y)_z = (x, z) / (y, z)$, the network equations of Table 2 can be written

$$m = \begin{pmatrix} m, j \\ k, j \end{pmatrix} k + \begin{pmatrix} m, k \\ j, k \end{pmatrix} j \quad (10.1)$$

$(k, j, m, n = E_1, E_2, I_1, I_2)$ that is, any one-to-one correspondence between the two sets of variables is allowed, giving six sets of two equations each. Differentiating m at constant n and differentiating n at constant m yields,

$$(m, n) (j, k) = (m, k) (j, n) - (m, j) (k, n) \quad (10.2)$$

$$\text{Consistency requires } (x, y) = -(y, \bar{x}) \quad (10.3)$$

$$\text{For a passive network } (E_1, I_1) = (I_2, E_2) \quad (10.4)$$

$$\text{For a symmetrical network } (E_1, I_2) = (I_1, E_2) \quad (10.5)$$

$n=3$ If $(\partial z / \partial y)_{u, v} = (z, u, v) / (y, u, v)$ Then the network equations of a six-terminal network become:

$$1 = \begin{pmatrix} i, j, k \\ i, j, k \end{pmatrix} i + \begin{pmatrix} i, k, i \\ j, k, i \end{pmatrix} j + \begin{pmatrix} i, i, j \\ k, i, j \end{pmatrix} k \quad (10.6)$$

giving twenty sets of three equations each.

$$(1, m, n, i, j, k = E_1, E_2, E_3, I_1, I_2, I_3)$$

again any one-to-one correspondence between the two sets of variables is allowed, giving twenty sets of three equations each.

Differentiating l at constant m and n and differentiating m at constant l and n , we obtain,

$$(i, j, k)(l, m, n) = (l, j, k)(i, m, n) + (l, k, i)(j, m, n) + (l, i, j)(k, m, n) \quad (10.7)$$

Consistency requires

$$(x, y, z) = -(z, y, x) \text{ for adjacent symbols} \quad (10.8)$$

$$\begin{aligned} \text{For a passive network } (E_1, I_1, E_3) &= (I_2, E_2, E_3) \\ (E_1, I_1, I_3) &= (I_2, E_2, I_3) \\ (1, 2, 3 \text{ cyclic}) & \end{aligned} \quad (10.9)$$

$$\begin{aligned} \text{For a symmetrical network } (E_1, I_2, I_3) &= (I_2, E_1, I_3) \\ (E_1, I_2, E_3) &= (I_2, E_1, E_3) \\ (1, 2, 3 \text{ cyclic}) & \end{aligned} \quad (10.10)$$

Example Express $M_{12} = (I_1, E_1, I_3) / (E_2, E_1, I_3)$ in terms of derivatives having (E_1, E_2, E_3) as independent variables. From (10.7) and (10.8)

$$(I_1, E_1, I_3)(E_1, E_2, E_3) = (E_3, E_1, I_3)(I_1, E_1, E_2) + (E_3, I_1, E_1)(I_3, E_1, E_2)$$

Dividing both sides by (E_1, E_2, E_3) twice yields

$$M_{12} L_{33} = L_{12} L_{33} - L_{12} L_{32}$$

$$\begin{aligned} \text{Where } L_{33} &= (I_3, E_1, E_2) / (E_3, E_1, E_2) \\ L_{12} &= (I_1, E_3, E_1) / (E_2, E_3, E_1) \\ L_{13} &= (I_1, E_1, E_2) / (E_3, E_1, E_2) \\ L_{32} &= (I_3, E_3, E_1) / (E_2, E_3, E_1) \end{aligned}$$

The result required is thus: $M_{12} = L_{12} - (L_{13} L_{32} / L_{33})$
A summary of relations between L parameters and M parameters with the corresponding flow graphs is given in Table 11.

$n > 3$ The number of network equations corresponding (10.1) and (10.6) are given by $(2n)! / (n!)^2$ sets of n equations each. The number of dynamic characteristics will thus be $n^2 (2n)! / (n!)^2$ or 2, 24, 180, 1120, 6300, for $n=1, 2, 3, 4, 5$ respectively. Extension of the method given above to networks with $n > 3$ is straight forward and, although the calculation becomes exceedingly tedious, it presents a considerable saving in time over solving the equations explicitly by conventional methods. The case of $n=4$ finds application in the theory of thermoelectric and thermomagnetic effects¹³, but such an analysis is considered beyond the scope of this paper.

Summary

A mathematical technique is developed for a systematic and rapid analysis of the functional relationships between variables (voltages and current) of a network. For the four-terminal network six anti-commuting quantities, referred to as 'deviation' are defined so that each 'deviation' correlates any two variables of the system. Elementary rules for a calculus of 'deviations' are developed and are represented graphically by flow graphs, i.e. diagrams exhibiting the topological

properties of feedback systems.

Conditions for (i) uniqueness, (ii) reciprocity, (iii) symmetry and (iv) stability of a network are derived using the calculus of 'deviations' and are in part based on arguments adapted from recent developments in the thermodynamics of irreversible processes, power dissipation taking the place of entropy production. Other analogies between electrical networks and thermodynamic systems as well as mechanical systems are used to show that concepts from circuit theory (such as feedback or reciprocity) and the techniques (in particular, those developed) can be used to advantage in other fields of engineering.

With the aid of the calculus of deviations, relationships between dynamic characteristics, including hybrid parameters and feedback parameters, are investigated systematically.

Numerous examples pertaining to passive, unilateral and active (in particular transistors) networks illustrate the power and directness of these techniques. The method is also useful in conjunction with logarithmic techniques, in the synthesis of two or more networks and in the analysis of 'multi-terminal' networks.

References

1. Shea, R.F., (editor), "Principles of Transistor Circuits", Wiley 1953.
2. Peterson, L.C., Equivalent Circuits of Linear Active Four-Terminal Networks, BSTL, Vol. 27, October 1948.
3. Giacoletto, L.J., Terminology and Equations for Linear Active Four-Terminal Networks including Transistors, RCA Review, March 1953.
4. Klein, W., Die Ersatzbilder des nicht umkehrbaren Vierpols, Archiv der Elek. Uebertragung, May 1952.
5. Golden, N.J., Linear Applications of Transistors, Proceedings 1953, Transistor Short Course, Pennsylvania State College.
6. Knausenberger, G., Equivalent Circuits and Matrices for Transistor Circuits, Proceedings 1953 Transistor Short Course, Pennsylvania State College.
7. Shaw, N.A., Derivation of Thermodynamical Relations for a Simple System, Phil. Trans. Roy. Soc. (London) A234, 249 (1935).
8. Margenau, H. and Murphy, G.M., "The Mathematics of Physics and Chemistry", Van Nostrand 1943. Chapter I.
9. Nordheim, L. and Fuss, E., Die Hamilton-Jacobische Theorie der Dynamik, Vol. V. Handbuch der Physik. Springer 1927.
10. Whittaker, E.T., "Analytic Dynamics", Cambridge University Press 1937, Chapter IX and X.
11. Dirac, P.A.M., "Principles of Quantum Mechanics", Oxford University Press 1944. Section 25-30.
12. Rojansky, V., "Introduction to Quantum Mechanics", Prentice-Hall, 1938. Chapter II and IX.
13. deGroot, S.R., Thermodynamics of Irreversible

Processes, Interscience Publishing Company, New York 1951.

14. Lord Rayleigh (J.W. Strutt), *Phil. Mag.* **26**, 776 (1913).

15. Onsager, L. and Machlup, S. *Fluctuations and Irreversible Processes.* *Phys. Rev.* **91**, 1505 (1953).

16. Mason, S.J., *Feedback Theory—Some Properties of Signal Flow Graphs.* *Proc. I.R.E.*, Vol. 41, 9, 1144 (1953) and Ph.D. Thesis, M.I.T., 1951.

17. Brown, G.S., and Campbell, D.P., "Principles of Servomechanisms", Wiley 1948.

18. Ryder, R.M., (Co-author), *Transistor Devices, Notes of Lectures*, Published by the University of Illinois, 1953.

19. Chu, G.Y., *A New Equivalent Circuit for Junction Transistors*, *Convention Record of the I.R.E.*, March 1954.

20. Bode, H.W., "Network Analysis and Feedback Amplifier Design", D. Van Nostrand Company, New York 1945.

21. Mishkin, E., *Feedback Techniques Applied to the Single Phase Transformer*, *Proc. A.I.E.E.*, in Press.



Fig. 1 - Four-terminal network.

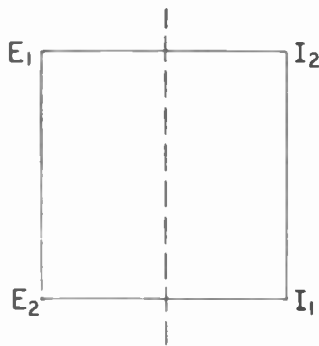
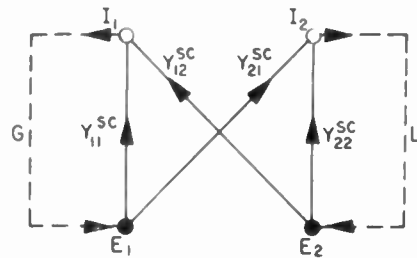
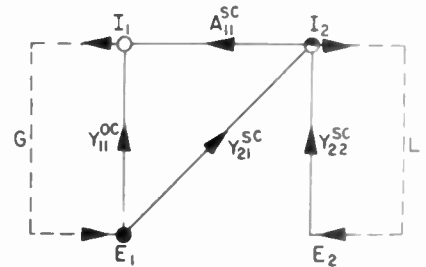


Fig. 2 - Symmetry properties of passive network.



(a)



(b)

Fig. 3 - Flow graphs of equivalent dynamic characteristics.

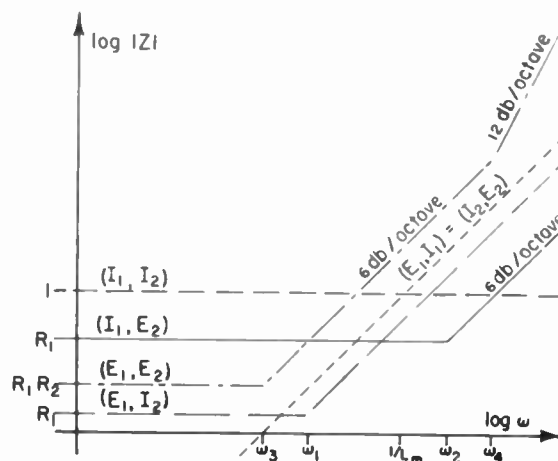


Fig. 4 - Dynamic characteristics of the transformer: Asymptotes for log-frequency versus log-magnitude plot.

Table 1

Open-Circuit and Short-Circuit Parameters

Dynamic Characteristics	Impedance: Z		Admittance: Y		Amplification: A	
	Driving point: Z_{11}	Transfer: Z_{12}	Driving point: Y_{11}	Transfer: Y_{12}	Driving point: A_{11}	Transfer: A_{12}
Open Circuit (superscript ∞) Current Constant	$\left(\frac{\partial E_1}{\partial I_1}\right)_{I_2}$	$\left(\frac{\partial E_1}{\partial I_2}\right)_{I_1}$	$\left(\frac{\partial I_1}{\partial E_1}\right)_{I_2}$	$\left(\frac{\partial I_1}{\partial E_2}\right)_{I_2}$	$\left(\frac{\partial E_1}{\partial E_2}\right)_{I_1}$	$\left(\frac{\partial E_1}{\partial E_2}\right)_{I_2}$
Short Circuit (superscript sc) Voltage Constant	$\left(\frac{\partial E_1}{\partial I_1}\right)_{E_2}$	$\left(\frac{\partial E_1}{\partial I_2}\right)_{E_2}$	$\left(\frac{\partial I_1}{\partial E_1}\right)_{E_2}$	$\left(\frac{\partial I_1}{\partial E_2}\right)_{E_1}$	$\left(\frac{\partial I_1}{\partial I_2}\right)_{E_1}$	$\left(\frac{\partial I_1}{\partial I_2}\right)_{E_2}$

Table 3. Analogues

Networks	Thermodynamics	Mechanics
Voltage E_1 Voltage E_2 Current I_1 Current I_2	Absolute Temperature T Absolute Pressure P Specific Entropy S Specific Volume V	Energy H Momentum P Time Coordinate T Displacement Coordinate Q
Uniqueness $E_2 = f(E_1, I_1)$	Equation of State $P = f(V, T)$	Equation of Motion $Q = f(H, T)$
Reciprocity This Paper $(E_1, I_1) = (I_2, E_2)$	Reversibility Maxwell's Equations $(T, S) = (V, P)$	Action and Reaction Hamilton's Equations $(H, T) = (Q, P)$
Principle of Least Dissipation	Thermodynamic Equilibrium	Principle of Least Action

Table 2

Standard Circuit Parameters for Four-Terminal Network

Network Equations	Flow Graphs ● Independent Variable ○ Dependent Variable	Passive Network	Symmetrical Network	Circuit Determinant
$E_1 = \frac{(E_1, I_2)}{(I_1, I_2)} I_1 + \frac{(E_1, I_1)}{(I_2, I_1)} I_2$ $E_2 = \frac{(E_2, I_2)}{(I_1, I_2)} I_1 + \frac{(E_2, I_1)}{(I_2, I_1)} I_2$		$z_{12} = z_{21}$ $(E_1, I_1) = (I_2, E_2)$	$z_{11} = z_{22}$ $(E_1, I_2) = (I_1, E_2)$	$ z = z_{11}z_{22} - z_{12}z_{21}$ $ z = \frac{(E_1, E_2)}{(I_1, I_2)}$
$I_1 = \frac{(I_1, E_2)}{(E_1, E_2)} E_1 + \frac{(I_1, E_1)}{(E_2, E_1)} E_2$ $I_2 = \frac{(I_2, E_2)}{(E_1, E_2)} E_1 + \frac{(I_2, E_1)}{(E_2, E_1)} E_2$		$y_{12} = y_{21}$ $(E_1, I_1) = (I_2, E_2)$	$y_{11} = y_{22}$ $(E_1, I_2) = (I_1, E_2)$	$ y = y_{11}y_{22} - y_{12}y_{21}$ $ y = \frac{(I_1, I_2)}{(E_1, E_2)}$
$I_1 = \frac{(I_1, I_2)}{(E_1, I_2)} E_1 + \frac{(I_1, E_1)}{(I_2, E_1)} I_2$ $E_2 = \frac{(E_2, I_2)}{(E_1, I_2)} E_1 + \frac{(E_2, E_1)}{(I_2, E_1)} I_2$		$g_{12} = -g_{21}$ $(E_1, I_1) = (I_2, E_2)$	$ g = 1$ $(E_1, I_2) = (I_1, E_2)$	$ g = g_{11}g_{22} - g_{12}g_{21}$ $ g = \frac{(I_1, E_2)}{(E_1, I_2)}$
$E_1 = \frac{(E_1, E_2)}{(I_1, E_2)} I_1 + \frac{(E_1, I_1)}{(E_2, I_1)} E_2$ $I_2 = \frac{(I_2, E_2)}{(I_1, E_2)} I_1 + \frac{(I_2, I_1)}{(E_2, I_1)} E_2$		$h_{12} = -h_{21}$ $(E_1, I_1) = (I_2, E_2)$	$ h = 1$ $(E_1, I_2) = (I_1, E_2)$	$ h = h_{11}h_{22} - h_{12}h_{21}$ $ h = \frac{(E_1, I_2)}{(I_1, E_2)}$
$E_1 = \frac{(E_1, I_2)}{(E_2, I_2)} E_2 + \frac{(I_1, E_2)}{(I_2, E_2)} I_2$ $I_1 = \frac{(I_1, I_2)}{(E_2, I_2)} E_2 + \frac{(I_1, E_2)}{(I_2, E_2)} I_2$		$ a = 1$ $(E_1, I_1) = (I_2, E_2)$	$a_{11} = -a_{22}$ $(E_1, I_2) = (I_1, E_2)$	$ a = a_{11}a_{22} - a_{12}a_{21}$ $ a = \frac{(E_1, I_1)}{(I_2, E_2)}$
$E_2 = \frac{(E_2, I_1)}{(E_1, I_1)} E_1 + \frac{(E_2, I_1)}{(I_1, E_1)} I_1$ $I_2 = \frac{(I_2, I_1)}{(E_1, I_1)} E_1 + \frac{(I_2, E_1)}{(I_1, E_1)} I_1$		$ b = 1$ $(E_1, I_1) = (I_2, E_2)$	$b_{11} = -b_{22}$ $(E_1, I_2) = (I_1, E_2)$	$ b = b_{11}b_{22} - b_{12}b_{21}$ $ b = \frac{(I_2, E_2)}{(E_1, I_1)}$

Table 4. Flow Graphs for Devices

Device	Transformer		Triode		Junction Transistor	
	Admittance	Impedance	Ground Grid	Ground Cathode	Ryder (18)	Chu (19)
References	Mishkin (21)		Knausenberger (6)		Ryder (18)	Chu (19)
Wiring Diagram						
Subscripts: 1 - input 2 - output 0 - ground						
Flow Graph						
Equivalent Variables	$I_1 - I_2$ $E_1 - E_2$	$I_p - I_s$ $E_p - E_s$	$I_o - I_p$ $E_o - E_p$	$I_g - I_p$ $E_s - E_p$	$I_e - I_o$ $E_e - E_o$	$I_e - I_o$ $E_e - E_o$
Subscripts	p - primary circuit s - secondary circuit m - magnetic circuit		g - grid p - plate o - cathode		b - base o - collector e - emitter	
Deviations	(I_1, I_2) (I_1, E_1) (I_1, E_2) (E_1, I_2) (E_2, I_2) (E_1, E_2)	1 $-Z_m$ $Z_m + Z_2$ $Z_m + Z_1$ $-Z_m$ $Z_1 Z_2 + Z_1 Z_m + Z_2 Z_m$	1 Z_g $(Z_g + Z_o)(1 + \mu) + Z_p$ Z_g $Z_g(1 + \mu)$ $Z_g Z_p + Z_g Z_o (1 + \mu)$	$Y_p Y_g$ $-$ $Y_g + Y_g Z_o (Y_p + g_m)$ $-Y_p$ $-g_m$ $1 + Z_o (Y_p + g_m)$	g_o $r'_b g_o + \mu_{bo} + \mu_{eo}$ 1 $g_o (r'_b + r'_e) - \mu_{eo} + \mu_{bo}$ a $r'_e + r'_b (1 - a)$	$Y_e Y_o \tanh \theta$ $Y_o / \cosh \theta$ Y_e Y_o $Y_e / \cosh \theta$ $\tanh \theta$
Circuit Connections	0 1 2 m p s	0 1 2 m p s	0 1 2 g o p	0 1 2 o g p	0 1 2 b e o	0 1 2 b e o
Remarks	I_m - magnetomotive force E_m - flux linkages Z_m - generalized reluctance N - transformer winding $N_1 = N_2 = 1$		μ - voltage amplification g_m - mutual conductance Y_p - plate conductance		μ - voltage amplification a - current amplification r'_b - majority current base resistance g_o - collector conductance Equivalent characteristic impedances: $Z_o = 1/Y_o$ $Z_e = 1/Y_e$	

Table 5.

Flow graphs of Equivalent Circuits for the Transistor

$$(I_1, I_2) = 1 \quad (V_1, V_2) = r_b r_c + r_c r_e + r_e r_b - r_b r_m$$

<p><u>Driving Diagram</u> Subscripts 1 - source 2 - load 0 - ground</p>						
<u>Circuit Connection</u>	<p>1 0 2 ε b o</p>	<p>1 0 2 b ε o</p>	<p>1 0 2 b o ε</p>	<p>1 0 2 ε o b</p>	<p>1 0 2 o ε b</p>	<p>1 0 2 o b ε</p>
<p><u>Flow Diagram</u> Subscripts b - base ε - emitter o - collector</p>						
<u>Deviations</u>	<p>(V_2, I_2) (V_1, I_2) (I_1, V_2) (I_1, V_1)</p> $\begin{matrix} r_b + r_m \\ r_b + r_e \\ r_b + r_o \\ r_b \end{matrix}$	$\begin{matrix} r - r_m \\ r_b + r_e \\ r_o + r_e - r_m \\ r_e \end{matrix}$	$\begin{matrix} r_o \\ r_b + r_o \\ r_o + r_e - r_m \\ r_o - r_m \end{matrix}$	$\begin{matrix} r_o - r_m \\ r_o + r_e - r_m \\ r_b + r_c \\ r_c \end{matrix}$	$\begin{matrix} r_e \\ r_o + r_e - r_m \\ r_b + r \\ r_e - r_m \end{matrix}$	$\begin{matrix} r_b \\ r_b + r_o \\ r_b + r_e \\ r_b + r_m \end{matrix}$

Table 6. Energy Converters

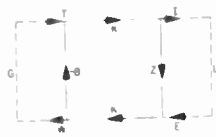
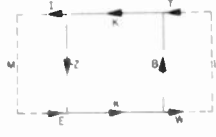

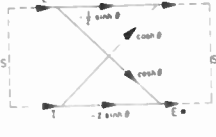

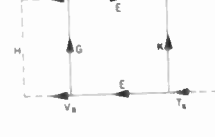
Generator	Motor	Pump	Transmission Line	Electromagnetic Wave	Thermoelectric Effect
					
T-Generator Torque W-Angular Speed I-Armature Current E-Back EMF	I-Armature Current E-Back EMF T-Motor Torque W-Angular Speed	T-Torque of Pump W-Angular Speed Q-Flow(ft ³ /sec/rad) P-Pressure(lb/ft ²)	E-voltage at sender (S) I-current at sender (S) E*-voltage at receiver (S*) I*-current at receiver (S*)	E-electric field in S H-magnetic field in S E*-electric field in S* H*-magnetic field in S*	T _x -temperature gradient TS _x -heat flow E _x -voltage gradient I _x -current flow
G - Generator L - Load	M - Motor L - Load	M - Prime Mover L - Load	S - Sender S* - Receiver	S - Stationary System S*-Moving System	H - Heat Source L - Electrical Load
B - friction and inertia Z - armature impedance	Z - armature impedance B - friction and inertia	B - friction and inertia Z - leakage(ft ⁵ /lb/sec)	Z - characteristic impedance	c - velocity of light	G - electrical conductivity TK - thermal conductivity
K - Generator constant	K - Motor constant	K - Flow constant	theta - (attenuation constant x length of line) cosh ² theta + sinh ² theta = 1	v-relative velocity of S with respect to S* theta = tanh (v/c)	E-thermoelectric power T - temperature

Table 7. Two-Terminal Flow Graphs: Elementary Operations

MULTIPLICATION		\equiv	
ADDITION		\equiv	
RECIPROCAL		\equiv	
$\frac{1}{R} = \sum_{N=0}^{\infty} \circ^N = \frac{1}{1 - \circ}$		\equiv	

Table 8. Three-Terminal Flow Graphs: Equivalent Junctions

	Contributive Junction \circ	Distributive Junction \bullet		
PATH INVERSION	 (I)	 (II)	 (III)	 (IV)
CASCADE CONVERSION	 (V)	 (VI)	 (VII)	 (VIII)
FEEDBACK CONVERSION	 (IX)	 (X)	 (XI)	 (XII)

Table 9. Hybrid Parameters For Four-Terminal Flow Graphs

Flow Graph						
Junctions: ○ Contributive ● Distributive ⊖ Multiple						
(I_1, I_2)	1	1	1	1	1	1
(E_1, I_2)	z_{11}	z_{11}	$h_{11} - h_{12} z_{22} h_{21}$	$h_{11} + z_{12} h_{21}$	z_{11}	$-b_{22} z_{12}$
(I_1, E_1)	z_{12}	$-z_{11} \epsilon_{12}$	$h_{12} \epsilon_{22}$	z_{12}	$a_{12} - z_{11} a_{22}$	z_{12}
(E_2, I_2)	$z_{11} \epsilon_{21}$	$\epsilon_{21} z_{11}$	$-z_{22} h_{21}$	$-z_{22} h_{21}$	z_{21}	$b_{12} - b_{11} z_{12} b_{22}$
(I_1, E_2)	$\epsilon_{22} + z_{12} \epsilon_{21}$	$\epsilon_{22} - \epsilon_{21} z_{11} \epsilon_{12}$	z_{22}	z_{22}	$-a_{22} z_{21}$	$z_{12} b_{11}$
(E_1, E_2)	$z_{11} \epsilon_{22}$	$z_{11} \epsilon_{22}$	$h_{11} z_{22}$	$h_{11} z_{22}$	$-a_{12} z_{21}$	$-z_{12} b_{12}$

Table 10. Feedback Parameters for Four-Terminal Flow Graphs

Flow Graph						
Junctions: ○ Contributive ● Distributive ⊖ Multiple						
(I_1, I_2)	$1 - \epsilon_{21} h_{12}$	$1 - h_{11} y_{12} \epsilon_{22} y_{21}$	$1 - b_{12} y_{22} a_{12} y_{11}$	$1 - h_{21} \epsilon_{22} y_{12}$	$1 + h_{11} y_{12} \epsilon_{21}$	$1 - h_{21} \epsilon_{12}$
(E_1, I_2)	h_{11}	h_{11}	$-b_{12} y_{22} a_{12}$	h_{11}	h_{11}	h_{11}
(I_1, E_1)	$\epsilon_{22} h_{12}$	$-h_{11} y_{12} \epsilon_{22}$	a_{12}	$-\epsilon_{22} y_{12} h_{11}$	$-\epsilon_{22} y_{12} h_{11}$	$-\epsilon_{12} h_{11}$
(E_2, I_2)	$h_{11} \epsilon_{21}$	$-\epsilon_{22} y_{21} h_{11}$	b_{12}	$-h_{21} \epsilon_{22}$	$\epsilon_{21} h_{11}$	$-h_{21} \epsilon_{22}$
(I_1, E_2)	ϵ_{22}	ϵ_{22}	$-a_{12} y_{11} b_{12}$	ϵ_{22}	ϵ_{22}	ϵ_{22}
(E_1, E_2)	$h_{11} \epsilon_{22}$	$h_{11} \epsilon_{22}$	$-a_{12} b_{21}$	$h_{11} \epsilon_{22}$	$h_{11} \epsilon_{22}$	$h_{11} \epsilon_{22}$

Table 11.

Relations between two sets of dynamic characteristics for circuit with three currents and three voltages.

Independent Variables ●	E_1 E_2 E_3	E_1 E_2 E_3
Dependent Variables ○	I_1 I_2 I_3	I_1 I_2 I_3
<p>L - Parameters</p> $L_{11}=(I_1, E_2, E_3)/L$ $L_{12}=(I_1, E_3, E_1)/L$ $L_{13}=(I_1, E_1, E_2)/L$ $L_{21}=(I_2, E_2, E_3)/L$ $L_{22}=(I_2, E_3, E_1)/L$ $L_{23}=(I_2, E_1, E_2)/L$ $L_{31}=(I_3, E_2, E_3)/L$ $L_{32}=(I_3, E_3, E_1)/L$ $L_{33}=(I_3, E_1, E_2)/L$ $L=(E_1, E_2, E_3)$		
<p>M - Parameters</p> $M_{11}=(I_1, E_2, I_3)/M$ $M_{12}=(I_1, I_3, E_1)/M$ $M_{13}=(I_1, E_1, E_2)/M$ $M_{21}=(I_2, E_2, I_3)/M$ $M_{22}=(I_2, I_3, E_1)/M$ $M_{23}=(I_2, E_1, E_2)/M$ $M_{31}=(E_3, E_2, I_3)/M$ $M_{32}=(E_3, I_3, E_1)/M$ $M_{33}=(E_3, E_1, E_2)/M$ $M=(E_1, E_2, E_3)$		

SOME TECHNIQUES FOR NETWORK SYNTHESIS

by

George L. Matthaei
Division of Electrical Engineering
University of California
Berkeley, California

Summary

A method is presented for realization of any minimum-phase transfer function in a c-r (i.e., constant-resistance) ladder. Such networks have applications similar to those of conventional RLC, c-r bridged-T's but possess the advantages of more flexibility in design, fewer elements, and often less loss.

C-r networks using two pole and zero arm immittances are particularly easy to design. Then simple formulas can guide the designer to networks without unnecessary elements. Straight-forward techniques are presented for realization of the arm immittances. The realization process is facilitated by rules for making RLC continued-fraction expansions by "forward" and "reverse" division.

Introduction

We shall first outline techniques which taken together provide an organized system for realizing any second-order immittance function without the use of coupled coils. Herein, the term, "second-order immittance", will be used to refer to any impedance or admittance function having two poles and zeros. As Bode shows,¹ such immittances can be used as basic building blocks in the synthesis of "constant-resistance" networks having transfer functions of arbitrary complexity. In the latter part of this paper a constant-resistance network which is apparently new will be described.

Three Classes of Second-Order Immittances

Any second-order immittance function can be expressed in the form:

$$F(p) = \frac{gp^2 + ap + b}{p^2 + ep + d} \quad (1)$$

where g , a , b , e , and d are constant coefficients and p is the complex frequency variable $p = \sigma + j\omega$. Brune shows that for such an immittance to be realizable in a passive network, $F(p)$ must be what he calls a positive-real (abbreviated p-r) function.² It can be shown that for this relatively simple case the conditions for p-r character may be reduced to:

- A. All non-zero coefficients must be real and positive.
- B. $\text{Re } F(j\omega) \geq 0$ for all $p = j\omega$. (2)

It can also be shown that if condition A is satisfied, condition B will be satisfied if and only if

$$b - ae + dg \leq 2 \sqrt{gbd} \quad (3)$$

The ease with which any given $F(p)$ can be realized depends very much on the nature of its $\text{Re } F(j\omega)$ characteristic. In general, we may divide p-r, $F(p)$ functions into three classes:

Class I. $\text{Re } F(j\omega)|_{\min}$ occurs at $\omega = 0$ as shown in Fig. 1. $F(p)$ can be shown to be a p-r, Class I function if p-r condition A is satisfied and

$$g > \frac{b}{d} \text{ while } \frac{ad - be}{gd - b} \geq \frac{d}{e} \quad (4)$$

Class II. $\text{Re } F(j\omega)|_{\min}$ occurs at $\omega = \infty$ as shown in Fig. 2. $F(p)$ can be shown to be a p-r, Class II function if p-r condition A is satisfied and

$$g < \frac{b}{d} \text{ while } 0 < \frac{b - gd}{a - ge} \leq e \quad (5)$$

Class III. $\text{Re } F(j\omega)|_{\min}$ occurs at a finite value of ω as shown in Fig. 3. $F(p)$ will be a p-r, Class III function if: p-r condition A is satisfied, while (3) is satisfied, while (4) and (5) are not.

Some special cases are often of interest. If in (1) $b = 0$, and condition A is satisfied then either (3) or (4) will yield

$$\frac{a}{g} \geq \frac{d}{e} \quad (6)$$

In this case if (6) is satisfied, $F(p)$ is a Class I function which satisfies (2) with an equal sign at $\omega = 0$. Herein we shall refer to immittance functions that satisfy (2) with an equal sign at some frequency ω as "minimum real part" functions.

If $g = 0$ in (1) and condition A is satisfied, then either (3) or (5) will yield

$$\frac{b}{a} \leq e \quad (7)$$

In this instance if (7) is satisfied, $F(p)$ is a p-r, Class II, minimum real part function.

$$\text{When } F(p) = \frac{gp^2 + b}{ep} \quad (8)$$

$F(p)$ is a reactance function and will be p-r if and only if p-r condition A is satisfied. Eq. (2) is satisfied with an equal sign for all $j\omega$ in this case, thus (8) is simultaneously a Class I, II, and III function.

The general one-pole and -zero function

$$F(p) = \frac{ap + b}{ep + d} \quad (9)$$

also needs only to satisfy p-r condition A in order to be p-r. This function can never be Class III, but will be Class I if $a/e > b/d$ and Class II if $a/e < b/d$.

Continued Fraction Expansions

The input impedance function of the ladder network of Fig. 4 may be represented by the continued-fraction expansion

$$Z_{in} = Z_1 + \frac{1}{Y_2 + \frac{1}{Z_3 + \frac{1}{Y_4 + \dots \text{etc.}}}} \quad (10)$$

Also the input admittance of the network in Fig. 5 may be analogously represented by

$$Y_{in} = Y_1 + \frac{1}{Z_2 + \frac{1}{Y_3 + \frac{1}{Z_4 + \dots \text{etc.}}}} \quad (11)$$

Some kinds of immittances are very easily realized in these forms by making a continued-fraction expansion by use of what shall herein be referred to as forward and reverse division.

Let

$$F_0(p) = \frac{q_n p^n + q_{n-1} p^{n-1} + \dots + q_1 p + q_0}{h_k p^k + h_{k-1} p^{k-1} + \dots + h_1 p + h_0} \quad (12)$$

be an immittance of arbitrary complexity. We shall define forward division as long division carried out in the manner indicated by:

$$h_k p^k + \dots + h_1 p + h_0 \sqrt{q_n p^n + \dots + q_1 p + q_0} \quad (13)$$

Similarly reverse division will be defined as long division carried out in the manner indicated by:

$$h_0 + h_1 p + \dots + h_k p^k \sqrt{q_0 + q_1 p + \dots + q_n p^n} \quad (14)$$

It should be noted that for (12) to be p-r, the highest powers in the numerator and denominator polynomials cannot differ by more than one, and likewise for the lowest powers.

An important kind of continued-fraction expansion can be made by use of the following operations involving forward and reverse division:

1. If $q_0 > 0$ but $h_0 = 0$, then $F_0(p)$ has a pole at the origin, and it can be removed by one step of reverse division. The result will be of the form

$$F_0(p) = \frac{q_0}{h_1 p} + F_1(p), \quad (15)$$

where $F_1(p)$ will not have a pole at the origin and will be p-r if $F_0(p)$ is p-r.

2. If $q_0 = 0$, but $h_0 > 0$, then $F_0(p)$ has a zero at the origin. In this case the function is inverted to make the zero at the origin become a pole and then operation 1 is applied to remove the pole. In this case the result is of the form

$$F_0(p) = \frac{1}{1/F_0(p)} = \frac{1}{\frac{h_0}{q_1 p} + F_2(p)}, \quad (16)$$

where $F_2(p)$ will not have a pole at the origin and will be p-r if $F_0(p)$ is p-r.

3. If terms g_0 and h_0 are present and $F_0(p)$ is a Class I p-r function, then a step of reverse division will remove a constant equal to $\text{Re } F(j\omega) \Big|_{\min.} = \text{Re } F(j0) = F(0)$, and the remainder function $F_3(p)$ will be a minimum real-part p-r function. The result is of the form

$$F_0(p) = \frac{q_0}{h_0} + F_3(p). \quad (17)$$

If $F_0(p)$ is not a Class I function but its reciprocal is, reverse division may be applied to $1/F_0(p)$ to give

$$F_0(p) = \frac{1}{1/F_0(p)} = \frac{1}{\frac{h_0}{q_0} + F_{3A}(p)}. \quad (18)$$

The remainder function $F_3(p)$ or $F_{3A}(p)$ will have a zero at the origin so can be broken down further by operation 2.

4. If $n = k + 1$, then $F_0(p)$ has a pole at infinity which can be removed by forward division to give

$$F_0(p) = \frac{q_n p}{h_k} + F_4(p). \quad (19)$$

$F_4(p)$ will not have a pole at infinity and will be p-r if $F_0(p)$ is p-r.

5. If $n = k - 1$, then $F_0(p)$ has a zero at infinity. In this case the function is inverted to make the zero a pole, and then operation 4 is applied to give

$$F_0(p) = \frac{1}{\frac{h_k p}{q_n} + F_5(p)}. \quad (20)$$

$F_5(p)$ will not have a pole at infinity and will be p-r if $F_0(p)$ is p-r.

6. If $F_0(p)$ is a Class II function and $n = k$, then a step of forward division will remove a constant equal to $\text{Re } F_0(j\omega) \Big|_{\min.} = \text{Re } F(j\infty) = F(\infty)$, and the remainder function $F_6(p)$ will be a minimum real-part function. The result is of the form

$$F_o(p) = \frac{q_n}{h_k} + F_6(p) \quad (21)$$

If $F_o(p)$ is not Class II but its reciprocal is, forward division may be applied to $1/F_o(p)$ to give

$$F_o(p) = \frac{1}{\frac{h_k}{q_n} + F_{6A}(p)} \quad (22)$$

The remainder function $F_6(p)$ or $F_{6A}(p)$ will have a zero at infinity so can be broken down further by operation 5.

By use of these six operations numerous immittance functions can be completely broken into a continued fraction of simple p-r terms. These terms can then be identified as series impedances or shunt admittances of a ladder network in accordance with eq. (10) or (11). Many readers will recall that Cauer's continued-fraction method for synthesis of RC, RL, and LC immittances utilizes what amounts to these same operations.⁴ Not so widely recognized is the fact that these operations are also helpful for synthesis of RLC networks. The necessary condition which makes it possible to completely expand an immittance this way is that after each step of forward or reverse division, the remainder function or its reciprocal must be a Class I or II p-r function. If the remainder function and its reciprocal are both Class III, additional techniques must be introduced in order to break down the remainder function.

Synthesis of Second-Order Immittances

The simplest network which can represent a second-order immittance will be determined by the properties of the given immittance and will vary considerably. The realization methods about to be described provide a straight-forward approach for realizing any second-order p-r function.

Type A Realization

Any Class I or II $F(p)$ can be realized quickly by making a continued-fraction expansion by use of forward and reverse division. Such realizations will be referred to as Type A realizations. Consider the example:

$$F(p) = \frac{2p^2 + 5p + 2}{p^2 + 2p + 2} \quad (23)$$

Eq. (4) shows this to be a Class I function. Applying reverse division (operation 3) gives

$$F(p) = 1 + \frac{p^2 + 3p}{p^2 + 2p + 2}, \quad (24)$$

where the remainder is a Class I function with a zero at the origin. By inversion and reverse division (operation 2):

$$F(p) = 1 + \frac{1}{\frac{2}{3p} + \frac{4/3 + p}{3 + p}} \quad (25)$$

As indicated in connection with eq. (9), the re-

mainder is easily seen to be Class I. By reverse division (operation 3):

$$F(p) = 1 + \frac{1}{\frac{2}{3p} + \frac{4/9 + \frac{5}{9}p}{3 + p}} \quad (26)$$

Inverting the remainder and dividing gives

$$F(p) = 1 + \frac{1}{\left(\frac{2}{3p} + 4/9\right) + \frac{1}{\left(\frac{27}{5p} + \frac{9}{5}\right)}} \quad (27)$$

If $F(p)$ is construed to be an impedance, we may associate (27) with (10) and Fig. 4, and the network will be as shown in Fig. 6A. If $F(p)$ were an admittance, we should associate (27) with (11) and Fig. 5 to get the "reciprocal" network to that of Fig. 6A.

If the remainder of (25) is inverted it will become Class II. A different realization can be obtained by performing this inversion and then finishing the expansion by use of forward-division operations. The result is

$$F(p) = 1 + \frac{1}{\frac{2}{3p} + \frac{1}{1 + \frac{1}{\frac{3p}{5} + \frac{1}{5/4}}}} \quad (28)$$

Fig. 6B shows the corresponding network if $F(p)$ is an impedance.

Type A realizations will require no more than five elements. Some will require less; for example function (8) needs only two elements.

Type B Realizations

Some $F(p)$ are Class III as they stand, but become Class I or II when inverted. The function $1/F(p)$ will be Class I if

$$g < b/d \text{ and } \frac{ad - be}{gd - b} \geq \frac{b}{a}, \quad (29)$$

while it will be Class II if

$$g > b/d \text{ and } 0 < \frac{b - gd}{a - ge} \leq \frac{a}{g}. \quad (30)$$

To get a network, we may start with

$$F(p) = \frac{1}{1/F(p)}, \quad (31)$$

and then expand $1/F(p)$ just as in a Type A realization. Due to (31), what would normally be the first term of the expansion will be missing. This simply means that the network of Fig. 4 would have $Z_1 = 0$ while Fig. 5 would have Y_1 missing.

"Type B" realizations also require no more than five elements. It is interesting to note that in some cases where $F(p)$ is Class I or II, $1/F(p)$ will be Class II or I, respectively. In

such instances there will often be a total of four, distinctly different, five-element equivalent circuits of which two can be obtained by the Type A method and the other two by the Type B.

Type C Realization

If neither $F(p)$ nor its reciprocal is a Class I or II function, then the Type A and B procedures fail. Any Class III function can be realized by Brune's method, but his method requires unity-coupled coils in order to realize a Class III function of the form of eq. (1).² If unity-coupled coils are to be excluded, the simplest procedure appears to result from breaking the Class III function into the sum of a Class I, minimum real-part function plus a Class II, minimum real-part function. This gives

$$F(p) = \frac{gp^2 + tp}{p^2 + ep + d} + \frac{(a-t)p + b}{p^2 + ep + d} \quad (32)$$

It can be shown that both terms in (32) will be p-r if p-r condition A is satisfied and

$$gd/e \leq t \leq a - b/e \quad (33)$$

Consider the example

$$F(p) = \frac{p^2 + 4p + 7}{p^2 + 2p + 1} \quad (34)$$

In this case (33) can be satisfied on both sides by an equal sign if $t = 1/2$. Eq. (32) becomes

$$F(p) = \frac{p^2 + 0.5p}{p^2 + 2p + 1} + \frac{3.5p + 7}{p^2 + 2p + 1} \quad (35)$$

The two terms in (35) can be expanded in continued fractions to give:

$$F(p) = \frac{2}{p} + \frac{1}{1 + \frac{1}{2p}} + \frac{1}{3.5p} + \frac{1}{3.5p + \frac{1}{(1/7)}} \quad (36)$$

Eq. (36) represents two ladder networks connected together. If $F(p)$ is construed to be an impedance, the realization is as shown in Fig. 7.

Observe that Fig. 7 represents two, three-element ladders connected in series. In most cases it will only be possible to satisfy (33) with an equal sign on one side. Then one of the ladders will have four elements while the other will have only three. In general it will be found that this type of realization will require no more than seven elements. It can be shown that if $F(p)$ fails to qualify for Type C realization, then $1/F(p)$ also fails to qualify.

Type D Realization

If $F(p)$ satisfies p-r condition A and eq. (3), but doesn't satisfy the conditions for Type A, B, or C realization, then what will herein be called a Type D realization appears to be necessary if unity-coupled coils are to be avoided.

The first step in this realization is to subtract a constant α_1 (see Fig. 3) from $F(p)$ so that the result will be a minimum real-part function $F'(p)$. The constant α_1 can be found by finding the smallest root $\alpha_k = \alpha_1$ of

$$(U^2 - 4d^2)\alpha_k^2 + (4gd^2 + 4bd - 2UW)\alpha_k + (W^2 - 4gdb) = 0 \quad (37)$$

which will satisfy

$$x_k = \frac{W - \alpha_k U}{2g - 2\alpha_k} > 0 \quad (38)$$

where

$$W = b - ae + dg,$$

$$\text{and } U = 2d - e^2,$$

the a , b , e , d , and g parameters being those of (1). If there are two α_k roots of (37) that satisfy (38), then the larger one is α_2 in Fig. 3 while the smaller one is α_1 . The frequency at which the minimum or maximum of $\text{Re } F(j\omega)$ occurs can be found from

$$j\omega_k = j\sqrt{x_k} \quad (39)$$

where ω_1 will be the frequency of the minimum point and ω_2 will be the frequency of the maximum; x_k for $k = 1, 2$ being given by (38).

Now the minimum real-part function $F'(p)$ may be expressed as

$$F'(p) = F(p) - \alpha_1 \quad (40)$$

$$= \frac{g'p^2 + a'p + b'}{p^2 + e'p + d'} \quad (41)$$

Application of either forward or reverse division to (41) will give a non-p-r remainder; so instead, from (41) we shall form the equivalent function

$$F'(p) = \frac{F'(p)(p+K)}{(p+K)} = \frac{Jp^2 + J\omega_1^2}{p^3 + (e' + K)p^2 + (d' + Ke')p + d'K} + \frac{g'p^3 + (a' + g'K - J)p^2 + (b' + a'K)p}{p^3 + (e' + K)p^2 + (d' + Ke')p + d'K} \quad (42)$$

where

$$K = \frac{e'\omega_1^2}{d' - \omega_1^2} \quad \text{and} \quad J = \frac{b'K}{\omega_1^2}$$

Both terms of (42) are p-r; the first having a zero at infinity, the second having a zero at the origin, and both having zero real part at ω_1 . The circuit is now obtained by expanding both terms of (42) in continued-fraction expansions, starting

the first term with inversion and forward division (operation 5) and the second term with inversion and reverse division (operation 2).

Fig. 8 shows a realization for the impedance $F(p) = Z(p) = \frac{1.1 p^2 + 0.813 p + 0.29394}{p^2 + 2p + 2}$. (43)

Note that for this impedance case, α_1 is realized as a series resistance and the continued-fraction expansions are realized as two ladders connected in series. It is interesting to note that this same network can be obtained by the method of Bott and Duffin.⁵ The technique described herein, however, has the advantage of eliminating the work of finding the real, positive root of a third-degree polynomial.³ Type D realizations will require no more than nine elements. If (43) had been an admittance, the structure would have taken the form of a conductance and two ladders all connected in parallel.

The Ranges of Realization

To give a better insight into when these different kinds of realization are possible, Fig. 9 illustrates the various ranges of zero locations of $F(p)$ eq. (1), when the poles are located at $p = -2 \pm j1$. Only the second quadrant of the p -plane is shown since no zeros can occur in the first or fourth quadrants and the second and third quadrants are symmetrical with respect to the real axis. With these given poles, $F(p)$ will be p - r if it has a real, positive constant multiplier and the zeros occur in conjugate pairs such that the second-quadrant zero lies within the outer curved contour. If the zero lies within the cross-hatched region marked A_I (45 deg. cross-hatch lines falling from left to right), then the function is Class I and can be realized with a Type A realization. If the zero is in the A_{II} region (45° cross-hatch rising from left to right), $F(p)$ is a Class II function and Type A realization is again possible. Note that the region of Type B realization (horizontal cross-hatching) overlaps the Type A region in places. Where they overlap, both $F(p)$ and its reciprocal are Class I or II functions; and where they do not overlap, either $F(p)$ or its reciprocal is a Class III function. If the second-quadrant zero lies within the outer contour but outside of the regions of Type A or B realization, then both $F(p)$ and its reciprocal are Class III p - r functions.

The region of Type C realizability overlaps all of the region of Type A realization and part of the region of Type B realization; however, since more elements are required for Type C, one would probably want to use it only when the zero lies in one of the unshaded regions marked C. The region of Type D realization overlaps part of the region of Type C realization and part of the region of Type B; but again since Type D realization requires the most elements of all, it would usually be undesirable unless the second-quadrant zero lies in one of the vertically cross-hatched regions marked D. If $F(p)$ has simple zeros on

the real axis, any of the Type A, B, C, or D realizations may be possible depending on the relative locations of the zeros. It is interesting to note that the boundaries of the various regions of realization in Fig. 9 all have geometric symmetry with respect to the circle about the origin which passes through the poles.

Synthesis of a Constant-Resistance Ladder

Let us now consider the design of a ladder network having a constant-resistance input and a prescribed, minimum-phase transfer function. We shall stipulate that the network is to be driven by a generator with a one-ohm internal resistance, and the ladder impedance is to match this. The transfer function may be written as

$$T(p) = A_1 \frac{H_1}{H_2} = A_1 T_m(p) = \frac{\text{Input Voltage}}{\text{Output Voltage}} \quad (44)$$

where H_1 and H_2 are polynomials, A_1 is a constant, and $T_m(p)$ will be defined later. We shall temporarily stipulate that $T(p)$ is a non-minimum real part p - r function; however as we shall see, this stipulation is easily removed.

If the generator has an internal resistance of one ohm and the network has an input resistance of one ohm, then the zero-impedance voltage generator in a Thevenin equivalent circuit will see the resistance:

$$Z_{in} = 2 = 2 \frac{H_1}{H_1} = 1 + \frac{H_1}{H_1} \quad (45)$$

Here one should note that the zeros of an impedance will be natural modes of vibration if it is driven by a zero-impedance generator; the poles will be natural modes of vibration if the impedance is driven from an infinite-impedance source. Introducing H_1 into both the numerator and denominator of (45) will give the circuit natural modes of vibration corresponding to the zeros of H_1 when Z_{in} is driven by either a zero internal impedance source or an infinite internal impedance source. It can be shown that the zeros of (44) will be natural modes of vibration, hence, specifying (45) in the way that we have will give (44) the proper zeros.

Now let us define

$$\nu_1 = \text{Re} \left. \frac{H_1(j\omega)}{H_2(j\omega)} \right|_{\min} > 0 \quad (46)$$

and

$$\nu_2 = \text{Re} \left. \frac{H_2(j\omega)}{H_1(j\omega)} \right|_{\max} > 0 \quad (47)$$

Both ν_1 and ν_2 will be greater than zero since $T(p)$ was stipulated to be a non-minimum real part p - r function. Now the function (45) may be expressed as the continued fraction

$$Z_{in} = 1 + \left(\frac{K_2 H_1 - H_2}{K_2 H_1} \right) + \frac{1}{\left(\frac{K_2 H_1 - K_1 K_2 H_2}{H_2} \right) + (K_1 K_2)} \quad (48)$$

$$= 1 + Z_1 + \frac{1}{Y_2 + G_L}, \quad (49)$$

where each term of the expansion will be p-r if K_1 and K_2 are constants such that

$$0 < K_1 \leq \nu_1 \quad (50)$$

$$\text{and } K_2 \geq \nu_2. \quad (51)$$

A circuit for expansions (48) and (49) is shown in Fig. 10. The transfer function of this circuit has the zeros of (44) due to (45). The frequencies of infinite loss (poles of eq. (44)) are caused by the signal being shorted out by the shunt branch Y_2 at frequencies corresponding to the zeros of H_2 . The series branch Z_1 serves as what Guillemin and others have called a "zero-shifting" branch, and it will not cause any points of infinite loss.^{3,6} Thus the points of infinite loss will be determined by the poles of Y_2 ; and these are the same as the poles of (44), as they should be. From such reasoning it can be seen that the network in Fig. 10 has the transfer function (44), at least within a constant multiplier. By carrying out the synthesis in a similar but dual manner, the constant-resistance network in Fig. 11 is obtained. This network also has the transfer function (44) within a constant multiplier.

Since the input impedance of the transmission network in Fig. 10 (or in Fig. 11) is a constant resistance, the impedance level of one section of this type can be adjusted so that its input will serve as the proper constant load resistance for another section. In this manner any number of such simple sections can be designed separately and then cascaded together to give a ladder network whose transfer function is the product of the section transfer functions (or the product of the section transfer functions within a constant multiplier, depending on how the transfer function is defined). Thus as with other constant-resistance networks,¹ the realization of a complicated transfer function can be greatly simplified by carrying out the realization in small parts. It should be noted that the transfer function for each component section must be a non-minimum real part p-r function, but the overall transfer function need not be p-r! The process then is to select poles and zeros for each section of the ladder so that the section transfer functions will be non-minimum real part p-r. If it should be impossible to factor the overall transfer function into a complete set of p-r factors, additional factors of the form $(p - p_1)/(p - p_1)$

can always be introduced so as to make p-r factorization possible. Thus any minimum phase transfer function can be realized within a flat loss factor in a constant-resistance ladder.

If the overall transfer function is factored so that each section transfer function has either one or two poles and zeros, then the branch immittances of the sections will have only one or two poles and zeros and the synthesis techniques previously mentioned can be used to carry out the design in a straight-forward manner.

Ladder and Bridged-T Flat Loss Comparison

Since these networks will often have unequal terminations, let us define our transfer function in terms of the voltage available at the load conductance G_L (or resistance R_L). Thus,

$$P_{g \text{ Avail.}} = \frac{E_g^2}{4 R_g} = P_{L \text{ Avail.}} = G_L E_L^2 \text{ Avail.}, \quad (52)$$

where $P_{g \text{ Avail.}}$ is the available power of the generator, E_g and R_g are as defined in Fig. 10, $P_{L \text{ Avail.}}$ is the power available at the load, and $E_L \text{ Avail.}$ is the voltage which this generator would cause across the given load G_L if the generator and load were perfectly matched by an ideal transformer. From (52),

$$E_L \text{ Avail.} = \frac{1}{2\sqrt{R_g G_L}} E_g. \quad (53)$$

It appears logical to define the transfer function then as

$$T(p) = \frac{E_L \text{ Avail.}}{E_L}, \quad (54)$$

the ratio of the available load voltage to the delivered load voltage. Eq. (54) is essentially an input over output ratio as is (44). Since the delivered load voltage cannot exceed the available voltage,

$$\left| T(j\omega) \right| = \left| \frac{E_L \text{ Avail.}}{E_L} \right| \geq 1. \quad (55)$$

Any transfer function (54) that satisfies (55) with an equal sign at some steady-state frequency $j\omega$ is a minimum-loss transfer function.

Let us define

$$T_m(p) = \frac{H_1}{H_2} \quad (56)$$

as a minimum-loss transfer function. Then A_1 in (44) is a constant factor equal to or greater than one, which indicates the flat loss of the network. For the networks of Figs. 10 or 11, A_1 is computed to be

$$A_1 = \sqrt{\frac{K_2}{K_1}}. \quad (57)$$

The smallest value for A_1 (and the least flat loss) will be obtained when (50) and (51) are satisfied with equal signs to give:

$$A_1|_{\min.} = \sqrt{\frac{\nu_2}{\nu_1}} \quad (58)$$

Since any minimum-phase transfer function which can be realized in one conventional constant-resistance bridged-T section can also be realized in one constant-resistance ladder section,¹ the two have about the same realm of application, and it appears relevant to compare their attenuation factors. The smallest attenuation factor for the conventional constant-resistance bridged-T is computed to be:

$$A_2|_{\min.} = \frac{1}{\nu_1} \quad (59)$$

Since transfer function (56) has no flat loss, from (50) and (51) it can be seen that both ν_1 and ν_2 must be no greater than one. Hence the ladder $A_1|_{\min.}$ will always be less than the bridge-T $A_2|_{\min.}$ except when both methods of realization give minimum-loss transfer functions, i.e. except when $A_1|_{\min.} = A_2|_{\min.} = 1$.

Some Practical Considerations

Let us assume that the poles and zeros of a complicated transfer function are to be factored into groups and then realized in a chain of constant-resistance ladder sections. Let us suppose further that each section is to contribute two poles and zeros to the overall transfer function. It may be possible to group the poles and zeros so that most or all of the section transfer functions are Class I or II p-r functions. Each section that has a Class I or II transfer function will have a second-order, Class I or II immittances in both arms, and considerably fewer elements will be required than if Class III immittances were involved.

Let us suppose

$$\frac{gp^2 + ap + b}{p^2 + ep + d} = \text{an (Input/Output) ratio} \quad (60)$$

has been factored out to be realized as one section of the ladder. It can be shown that if (60) satisfies eqs. (4), then (60) and all of the arm immittances in Figs. 10 and 11 will be Class I functions. Then the corresponding minimum-loss transfer function (56) is

$$T_m(p) = \frac{d}{b} \left(\frac{gp^2 + ap + b}{p^2 + ep + d} \right) = \frac{E_L \text{ Avail.}}{E_L} \quad (61)$$

for which

$$\nu_1 = \nu_2 = 1 \quad (62)$$

If (60) satisfies eqs. (5), then (60) and the arm immittances in Figs. 10 and 11 will all be Class II functions. The corresponding mini-

imum-loss transfer function is

$$T_m(p) = \frac{1}{g} \left(\frac{gp^2 + ap + b}{p^2 + ep + d} \right) = \frac{E_L \text{ Avail.}}{E_L} \quad (63)$$

for which

$$\nu_1 = \nu_2 = 1 \quad (64)$$

Note from (58), (62), and (64), that if (60) is a Class I or II function, the section will require no flat loss.

If (60) is Class III, the multiplier required in order to form the minimum-loss transfer function $T_m(p)$ can be determined as described in the Appendix. This is of interest for pre-determining the flat loss of the section, but it is not a necessary step in the design procedure. Defining (60) as H_1/H_2 , the values of (46) and (47) can be determined by applying eqs. (37) and (38) to (60) and its reciprocal. The values of ν_1 and ν_2 obtained would not be the same as those computed from $T_m(p)$, and eqs. (57) to (59) would no longer hold, but the resulting network would be the same. If (60) is Class III, at least one of the arm immittances will be Class III.

Let us realize the Class I, p-r, minimum-loss function

$$T_m(p) = \frac{2.759 p^2 + 6.897 p + 5.000}{p^2 + 4.000 p + 5.000} \\ = \frac{2.759(p + 1.25 + j 0.5)(p + 1.25 - j 0.5)}{(p + 2 + j 1)(p + 2 - j 1)} \quad (65)$$

If in the design, (50) and (51) are satisfied with equal signs, the arm immittances will be minimum real-part functions, and the network will have no flat loss. Doing this and using the configuration in Fig. 10, the network is as shown in Fig. 12. It is interesting to note that this circuit uses only eight elements (excluding terminations) in order to realize the transfer function (65), while a conventional constant-resistance bridged-T would use ten, and a constant-resistance lattice would use sixteen.

If any of the branch immittances used are two pole and zero Class III functions and if (50) and (51) are satisfied with equal signs, either Type D realizations or impractical Brune realizations with unity-coupled coils will be required. The branch immittances would be minimum real-part functions and Type D realization would require eight elements for each Class III immittance. If eqs. (50) and (51) are satisfied by the inequality signs, then the branch immittances will not be minimum real-part functions. By doing this it will be possible sometimes to obtain Type C realizations (usually 7 elements), or Type B realizations (usually five elements) for the Class III immittance branches. Of course the price for the reduction in number of elements is additional flat loss.

If some cases the designer may wish to account for coil and condenser dissipation. If the circuit resulting when (50) and (51) are satisfied with an equal sign cannot be adapted to this, by using the inequality signs (thus increasing the flat loss) it will usually be possible to modify the realization to account for coil and condenser dissipation.

In some cases this method of synthesis will give unequal terminations. It should be noted that if the configuration in Fig. 10 gives a rise in impedance level from input to output, the circuit of Fig. 11 will yield the identical transfer function (54) with a drop in impedance level from input to output. Thus if a chain of sections with unequal terminations are to be connected together, some control can be exerted over the impedance levels at the ends by choosing between the sections of Fig. 10 and Fig. 11, or by using some of each.

When transfer functions are defined as in (54), then the overall transfer function for a chain of constant-resistance sections is exactly

$$T(p) = T_1(p) T_2(p) \dots T_n(p), \quad (66)$$

where the $T_k(p)$ are the transfer functions of the individual sections. One might at first think that if all of the component sections have minimum-loss transfer functions, then $T(p)$ would also have a minimum-loss transfer function. However, note that this is true only when the magnitudes of the individual transfer functions all have the value one at the same frequency $p = j\omega$. For this reason the designer will sometimes introduce extra loss by realizing a complicated transfer function in a chain of simple sections rather than in one complicated section. The design simplicity obtained will usually be worth the price, however.

A somewhat more complete discussion which indicates how the equations in this paper were derived will be found in Reference 7.

Appendix

Suppose that we have a Class III function

$$T_m(p) = B \left(\frac{p^2 + ap + b}{p^2 + ep + d} \right) \quad (67)$$

where a , b , e , and d are known but B is to be determined so that the minimum value of $|T_m(j\omega)|$ is one. The correct value of B can be determined by examining several trial values. The first is

$$B = d/b. \quad (68)$$

The second is

$$B = 1. \quad (69)$$

A possible third and fourth value may be obtained by forming the polynomial in Q :

$$(C^2 - 4b^2)Q^2 + (2DC + 4b^2 + 4d^2)Q + (D^2 - 4d^2) = 0, \quad (70)$$

where $C = 2b - a^2$ and $D = e^2 - 2d$; and then selecting roots $Q = Q_k$ which satisfy

$$y_k = \frac{(Q_k C + D)}{2(Q_k - 1)} = (\text{real, positive}). \quad (71)$$

For each Q_k that satisfies (70) and (71) there is a trial value

$$B = \sqrt{Q_k}. \quad (72)$$

The desired value of B is the largest value obtainable from (68), (69), or (72). Eq. (68) will be valid when $|T_m(j\omega)|_{\omega=0} = 1$ occurs at the origin, (69) when it occurs at infinity, and (72) when it occurs at

$$+j\omega_k = \pm j \sqrt{y_k}. \quad (73)$$

References

1. Bode, H. W., Network Analysis and Feedback Amplifier Design, Van Nostrand and Co., New York, 1945, Chapter 12.
2. Brune, O., "Synthesis of a Finite Two-Terminal Network whose Driving-Point Impedance is a Prescribed Function of Frequency", Journal of Math. and Physics, Vol. 10, pp. 191-236, 1931.
3. Guillemin, E. A., "Modern Methods of Network Synthesis", Advances in Electronics, Vol. III, Edited by L. Marton, Academic Press, Inc., New York, 1951, pp. 261-303.
4. Guillemin, E. A., Communication Networks, Vol. II, John Wiley and Sons, New York, 1935, pp. 198-216.
5. Bott, R., and Duffin, R. J., "Impedance Synthesis Without the Use of Transformers", Journal of Applied Physics, Vol 20, p. 816, August, 1949.
6. Guillemin, E. A., "A Note on the Ladder Development of RC-Networks," Proc. of the I.R.E., Vol. 40, pp. 482-485, April, 1952.
7. Matthaei, G. L., "Some Techniques for Network Synthesis", Research Report Series No. 60, Issue No. 103, December 1, 1953, Electronics Research Laboratory, Institute of Engineering Research, University of California, Berkeley 4, Calif.

Fig. 1

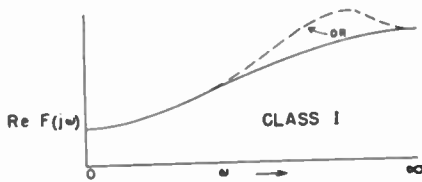


Fig. 2

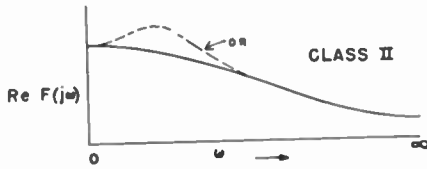


Fig. 3



Fig. 4

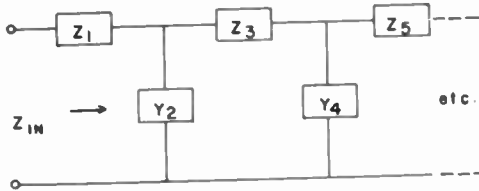


Fig. 5

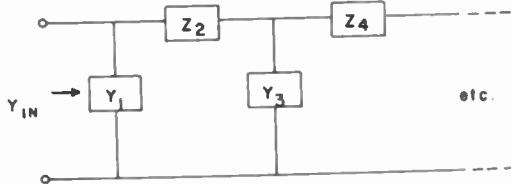


Fig. 6(a)

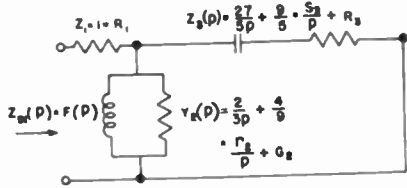


Fig. 6(b)

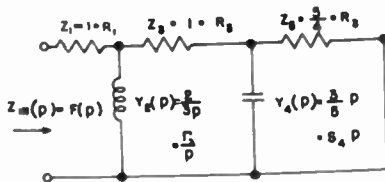


Fig. 7

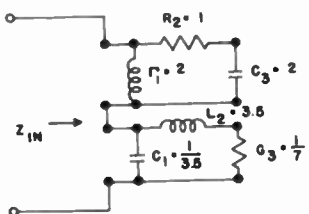


Fig. 8

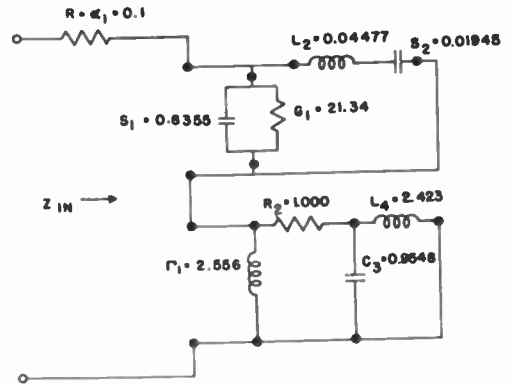


Fig. 9

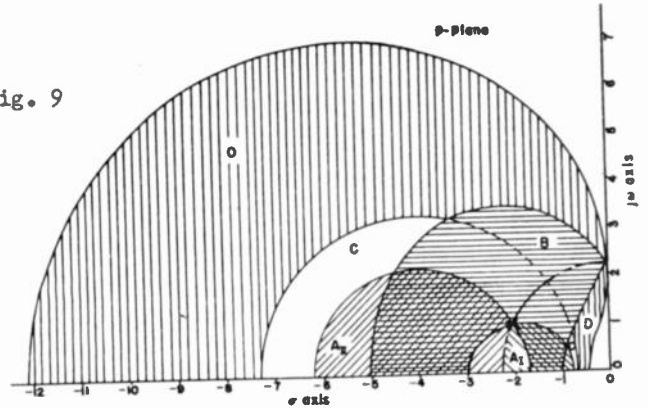


Fig. 10

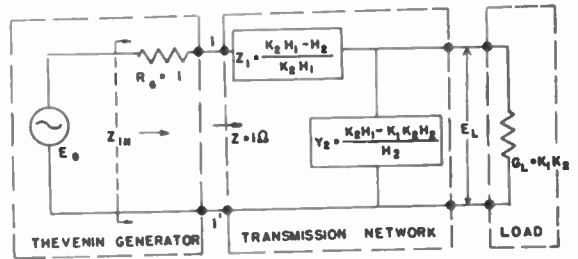


Fig. 11

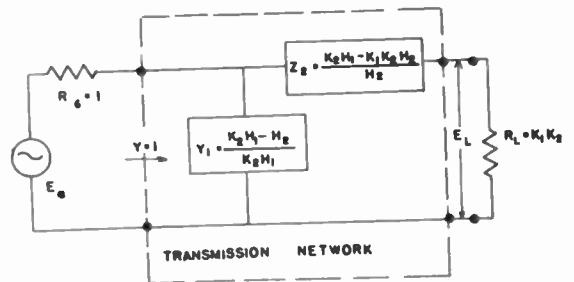
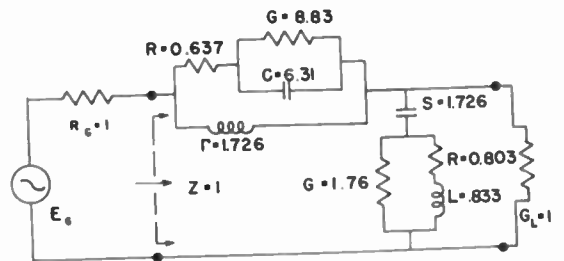


Fig. 12



AN ITERATIVE METHOD
FOR
R C LADDER NETWORK SYNTHESIS *

R. E. Scott and N. DeClaris
Research Laboratory of Electronics and Department of Electrical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts

Summary

An iterative procedure is described for synthesizing R C ladder networks from their pole and zero positions. Criteria for convergence are developed and an example is given.

- - - - -

The iterative procedure for network synthesis starts with a given set of poles and zeros. A network topology is chosen by inspection, which yields the required zeros and provides enough flexibility to determine the poles. Arbitrary element values are assumed and then each tuned circuit is "relaxed" in turn to give a correct resonance at one of the poles. For the general R L C circuit it is sometimes difficult to pick the correct network topology by inspection, and no general proof of the convergence has yet been offered. For the R C case where all the poles and zeros are on the negative real axis, the network configuration is known to be a ladder and hence the first difficulty is overcome. It can also be shown that if the relaxation is carried out in a prescribed way the process must converge.

The principal advantages of the iterative method over any of the conventional ones lie in the added flexibility which it gives the designer in the choice of the form of the network, and in the numerical simplicity of the computations which are self-checking at every stage in the procedure.

The Properties of R C Ladder Networks

The pertinent properties of R C ladder networks are summarized in the following theorems.

Theorem I:

The transfer impedance of an R C two-terminal-pair network is of the form,

$$Z_{12}(s) = K \frac{\prod_{i=1}^n (s + \beta_i)}{\prod_{i=1}^m (s + \alpha_i)}$$

where all the α_i 's are real, distinct, and positive and the β_i 's occur in complex conjugate pairs. Furthermore $n \leq m$.**

Corollary 1

If the R C network is of the form of a single ladder then both α_1 and β_1 are real, distinct and positive.

Theorem II

The driving point impedance of an R C network is of the form,

$$Z_{11}(s) = \frac{\prod_{i=1}^n (s + \delta_i)}{\prod_{i=1}^m (s + \alpha_i)}$$

where α_i and δ_i are real, distinct and positive and form an alternating sequence when arranged according to the order of their ascending magnitude beginning with α_1 and $m=n$ or $m=n+1$.

Theorem III

The sum of two polynomials with real and negative zeros which alternate along the real axis of the complex plane forms another polynomial whose zeros are also real and alternate with the zeros of the above polynomials.

Theorem IV

All the poles of a transfer impedance of a two-terminal-pair network are contained in the driving point impedance viewed from any pair of terminals.

* This work was supported in part by the Signal Corps, the Air Materiel Command, and the Office of Naval Research.

** There is an additional condition for the constant multiplier K. However, no use of this condition is made in this paper.

Iterative Procedures for RC Ladder Networks

A convenient procedure for realizing RC ladder networks consists of the following:

1. Arrange the zeros and poles in sequences of ascending order of magnitude and associate poles and zeros in pairs.
2. If the pole lies to the left of the zero it will be realized by a type 1 section. (see Fig. 1) If it lies to the right it will be realized by a type 2 section. (see Fig. 2)
3. Assume arbitrary values of the elements in the complete network. (Note: the shunt R's are redundant elements which give the flexibility required for a general case.)
4. Iterate the shunt R's one by one. Adjust the residues of the RC circuits to keep the Lagrangian energy functions negative at all the poles. This last condition is necessary to keep the resistors all positive.

The Series and Shunt Component Sections

The series section is shown in Fig. 1. Let Z_0 and Z_1 be the driving point impedances of two RC networks. In particular let,

$$Z_0 = \frac{\prod_{i=1}^n (s+p_i)}{\prod_{i=1}^n (s+q_i)} \quad Z_1 = \frac{\prod_{i=1}^n (s+m_i)}{\prod_{i=1}^n (s+n_i)} \quad (3)$$

with the additional condition that $q_i < n_i$.

After a few algebraic manipulations one obtains:

$$\frac{1}{Z_{11}} = \frac{\prod_{i=1}^n (s+n_i)}{\prod_{i=1}^n (s+m_i)} + \frac{1}{R_1} + \frac{(s+\alpha) \prod_{i=1}^n (s+q_i)}{(s+\alpha) \prod_{i=1}^n (s+p_i) + \alpha R_2 \prod_{i=1}^n (s+q_i)} \quad (4)$$

where:

$$\alpha = \frac{1}{R_2 C} \quad (5)$$

In accordance with Theorem III,

$$(s+\alpha) \prod_{i=1}^n (s+p_i) + \alpha R_2 \prod_{i=1}^n (s+q_i) = K \prod_{i=1}^n (s+d_i) \quad (6)$$

where there is at least one $d = d_0$, such that

$$d_0 > \alpha$$

The exact location depends upon R_2 . Finally the poles of $Z_{11}(s)$ are determined from the polynomial

$$\prod_{i=1}^n (s+d_i) \prod_{i=1}^n (s+m_i) + R_1 \left[\prod_{i=1}^n (s+n_i) \prod_{i=1}^n (s+d_i) + \prod_{i=1}^n (s+q_i) \prod_{i=1}^n (s+m_i) \right] \quad (7)$$

It can be shown that the above polynomial has real zeros a_i , one of which is $a_0 > \alpha$ the exact location depending upon R_1 . It can be stated therefore that a pole a_0 is constrained by R_1 to give $a_0 > \alpha$. The choice of R_2 is arbitrary,

and can be used to control the nature of the energy functions.

A similar analysis will show that for networks of the type shown in Fig. 2 R_1 constitutes the constraint on a pole a_0 such that,

$$a_0 < \alpha$$

where

$$\alpha = \frac{1}{R_2 C} \quad (8)$$

with R_1 setting the impedance level.

Convergence

For the transfer function of an RC network the system equations take the form

$$A_{ij} x + y = C_{ij} x \quad (9)$$

where

$$C_{ij} = \lambda_i \delta_{ij} \quad \delta_{ij} = \text{Kronacker Delta} \quad (10)$$

and x and y are the desired voltage and current functions.

This equation can be written as

$$(A_{ij} + \lambda_i \mathbb{1}) x + y = 0 \quad (11)$$

or

$$|A_{ij} - \lambda_i \mathbb{1}|_S = \beta_i = 0 \quad (12)$$

where $\mathbb{1}$ is the unit matrix.

It must be pointed out that due to its special form the matrix A_i has only the diagonal elements and elements adjacent to the diagonal. The element C_{ij} is independent of s and furthermore there exist β_i such that

$$A_{ij}(\beta_i) = 0 \quad i = 1, 2, \dots, n \quad (13)$$

Since the determinant

$$|A_{ij}(s)|$$

is of a polynomial in n power of s the β_i are the only possible roots and

$$|A_{ij}(\beta_i) + C_{ij}| = K_i \quad (14)$$

where K_i is a real number.

Equation 14 in expanded form constitutes n systems of n simultaneous and interrelated equations whose elements C_{ij} are necessarily the same in all systems.

In the iterative method an arrangement is made by which all the elements of the matrix (A_{ij}) in Eq. 13 are known with the exception of the element a_{ii} (diagonal). A first guess of these elements is made and one computes the error λ_1 as follows

$$\lambda_1^i = \frac{\Delta^i(\beta_1)}{\Delta_{II}^i(\beta_1)} \quad (15)$$

where $\Delta^i(\beta_1)$ is the determinant of the matrix and $\Delta_{II}^i(\beta_1)$ is the co-factor of the element a_{ii}^i . A new matrix is written for which:

$$a_{ii}^{i+1} = a_{ii}^i - \lambda_1^i \quad (16)$$

The superscript denotes the order of iteration. After n iterations (n being the total number of poles or $n+1$ the number of rows of the matrix A_{ij})

$$\Delta^{(n)} = \Delta_{ij}^{(n)} = a_{11}^{(o)} - \lambda_1^{(1)} \prod_{i=2}^{n+1} \left[a_{ii}^{(o)} - \lambda_i^{(1)} - \frac{a_{i-1,i}^2}{a_{i-1,i-1} - \lambda_{i-1}^{(1)}} \right] \quad (17)$$

and

$$\Delta_{11}^{(n)} = a_{11}^{(o)} - \lambda_1^{(1)} \prod_{i=3}^{n+1} \left[a_{ii}^{(o)} - \lambda_i^{(1)} - \frac{a_{i-1,i}^2}{a_{i-1,i-1} - \lambda_{i-1}^{(1)}} \right] \quad (18)$$

a_{ii}^o being the first guess element value. Therefore:

$$\lambda_1^{(2)} = \frac{a_{11}^{(o)} - \lambda_1^{(1)}}{a_{22}^{(o)} - \lambda_2^{(1)}} \left[a_{22}^{(o)} - \lambda_2^{(1)} - \frac{a_{12}^2}{a_{11}^{(o)} - \lambda_1^{(1)}} \right] \quad (19)$$

or:

$$\lambda_1^{(2)} = a_{11}^{(o)} - \lambda_1^{(1)} - \frac{a_{12}^2}{a_{22}^{(o)} - \lambda_2^{(1)}} \quad (20)$$

but,

$$\lambda_1^{(1)} = \left[a_{22}^{(o)} - \frac{a_{12}^2}{a_{11}^{(o)}} \right] \left[\frac{a_{11}^{(o)}}{a_{22}^{(o)}} \right] = a_{11}^{(o)} - \frac{a_{12}^2}{a_{22}^{(o)}} \quad (21)$$

and

$$\frac{a_{12}^{(o)2}}{a_{22}^{(o)} - \lambda_2^{(1)}} = \frac{a_{12}^2}{a_{22}^{(o)}} + \lambda_2^{(1)} \left[\frac{a_{12}^{(o)}}{a_{22}^{(o)}} \right]^2 + \dots \quad (22)$$

The above values of elements a_{ij} are computed for $s = -\beta_1$.

It is obvious that

$$\lambda_1^{(2)} = \lambda_2^{(1)} \left(\frac{a_{12}^{(o)}}{a_{22}^{(o)}} \right)^2 + \dots \quad (23)$$

The higher terms omitted since

$$a_{22}^{(o)} > a_{12}^{(o)}$$

Now

$$\lambda_2^{(1)} = \frac{a_{12}^{(o)} - a_{12}^{(o)2}/a_{11}^{(o)} - \lambda_1^{(1)}}{a_{33}^o} \quad \text{for } s = -\beta_2 \quad (24)$$

or

$$\lambda_2^{(1)} = \frac{a_{22}^{(o)}}{a_{33}^{(o)}} - \frac{a_{12}^2}{a_{33}^{(o)}(a_{11}^{(o)} - \lambda_1^{(1)})} \quad (25)$$

for

$$\frac{a_{12}}{a_{22}} > \frac{a_{22}}{a_{11}} > \frac{a_{12}^2}{a_{33}a_{11}} \quad (26)$$

the process converges since

$$\lambda_1^{(2)} > \lambda_1^{(1)}$$

Example

Given the transfer function

$$Z_{12}(s) = \frac{s}{(s+1)(s+3)}$$

find the network.

The poles and zeros arranged in a sequence of ascending order are:

$$\begin{array}{ll} \text{poles:} & -1 \quad -3 \\ \text{zero:} & 0 \quad \infty \end{array}$$

The pole (-1) is associated with the zero (0) and the pole (-3) with the zero at (∞) . Note that both section types must be used. And the network takes the form shown in Fig. 3.

Suppose that $4C_1 = C_2$. (Arbitrary condition.) As a first choice let:

$$C_2 = 1 \quad R_2 = \frac{1}{2}$$

Match the impedances for $s = -1$ at the section ab. That is,

$$Z_1(-1) = -Z_2(-1)$$

or

$$R_1 = -(-3) = 3$$

Now match the impedances at section cd, for $s = -3$.

$$Y_3(-3) = -Y_4(-3)$$

$$\frac{1}{R_2} = -(-\frac{12}{5}) = R_2 = \frac{5}{12}$$

Going back to section ab and with $s = -1$

$$R_1 = -(-\frac{23}{7}) = \frac{23}{7}$$

At section cd for $s = -3$

$$\frac{1}{R_2} = -(-3.975) \quad R_2 = .253$$

For $s = -1$ at section ab

$$R_1 = -(-3.664) = 3.664$$

After a few iterations

$$R_1 = 3.33 \quad R_2 = .4$$

$$C_2 = 1 \quad C_1 = \frac{1}{4}$$

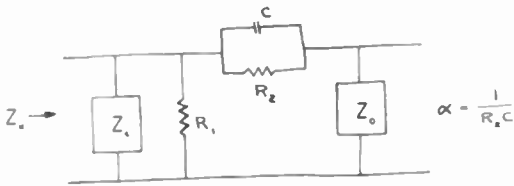


Fig. 1 - Series element network type I.

References

E.A. Guillemin: "Synthesis of R C Networks", Jour. Math. Phys., pp22-42, April 1949.

A. Fialkow and I. Gerst: "The Transfer Function of an R C Ladder Network," Jour. Math. Phys., pp 44-72, 1951

H. Ozaki: "Synthesis of R C-3 Terminal Network Without Ideal Transformers," Tech. Reports of the Osaka University, Vol. 3, No.60, 1953

R.E. Scott, and R.L. Blanchard: "An Iterative Method for Network Synthesis," Tech. Paper presented at IRE, NESCON, August 1953

G. Temple: "The General Theory of Relaxation Methods in Linear Systems," Proc. Roy. Soc., 169A, 1939, pp476-500

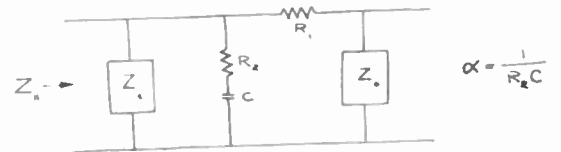


Fig. 2 - Shunt element network type II.

$$Z_{12}(s) = \frac{s}{(s+1)(s+3)}$$

Poles	-1	-3
Zeros	0	$-\infty$
Network:	I	II
Type		

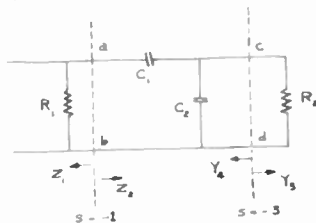


Fig. 3 - An example.

NETWORKS TERMINATED IN RESISTANCE
AT BOTH INPUT AND OUTPUT

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

Abstract

In this paper a lattice with a resistive termination at both input and output is realized. Any physically realizable transfer function--impedance, admittance, or dimensionless voltage ratio--may be realized by the method presented. The method used may be based on either of two previous procedures that realized open-circuited lattices. If the given transfer function has a numerator of lower degree than its denominator, then an even more practical termination may be obtained at both input and output, namely, one that has a shunt capacitance in addition to the resistive termination. The lattice arms contain no mutual inductance and may always be designed to have no pure inductances, that is, every inductance present in an arm has an associated series resistance.

Introduction

The problem of realizing a general transfer function by a lattice network is an important one in modern network design, and has been treated in a number of papers.¹⁻⁵ Only one¹ of the available procedures, however, demonstrates how to obtain the practical and desirable form of lattice that has resistance terminations at both input and output. Here a different method is presented for realizing a general transfer impedance, transfer admittance, or dimensionless voltage ratio as a lattice terminated at both ends in resistance. No mutual inductance is necessary for realizing the lattice arms and every inductance may be designed to possess an associated series resistance, so that in building the network lossy coils may be used. Furthermore, it is often desirable from a practical point of view to obtain a network with a shunt capacitance at the input and output terminals; it is clear that the physical realization of such a network requires that the numerator of the transfer function be of lower degree than the denominator. When this condition holds, the method of this paper achieves the desired capacitance at both input and output terminals.

The first part of the method may use either one of the two procedures^{4,5} for realizing an open-circuited lattice. The procedure explained in this paper starts with the technique of partial fraction expansion;⁴ it will be clear to the reader how the other method⁵ may be similarly adapted. Only a brief summary of the partial fraction method is given in order to establish definitions and necessary formulas; for further details the reader is referred to the reference.

After the realization of the open-circuited lattice it is necessary to remove a constant from each of the lattice arms. To show this is always

possible a discussion is given of the variation along the j axis of the real part of the lattice-arm driving-point functions. Then the explanation of the synthesis procedure is completed and an illustrative example is worked out.

Realization of Open-Circuited Lattice

It was previously shown⁴ that a general transfer impedance, which can always be written within a multiplicative constant in the form

$$Z_{12} = \frac{E_2}{I_1} = \frac{p(s)}{q(s)} \quad (1)$$

$$= \frac{s^m + a_{m-1}s^{m-1} + \dots + a_1s + a_0}{s^n + b_{n-1}s^{n-1} + \dots + b_1s + b_0}, \quad (m \leq n+1)$$

may be realized as the open-circuited lattice shown in Fig. 1 for which

$$Z_{12} = \frac{1}{2}(Z_b - Z_a). \quad (2)$$

If the poles of Z_{12} are simple and its numerator is of lower degree than the denominator, the partial fraction expansion of the impedance of each of the lattice arms has the form

$$Z = \sum_{\mu=1}^n \frac{k_{\mu}}{s-s_{\mu}}. \quad (3)$$

Considering complex conjugate poles as combined into one term, we find that each of the partial fraction terms has the significant positive real characteristic, that is, the terms are separately realizable by inspection. Thus the two types of terms that occur are given by

$$z_1 = \frac{k_1}{s-s_1} \quad (4)$$

$$= \frac{a_1}{s+a}$$

and

$$z_2 = \frac{k_2}{s-s_2} + \frac{\bar{k}_2}{s-\bar{s}_2}$$

$$= \frac{a_2 + j\beta_2}{s + \sigma_2 - j\omega_2} + \frac{a_2 - j\beta_2}{s + \sigma_2 + j\omega_2} \quad (5)$$

$$= \frac{2a_2(s+d_2)}{s^2 + \sigma_2 s + |s_2|^2}$$

where a , a_1 , a_2 , d_2 , and σ_2 are real and positive constants, and d_2 is not greater than $2\sigma_2$. These terms are immediately realizable in the forms shown in Figs. 2 and 3, and the complete lattice has arms containing a series connection of such networks.

When $m = n$ one or both of the expansions for the lattice arms contains a constant term, and when $m = n+1$ at least one of the arms will contain a pole at infinity. Corresponding to these terms a series resistance and a series inductance, respectively, will be present in the lattice arms. For a Z_{12}

that possesses multiple poles the method of realization explained in reference 4 introduces a constant term into each of the lattice arms. We shall see that this precludes obtaining a shunt capacitance at both input and output even when $m < n$; however, the method of reference 5 permits the desired capacitance to be obtained for this case.

Variation of the Real Part of Z

We now show that the real part of $Y = \frac{1}{Z}$ for

$s = j\omega$, denoted hereafter by $\text{Re}[Y(j\omega)]$, has no zeros for all real values of ω including infinity, where Z represents the form of the driving-point impedance of each of the lattice arms; that is, the lattice arms have nonminimum-conductive driving-point admittances. As a result a conductance may always be removed from each of the lattice arms without destroying the positive real quality of its driving-point function.

Since z_1 and z_2 , given respectively in (4) and (5), represent driving-point impedances, their real parts along the j axis are never negative. It is furthermore clear from inspection of (4) that the real part of z_1 is nonzero at the origin and decreases monotonically to a zero value at infinite frequency. Similarly, for terms of the form of z_2 , inspection of (5) shows that $\text{Re}[z_2(j\omega)]$ is also finite and nonzero at the origin and has a zero value at infinite frequency, though its intermediate variation is not monotonic. It, too, possesses no zero in the real part for finite frequencies. If we consider the given Z_{12} as a proper fraction with simple poles, then each of the lattice arms is of the form given by (3) and the real part of Z is the sum of the real parts of the two types of terms considered above. Suppose we now write

$$Z = \frac{m_1 + n_1}{m_2 + n_2} \quad (6)$$

where m_1 and n_1 represent respectively the even and odd parts of the numerator, while m_2 and n_2 play the same roles for the denominator. Then

$$\text{Re}[Z(j\omega)] = \left. \frac{m_1 m_2 - n_1 n_2}{m_2^2 - n_2^2} \right|_{s=j\omega} \quad (7)$$

and the above reasoning yields the conclusion that the numerator ($m_1 m_2 - n_1 n_2$) possesses no zeros for real ω and is therefore always positive. The total function $\text{Re}[Z(j\omega)]$ has a zero at infinity.

As for the admittance $Y = \frac{1}{Z}$, its real part is given by

$$\text{Re}[Y(j\omega)] = \left. \frac{m_1 m_2 - n_1 n_2}{m_1^2 - n_1^2} \right|_{s=j\omega} \quad (8)$$

from which we note that it has the same numerator as $\text{Re}[Z(j\omega)]$. Therefore $\text{Re}[Y(j\omega)]$ is always positive and nonzero except possibly at infinite ω . But it is also nonzero at infinity, for the degree of the denominator will be greater than that of the numerator in (8) only when Z possesses no other terms except one or more of the form of z_2 in each of which the constant d is equal to 2σ . Since it is always possible⁴ to make d less than 2σ , we may state the conclusion: the $\text{Re}[Y(j\omega)]$ is always positive and nonzero for all (finite and infinite real) values of ω .

Restricting the discussion to a proper fraction containing only simple poles represents no loss in generality, for the same conclusion applies in the other cases. If multiple poles are present in the given transfer impedance Z_{12} , a term whose real part is positive and nonzero for all ω is added to the lattice arm impedances. If in Z_{12} the degree of p is equal to the degree of q , a constant is added to one or both of the lattice arm impedances. Finally, if the degree of p exceeds that of q , none of the transfer functions is physically realizable with a resistance termination at both input and output, as is demonstrated below.

It has been shown³ that a transfer voltage ratio is not physically realizable if the degree of its numerator is greater than the degree of its denominator, that is, if a pole at infinity is present. But we desire networks terminated in resistance at both input and output. For such networks the same rational function within a constant multiplier represents the transfer voltage ratio, the transfer admittance and the transfer impedance. Thus all three types of transfer functions are unrealizable in the form of the desired network if the degree of the numerator exceeds that of the denominator. Another way of seeing this is to note that if an open-circuited lattice is synthesized whose transfer impedance is given by such an improper rational fraction, then at least one of the impedances of the lattice arms must have a pole at infinity. Consequently a conductance cannot be removed from the corresponding admittance because its real part will have a zero at infinite frequency.

Completion of the Synthesis Procedure

The open-circuited lattice that has been derived may now be converted to the desired form. We have seen that the real part of each of the lattice-arm admittances will have one or more positive non-zero minima; we now determine the smallest minimum of both admittances and denote them respectively by G_a and G_b . It is then possible to obtain an equivalent lattice⁵ by removing from each of the arms a conductance of value less than the smaller of G_a and G_b and placing it in parallel with the input and output terminals of the lattice. This transformation, shown in Fig. 4, thus yields the desired

resistance terminations for Z_{12} . To obtain $K = \frac{E_2}{E_1}$ and $Y_{12} = \frac{I_2}{E_1}$ requires merely an application of Thevenin's theorem to the input; this yields the network of Fig. 5 for which:

$$\begin{aligned} K &= \frac{E_2}{E_1} \\ &= \frac{GE_2}{I_1} \\ &= GZ_{12} \end{aligned} \quad (9)$$

and

$$\begin{aligned} Y_{12} &= \frac{I_2}{E_1} \\ &= \frac{GE_2}{E_1} \\ &= GK \\ &= G^2 Z_{12} \end{aligned} \quad (10)$$

It is clear from the above equations that the constant gain factor achieved for the transfer voltage ratio is directly proportional to G . This makes it desirable, if one is interested in gain, to remove as large a conductance as possible from the arms. However, one may be more interested in using low- Q coils for the realization of the lattice arms, which problem we discuss below; in this case it is necessary to retain a large conductance in each of the lattice arms.

For realizing the remainder of the lattice arms, that is, the admittances Y'_a and Y'_b in the network of Fig. 5, we may use the Bott-and-Duffin procedure.⁶ This yields a network containing pure inductances but no mutual inductance. However, we desire that every inductance possess an associated series resistance; to achieve this we substitute a new variable ($s-h$) for s before using the Bott-and-Duffin method, that is, we make use of the technique of predistortion introduced by Darlington.⁷

Predistortion requires that for each arm admittance we first determine the equation of the curve in the left half of the complex plane that

represents the locus on which the admittance has a zero real part. For example, working with the series arm,

$$\begin{aligned} Y_1 &= Y'_a + (G_a - G) \\ &= \frac{u_1(\sigma, \omega) + jv_1(\sigma, \omega)}{u_2(\sigma, \omega) + jv_2(\sigma, \omega)}, \end{aligned} \quad (11)$$

we obtain the curve

$$\begin{aligned} \text{Re} [Y_1] &= 0 \\ &= u_1 u_2 + v_1 v_2 \equiv f(\sigma, \omega) = 0. \end{aligned} \quad (12)$$

Considering σ as an implicit function of ω given by $f(\sigma, \omega)$ and evaluating the derivative

$$\frac{d\sigma}{d\omega} = - \frac{\frac{\partial f}{\partial \omega}}{\frac{\partial f}{\partial \sigma}} \quad (13)$$

we find the smallest minimum value of σ , that is, the point at which the curve is closest to the j axis. To each of the zeros and poles of Y_1 we may now add the positive constant, which is chosen less than or equal to this minimum distance, without destroying the positive real quality of Y_1 . Then, after realization of the arm by the Bott-and-Duffin procedure, the network obtained is corrected for the predistortion: for every L a series combination of L and a resistance of Lh ohms is substituted, while every C is replaced by a parallel combination of C and a conductance of Ch mhos. A similar procedure is followed for the diagonal arm.

Finally, if the given transfer function is a proper fraction, it is clear that the admittances of both of the lattice arms will possess a pole at infinity and a corresponding shunt capacitance in their network representations. Thus, a capacitance may be removed from each of the arms yielding an equivalent lattice with a shunt capacitance at the input and output terminals.

The steps in the synthesis procedure may now be summarized as follows:

1. Realize the given function as an open-circuited lattice by the method of reference 4 or reference 5.
2. Obtain an equivalent lattice with a shunt conductance at the input and output terminals. If the degree of the numerator of the given transfer function is lower than that of the denominator, also remove a shunt capacitance from each of the lattice arms.
3. Predistort each of the remaining lattice admittances as explained above. Then realize each arm by the Bott-and-Duffin procedure, after which

the networks obtained are corrected for the predistortion.

L. If necessary use Thevenin's theorem on the input to obtain the given type of transfer function.

Illustrative Example

To demonstrate the complete procedure, we realize the nonminimum-phase voltage ratio

$$K = \frac{E_2}{E_1} = \frac{s^2 - s - 12.4}{10s^2 + 46s + 60}$$

as a resistance-terminated lattice.

First we represent the above function as the transfer impedance of an open-circuited lattice. Using the method of reference 4 we find

$$Z_b = \frac{s^2 + 6s + 10}{5s^2 + 23s + 30}$$

$$Z_a = \frac{7s + 22.4}{5s^2 + 23s + 30}$$

The series impedance Z_a is immediately realizable by inspection as

$$\frac{1}{Z_a} = \frac{5}{7s} + 1 + \frac{7.6}{7s + 22.4}$$

Upon attempting to remove a conductance of one mho from Y_b we obtain the positive real remainder

$$Y_b - 1 = \frac{4s^2 + 17s + 20}{s^2 + 6s + 10}$$

Thus this conductance may be removed from the lattice arms to yield an equivalent resistance-terminated lattice. The remainder of the diagonal arm admittance has a zero real part curve whose closest point to the j axis occurs at $\sigma = -2$. Substituting the new variable $(s-2)$ for s in the remainder, we obtain

$$Y' = \frac{4s^2 + s + 2}{s^2 + 2s + 2}$$

which we now realize by the Bott-and-Duffin procedure⁸ as the network shown in Fig. 6. This network is then corrected for the predistortion.

Now applying Thevenin's theorem to the input of the lattice thus obtained, we finally realize the lattice shown in Fig. 7.

Conclusion

Any realizable transfer voltage ratio, transfer admittance or transfer impedance may be realized by the method presented in this paper as a lattice terminated in resistance at both its input and output terminals. No mutual inductance is necessary and each inductance has an associated series resistance so that low-Q coils may be used in building the network. When the transfer function is a proper fraction, then a shunt capacitance may be obtained at both the input and output terminals.

References

1. H. W. Bode, "Network Analysis and Feedback Amplifier Design," D. Van Nostrand Co., New York, N. Y., 1945.
2. E. A. Guillemin, "Summary of Modern Methods of Network Synthesis," Advances in Electronics, vol. 3, Academic Press, New York, N. Y., 1951.
3. R. Kahal, "Synthesis of the Transfer Function of 2-Terminal Pair Networks," A. I. E. E. Proceedings, vol. 71, 1952.
4. L. Weinberg, "RLC Lattice Networks," Proc. I. R. E., pp. 1139-1144; September, 1953.
5. L. Weinberg, "A General RLC Synthesis Procedure," Convention Record of the I. R. E., 1953; also Proc. I. R. E., Feb., 1954.
6. R. Bott and R. J. Duffin, "Impedance Synthesis Without Use of Transformers," Jour. Appl. Phys., p. 816; August, 1949.
7. S. Darlington, "Synthesis of Reactance Four-Poles," Jour. Math. Phys., pp. 257-353; September, 1939.
8. For details on the realization of this driving-point function, see E. A. Guillemin, "A Summary of Modern Methods of Network Synthesis," Advances in Electronics, vol. III, 1951, where the dual of the admittance is synthesized.

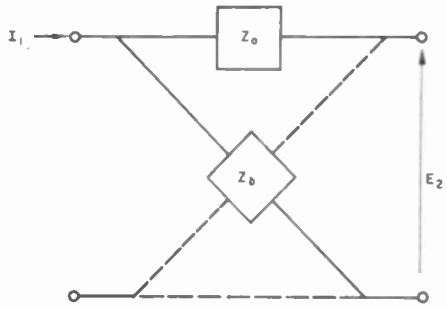


Fig. 1
Open-circuited lattice network characterized by $Z_{12} = \frac{1}{2}(Z_b - Z_a)$.

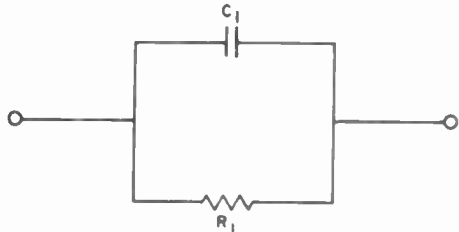


Fig. 2
Form of realization of a term containing a real pole, where $z_1 = \frac{1}{C_1(s + \frac{1}{R_1 C_1})}$.

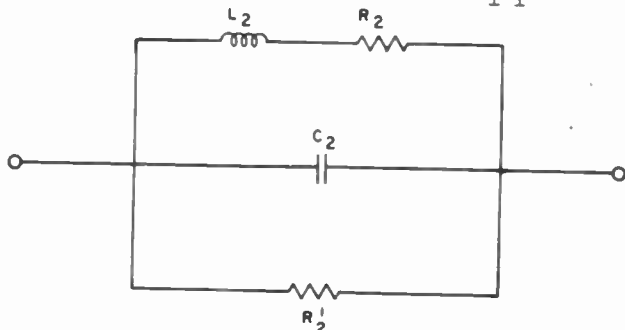


Fig. 3
Form of realization of a pair of terms containing complex conjugate poles, where

$$z_2 = \frac{(s + \frac{R_2}{L_2})}{C_2 \left[s^2 + \left(\frac{R_2}{L_2} + \frac{1}{R'_2 C_2} \right) s + \frac{R_2}{L_2 R'_2 C_2} + \frac{1}{L_2 C_2} \right]}$$

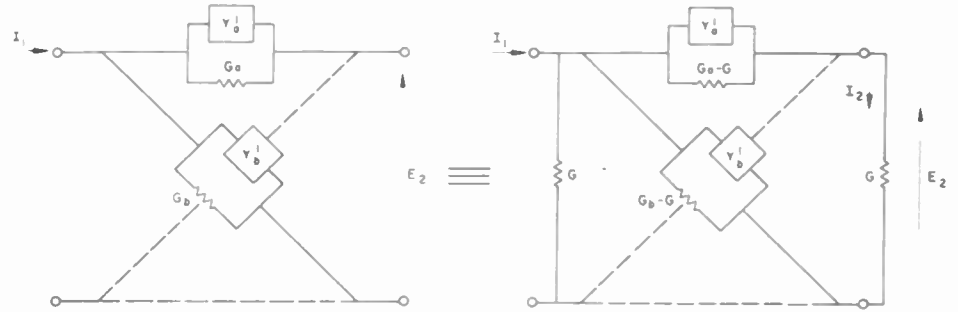


Fig. 4
Transformation yielding a lattice terminated in resistance at both input and output.

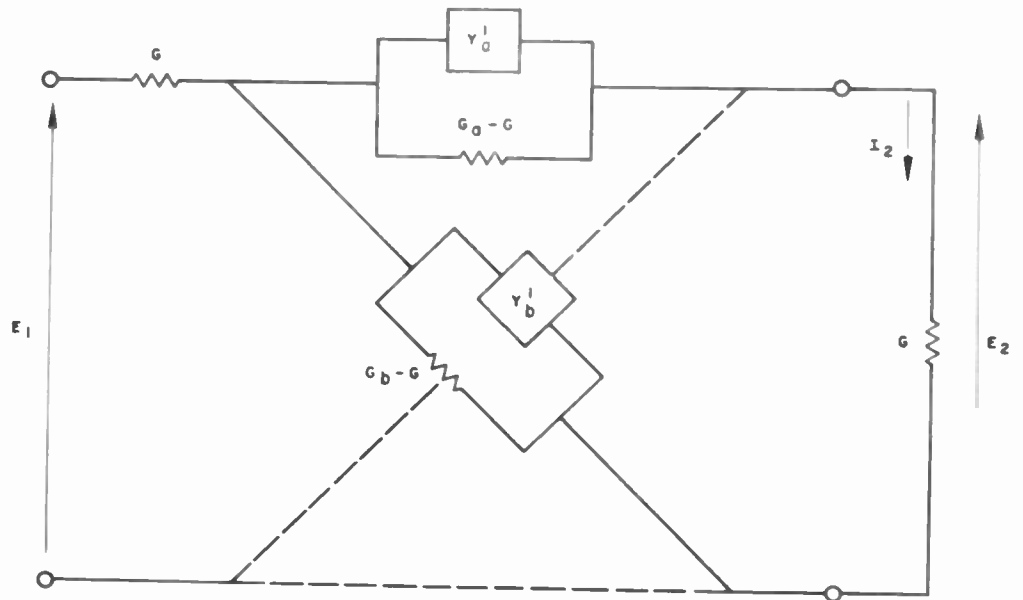


Fig. 5
Network obtained by application of Thevenin's theorem to resistance-terminated lattice of Fig. 4.

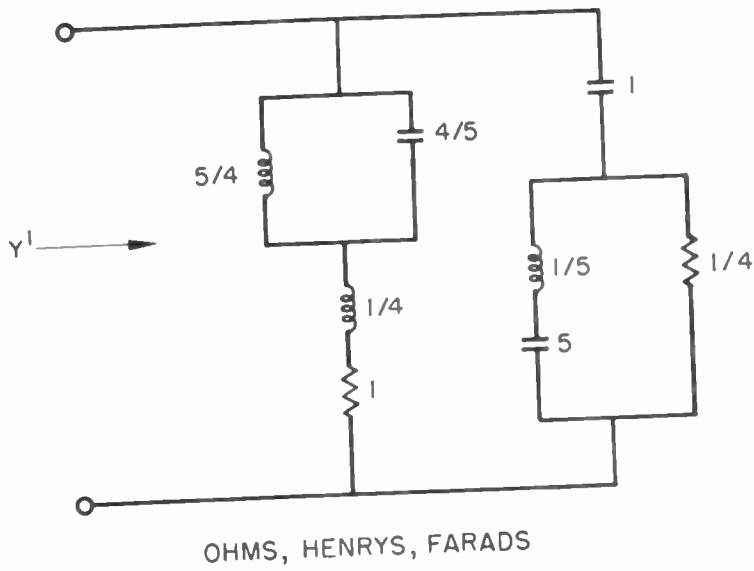


Fig. 6
Realization of the admittance

$$Y' = \frac{4s^2 + s + 2}{s^2 + 2s + 2}$$

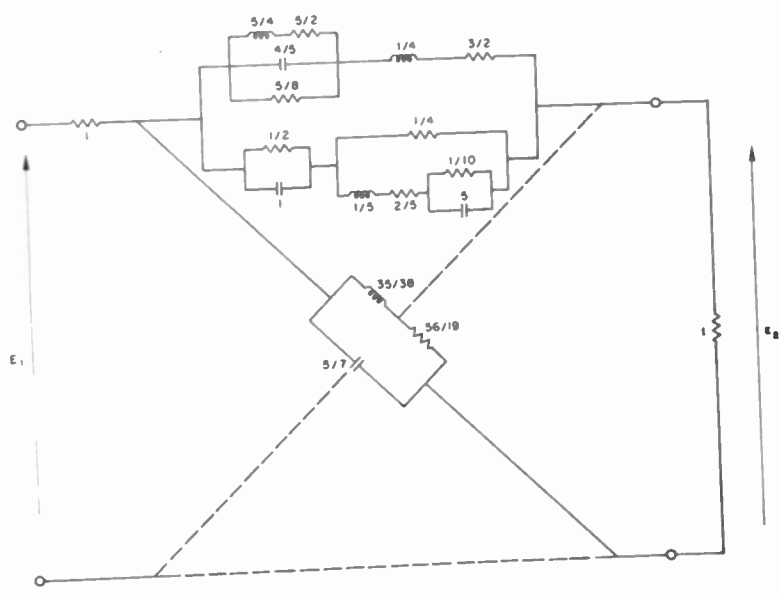


Fig. 7
Lattice network characterized by the given transfer voltage ratio

$$K = \frac{s^2 - s - 12.4}{10s^2 + 46s + 60}$$

APPROXIMATING BAND-PASS ATTENUATION AND PHASE FUNCTIONS

V. H. Grinich
Stanford Research Institute
Stanford, California

Abstract

A problem that often arises in synthesis is to obtain transfer functions that give an approximation to assigned attenuation and phase functions over a stated frequency interval. In this paper approximation in two different senses is discussed: (1) derivative matching (Taylor), and (2) nearly equal-ripple (quasi-Chebyshev). The first type is used as an intermediate step in the second.

An existing method using Chebyshev polynomial series gives nearly equal-ripple approximations to an assigned attenuation function in both the "low-pass" and "band-pass" intervals.¹ This method can be used to approximate phase alone or phase and attenuation simultaneously but only in the low-pass interval. By stressing the concept of conformal mapping this method can be extended to the band-pass case.

This paper gives the extension for obtaining the solution to these problems (approximation of phase alone or simultaneous phase and attenuation) in the band-pass interval. The procedure uses an elliptic function conformal mapping. In the cases of all-pass linear phase or linear phase coupled with constant attenuation the nearly equal-ripple approximations are shown to be simply related to the Taylor approximation.

I. Introduction

In applying network theory to practice, we usually find that the work is divided into these three well-known parts: (a) studying the demands placed on the network function in the light of what is physically realizable, (b) finding an approximation to these demands that will yield a physical and economical network and (c) realizing the network in terms of its structure and element values. In this paper the methods discussed are useful in the approximation problem, part (b), when the demands, part (a), are made in terms of the usual steady state characteristics over the "band-pass" interval.

For the case where the specified interval includes zero frequency, the low-pass case, the approximation problem has been fully treated by Darlington.¹ Approximation of band-pass attenuation functions can also be accomplished by this technique. However, this method is not applicable to approximating band-pass phase or the simultaneous approximation of band-pass attenuation and phase.*

* If the bandwidth ratio is sufficiently close to unity, i.e. the narrow band case, the low-pass techniques may be applied to band-pass problems by the usual low-pass to band-pass transformation methods. If the bandwidth ratio is too large then the methods described in this paper can be used.

The types or senses of approximation that we consider are two: first, the Taylor, and second, the quasi-Chebyshev approximations. The Taylor approximation is mainly useful as a means for getting the second type, the quasi-Chebyshev approximation. In the Taylor or power series approximation a single point on the frequency axis is selected at which the value of the network function and as many derivatives of it as possible are set equal to the corresponding values in the prescribed function. Thus, by means of derivative matching we get a "maximally-flat" error curve that represents the difference between the prescribed function, say \bar{y} , and the approximating network function, say y .

In the true Chebyshev, or equal ripple approximation, the maximum magnitude of the error curve in the useful (non-zero) frequency interval is made a minimum. As a result of this, the error curve ripples about zero with positive and negative peak values that are all equal in magnitude. Instead of trying to find an exact solution (which has been obtained for a few special functions) we seek a certain type of nearly equal-ripple solution that we shall refer to as a quasi-Chebyshev approximation in this paper. We may define this type as a solution obtained by the following method.

The prescribed function \bar{y} is expanded into a series of functions where the individual functions E_i oscillate about a constant value in an equal-ripple fashion. Thus

$$\bar{y} = \sum_{i=0}^{\infty} \bar{a}_i E_i \quad (1)$$

Then a network or approximating function y is obtained in which

$$y = \sum_{i=0}^{\infty} a_i E_i \quad (2)$$

and wherein as many a_i are set equal to the corresponding \bar{a}_i as possible.

For the low-pass interval and for band-pass attenuation functions, the functions E_i are Chebyshev polynomials. For the approximation of band-pass phase or phase and attenuation, this paper presents methods where the functions E_i turn out to be a set of elliptic functions.²

However, instead of defining the quasi-Chebyshev approximation in this manner, which becomes too complicated for the bandpass case, we can use the following alternate definition that applies to both the low-pass (Chebyshev polynomial series method) and band-pass cases. The definition proceeds as follows:

The useful interval of frequencies along the ω -axis in the p -plane is transformed into a circle in another plane related to the p -plane by a con-

formal mapping. With the center of this circle as the expansion point a Taylor (power) series expansion is made which gives the desired network transmission function \bar{y} on the circumference of the circle.* The approximating network function y is then constructed that matches as many coefficients as possible of the Taylor series of \bar{y} in this transformed plane. As a convenient introduction for the band-pass problem, we briefly review Darlington's work with this later definition in mind.

II. Conformal Mappings-Attenuation And Phase Invariant Transformations

If a conformal mapping is made relating the p -plane to another plane, say the c -plane, then the attenuation and phase functions in the p and c -planes have the same values at the point p and the corresponding point in the c -plane.² This concept is used when mappings which have certain helpful symmetries are utilized in a particular problem.

In the type of conformal mapping used here, the set of singularities from the p -plane are mapped into another plane where the original set of singularities is increased into two or more sets. These sets are always symmetrical about some line in the mapped plane.⁴ Removing one or more of the sets of these singularities results in a simplification of the synthesis problem. The resulting attenuation and phase functions are still readily related to the original and hence we refer to this removal process as an attenuation and phase invariant transformation.

II.A. Low-Pass Interval Mappings and Transformations

We can illustrate the above by first considering the mapping used by Darlington for the low-pass case shown in Fig. 1. Here the p -plane (with its useful interval extending from -1 to $+1$ along the ω -axis) is transformed over into the Z -plane by the relation

$$p = \frac{1}{2}(Z - 1/Z). \quad (3)$$

Thus, the useful interval lies along the circumference of the circle of unit radius centered on the origin of the Z -plane. We then use the further mapping from the Z -plane to the W -plane where $Z = e^W$. Here the useful interval is repeated periodically along the Ψ -axis, the axis of imaginaries in the W -plane.

Figure 2 shows examples of network singularities in the p -plane which are to be transformed into the Z and W -planes. In Fig. 2(a) the rati-

al function $R(p)$ is the actual network function. If only attenuation is to be considered, the "power" function $R(p)R(-p)$ in Fig. 2(b) contains only information relating to attenuation with zero phase along the ω -axis. If only phase is under consideration, the "all-pass" function, $R(p)/R(-p)$, is useful since attenuation is constant along the ω -axis.

Figure 3(a) shows the singularities in the Z -plane that result from transforming from the p to the Z -plane the singularities shown in Fig. 2(b). Figure 3(b) shows the result after the singularities are mapped into the W -plane. We now note that removal of all the singularities in the left half-plane of the W -plane only effects the attenuation by reducing it to one-half of its initial value. This transformation or removal of singularities introduces an extraneous phase function along the Ψ -axis. This phase is not to be confused with the phase function of the original network transfer function shown in Fig. 2(a). The remaining singularities in the W -plane are now mapped back into the Z -plane.

The transfer and transmission functions that are obtained after the removal of the singularities are referred to as the reduced functions. All the singularities are now external to the unit circle, $|Z| = 1$, as shown in Fig. 3(a). This clearing out of singularities from the circle, $|Z| = 1$, is an important step since it allows us to expand the reduced attenuation function into a power series about the origin in the Z -plane with assurance that the power series will converge in the unit circle. Since the singularities of Fig. 3(c) are symmetrical about the axis of imaginaries in the Z -plane, the power series expansion for the reduced transmission function contains only even order terms in Z .

Figure 4 pertains to the phase invariant transformation. Figure 4(a) is the result of transforming singularities from Fig. 2(c) into the Z -plane. These singularities are then mapped into the W -plane as shown in Fig. 4(b). There we note that removal of the left half-plane singularities only reduces the phase by a factor of one-half everywhere along the useful interval. After this removal of singularities there is an extraneous attenuation obtained at the useful interval. These right half-plane singularities are now mapped over into the Z -plane as in Fig. 4(c). Again the unit circle is clear of singularities and a power series expansion about $Z = 0$ is in order. The singularities of Fig. 4(c) are anti-symmetrical about the imaginary axis of the Z -plane and the power series expansion for the reduced transmission function contains only odd order terms in Z .

As Darlington shows power series expansions for the assigned attenuation and phase functions can be readily obtained. Network functions can then be found so that in terms of Z -plane analysis the assigned and network power series have identical coefficients up to a certain degree set by the number of singularities in the network function.

* A transfer function is expressed as a rational fraction in terms of which the numerator and denominator factors are poles (natural modes) and zeros. The corresponding transmission function is obtained by taking the logarithm of the transfer function.

Figure 5(a) shows the result of multiplying the transfer functions contained in Fig. 3(c) and Fig. 4(c). The transmission function that results from taking the logarithm of the rational fractions in Fig. 5(a) contains both even and odd terms. Since we know that in the power series for the transmission function of Fig. 5(a) the even (odd) terms evaluated along the useful interval give true attenuation (phase) the extraneous term can be ignored. Hence the rational function shown in Fig. 5(a) can be used in lieu of that shown in Fig. 5(b) which contains all the singularities transformed from the p -plane.

We can restate the ideas in the above paragraph in the following manner. (The following method is more useful since it can be extended to the case of attenuation and phase invariance in the band-pass interval.) First we may consider that the assigned attenuation and phase, $\bar{a} + j\bar{\beta}$ are due to an infinite network (reference 1, p. 640 footnote). On removing the singularities of this network that lie in the unit circle of the Z -plane, we have the resulting reduced transmission function " \bar{a} " + j " $\bar{\beta}$ " along the useful interval in the Z -plane.* The reduced network that corresponds to the actual network with the unit circle clear of singularities is found so that its transmission function " a " + j " β " approximates " \bar{a} " + j " $\bar{\beta}$ " in the Taylor sense at $Z = 0$. Thus on reconstructing the entire (actual) network we end up with the quasi-Chebyshev approximation of $\bar{a} + j\bar{\beta}$, since the extraneous terms are removed.

II.B. Band-Pass Interval Mappings and Transformations

In the band-pass case we require different conformal mappings and further attention must be given to the attenuation and phase invariant transformations that are used. Figure 6 shows the geometry of the mapping that we use in the following discussion.⁵

For the band-pass case the useful range consists of two intervals along the ω -axis. These have been normalized so that we are interested in frequencies that satisfy the relationship $\sqrt{k'} \leq |\omega| \leq 1/\sqrt{k'}$. If we always make sure that the network functions we deal with have with every complex singularity its complex conjugate, then considerations relating to the attenuation and phase in the interval for plus frequencies will automatically give the proper values for the negative frequency interval. Thus, in what follows, we shall refer to the useful interval along the plus ω -axis as the useful interval.

In order to transform the useful interval into a circle we use the intermediate elliptic function mapping in the w -plane

$$p = \sqrt{k'} \operatorname{tn}(w, k). \quad (4)$$

Thus, in the w -plane we have an infinity of rectangular "cells," of dimension $2K$ by $2K'$, extending into both directions. One cell is shown in Fig. 6(b). Corresponding to each cell is one Riemann surface of the p -plane. By means of the exponential function mapping,

$$z = e^{\frac{j\pi w}{K}}, \quad (5)$$

the horizontal straight lines representing the useful interval in the w -plane are mapped into circles concentric about the origin in the z -plane. In what follows the annulus that has the inner and outer radius a and a^{-1} , respectively is referred to as the unit annulus.* This is shown in Fig. 6(c).

Figure 7 pertains to the attenuation invariant transformation for the band-pass case. In Fig. 7(a) we show a typical transfer function in the p -plane being modified for the situation where we are interested in attenuation only. Thus the singularities are symmetrical about the ω -axis. The singularities are mapped into the w -plane [Fig. 7(b)] and at this point we see that if we remove all singularities above the horizontal line $w = jK'$, the attenuation will be reduced everywhere along the line $w = jK'$ by a factor of one-half, but it will still have the same functional form. When we further map these singularities into the z -plane [Fig. 7(c)] the circle of radius a centered on the origin of the z -plane is now free of singularities. Hence, we can, as in the lowpass case, expand the attenuation into a power series and obtain a representation along one mapping of the useful interval, namely, the circle of radius a in the z -plane. As before, an extraneous phase term has been introduced, but since we see that reconstructing the entire function gives the true attenuation function, it can be ignored. The singularities inside the unit annulus give a contribution to the attenuation $u_{(1)}$ that is evaluated along the circle of radius a . Using a power series representation we can write

$$u_{(1)} = \operatorname{Re} \sum_{i=1}^{\infty} a_i z^i. \quad (7)$$

If we investigate the attenuation contributed by the singularities that are outside the circle of radius a , we find that the contributions due to the singularities in successive annuli all add. Hence, if we call the term in the attenuation due to the poles and zeros in the annulus, which has inner and outer radii $a^{-\nu+2}$ and $a^{-\nu}$, respectively $u_{(\nu)}$, we can express their contribution to the attenuation as

$$u_{(\nu)} = \operatorname{Re} \sum_{i=1}^{\infty} a_i (a^{-2\nu} z)^i \quad (8)$$

Thus, the total attenuation due to all singularities between radius a and infinity can be written in the following fashion:

$$a = \operatorname{Re} \sum_{i=1}^{\infty} u_{(\nu)} = \operatorname{Re} \sum_{i=1}^{\infty} \frac{a_i z^i}{1 - a^{-2\nu} z^i}. \quad (c)$$

* The parameter q is defined by

$$q = e^{-\frac{\pi K'}{K}}. \quad (6)$$

* The quote marks are used to indicate the presence of extraneous phase and extraneous attenuation which occur on evaluating the reduced functions along the useful interval.

We next consider the problem of phase in the bandpass case. Here we use the "all-pass" function shown in the p-plane in Fig. 8(a). Mapping these to the w-plane Fig. 8(b), we can again remove all singularities above the horizontal line $w = jK'$ and obtain a reduction of phase slope in the useful interval by a factor of one-half. We are concerned with phase slope instead of phase since a closer investigation shows that now the average phase over the interval is not under our control.² We again map the remaining singularities into the z-plane where we can write out the phase contribution due to the unit annulus in Fig. 8(c) in the following manner:

$$\beta(1) = \text{Re} \sum_{i=1}^{\infty} b_i z^i \quad (10)$$

Summing up the contributions due to all rings, where the ν th annulus gives the contribution

$$r(\nu) = (-1)^{\nu+1} \text{Re} \sum_{i=1}^{\infty} b_i [q^{2(\nu-1)} z]^i \quad (11)$$

we have the following relation for the total phase:

$$\beta = \sum_{\nu=1}^{\infty} \bar{\beta}(\nu) = \text{Re} \sum_{i=1}^{\infty} \frac{b_i z^i}{1 + q^{2i}} \quad (12)$$

Again, we note that due to these removal of singularities effects, the resulting transmission function has both a phase and an extraneous attenuation.

In order to be able to apply the same formulae that are used in band-pass Taylor approximations, a further mapping is useful. This relates the s-plane of Fig. 9 to the z-plane by the transformation*

$$s = j \frac{1-z}{1+z} \quad (13)$$

We note that if we consider the area inside the two circles to be "cut-out" of the s-plane (consider only the portion of the z-plane that is the unit annulus as being of interest) then there is a one-to-one correspondence between the s-plane and the p-plane. Furthermore, the quadrants are in the same order in both planes, which is helpful in studying realizability.

Moreover, since the coefficients for $\bar{u}(1)$ and $\bar{\beta}(1)$ can be readily obtained from those of \bar{u} and $\bar{\beta}$, the unit annulus branch of the transformation need only be considered. Hence, in place of expanding in terms of z, we can expand our desired functions in a power series about the points $s = j1$. The useful interval becomes a circle of radius $2q/(1-q^2)$ centered on the point $s = j(1+q^2)/(1-q^2)$.

* The overall mapping relation between the p and s-planes can be written in the one formula

$$p = \sqrt{k'} \tan \left\{ (jk'/r) \ln \left[\frac{(j+s)/(j-s)}{k} \right], k \right\} \quad (14)$$

With respect to simultaneous attenuation and phase invariance we can rephrase the last paragraph of Section II. A in terms of the s-plane analysis. First the singularities of Fig. 10(a) gives the attenuation $\bar{u}(1)$ the real part of which corresponds to $\alpha(1)$ along the useful interval. Similarly in Fig. 10(b) we obtain the phase $j\bar{\beta}(1)$, the imaginary part of which is $j\nu(1)$ along the useful interval. We again introduce the concept of the infinite network that produces the assigned transmission function \bar{u} and $\bar{\beta}$ in the useful interval. Carrying out the removal of singularities of the infinite network so that the unit annulus and its various mappings becomes clear of singularities results in the transmission function $\bar{u}(1) + j\bar{\beta}(1)$. Matching coefficients for the power series of $\bar{u}(1) + j\bar{\beta}(1)$ and $\bar{u}(1) + j\bar{\beta}(1)$ at the point $s = j1$ gives the quasi-Chebyshev approximation in the band-pass case [as in Fig. 10(c)].

III. Taylor Approximations For Band-Pass Attenuation And Phase Functions

In this section we study band-pass Taylor approximation techniques that must be used to obtain bandpass quasi-Chebyshev approximations. We limit ourselves to discussing all-pole transfer functions. However, other transfer functions can be handled by similar methods.²

As a first step we write out the assigned phase $\bar{\beta}$ in a Taylor series about the point $\omega = 1$. Thus

$$\bar{\beta} = \bar{\beta}_0 + \bar{\beta}_1 (\omega-1) + \bar{\beta}_2 (\omega-1)^2 + \dots \quad (15)$$

The expression of the reciprocal of the network transfer function in terms of p is

$$e^{\alpha} + j\beta = 1 + e_1 p + e_2 p^2 + \dots = P_n(p) \quad (16)$$

Splitting $P_n(p)$ into its even and odd parts we then have

$$\tan \beta = \frac{\text{Odd}[P_n(p)]}{j \cdot \text{Even}[P_n(p)]} = \frac{O(\omega)}{E(\omega)} = \frac{N(x)}{D(x)} \quad (17)$$

where we have let $p = j\omega$ and $\omega-1 = x$.

As an example let us consider the case $n = 2$. Hence

$$\tan \beta = \frac{e_1 x + e_2}{-x^2 - 2x + e_2 - 1} \quad (18)$$

If we write $\bar{\beta} = (\bar{\beta}_0) + (\bar{\beta}_1 x + \bar{\beta}_2 x^2 + \dots)$ and making use of the trigonometric identity concerning the tangent of the sums of two angles, we have

$$\tan \beta = \frac{\bar{t}_0 + \bar{\beta}_1 x + \bar{\beta}_2 x^2 + O(x^3)}{1 - \bar{t}_0 \bar{\beta}_1 x + \bar{\beta}_2 x^2 + O(x^3)} \quad (19)$$

where $\bar{t}_0 = \tan \bar{\beta}_0$ and $O(x)$ means terms of order x^3 and higher. (In case $\bar{\beta}_0 = \frac{\pi}{2}$ we can rewrite our equation in terms of the cotangent.)

Now both $\tan \bar{\beta}$ and $\tan \beta$ could be expanded into power series in x and as many of the coefficients of like powers of x could be set equal in order to get the Taylor approximation. In place

of this operation which gives nontractable equations for the coefficients of the polynomial, $P_n(p)$, we keep the forms shown in Eq. 18. Hence

$$\frac{e_1 x + e_2}{-x^2 - 2x + e_2 - 1} \cong \frac{\bar{t}_0 + \bar{\beta}_1 x + \bar{\beta}_2 x^2 + 0(x^3)}{1 - \bar{t}_0 \bar{\beta}_1 x + \bar{\beta}_2 x^2 + 0(x^3)} \quad (20)$$

where \cong means approximates. Cross multiplying and equating coefficients of like powers of x gives a set of linear equations in term of the polynomial coefficients, e_1 and e_2 .

If the phase β_0 and phase slope β_1 are specified at band center ($\omega=1$), the error is of order x^2 and the coefficients are

$$e_1 = \frac{2\bar{t}_0^2}{\bar{\beta}_1 - \bar{t}_0(1 - \bar{t}_0\bar{\beta}_1)} \quad (21)$$

$$e_2 - 1 = \frac{2\bar{t}_0^2}{\bar{\beta}_1 - \bar{t}_0(1 - \bar{t}_0\bar{\beta}_1)}$$

For certain applications, either $\bar{\beta}_0$ or $\bar{\beta}_1$ is not specified. In this case, the error term may be raised to the order of x^3 . The coefficients are found by solving the following set of equations:

$$e_1 - \bar{t}_0(e_2 - 1) + 0 = 0$$

$$(1 - \bar{t}_0\bar{\beta}_1)e_1 - \bar{\beta}_1(e_2 - 1) + 2\bar{t}_0 = 0 \quad (22)$$

$$-\bar{t}_0(\bar{\beta}_2 + \bar{\beta}_1)e_1 - \bar{\beta}_2(e_2 - 1) + 2\bar{\beta}_1 + \bar{t}_0 = 0$$

For a solution the determinant formed by the coefficients of e_1 , $e_2 - 1$ and the remaining column must be zero. Hence

$$(\bar{\beta}_1 + 2\bar{\beta}_2)\bar{t}_0^3 + (1 - 2\bar{\beta}_1^2)\bar{t}_0^2 + (\bar{\beta}_1 - \bar{\beta}_2)\bar{t}_0 - 2\bar{\beta}_1 = 0. \quad (23)$$

If we insert the specified values we can solve for the real solutions of the unspecified phase term, that is, for \bar{t}_0 or $\bar{\beta}_1$. For each solution we must further test for realizability, etc.

If neither $\bar{\beta}_0$ nor $\bar{\beta}_1$ are specified, as in delay equalizer design, it may be possible to match one more term. Methods for handling this case are discussed in reference (2).

For the case where phase and attenuation are to be considered, we write

$$e^{2u} = |F_n(p)|^2 = N^2 + D^2 = D^2(1 + \tan^2 \beta), \quad (24)$$

Thus
$$e^u = \frac{D}{\cos \beta}. \quad (25)$$

On letting $u = u_0 + u_1 x + u_2 x^2 + \dots$, we have the following expression,

$$e^{u-a_0} = \frac{D}{\sin \beta} \frac{\cos \beta_0}{D_0} \quad (26)$$

where $D_0 = D(x)$ at $(x) = 0$.

Expanding the exponential $e^{\bar{u}-\bar{u}_0}$ and equating it to the right-hand side of Equation (26) we can equate like coefficients of x and obtain a set of linear equations for the coefficients e_i .

We must necessarily specify coefficients of the power series expansion of the phase up to the highest specified coefficient desired in the attenuation characteristics. On combining this set

of linear simultaneous equations due to attenuation specifications with the set due to phase demands, a solution for the coefficients e_i can be obtained.

For certain applications β_0 or β_1 or β_0 and β_1 are not specified. In these cases, techniques similar to those in phase only approximations can be used to increase the accuracy of the approximation at the expense of additional computations.

It should be stressed at this point that the major difficulties in getting quasi-Chebyshev solutions lie in the Taylor part of the problem. The numerical work is straightforward in the case all phase terms are specified, but when higher order matching is desired by leaving β_0 or β_1 unspecified, the required computational work increases rapidly since real roots of auxiliary polynomials such as in Equation (23) must be found.

IV. Examples of Quasi-Chebyshev Approximation

As a recapitulation we shall design a wide-band phase discriminator (a phase-difference network in which the phase difference varies linearly with frequency) that is to be R-C realizable.

One proceeds in the following manner:

1. Translate the data into the commensurate Taylor approximation.
2. Set up the simultaneous equations relating the coefficients of the network polynomial to the phase slope and the mid-band phase. Since the phase slope is unspecified, we first obtain the auxiliary polynomial in p_1 and solve for its real roots.
3. Solve the set of linear simultaneous equations for the n coefficients of the network polynomial.
4. Find the polynomial that gives an R-C realizable network.
5. Solve for the roots of the polynomial in the s -plane.
6. Transform the singularities from the s -plane to the p -plane by means of the relations given in Eq. (14).

In order to do step (1) for a linear all-pass phase difference in the useful interval, we first set $2\bar{p} = 2k\omega$. Then from the transformation of Eq. (4) [see Fig. 6(b)] we can write for the useful interval

$$\omega = \frac{\sqrt{k'}}{\operatorname{dn}(u, k)} \quad (27)$$

Letting

$$\psi = \frac{m u}{h}, \quad (28)$$

we have

$$\frac{1}{\operatorname{dn}(u, k)} = \frac{\pi}{2kk'} + \frac{2i\pi}{\pi k'} \sum_{i=1}^{\infty} \frac{(-q)^i}{1+q^{2i}} \cos(i\psi). \quad (29)$$

Since the constant term is not under control, we henceforth omit it. Hence by means of the re-

relationships given in the appendix one obtains

$$\bar{\beta}(1) = \frac{2\pi k_0}{k k'} \sum_{i=1}^{\infty} (-q)^i \cos(i\psi). \quad (30)$$

Converting to the variable z where $z = qe^{i\psi}$ gives

$$\bar{\beta}(1) = \frac{2\pi k_0}{k k'} \sum (-q)^i \quad (31)$$

where we understand that the phase invariant transformation is to be applied. Hence

$$\bar{\beta}_i = \frac{2\pi k_0}{k \sqrt{k'}}. \quad (32)$$

For the one-half, all-pass phase in the s -plane we write

$$1/2 \bar{\beta}(1) = \bar{\beta}_0 + \bar{\beta}_1 x + \bar{\beta}_2 x^2 + \dots \quad (33)$$

Leaving $\bar{\beta}_0$ unspecified we have by the relationship in Eq. (30),

$$\bar{\beta}_1 = \frac{\pi k_0}{2k \sqrt{k'}} \quad (34)$$

$$\bar{\beta}_i = 0, i > 1, \quad (35)$$

which completes step (1).

Equations (34) and (35) contain the interesting result that the reduced phase in the s -plane is a Taylor approximation to linear phase at $s = j\Omega$. Hence any band-pass Taylor approximations for linear phase can be reused to get a quasi-Chebyshev approximation in the p -plane. To do the work of step (2), we substitute in the equations given in Section III. Letting $\beta_0 = 45^\circ$, we solve for the real roots of the polynomial in p_1 . The values obtained are $p_1 + 0.3090$ and $+0.8090$. Solving for the values of e_1 and e_2 (step 3), we find that $p_1 = +0.3090$ gives a R-C realizable network (step 4) with $e_1 = -1.236$ and $e_2 = -0.236$. This has roots at $s = -0.1681$ and $s = 1.4040$ (step 5). Transforming these points back through the z and w -planes to the p -plane (step 6), we have for the bandwidth ratio $1/k' = 5.76$ (modular angle $\theta = 80^\circ$), roots at $p = -0.142$ and 1.485 . In the "all-pass" phase-difference function the opposite type of singularities occurs at the point $p = +0.142$ and -1.485 .

Figure 11 shows the final result obtained in the design of a wide-band phase discriminator. The lower curve shown is the deviation from a linear phase characteristic.

We next consider the quasi-Chebyshev approximation to an ideal filter. ($\bar{a}(\omega) = \text{constant}$ and $\bar{\beta}(\omega) = \bar{k}(\omega)$) Using the results in the appendix, Equations (36) - (38) we find that for $\bar{a} - \bar{a}_0 = 0$, requires that $\bar{A}_i = \bar{\beta}_i = 0$. Hence by Equation (44) we find that the \bar{a}_i for the intermediate Taylor approximation of \bar{a} are also zero. Thus, the band-pass quasi-Chebyshev approximation to an ideal filter can be obtained directly from the results of a Taylor approximation to an ideal filter since our preceding example has also shown that linear phase in one approximation remains linear in the other.

Applying the method to the case where the network function is an all-pole function with 4 poles

gives the result shown in Fig. 12. Although the Taylor approximations in this case guarantee only an error of the order of the third degree in phase, the quasi-Chebyshev phase approximation has an error of the fourth degree (i.e., there are four points of zero deviation from linear phase). This is due to the fact that the fourth order term is much larger than the third order term in the Taylor approximation, and this effect is carried over into the quasi-Chebyshev approximation.

Figure 13 shows a quasi-Chebyshev approximation to a linear phase characteristic. The network function is an all-pole function with 4 natural modes. Various methods proposed for successive approximation techniques could be used to further improve these solutions if the ripple errors were considered excessive at any point.

V. Conclusions

A technique has been shown for approximating an arbitrary phase or phase and attenuation characteristic in the band-pass interval. A nearly equal-ripple (quasi-Chebyshev) solution is obtained. The technique described requires the use of (1) conformal mappings that put the useful frequency interval onto a circle, (2) certain attenuation and phase invariant transformations, and (3) a Taylor approximation.

Although one of the mappings is done by means of elliptic functions, the elliptic function manipulations are of minor importance in the numerical aspects of the problem. Rather the familiar Taylor series approximation is the part wherein most of the computation lies. As a result, the band-pass quasi-Chebyshev technique of approximation is a useful method for obtaining a solution to many synthesis problems.

Acknowledgment

The author is indebted to Dr. D. F. Tuttle, Jr., who acted as advisor during this research. The work reported was done at Stanford University under Office of Naval Research Contract N6onr251 Task 7 (NRO73 360), jointly sponsored by the U. S. Army Signal Corps, the U. S. Air Force, and the U. S. Navy (Office of Naval Research).

Appendix

In order to determine the coefficients for the expansions in the various planes the following formulae pertaining to attenuation and phase are used.

The assigned attenuation is first expanded into a Fourier series in terms of the variable ψ where $\psi = (m\pi)/k$. Hence

$$\bar{a} = \sum \bar{A}_i \cos i\psi. \quad (36)$$

For many functions the corresponding Fourier series expansions in the ψ variable have been given.⁶ If the desired functions have no simple analytical expressions, then trigonometric inter-

polation methods can always be used.⁷

The attenuation in terms of the z -plane coefficients can be written for $z = qe^{j\psi}$ as

$$\bar{a} = \sum \frac{\bar{a}_i q^i}{1 - q^{2i}} \cos i\psi. \quad (37)$$

Hence

$$\bar{a}_i = -(q^i - q^{-i}) \bar{a}_i. \quad (38)$$

For phase the analogous relationships are

$$\bar{b} = \sum \bar{b}_i \cos i\psi = \sum \frac{\bar{b}_i q^i}{1 + q^{2i}} \cos i\psi \quad (39)$$

and hence

$$\bar{b}_i = (q_i + q^{-i}) \bar{b}_i. \quad (40)$$

In order to get the coefficients for the Taylor series expansions at the point $s = j1$, we have the following relationship where $\bar{y}'(1)$ corresponds to either $\bar{a}'(1)$ or $\bar{b}'(1)$:

$$y'(1)(z) = \sum \bar{c}_i q^i \cos i\psi = \sum \bar{c}_i z^i \quad (41)$$

where the second equal sign assumes that attenuation and phase invariance have been considered. Now let $\omega - 1 = x$ and furthermore let

$$y'(1)(z|x) = y'(1)(x). \quad (42)$$

If we expand $y'(1)(x)$ into a power series

$$y'(1)(x) = \sum \bar{d}_i x^i, \quad (43)$$

it can be shown that the following relationships hold between the \bar{c} and the \bar{d} coefficients:²

$$\begin{aligned} \bar{d}_1 &= -1/2 \bar{c}_1 \\ \bar{d}_2 &= 1/4 (\bar{c}_1 + \bar{c}_2) \\ \bar{d}_3 &= -1/8 (\bar{c}_1 + 2\bar{c}_2 + \bar{c}_3) \\ \bar{d}_4 &= 1/16 (\bar{c}_1 + 3\bar{c}_2 + 3\bar{c}_3 + \bar{c}_4) \\ \bar{d}_i &= (-1/2)^i [\bar{c}_1 + \binom{i}{1} \bar{c}_2 + \dots + \binom{i}{i-r} \bar{c}_r + \dots + \bar{c}_i] \end{aligned} \quad (44)$$

where $\binom{i}{r} = \frac{i!}{(i-r)!r!}$.

References

1. S. Darlington, "Network Synthesis Using Chebyscheff Polynomial Series," BSTJ, Vol. 31 (July 1952), pp. 613-665.
2. V. H. Grinich, "On the Approximation of Arbitrary Phase-Frequency Characteristics," Tech. Rep. No. 61, Electronics Research Lab., Stanford University, May 1, 1953. (Also dissertation for the Ph.D. degree, Stanford Univ., June 1953).
3. S. Darlington, "The Potential Analogue Method of Network Synthesis," BSTJ, Vol. 30 (April 1951), pp. 315-365.
4. E. Weber, "Electro-Magnetic Fields," Vol. 1, (Wiley 1950), pp. 301-310.
5. G. L. Matthaei, "Conformal Mappings for Filter Transfer Function Synthesis," Proc. IRE, Vol. 41, No. 11 (Nov. 1953), pp. 1658-1664.
6. E. T. Whittaker and G. N. Watson, Modern Analysis (4th ed., Cambridge University Press, 1940), p. 508.
7. C. Lanczos, "Trigonometric Interpolation of Empirical and Analytic Functions," Jour. of Math. and Physics, Vol. 17 (1938-39), pp. 123-199.

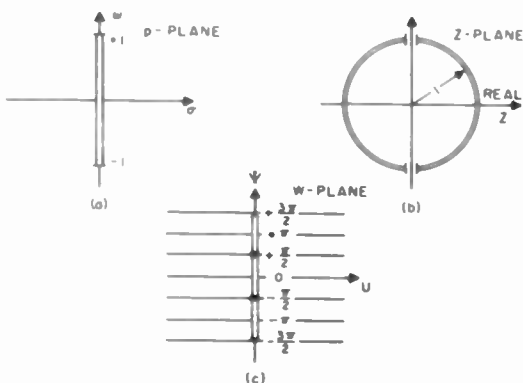


Fig. 1
Conformal mappings for the low-pass interval.

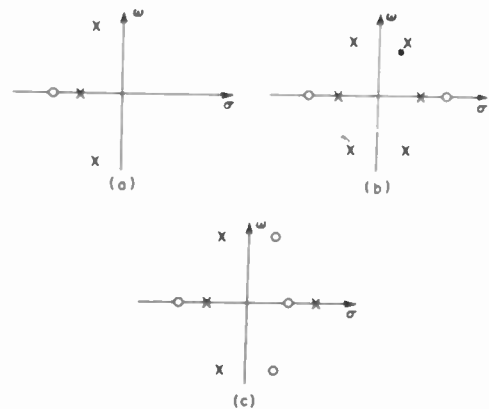


Fig. 2
Network functions in the p-plane.

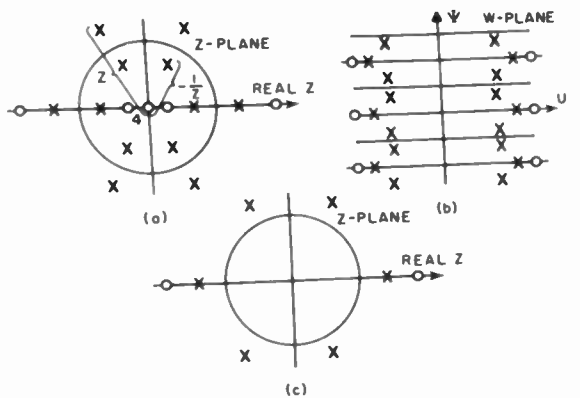


Fig. 3
Attenuation invariant transformation in the Z-plane.

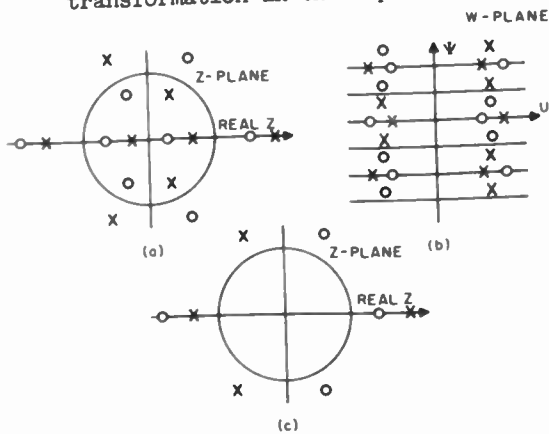


Fig. 4
Phase invariant transformation in the Z-plane.

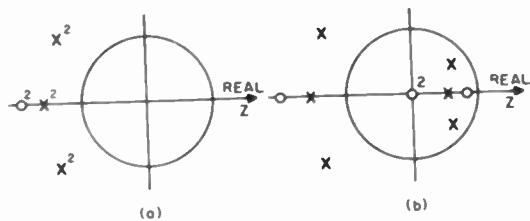


Fig. 5
Attenuation and phase invariance in the Z-plane.

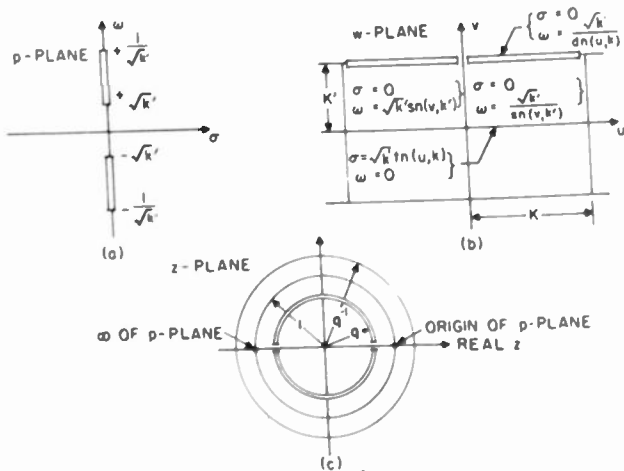


Fig. 6
Band-pass mappings.

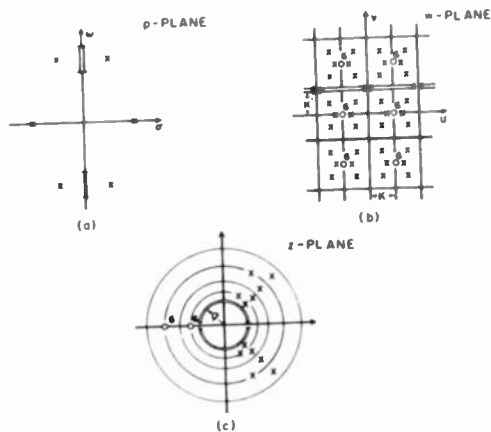


Fig. 7
Band-pass attenuation invariant transformation.

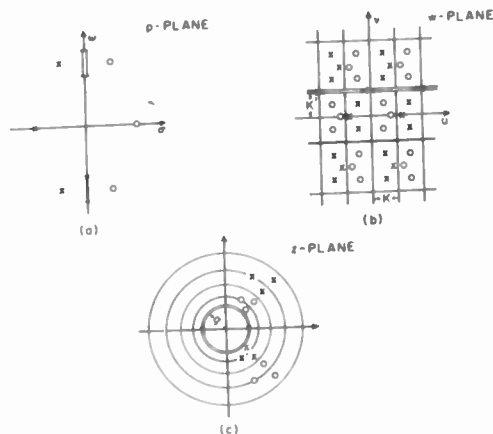


Fig. 8
Phase invariance in the Z-plane.

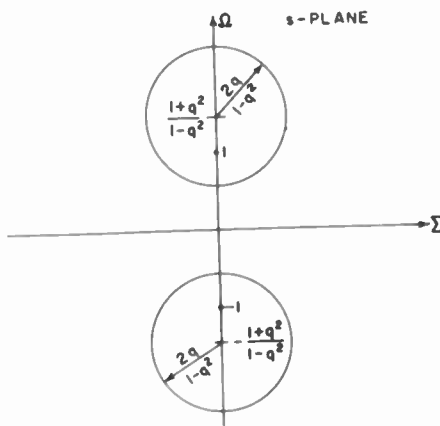


Fig. 9
The geometry of the s-plane.

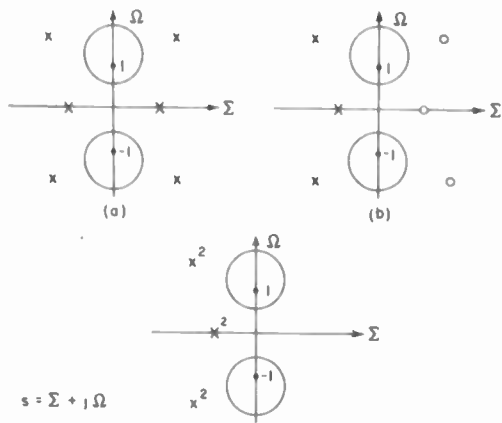


Fig. 10
Attenuation and phase invariance in the s-plane.

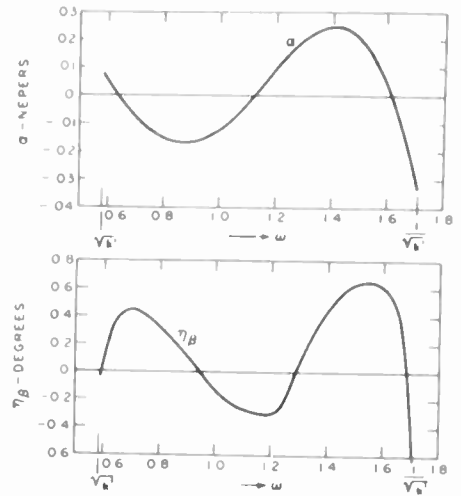


Fig. 12
Quasi-Chebyshev Approximation for band-pass linear phase and constant attenuation.

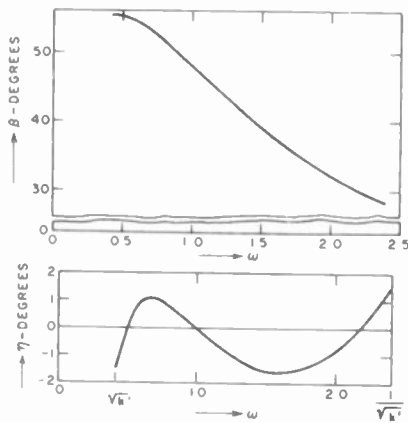


Fig. 11
Wide-band phase discriminator phase response.

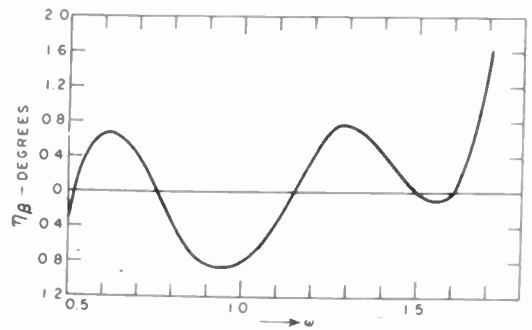


Fig. 13
Quasi-Chebyshev Approximation to linear phase (error curve).

AN APPLICATION OF MODERN NETWORK SYNTHESIS TO THE DESIGN OF CONSTANT-TIME-DELAY NETWORKS WITH LOW-Q ELEMENTS

Leo Storch
Hughes Aircraft Company
Culver City, California

Summary

The design of lumped-constant delay networks has been dominated by image-parameter filters, with particular attention devoted to linearizing their phase-angle versus frequency curve in the case of the m -derived or more complex sections. This type of design is characterized by the assumption that all reactors are lossless and by the cascade-connection of several alike sections in order to obtain a desired over-all bandwidth-delay product. The low-pass filters usually turn out to be of the bridged-T type with negative mutual inductance, a capacitor in the bridge arm, and maybe added shunt capacitors across input and output terminals. They appear to have poor pulse and square-wave response on account of an additional high pass band with an uncontrolled phase characteristic. References 1 to 4 will orient the reader in this field, with no slight intended to the numerous papers not mentioned here.

With the advent of modern network synthesis, a more flexible approach has become feasible. One can afford to concentrate on the over-all transfer function and to synthesize subsequently a suitable network realization, rather than being limited to adapting one of the relatively small number of tractable network sections to a specific application and achieving the desired over-all effect by cascading several sections on an image-impedance basis. In Part I of this paper, a realizable low-pass approximation of the ideal distortionless delay operator ke^{-pt_0} ($t_0 =$ delay time) will be discussed. It is a "best" approximation in the sense that it produces "maximally flat" delay in the frequency domain and accordingly a rather symmetrical impulse response in the time domain, which is centered about $t = t_0$ and approaches the Gaussian curve in the limit. This transfer function will be synthesized in Part II of the paper in the form of a passive low-pass ladder network. A resistor will be associated with each reactor, which can be chosen to correspond to a rather low Q at the reference frequency $\omega_0 = \frac{1}{t_0}$. These resistive elements are taken into account as part of the synthesis procedure, so that their presence causes merely a fixed loss but no distortion of the desired transfer function. In Part III, the

circuit for the ninth-degree approximation (Figure 6) is used as an illustration of the method and oscillograms of its square-wave response (Figure 7) are compared with those of a conventional bridged-T network of comparable delay-bandwidth product (Figure 8) but approximately twelve times as much coil weight and volume.

Part I: A Realizable Low-Pass Approximation of the Ideal Delay Transfer Function ke^{-pt_0}

An ideal delay network would produce an output which is exactly like the input but shifted along the time axis in the positive direction by t_0 , the desired delay time. A network which had the property that $f_{out}(t) = f_{in}(t-t_0)$ would have to possess e^{-pt_0} as a transfer function and a δ -function at $t = t_0$ as an impulse response. This follows from a theorem in Laplace transform theory, that if

$$f_{in}(t) = \mathcal{L}^{-1} [F_{in}(p)], \text{ where } f_{in}(t) \equiv 0 \text{ when } t < 0, \text{ then } f_{out}(t) = f_{in}(t-t_0) \text{ if, and only if,}$$

$$f_{out}(t) = \mathcal{L}^{-1} [F_{in}(p) \cdot e^{-pt_0}] = \mathcal{L}^{-1} [F_{out}(p)],$$

and from the definition of "transfer function" as $\frac{F_{out}(p)}{F_{in}(p)}$. Relaxing this requirement some-

what, an ideal distortionless delay network is characterized by the transcendental transfer function ke^{-pt_0} , since the positive real constant "k" is responsible merely for a fixed gain or loss without introducing distortion in its usual meaning⁶. The prominent property of a distortionless delay network is the linear relationship of phase-angle versus real frequency ($p = jw$) for the infinite range $0 < w < \infty$.

But a finite, linear, passive, lumped-constant network is never associated with a transfer function outside the class of rational functions of $p = \sigma + jw$ and expressible as $T_{(p)} = \frac{P(p)}{Q(p)}$,

$$T_{N(p)} = \frac{a_0}{Q_{N(p)}} = \frac{a_0}{p^N Q_N\left(\frac{1}{p}\right)} = \frac{a_0}{\sum_{r=0}^N a_r p^r} \quad (11)$$

where the coefficients are positive integers

$$a_r = \frac{(2N-r)!}{2^{N-r} r! (N-r)!} = \binom{2N-r}{r} \cdot [1 \cdot 3 \cdot 5 \cdot \dots \cdot (2N-1-2r)] \quad (12)$$

and in particular

$$a_0 = \frac{(2N)!}{2^N (N)!} = 1 \cdot 3 \cdot 5 \cdot 7 \cdot \dots \cdot (2N-1)$$

It may be reassuring to compare $\frac{Q_{N(p)}}{a_0}$

with the McLaurin series of e^p , i. e., $e^p = \sum_{r=0}^{\infty} \frac{1}{r!} p^r$. If a_r in (12) is expanded, then it is seen that ($r > 1$)

$$\frac{a_r}{a_0} = \frac{(N-1)(N-2)(N-3)\dots(N-r+1)}{r!(N-\frac{1}{2})(N-1)(N-\frac{3}{2})\dots(N-\frac{r-1}{2})}$$

Since both the numerator and the denominator

contain "r" terms, $\frac{a_r}{a_0} \rightarrow \frac{1}{r!}$ as $N \rightarrow \infty$, so

that $\frac{Q_{N(p)}}{a_0}$ equals the McLaurin series of e^p

in the limit. Two questions are likely to arise in the reader's mind: (a) why not use the straightforward McLaurin series in the first place; (b) how would the properties of a network synthesized on the basis of the McLaurin series differ from those of a network based on (11) and (12). These questions can be forestalled by the simple statement that a polynomial formed from the first N terms of the McLaurin series of e^p is not a Hurwitz polynomial when N is larger than 4.¹⁵ Therefore, it is not eligible as the denominator of a realizable transfer function for N larger than 4, which includes the majority of applications.

3. Delay and Loss Properties of $T_{N(p)}$

The study of the amplitude and phase characteristics of the transfer function (11) and the choice of a suitable N are greatly simplified by taking advantage of the relationship between Bessel polynomials and the

spherical Bessel functions $\sqrt{\frac{\pi}{2x}} J_{\pm(\gamma + 1/2)}(x)$ for which reasonably comprehensive tables

have been published¹⁶. In this context, it is necessary to deal with real frequencies, i. e., $p = j\omega t_0 = ju$. Inasmuch as it can be shown¹⁷ that

$$y_N\left(\frac{1}{ju}\right) = j^{-N} \sqrt{\frac{\pi u}{2}} e^{ju} \left[(-1)^N J_{-N-1/2}(u) \cdot j^{-j} J_{N+1/2}(u) \right] \quad (13)$$

and since $Q_N(ju) = (ju)^N y_N\left(\frac{1}{ju}\right)$ by (10) and (11), therefore,

$$T_{N(ju)} = \frac{a_0 \cdot e^{-ju}}{u^{N+1} \sqrt{\frac{\pi}{2u}} \left[(-1)^N J_{-N-1/2}(u) \cdot j^{-j} J_{N+1/2}(u) \right]} \quad (14)$$

where $a_0 = \frac{(2N)!}{2^N \cdot N!}$ as in (12).

One is now in the pleasant position of being able to study and explore the properties of the transfer function merely by looking up values in published tables and multiplying them by u^{N+1} rather than having to devote time and energy to the numerical evaluation of N'th degree polynomials. Expression (14) has the additional merit of displaying explicitly the distortion with respect to the ideal delay operator e^{-ju} . From

$$T_{N(ju)} = A e^{-j\theta} = (\text{Distortion factor}) \cdot e^{-ju} = (A e^{j\epsilon}) \cdot e^{-ju} \quad (15)$$

it follows that phase distortion as measured by the angle of lag "ε" with respect to the ideal value of "u" radians is given by

$$\epsilon = \tan^{-1} \frac{J_{N+1/2}(u)}{(-1)^N J_{-N-1/2}(u)} \quad (16)$$

and amplitude distortion is given by

$$A = \frac{a_0}{u^{N+1} \left\{ \frac{\pi}{2u} \left[J_{-N-1/2}^2(u) + J_{N+1/2}^2(u) \right] \right\}^{1/2}} \quad (17)$$

The loss $L = -20 \log A$ db tends to the Gaussian form

$$L \rightarrow \frac{10u^2}{(2N-1) \ln 10} \text{ db}$$

with increasing N, which corresponds to an approximate 3 db bandwidth of $u_{3\text{db}} \approx \sqrt{(2N-1) \ln 2}$. These approximations are serviceable for values of N larger than 3.¹⁹

There are two methods of defining delay which are in common use, phase delay $t_{ph} = \frac{\theta}{\omega}$, and group delay $t_g = \frac{d\theta}{d\omega}$. In the region of nearly constant delay, they are approximately equal. For the chosen transfer function (14) as derived from (11):

$$t_{ph} = \frac{\theta}{\omega} = t_o \left(1 - \frac{\epsilon}{u}\right) \quad (18)$$

and

$$t_g = \frac{d\theta}{d\omega} = t_o \left(1 - \frac{d\epsilon}{du}\right) = t_o \left[1 - \frac{1}{u^2} \frac{1}{\left[\frac{\pi}{2u} (J_{-N-1/2}^2(u) + J_{N+1/2}^2(u)) \right]} \right] \quad (19)$$

where t_o is the nominal or zero-frequency delay. Expression (19) is useful for quantitative purposes, but a more vivid impression of the delay characteristic is obtained by expanding the distortion term in (19) in a McLaurin series, which certainly converges in the vicinity of $u = 0$:

$$t_g = t_o \left[1 - \frac{u^{2N}}{a^2} + \frac{u^{2N+2}}{(2N-1)a^2} - \frac{2(N-2)}{(2N-1)^2(2N-3)a^2} u^{2N+4} \dots \right] \quad (20)$$

The absence of the terms in $u^2, u^4, \dots, u^{2N-2}$ indicates that the first $N-1$ derivatives of t_g are zero at $u = 0$. This explains why the delay network synthesized on the basis of (11) has been characterized as "maximally flat"¹⁸⁻¹⁹, in analogy to the usage of this term in insertion loss design²⁰. The delay is closely equal to its D. C. value t_o up to a certain upper frequency which is a function of N , and then it declines steadily and smoothly as the frequency increases to infinity.

With equations (14) to (19) representing a complete description of the pertinent characteristics in compact form and in terms of tabulated functions, it is hardly a chore to evaluate the loss and delay distortion versus frequency for a given N and t_o , or to establish the required value of N for specified attenuation or delay distortion versus a given delay-bandwidth product. The phase-angle, group delay, and loss characteristics of $T_{9(j\omega)}$ are drawn in Figure 1. This particular value of N has been chosen for illustration because the circuit realization and response for $T_{9(j\omega)}$ are described in Part III of the paper.

4. The Impulse Response Corresponding to $T_{N(p)}$.

The transfer function may be written as

$$T_{N(p)} = e^{\Gamma(p)} = e^{\gamma_2 p^2 + \gamma_4 p^4 + \gamma_6 p^6 + \dots + \gamma_1 p + \gamma_3 p^3 + \gamma_5 p^5 + \dots} \quad (21)$$

where the exponent $\Gamma(p)$ is expanded in a McLaurin series. This is permissible since $T_{N(p)}$ is a low-pass transfer function and analytic in the vicinity of the origin and up to the boundary set by the radius of convergence. Since for real frequencies $p = ju$ the odd powers in the exponent of (21) determine the phase angle $(-\theta)$ [see (15)],

$$t_g = \frac{d\theta}{d\omega} = t_o \left[-\gamma_1 + 3\gamma_3 u^2 - 5\gamma_5 u^4 \dots \right] \quad (22)$$

must agree with (20). Therefore: $\gamma_1 = -1$, and $\gamma_3 = \gamma_5 = \gamma_7 = \dots = \gamma_{2N-1} = 0$. Let $f(t/t_o)$ be the impulse response corresponding to $T_{N(p)}$ and expand the exponential underneath the integral sign in the Laplace-transform equation, where t_o has been inserted explicitly on the right-hand side of (23):

$$T_{N(p)} \cdot e^{p} = \int_0^{\infty} f(t/t_o) e^{-p(t-t_o)} dt/t_o = \sum_{k=0}^{\infty} \left[\int_0^{\infty} f(t/t_o) \cdot (t-t_o)^k dt/t_o \right] (-p)^k / k! \quad (23)$$

If

$$T_{N(p)} \cdot e^{-\gamma_1 p} = T_{N(p)} \cdot e^p = e^{\gamma_2 p^2 + \gamma_4 p^4 + \gamma_6 p^6 + \dots + \gamma_{2N+1} p^{2N+1} + \gamma_{2N+3} p^{2N+3} + \dots} = e^{\Gamma_c(p)}$$

is also expanded in a power series about $p = 0$, $T_{N(p)} \cdot e^p = 1 + \Gamma_c(p) + 1/2 \Gamma_c^2(p) + 1/3! \Gamma_c^3(p) + \dots$, and set equal to (23), it follows that the first N odd moments of the impulse response with respect to the mean $t = t_o$ are zero, i. e.,

$$\int_0^{\infty} f(t/t_0) (t-t_0)^{2m-1} dt = 0 \text{ for } m = 1, 2, \dots, N.$$

Consequently, the impulse response is symmetrical about $t = t_0$ to the extent that the 3rd, 5th, 7th . . . and $(2N-1)$ st central moments vanish (the 1st central moment is zero by definition). Perfect symmetry is approached as $N \rightarrow \infty$, since all odd central moments tend to zero as $N \rightarrow \infty$.

An independent check is available for the above statements. If the roots of $Q_{N(p)}$ are introduced, $Q_{N(p)} = \prod_{r=1}^N (p - \rho_r)$, then

$$\Gamma(p) = \sum_{k=1}^{\infty} \gamma_k p^k = \ln T_{N(p)} \\ = \ln a_0 - \sum_{r=1}^N \ln (p - \rho_r) \quad (24)$$

After each term $\ln (p - \rho_r)$ is expanded in a McLaurin series, it follows that:

$$\gamma_k = \frac{1}{k} \sum_{r=1}^N (1/\rho_r)^k \quad (25)$$

Burchnall²¹ has shown that when $Q_{N(p)}$ is given by (11) and (12), then

$$\sum_{r=1}^N 1/\rho_r = -1, \quad \sum_{r=1}^N (1/\rho_r)^{2m-1} = 0, \quad (26)$$

$$m = 2, 3, 4, \dots, N.$$

This is a welcome corroboration that the 3rd, 5th, 7th . . . and $(2N-1)$ st central moments are equal to zero. Furthermore, $Q_{N(p)}$ is a unique solution, except for a constant multiplier, if the maximally flat delay property in accordance with (26) is postulated.

An exact description of the impulse response would require a knowledge of all the roots of $Q_{N(p)}$, which can only be acquired by solving for the roots of an N 'th degree polynomial for each specific value of N . However, a certain amount of general information about the nature of the roots is available²¹:

- (a) all the roots of $Q_{N(p)}$ are simple
- (b) $Q_{N(p)}$ has only one real zero when N is odd;

if ρ_i^* is the real zero of $Q_{i(p)}$, then $(-1) = \rho_1^* > \rho_3^* > \rho_5^* \dots > -\infty$.

- (c) $Q_{N(p)}$ has no real zero when N is even
- (d) $Q_{N(p)}$ and $Q_{N+1(p)}$ have no root in common
- (e) all roots of $Q_{N(p)}$ lie outside the unit circle for $N > 1$.
- (f) since $a_0 = a_1$: $\sum_{r=1}^N 1/\rho_r = -1$
- (g) since $a_{N-1} = \binom{N+1}{2}$: $\sum_{r=1}^N \rho_r = (-1) \cdot \binom{N+1}{2}$,
and $\prod_{r=1}^N \rho_r = (-1)^N \cdot a_0 = (-1)^N \cdot 1 \cdot 3 \cdot 5 \cdot 7 \dots (2N-1)$

The roots of $Q_{N(p)}$ have been tabulated in reference 19 for N from 1 to 9 and an approximation of the impulse response has been described there, which tends to a Gaussian curve for large N .

The calculated impulse response corresponding to $T_9(p)$ is plotted in Figure 2.

5. A Class of Transfer Functions for "Maximally Flat" Delay Networks.

Although Part II of this paper is concerned with the realization of $T_{N(p)}$ exclusively, it is worthwhile to mention that a whole class of realizable transfer functions can be deduced immediately from $T_{N(p)}$, each member of which furnishes maximally flat delay. This class is given by:

$$S_{M, N(p)} = \frac{Q_{N(0)}}{Q_{M(0)}} \cdot \frac{Q_{M(-xp)}}{Q_{N(p)}}, \quad 0 \leq M \leq N$$

The total zero-frequency delay equals $(1+x)t_0$, where "x" is chosen so as to equalize the useful frequency range of $Q_{M(-xp)}$ and of $Q_{N(p)}$. $S_{M, N(p)}$ appears to have a smaller total delay in general than $T_{M+N(p)}$, but its rate of transmission loss is lower in comparison. As M approaches N , "x" approaches "1" and the loss bandwidth continues to increase. Finally, $M = N$ corresponds to an all-pass network with a zero-frequency delay of $2t_0$. The only member of this class which is a minimum-phase transfer function is $S_{0, N(p)} = T_{N(p)}$. All others

possess zeros in the right half of the p-plane.

The loss and delay characteristics of $S_{M,N(p)}$ can be obtained by combining

$$L = 20 \log (A_M - A_N),$$

$$t_{ph} = t_o \left[1 + x - (\epsilon_M + \epsilon_N) \right], \text{ and}$$

$$t_g = t_o \left[1 + x - \left(\frac{d\epsilon_M}{du} + \frac{d\epsilon_N}{du} \right) \right]$$

with (16) and (17).

If the realization of $S_{M,N(p)}$ is attempted by the method of Part II, it is necessary first to multiply numerator and denominator by a Hurwitz polynomial which makes the over-all numerator an even function of the frequency variable used in the LC development. If this variable is 'p' (i.e., $d = 0$), then the multiplier is $Q_{M(xp)}$ and the class of modified transfer functions is given by:

$$\bar{S}_{M,N(p)} = \frac{Q_{N(o)}}{Q_{M(o)}} \cdot \frac{Q_{M(-xp)} Q_{M(xp)}}{Q_{N(p)} Q_{M(xp)}}, \quad 1 \leq M \leq N.$$

Part II: The Ladder Network Realization of $T_{N(p)}$

In Part I a rather convincing case has been made with regard to the suitability of $T_{N(p)}$ as the transfer function of a low-pass lumped-parameter delay network. Also fairly simple expressions have been set up in terms of tabulated functions to guide the choice of N in meeting given requirements regarding delay, bandwidth, and distortion. It is now necessary to find a suitable configuration of specific resistances, inductances, and capacitances, which actually possesses $T_{N(p)}$ as its transfer function. It is particularly desired that the network be unbalanced and that coils with only low Q's be required in its construction. Since the source of the signal is likely to be a vacuum tube, i.e., a high-impedance (constant-current) generator, the emphasis will be placed on realizing $T_{N(p)}$ as a transfer impedance, so that $T_{N(p)}$ will equal the ratio of output voltage to input current of the resulting network. The method of designing a network so that $T_{N(p)}$ is represented by the ratio of output voltage to input voltage is quite analogous and is useful when dealing with a low-impedance (constant-voltage) generator.

1. Network Synthesis on a Quasi-Reactive Basis.

The choice of a constant as the numerator in the transfer function (11) opens the way to an uncomplicated method of synthesis as well as to a very attractive network configuration. The key to the dominating position occupied by a constant in the numerator is the fact that it makes the network realizable as a ladder structure and that the numerator retains its evenness in the frequency variable under the transformation $p = s-d$ ²², where "s" is an auxiliary frequency variable and "d" a constant. The combination of these two properties sufficiently makes it possible to develop the network on a quasi-reactive basis in the auxiliary "s" domain by the straightforward Cauer process²³, but to arrive at a final network which has a uniform loss component attached to each reactance. The loss components may be proportioned so as to reduce the required Q's of the coils to rather low values at the expense of a fixed transmission loss. It goes without saying that the successful restriction of the actual synthesis process to the 2-element LC case is rather attractive as compared to a general RLC method of synthesis.

The transfer impedance of a two-terminal-pair, finite, linear, bilateral network, terminated by a normalized load resistance of 1 ohm, is easily derived from the equivalent T-section (Figure 3) as:

$$Z_{tr} = \frac{V_2}{I_1} = \frac{z_{12}}{1 + z_{22}} \quad (27)$$

Let "d" be chosen such that the shift of the real-frequency axis to the left, implied in the translation $p = s-d$, does not carry it past any of the roots of $Q_{N(p)}$, and let the transformation $p = s-d$ be carried out in (11) obtaining:

$$\begin{aligned} T_{N(s)}^* &= \frac{a_o}{Q_{N(s)}^*} = \frac{a_o}{G_{N(s)}^* + H_{N(s)}^*} \\ &= \frac{a_o}{\sum_{r=0}^N a_r^* s^r} \end{aligned} \quad (28)$$

Here $G_{N(s)}^*$ and $H_{N(s)}^*$ represent the even and odd parts, respectively, of $Q_{N(s)}^*$. Since the numerator is even in "s", (28) may be realized within a real multiplier as the transfer impedance of a low-pass, reactive network as follows. Let the constant in the numerator be changed to a_o^* , in order to make the D.C. value ($s=0$) of (28) equal to "1" as required for the transfer impedance of a low-pass reactive network when

ideal transformers are excluded. After re-arranging the terms, so that

$$Z_{tr}^* = \frac{\frac{a_0^*}{H_N^*(s)}}{1 + \frac{G_N^*(s)}{H_N^*(s)}} \quad (29)$$

the following identifications are prominent in view of (27):

$$z_{12}^* = \frac{a_0^*}{H_N^*(s)} \quad (30)$$

$$z_{22}^* = \frac{G_N^*(s)}{H_N^*(s)} \quad (31)$$

Surely, $z_{22}^* = \frac{G_N^*(s)}{H_N^*(s)}$ is a driving-point reactance function inasmuch as $Q_N^*(s)$ is a Hurwitz polynomial by hypothesis (see Part I, section 1).

Furthermore, $z_{12}^* = \frac{a_0^*}{H_N^*(s)}$ is an odd function of "s", its poles are also poles of $z_{22}^* = \frac{G_N^*(s)}{H_N^*(s)}$, and the numerator is of lower degree than the denominator. Therefore, the Z_{tr}^* of (29) is realizable by a reactive network.

1.1 The Ladder Development

Synthesis by means of Cauer's straightforward development of a reactive ladder network²³ is particularly applicable, since the structure of the numerator is so simple. This method does not require finding the roots of the N'th degree equation $Q_N^*(s) = 0$ and it leads to a network which requires only N reactances for an N'th degree transfer function (11), which is the smallest possible number²⁴. The principal step required by this method is the expansion of the single reactance function z_{22}^*

$= \frac{G_N^*(s)}{H_N^*(s)}$ in a continued fraction in "s", proceeding as explained in connection with (2) and remembering that $\psi(s) = z_{22}^*$ when N is even and $\psi(s) = \frac{1}{z_{22}^*}$ when N is odd. All the coefficients

a_j^* ($j = 1, 2, \dots, N$) of the continued fraction must be non-zero and positive. If zero or negative terms should appear, it would indicate that "d" has been chosen too large or, much more likely, that not enough significant figures have been retained in calculating the a_r^* 's and in the subsequent computations, which may happen when N is quite large. It follows immediately that the circuit diagram of z_{22}^* is a ladder network (Figure 4) with inductances of values a_{2k}^* in the series arms and capacitances of values a_{2k-1}^* in the shunt arms ($k = 1, 2, \dots$)¹¹. The computations of the component values may be

$$\text{checked by } \sum_k a_{2k-1}^* = \frac{a_1^*}{a_0^*} \text{ and } \prod_{k=1}^N a_k^* = \frac{1}{a_0^*}.$$

A degree-reducing, common factor of $G_N^*(s)$ and $H_N^*(s)$ would have to be an even polynomial in "s", which would possess roots on the real-frequency axis $\text{Re}[s] = 0$ or in the right half of the s-plane. But such a factor must be ruled out, since it would also be contained in $Q_N^*(s) = G_N^*(s) + H_N^*(s)$ and thereby violate the condition that $Q_N^*(s)$ must be a Hurwitz polynomial. The ladder development of Figure 4 actually corresponds to (31) and not just to some sub-multiple thereof.

If a pair of input terminals is put across C_1 , for N odd or even, then z_{12}^* has no finite zeros of transmission, since none of the network arms is resonant at a finite, non-zero frequency; furthermore, z_{22}^* can have at most a pole at $s = \infty$ in addition to the poles of z_{12}^* , which occurs when N is even and is due to the numerator being higher in degree by "1" than the denominator. Therefore, $z_{12}^* = \frac{a_0^*}{H_N^*(s)}$ as required by (30). All that remains to be done is to connect a 1-ohm load resistance across the z_{22}^* -terminals (output terminals), in order to complete the network realization of Z_{tr}^* of (29).

Nothing has been said so far about z_{11}^* , because it does not affect the transfer impedance at all. It will be shown, however, in section 9, that the input impedance Z_{in} in Figure 5 is finite for all real frequencies. Since this means that the residue condition applicable to LC networks²⁵ is automatically fulfilled with the equal sign in this synthesis procedure, z_{11}^* is uniquely determined.

1.2 Final Network

To remove the effect of the initial frequency transformation $p = s-d$, the network arms $s \cdot L_{2k}$ and $s \cdot C_{2k-1}$ must be replaced by $L_{2k}(p+d)$ and $C_{2k-1}(p+d)$, respectively; i. e., a resistor $d \cdot L_{2k}$ must be placed in series with the inductance L_{2k} and a resistor $\frac{1}{d \cdot C_{2k-1}}$ must be shunted across the capacitor C_{2k-1} , where $k = 1, 2, 3, \dots, \frac{N}{2}$ (for N even) or $\frac{N+1}{2}$ (for N odd). By (11), (28), and (29), the resulting network of Figure 5 has the transfer impedance

$$Z_{tr} = \frac{a_o^*}{a_o} T_{N(p)} \quad (32)$$

which equals $T_{N(p)}$ except for a constant multiplier. The price which has to be paid for the insertion of the resistive components is the fixed loss corresponding to $k = \frac{a_o^*}{a_o}$, which may be calculated from $a_o^* = Q_{N(p)}|_{p=-d}$. For small values of "d" ($d \ll 1$), the loss is somewhat less than $20 \log \frac{1}{1-d}$ db. Finally, the normalizations are removed by multiplying all L_{2k} 's by $t_o \cdot R_L$, all C_{2k-1} 's by t_o/R_L , and all R 's by R_L , if a nominal delay t_o (in seconds) and a load impedance level R_L (in ohms) are desired. The new impedance level introduces a factor R_L in (32), i. e.,

$Z_{tr} = (a_o^* R_L / a_o) T_{N(p)}$. If Q has its conventional meaning, then $Q = \frac{1}{d}$ at $\omega_o = \frac{1}{t_o}$ for all network arms and varies linearly with frequency.

2. Illustrative Example: The Realization of $T_9(p)$

The effect of increasing an even N by "1" is to add a capacitor across the output terminals without changing the number of coils in the network. One will usually, therefore, prefer an odd N in order to maximize the delay-bandwidth product for a given number of coils, usually bulkier and more expensive than capacitors. Let $N = 9$, which possesses a delay-bandwidth product suitable for an application the writer had to deal with, and let $t_o = 1.25$ milliseconds, $R_L = 4000$ ohms. The loss (with respect to D.C.) and delay properties of $T_9(j\omega)$ are shown in Figure 1, the "u" scale of which may be

converted to a direct frequency scale corresponding to $t_o = 1.25$ milliseconds by multiplying each value of "u" by 128 cps.

First, the a_r 's of (12) have to be calculated for $N = 9$. This may be done with the help of factorial tables²⁶, or by a direct method after writing out the terms and canceling common factors in numerator and denominator. The polynomial $Q_{N(p)}$ may also be obtained by recursion [see (9)]:

$$Q_{k(p)} = (2k-1)Q_{k-1(p)} + p^2 Q_{k-2(p)} \quad (33)$$

with $Q_{-1} = 1$, $Q_0 = 1$; but this method is not efficient for large N 's. One obtains for $Q_9(p)$:

$$\begin{aligned} Q_9(p) = \sum_{r=0}^9 a_r p^r = & p^9 + 45p^8 + 990p^7 + 13,860p^6 \\ & + 135,135p^5 + 945,945p^4 + 4,729,725p^3 \\ & + 16,216,200p^2 + 34,459,425p \\ & + 34,459,425. \end{aligned} \quad (34)$$

Secondly, a value of "d" is chosen on the basis of the coils to be used or of the fixed loss that can be tolerated. The upper limit on "d", imposed by the condition of physical realizability, is not likely to be a hindrance in practice, since it is 1.5 for $N = 2$ and higher for larger N 's; e. g., $d_{max} = 6.29$ for $N = 9$. The loss, of course, increases as "d" grows larger. A preliminary analysis indicates that for $t_o = 1.25$ milliseconds and $R_L = 4000$ ohms, which were required values for the particular application, the maximum required inductance will be about 1.5 henry. If a toroidal molybdenum-permalloy dust core of 0.8 inches outside diameter is decided upon, the minimum Q among the coils equals 4 at $f_o = 128$ cps ($f_o = \frac{1000}{2\pi \cdot 1.25} = 128$ cps).

Therefore, the value of "d" may be fixed at 0.25, which corresponds to a fixed loss of 2.2 db for $N = 9$.

Substituting $p = s-0.25$ in (34), the a_r 's may be obtained by means of synthetic division. The large number of significant figures is due to machine calculation and not indicative of required accuracy. Expanding the ratio of the odd to the even part

$$\begin{aligned} \psi(s) = & (s^9 + 902.25s^7 + 115,605.492s^5 \\ & + 3,864,041.03s^3 + 27,181,586.9s) \div \\ & (42.75s^8 + 12204.9375s^6 + 789,490.775s^4 \\ & + 13,003,312.9s^2 + 27,181,586.9) \end{aligned}$$

in a continued fraction, introducing the loss components, and removing the normalizations, the network of Figure 6 is obtained. It may be worth noticing that the spread in the values of the elements is not too large and that a generator impedance higher than 29,200 ohms can be accommodated without affecting the exact transfer characteristic.

The ladder configuration makes it possible to dispense with stringent tolerances for the network components, as required for the conventional bridged-T image-parameter delay networks. Even so, the measured amplitude and phase response are indistinguishable from the calculated curves of Figure 1. The time-domain behavior can be judged by Figure 2 and the oscillograms of Figure 7. For each square-wave input of Figure 7A, the response of the maximally flat delay network of Figure 6 is shown in Figure 7B, and that of a bridged-T delay network* (Figure 8) of comparable delay and linear phase-shift bandwidth is shown in Figure 7C. The superior square-wave response of the maximally flat delay network is rather pronounced. The more than 12:1 reduction in weight and volume achieved by means of the present design may be of considerable interest in view of today's prevalent need for miniaturization; the total weight of the coils is only 2 ounces for the maximally flat delay network of Figure 6, but it exceeds 1-1/2 pounds for the bridged-T delay network of Figure 8 (coil Q's are about 30 at 128 cps), with a similar ratio for the relative volumes.

A comparison may also be made with a 4-section constant-k filter delay network, which has the same general configuration as the network of Figure 6. Its t_g would differ from t_o by one percent at somewhat less than $u = 1.3$, assuming even infinite Q's for the coils. The same error occurs for the maximally flat delay network of Figure 6 not until $u = 6.08$ and even its loss is not down by 3 db from the D. C. value until "u" reaches 3.38. The only debit that can be charged against the maximally flat delay network, in competing with the constant-k filter, involves cascade-connection, which is discussed further in section number 3. Incidentally, the coils and capacitors of the constant-k filter are much larger for the same R_L and t_o , being 5 henry for the coil and 0.31 μ f for the capacitors per section.

These comparisons are a measure of the improvements achieved by transfer-function synthesis as compared to classical design of a network with a similar configuration.

*Designed by J. E. Taber of the Hughes Research and Development Laboratories.

3. Cascading of Delay Networks of the Type of Figure 6.

There are applications which require several equal or unequal increments of delay rather than a single fixed value. When the over-all fixed transmission loss is not excessive, the question arises how maximally flat delay networks, designed on the transfer impedance basis, can be cascaded without having to insert isolating amplifiers. All prerequisites for cascade-connection are fulfilled if the input impedance Z_{in} is equal to a constant resistance, preferably R_L , independent of frequency. This is not the case for the synthesis procedure of section number 2, which controls z_{12} and z_{22} but does not shape z_{11} , at the expense of additional elements, in order to maintain a constant input-impedance level. This method of synthesis does, however, produce an input impedance which is finite along the whole real-frequency axis and thereby paves the way for transforming the network into a "constant-resistance" filter.

If the transmission matrix for the network of Figure 5 is given by

$$\begin{pmatrix} V_1 \\ I_1 \end{pmatrix} = \begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{pmatrix} \cdot \begin{pmatrix} V_2 \\ I_2 \end{pmatrix}, \quad (35)$$

then

$$Z_{in} = \frac{\tilde{A} + \tilde{B}}{\tilde{C} + \tilde{D}}. \quad (36)$$

For a ladder network, the elements of the transmission matrix are entire rational functions of the impedances of the series arms and of the admittances of the shunt arms. Since these are $L_{2k} \cdot (p+d)$ and $C_{2k-1} \cdot (p+d)$ in Figure 5, \tilde{A} , \tilde{B} , \tilde{C} , and \tilde{D} are all polynomials in "p". Inasmuch as

$$Z_{tr} = \frac{V_2}{I_1} = \frac{1}{\tilde{C} + \tilde{D}} \quad (37)$$

in the present notation, the denominator of (36) cannot differ from the Hurwitz polynomial $Q_{N(p)}$ by more than a constant. It follows conclusively that Z_{in} is finite along the whole real-frequency axis, including $p = 0$ and $p = \infty$. This being the case, a driving-point impedance Z_{in}^C can be developed such that $Z_{in} + Z_{in}^C$ equals a constant for all real frequencies. A network arm connected in series with the input terminals does not alter the transfer impedance Z_{tr} of a network. Therefore, Z_{in}^C may be connected in

Bibliography

such a manner, in order to establish a fixed input-impedance level and enable cascading of the desired number of these networks, designed for equal or different amounts of delay.

While this method of input-impedance compensation is exact, it would appear to be rather wasteful in element for practical purposes. The number of elements needed to construct Z_{in}^c would equal the number used in the original network of Figure 6. No general approximate solution has been attempted, but a rather serviceable approximate compensating impedance has been worked out readily for the network realization of $T_{9(p)}$ (Figure 6). The added series arm, shown in Figure 9, and consisting only of a few elements, is sufficient for good compensation over the major portion of the pass-band. Several of the networks of Figure 9 have been operated in cascade with complete success, undistorted and properly delayed outputs having been taken from each junction point of two of these networks and at the termination.

It should be realized that the cascading of maximally flat delay networks increases the rate of transmission loss, since the loss curve per network is bell-shaped rather than roughly square. More precisely, a total of N reactances distributed equally among " n " networks, each of degree $\frac{N}{n}$ and contributing a delay $\frac{t_0}{n}$, has a 3 db bandwidth of

$$u_{3 \text{ db}} \approx \frac{1}{n} \sqrt{(2N - n) \ln 2} \quad (38)$$

but a network designed directly on the N 'th degree basis has a 3 db bandwidth of

$$u_{3 \text{ db}} \approx \sqrt{(2N - 1) \ln 2} \quad (39)$$

which is $n \sqrt{\frac{2N-1}{2N-n}}$ times larger than (38). It stands to reason that cascading normally would only be used when the additional tap-points are required. Otherwise, the network should be synthesized on the basis of the full N , since in this manner a greater useful delay-bandwidth product is obtained at the expense of more lengthy calculations.

Acknowledgements

The author is indebted to Dr. Louis Weinberg for stimulating discussions on network synthesis and to Mr. F. G. Rasmussen for making laboratory tests and taking the oscillograms, both of whom are with the Research and Development Laboratories, Hughes Aircraft Company.

- (1) U. S. Patent 2,250,461; 1941 (L. Batchelder).
- (2) M. H. Herb, C. W. Horton, F. B. Jones: "On the Design of Networks for Constant Time Delay," *Journal of Applied Physics*, v. 20, June 1949, pp. 616 - 620.
- (3) C. F. Floyd, R. L. Corke, H. Lewis: "The Design of Linear Phase Low-Pass Filters," *Proc. of the Institution of Elec. Engineers*, Part IIIA, 1952, pp. 777 - 787.
- (4) C. M. Wallis: "Design of Low-Frequency Constant Time Delay Lines," *A. I. E. E. Transactions*, Part I, 1952, pp. 135 - 140.
- (5) M. F. Gardner, J. L. Barnes: "Transients in Linear Systems," v. I; Wiley, New York, 1942, p. 236.
- (6) M. J. Di Toro: "Phase and Amplitude Distortion in Linear Networks," *Proc. I. R. E.* v. 36, January 1948, pp. 24 - 36.
- (7) H. W. Bode: "Network Analysis and Feedback Amplifier Design," van Nostrand, New York 1945, Chapters I and XI.
- (8) A. E. Guillemin: "Modern Methods of Network Synthesis," *Advances in Electronics*, v. III, Academic Press, New York, 1951, p. 275.
- (9) H. S. Wall: "Polynomials Whose Zeros have Negative Real Parts," *The Am. Mathematical Monthly*, v. 52, June-July 1945, pp. 308 - 322. A. E. Guillemin: "The Mathematics of Circuit Analysis," Wiley, New York, 1949, pp. 395 - 409.
- (10) W. Cauer: "Frequenzweichen konstanten Betriebswiderstandes," *E. N. T.*, v. 16, 1939, p. 116.
- (11) A. E. Guillemin: "Communication Networks," v. II, Wiley, New York, 1935, pp. 198 - 202.
- (12) G. N. Watson: "A Treatise on the Theory of Bessel Functions," Cambridge University Press, 1952, p. 303.
- (13) O. Perron: "Die Lehre von den Kettenbruechen," Teubner, Berlin, 1929, p. 5.
- (14) H. L. Krall, O. Fink: "A New Class of Orthogonal Polynomials: The Bessel Polynomials," *Am. Math. Soc., Transactions*, v. 65, January 1949, pp. 100 - 115; see p. 101.
- (15) K. E. Iverson: "The Zeros of the Partial Sums of e^Z ," *M. T. A. C.*, v. VII, July 1953, pp. 165 - 168.

(16) Mathematical Tables Project, N. B. S.: "Tables of Spherical Bessel Functions," 2 vols., Columbia University Press, New York 1947.

(17) Reference 14, p. 103.

(18) W. H. Huggins: "Network Approximation in the Time Domain," Air Force Cambridge Research Laboratories, Report E5048A, p. 34 et seq.

(19) W. E. Thomson: "Networks with Maximally Flat Delay," *Wireless Engineer*, v. 29, October 1952, pp. 255 - 263. Corrections: *ibid.*, p. 309.

(20) V. D. Landon: "Cascade Amplifiers with Maximal Flatness," *R.C.A. Review*, v. 5, 1941, pp. 347 - 362 and 481 - 497.

(21) J. L. Burchall: "The Bessel Polynomials," *Canad. J. of Math.*, v. 3, 1951, pp. 62 - 68.

E. Grosswald: "On Some Algebraic Prop-

erties of the Bessel Polynomials," *Am. Math. Soc., Transactions*, v. 71, Sept., 1951, pp. 197 - 210.

(22) S. Darlington: "Synthesis of Reactance 4-Poles," *J. of Math. and Physics*, v. 18, Sept. 1939, pp. 257 - 333. See Part IV, "Dissipative Reactance Networks." Reference 7, pp. 216 - 223.

(23) Reference 8, p. 286, et seq.

(24) Reference 11, p. 203.

(25) Reference 8, p. 274.

W. Cauer: "Theorie der linearen Wechselstromschaltungen," v. 1, Becker und Erler, Leipzig 1941, p. 373.

(26) N. B. S., Applied Math. Series 16: "Tables of $n!$ and $\Gamma(n+1/2)$ for the First Thousand Values of n ." U.S. Dept. of Commerce, 1951.

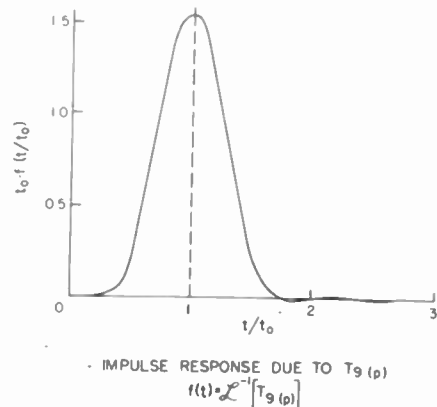
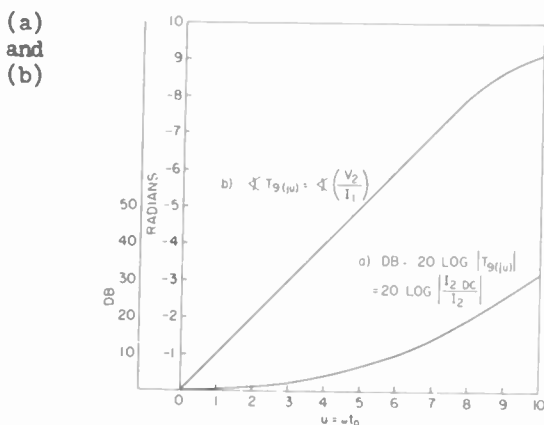


Fig. 2

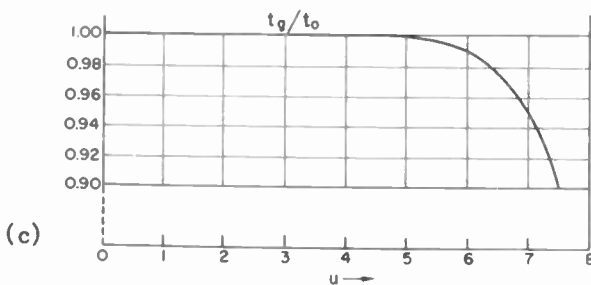


Fig. 1
 Transfer function $T_9(j\omega)$: (a) loss;
 (b) phase angle; (c) group delay.

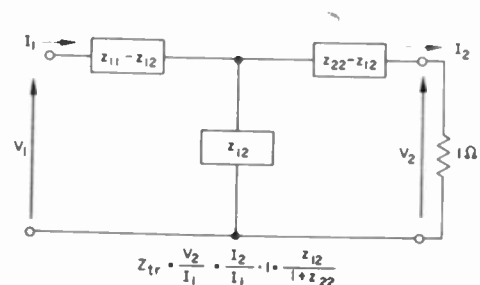


Fig. 3 - Equivalent-T network.

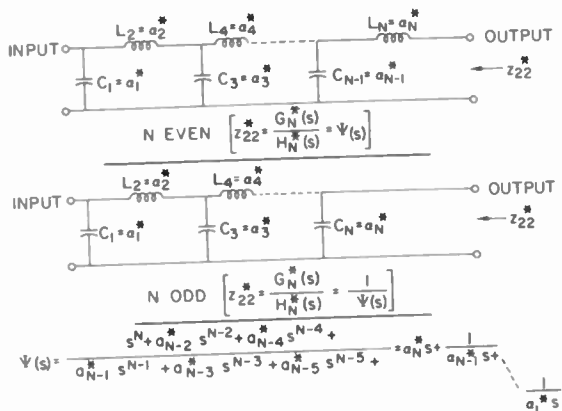
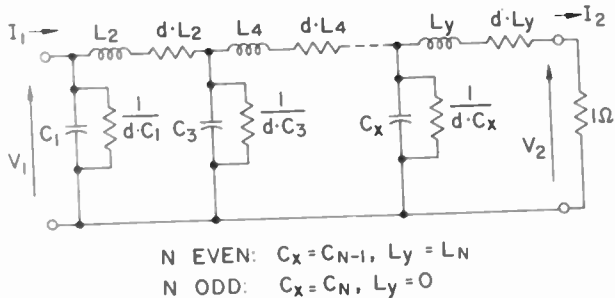


Fig. 4 - Ladder development of z_{22} .



MAXIMALLY FLAT DELAY NETWORK
 $Z_{tr} = \frac{V_2}{I_1} = \frac{a_0^*}{a_0} T_N(p)$

Fig. 5

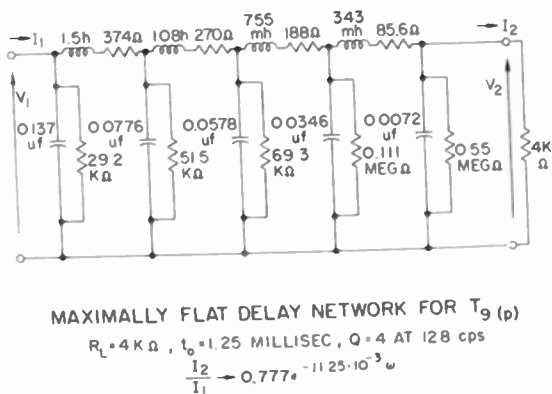


Fig. 6

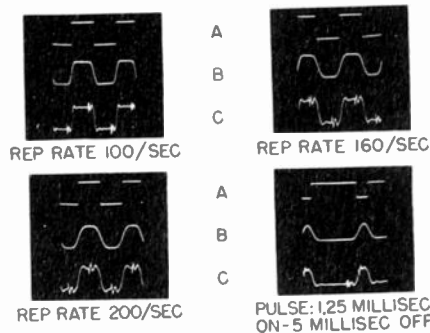


Fig. 7

(a) Square-wave response of 1.25 milliseconds delay networks; (b) maximally-flat; (c) bridged-T.

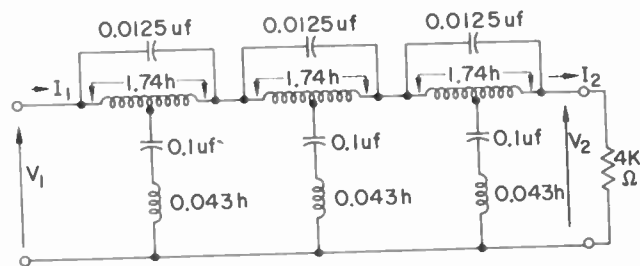
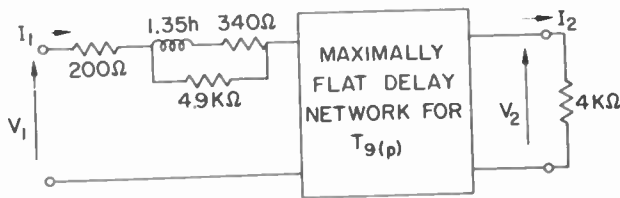


Fig. 8



MAXIMALLY FLAT DELAY NETWORK FOR
 CASCADE-CONNECTION

$R_L = 4K\Omega, t_0 = 1.25$ MILLISEC., $Q = 4$ AT 128 cps
 $\frac{V_2}{I_1} \approx \frac{I_2}{I_1} \rightarrow 0.777 e^{-j1.25 \cdot 10^{-3} \omega}$

Fig. 9

A TRANSISTOR ANALOG

R.D. Lohman
RCA Laboratories Division
Princeton, N.J.

Abstract

One approach to transistor circuit design is through the use of an equivalent circuit. The performance of the transistor under conditions where the equivalent circuit is valid may be predicted once the parameters of the equivalent circuit are known. However, at frequencies well within the useful operating range of junction transistors, the reactive parameters in the equivalent circuit become important and computation is very tedious. The same may be said of computations predicting the effect of variations in the equivalent circuit parameters.

The usefulness of the equivalent circuit concept may be extended by constructing a transistor analog. The analog employs a pentode vacuum tube as the active element of the equivalent circuit and variable resistances and capacitances as the passive elements. By adjusting the control voltages on the tube and choosing the proper values for the passive elements, the analog may be made to perform exactly like a transistor except for a transformation in impedance level and frequency scale.

There are certain advantages to be found in studying transistors by making measurements on the analog:

1. Circuit performance of

realizable transistors may be evaluated before the transistors are actually available.

2. The effect of varying the parameters of the equivalent circuit may be quickly determined and the information thus obtained may be used advantageously by device people.

3. Measuring equipment may be much simplified and stray capacities ignored if the analog is designed to operate at frequencies lower than those at which the transistor will operate.

The analog described in this paper is based on a modified equivalent circuit of a grounded emitter transistor developed by L.J. Giacoletto. The impedance level of the analog is 10 times that of the transistor in order that the relatively large (40,000 μ mhos at 1 ma. emitter current) mutual conductance of the transistor be represented by a single pentode tube. In addition, all capacities of the analog are 10 times those of the transistor so that data may be derived from the analog at frequencies lower by a factor of 100 than those at which the transistor will operate.

Measurements made on the analog have shown good agreement with those made on transistors.

Summary

The analogy between junction-transistor triodes and vacuum-tube triodes is sufficiently close that conventional vacuum-tube switching circuits such as multivibrators can be transistorized simply by changing the values of the circuit constants and replacing the vacuum tube with a junction transistor. The associated circuits will be at lower impedances and will operate at lower voltages, but the currents will remain of the same magnitude. Because of the alpha cutoff frequency, a transistorized switching circuit operates from dc to hundreds of kilocycles a second. Such measured characteristics of multivibrators as d-c potentials, gate widths, and repetition periods are in good agreement with theoretical values.

I. Introduction

Most switching circuits that have been designed with transistors have used point-contact types and are not very similar to switching circuits designed with vacuum-tube triodes¹. However, the analogy between junction-transistor triodes and vacuum-tube triodes is sufficiently close that many conventional vacuum-tube switching circuits (multivibrators and flip-flops in particular) can be transistorized by changing only the values of the circuit constants and not their configuration. The connections to the plate, grid, and cathode of the vacuum tube can be replaced by connections to the collector, base, and emitter, respectively, of the junction transistor. As alpha (the current gain from emitter to collector) approaches unity, this replacement becomes a better one, since collector current then approaches emitter current in value, corresponding to the equality of plate and cathode currents in the negatively biased vacuum-tube triode. Furthermore, if the transistors are n-p-n types, the power supply and waveform polarities of the vacuum-tube circuit are maintained in the transistorized version; for p-n-p types, the polarities are reversed.

Two junction-transistor circuits analogous to the familiar one-shot cathode-coupled multivibrator and the Eccles-Jordan flip-flop are described. Also described is a less conventional free-running multivibrator using junction-transistor triodes. This circuit has a square-wave output, whose repetition rate is controllable by a single RC time constant.

II. One-Shot Multivibrator

General

Figure 1 is a generalized schematic diagram of a one-shot emitter-coupled multivibrator using n-p-n junction transistors. It has been drawn to emphasize its similarity in configuration to the

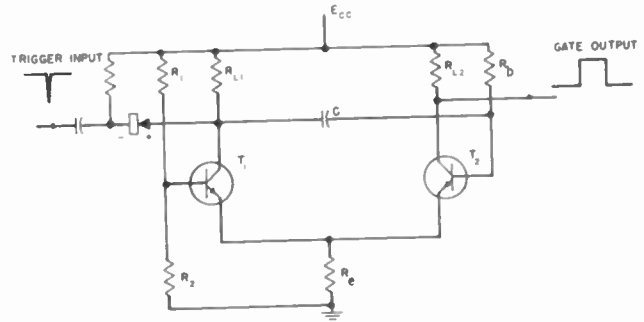


FIGURE 1. ONE-SHOT EMITTER-COUPLED MULTIVIBRATOR USING N-P-N JUNCTION TRANSISTORS

familiar one-shot cathode-coupled vacuum-tube circuit. Operation is quite similar to that of the vacuum-tube version. In the quiescent condition, transistor T₂ is on in the saturation region by the positive return of its base resistor (R_b). At the same time, the biasing network for transistor T₁, consisting of resistors R₁ and R₂, cuts off T₁ by maintaining its base sufficiently negative with respect to the common emitter potential, which is determined by the conduction of T₂. Corresponding to the vacuum-tube case, the circuit may be triggered (as shown in Figure 1) by a negative pulse on the collector of T₁ coupled through an isolating crystal diode. After the circuit is triggered, a rapid regeneration takes place until T₂ is turned off and T₁ is turned on. T₂ is held off until capacitor C has discharged sufficiently to bring the base of T₂ back approximately to the common emitter potential (now determined by the conduction of T₁). When this point is reached, regeneration takes place once more, and the circuit is restored to its initial condition, thus terminating a gate.

Gate Width

An approximate expression for gate width (T) may be derived as follows. Initially, since T₂ is on in the saturation region, the collector, base, and emitter of T₂ are all at about the same potential,

$$E_o = \frac{R_e E_{cc}}{R_e + \frac{R_{L2} R_b}{R_{L2} + R_b}} \quad (1)$$

(This condition is in contrast with the vacuum-tube case, in which a tube at zero bias still has appreciable plate-to-cathode voltage drop.) After triggering, when T₁ is turned on, its collector

drops in potential by

$$\Delta V_{c1} \approx \frac{R_{L1} R_b}{R_{L1} + R_b} \left[\frac{\alpha_1 R_2 E_{cc}}{R_e (R_1 + R_2)} - \frac{E_{cc} - E_o}{R_b} \right] \quad (2)$$

provided T_1 is in its active region, α_1 is its current gain (assumed constant), the impedances R_1 and R_2 are sufficiently low to maintain the base-to-ground potential of T_1 at $R_2 E_{cc} / (R_1 + R_2)$ and the base-to-emitter voltage drop in T_1 is neglected. This same drop, ΔV_{c1} (shown in Figure 2), appears on the base of T_2 . When the exponential

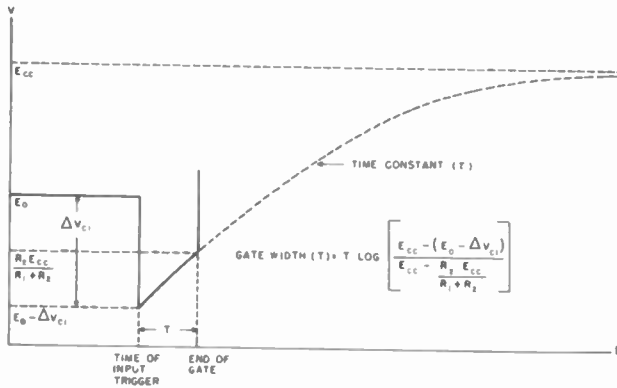


FIGURE 2. BASE-TO-GROUND VOLTAGE OF T_2 , ONE-SHOT MULTIVIBRATOR

waveform on the base of T_2 (which is heading for E_{cc} with the time constant $\tau = C [R_{L1} + R_b]$) passes through the new common emitter potential ($R_2 E_{cc} / [R_1 + R_2]$), the gate is terminated. Therefore the gate width is

$$T = \tau \log \left[\frac{E_{cc} + \Delta V_{c1} - E_o}{\frac{R_1 E_{cc}}{R_1 + R_2}} \right]$$

$$= C(R_{L1} + R_b) \log \left\{ \left[\frac{1}{1 + \frac{R_{L1}}{R_b}} \right] \left[\frac{1 + \frac{R_2}{R_1}}{1 + R_e \left(\frac{1}{R_{L2}} + \frac{1}{R_b} \right)} + \alpha_1 \left(\frac{R_2}{R_1} \right) \left(\frac{R_{L1}}{R_e} \right) \right] \right\} \quad (3)$$

$$\approx CR_b \log \left[\frac{1 + \frac{R_2}{R_1}}{1 + \frac{R_e}{R_{L2}}} + \left(\frac{R_2}{R_1} \right) \left(\frac{R_{L1}}{R_e} \right) \right],$$

$$R_b \gg R_{L1}, R_{L2}; \alpha_1 \approx 1 \quad (4)$$

Choice of Base Resistor R_b

There is a maximum permissible value of R_b to ensure initial operation of T_2 in its saturated region. Larger values of R_b shift the operating point of T_2 to its active region. Such operation is undesirable, because, for rapid recovery after the gate, capacitor C should discharge rapidly through the relatively low base-to-emitter impedance of T_2 when it is in the saturated region. This maximum value of R_b can be related to the other circuit constants.

Assume that T_2 is in its active region and make use of the relations for its collector current (I_{c2}) and emitter current (I_{e2}) derived by R. F. Shea² for single-battery operation.

$$I_{c2} = \frac{I_{c0} \left(1 + \frac{R_e}{R_b} \right) + \frac{\alpha_2 E_{cc}}{R_b}}{1 - \alpha_2 + \frac{R_e}{R_b}} \quad (5)$$

$$I_{e2} = \frac{I_{c2} - I_{c0}}{\alpha_2} \quad (6)$$

where I_{c0} is the collector current of T_2 for zero emitter current, and α_2 is the current gain of T_2 .

The maximum value of R_b is the value that just causes the collector-to-base voltage (V_{c2}) to drop to zero; therefore

$$V_{c2} = E_{cc} - I_{c2} R_{L2} - I_{e2} R_e = 0 \quad (7)$$

If the values of I_{c2} and I_{e2} from equations 5 and 6 are substituted into equation 7, the maximum value of R_b is found to be

$$(R_b)_{\max} = \frac{R_{L2} (\alpha_2 E_{cc} + I_{c0} R_e)}{(1 - \alpha_2) E_{cc} - I_{c0} (R_e + R_{L2})} \quad (8)$$

$$\approx \frac{\alpha_2 R_{L2}}{1 - \alpha_2}, I_{c0} \ll \frac{(1 - \alpha_2) E_{cc}}{R_e + R_{L2}} \quad (9)$$

Thus large values of R_b are permissible for high values of α_2 or I_{c0} . However, the value of I_{c0} must be small to prevent the gate width (T) from being reduced from the value given in equation 3. This is the case since negative emitter-to-base bias in T_2 , while reducing the emitter current to a small value in a good transistor, does not reduce the current flowing in the collector-base circuit much below the value of I_{c0} . Thus, in addition to the desired charging of capacitor C through the long time constant (τ) there will be an undesired charging by I_{c0} . Since I_{c0} varies with different transistors and temperatures, it is not a good practice to make the gate width a function of I_{c0} .

Values of R_b much smaller than $(R_b)_{\max}$ may be undesirable for at least two reasons.

1. Longer recovery time in terms of gate width will result because of the relatively smaller ratio of time constants for charging of capacitor C during the gate and after the gate.
2. Slower rise time of the collector waveform of T_2 will result because of increased minority-carrier storage in a more heavily saturated T_2 .

Experimental Circuit

An experimental circuit was constructed with the following values of constants (1% tolerance resistor values):

R_{L1} = 3300 ohms	R_1 = 3300 ohms
R_{L2} = 2200 ohms	R_2 = 2200 ohms
R_b = 33,000 ohms	E_{cc} = 15 volts
R_e = 2200 ohms	C = 0.1 μ f
	(measured value 0.106 μ f)

For the transistors used (Texas Instruments Type 201), $\alpha > 0.95$ and $I_{CO} < 10 \mu$ amp at room temperature. From equation 9, $(R_b)_{max}$ is equal to 42,000 ohms; thus, a value of 33,000 ohms is permissible. The theoretical value of gate width T from equation 3 is 1800 μ sec and from equation 4 is 2100 μ sec. Measurements made with 14 different transistors in the circuit gave a gate width of about 1460 μ sec \pm 10%. The discrepancy of about 20% between the measured value and the theoretical value of gate width from equation 3 may be attributable to the following factors.

1. The jump (ΔV_{C1}) was about 15% less than the theoretical value--chiefly because the value of α_1 used in equation 2 should be an average value, since α is not completely independent of collector current and voltage.
2. The several-tenths-of-a-volt potential difference between base and emitter of T_1 when T_1 is turned on were neglected in equations 2 and 3.

Equation 3 states that gate width T should be independent of collector supply voltage E_{cc} . For a pair of average transistors, the variation with E_{cc} that was found experimentally is shown in Table 1.

Table 1

E_{cc} in volts	T in μ sec
5	1270
10	1470
15	1550
20	1600

With $C = 0.005 \mu$ f (measured value 0.0046 μ f) instead of 0.1 μ f, substantially the same fractional agreement was found as that above with regard to a comparison of theoretical and measured

gate widths. (The theoretical value of gate width from equation 3 is 80 μ sec.)

III. Flip-Flop

General

Figure 3 shows a symmetrical flip-flop using p-n-p junction transistors. It has been drawn to emphasize its similarity in configuration to the

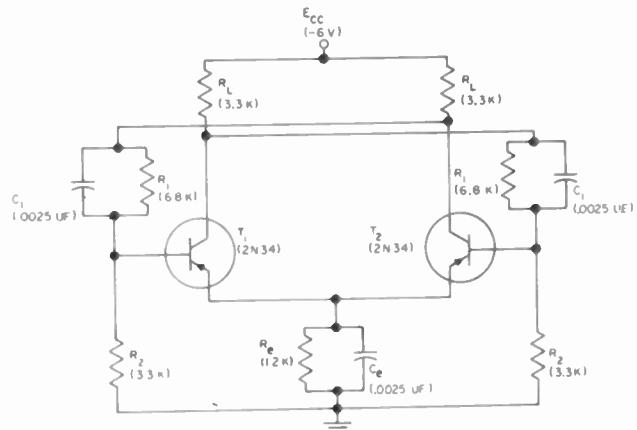


FIGURE 3. SYMMETRICAL FLIP-FLOP USING P-N-P JUNCTION TRANSISTORS

familiar Eccles-Jordan flip-flop vacuum-tube circuit. Operation is quite similar to that in the vacuum-tube version. There are two stable states-- T_1 on and T_2 off, and vice versa. Triggering (to be discussed later) from one state to the other is possible because of the cross-coupling capacitors (C_1) and the emitter bypass capacitor (C_e).

D-C Analysis

If several simplifying assumptions are made, a d-c analysis, useful from an engineering standpoint, can be made as follows. Assume identical ideal junction transistors--that is, when transistors are turned on in their active region, $\alpha = 1$ (therefore, base current = 0) and base-to-emitter voltage = 0. When transistors are turned off, all currents are equal to 0. Denote the current in the "on" transistor by I_o , the collector swing between "on" and "off" states by V_o , the base-to-emitter bias on the "off" transistor by V_b , and the collector-to-emitter voltage of the "on" transistor by V_c . The following equations can then be written:

$$I_o = \frac{R_2 E_{cc}}{R_e (R_L + R_1 + R_2)} \quad (10)$$

$$V_o = \frac{I_o R_L (R_1 + R_2)}{R_L + R_1 + R_2} \quad (11)$$

$$V_b = \frac{R_2 V_o}{R_1 + R_2} \quad (12)$$

$$V_c = \frac{(R_1 + R_2) E_{cc}}{R_L + R_1 + R_2} - I_o R_e - V_o \quad (13)$$

Simultaneous solution of equations 10 through 13 gives the following expressions for the resistances, R_L , R_e , R_1 , and R_2 as functions of E_{cc} , I_o , V_o , V_b , and V_c (this analysis applies equally well or better to the vacuum-tube flip-flop, provided the "on" tube does not draw grid current):

$$R_L = \frac{E_{cc} (V_o - V_b)}{I_o (V_o + V_c)} \quad (14)$$

$$R_e = \frac{V_o + V_c}{I_o \left(\frac{V_o}{V_b} - 1 \right)} \quad (15)$$

$$R_1 = \frac{E_{cc} (V_o - V_b)^2}{I_o [E_{cc} (V_o - V_b) - V_o (V_o + V_c)]} \quad (16)$$

$$R_2 = \frac{V_b E_{cc} (V_o - V_b)}{I_o [E_{cc} (V_o - V_b) - V_o (V_o + V_c)]} \quad (17)$$

Experimental Circuit

A practical circuit, based on equations 14 through 17 was designed as follows. Values of $E_{cc} = 6$ v (-6 v actually, since p-n-p transistors were used), $I_o = 1.2$ ma, $V_o = 3$ v, $V_b = 1$ v, and $V_c = 0$ v were chosen. Note that, since the transistors are not ideal, a greater-than-zero value of V_c is expected to ensure operation of the "on" transistor in the active region. From equations 14 through 17,

$$\begin{aligned} R_L &= 3,330 \text{ ohms} \\ R_e &= 1,250 \text{ ohms} \\ R_1 &= 6,670 \text{ ohms} \\ R_2 &= 3,330 \text{ ohms} \end{aligned}$$

The nearest 5% tolerance RMA values of

$$\begin{aligned} R_L &= 3,300 \text{ ohms} \\ R_e &= 1,200 \text{ ohms} \\ R_1 &= 6,800 \text{ ohms} \\ R_2 &= 3,300 \text{ ohms} \end{aligned}$$

were actually used. From equations 10 through 13, the corresponding theoretical values for I_o , V_o , V_b , and V_c are 1.23 ma, 3.06 v, 1.00 v, and -0.02 v respectively. Measurements on five RCA

type 2N34 and five each of Radio Receptor types RR 14H and RR 34H p-n-p junction transistors gave values of I_o (emitter current actually) = 1.0 to 1.04 ma, $V_o = 2.4$ to 2.7 v, $V_b = 0.5$ to 0.6 v and $V_c = 0.6$ to 0.9 v. The measured values of alpha for these fifteen transistors varied from about 0.95 to 0.98 at the manufacturer's rated operating points. Best agreement of measured and theoretical values was obtained with the higher-alpha units. All fifteen transistors had low I_{co} and emitter characteristics with a sharp knee within 0.1 v base-to-emitter potential. Thus, the 0.5 to 0.6 v value of V_b actually measured was more than enough to cut off the transistors.

Binary-Counter Operation

Figure 4 shows a triggering circuit (similar to that used with vacuum tubes) for operating the flip-flop as a binary counter. The crystal diode

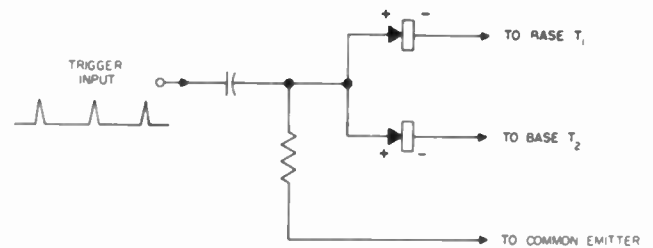


FIGURE 4. BINARY-COUNTER TRIGGERING CIRCUIT FOR FLIP-FLOP

connected to the base of the "off" transistor has a greater reverse bias than the crystal diode connected to the base of the "on" transistor. Thus the positive input trigger is routed to the "on" transistor and starts the regenerative action that ends in the other stable state of the flip-flop. The following trigger is similarly routed to the new "on" transistor and changes the flip-flop back to its original state, and so on, resulting in a binary-counter action. This description of triggering assumes that the proper values of capacitors C_1 and C_e (in Figure 3) have been chosen. Smaller values of these capacitors can be used when transistors with higher alpha cutoff frequencies* are used. Too-large values slow up the operation of the flip-flop. For the transistors used, which had alpha cutoff frequencies from 0.4 to 0.9 Mc, values of C_1 and C_e equal to 0.0025 μ f gave satisfactory triggering. The flip-flop connected as a binary counter operated at trigger repetition rates up to about 100 kc. Three identical flip-flop stages, in which the collector waveform of T_2 of the first stage was connected to the input of the second stage and so on, were successfully cascaded to make a scale of 8.

* Alpha cutoff frequency is defined as the frequency at which the magnitude of alpha is 3 db less than its low-frequency value.

IV. Free-Running Multivibrator

General

Figure 5 shows a free-running emitter-coupled multivibrator using n-p-n junction transistors. It is intended for application as an

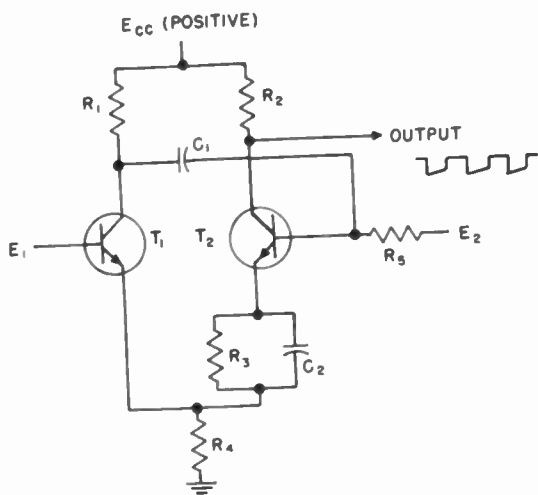


FIGURE 5. MULTIVIBRATOR SQUARE-WAVE GENERATOR USING N-P-N JUNCTION TRANSISTORS

approximate square-wave generator, whose repetition rate can be adjusted by adjusting the value of the single capacitor C_1 . Undoubtedly, the process of transistorizing described in Sections II and III could here be reversed to obtain an equivalent free-running cathode-coupled multivibrator using vacuum tubes.

The operation can be explained by referring to Figure 6, which shows the base-to-ground waveform of T_2 . Neglecting base-to-emitter potential

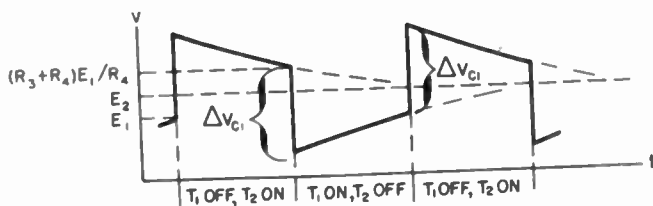


FIGURE 6. BASE-TO-GROUND VOLTAGE OF T_2 , SQUARE-WAVE GENERATOR

differences when a transistor is turned on in its active region (the intended region of operation for this circuit), it is seen that changes of state of the multivibrator take place when the base-to-ground potential of T_2 reaches a value of E_1 (T_2 off and T_1 on), or a value of $E_1(R_3 + R_4)/R_4$ (T_1 off and T_2 on). Capacitor C_2 is a small peaking capacitor and does not affect this reasoning. Through proper choice of the circuit constants,

the "off" periods of T_1 and T_2 can each be made equal to one time constant of the exponential waveform on the base of T_2 . This choice, in turn, should lead to minimum fractional time jitter of the output "square wave". These time constants are approximately equal to $C_1(R_1 + R_5)$, neglecting the effects of collector output impedance of T_1 and base input impedance of T_2 . Thus, if the above conditions are met, the theoretical repetition period of the output will be approximately $2C_1(R_1 + R_5)$.

Circuit-Constant Relationships

The relations necessary among the circuit constants to obtain the desired operation can be derived as follows. From Figure 6, two equations are obtained:

$$E_2 = \frac{E_1 + \frac{E_1(R_3 + R_4)}{R_4}}{2} \quad (18)$$

$$E_1 + \Delta V_{c1} - E_2 = \epsilon(E_2 - E_1) \quad (19)$$

where $\epsilon = 2.718$ and the jump in collector voltage of T_1 ,

$$\Delta V_{c1} \approx \alpha_1 \left(\frac{E_1}{R_4} \right) \left(\frac{R_1 R_5}{R_1 + R_5} \right) \quad (20)$$

if α_1 is the current gain of T_1 (assumed constant) and the base input impedance of T_2 is negligible. Solution of equations 18, 19, and 20 leads to relations for E_2 , R_3 , and $R_1 R_5 / (R_1 + R_5)$ as functions of E_1 , R_4 , and ΔV_{c1} :

$$E_2 = E_1 + \frac{\Delta V_{c1}}{1 + \epsilon} \quad (21)$$

$$R_3 = \frac{2R_4 \Delta V_{c1}}{E_1(1 + \epsilon)} \quad (22)$$

$$\frac{R_1 R_5}{R_1 + R_5} = \frac{R_4 \Delta V_{c1}}{\alpha_1 E_1} \quad (23)$$

To see whether there is enough middle-frequency gain to ensure regeneration when both T_1 and T_2 are turned on, consider that the loop is opened at the base of T_2 and an input signal e_1 is injected there. The ratio of e_2 , the resultant output signal at the collector of T_1 , to e_1 can be calculated approximately as follows. Neglecting R_2 , T_2 is simply a grounded collector stage (with approximately unity voltage gain to its emitter) feeding a grounded base stage T_1 , having a source impedance of R_3 and a load impedance of $R_1 R_5 / (R_1 + R_5)$. Thus the open-loop middle-frequency gain is:

$$\frac{e_2}{e_1} = \frac{\alpha_1 R_1 R_5}{R_3(R_1 + R_5)} = \frac{1 + \epsilon}{2} \quad (24)$$

making use of equations 22 and 23. Thus, for the desired circuit, $e_2/e_1 = 1.86$, which is greater than unity as required for regeneration.

Experimental Circuit

An experimental circuit, based on values of $E_1 = 7.5$ v, $R_4 = 5,600$ ohms, $\Delta V_{C1} = 7.5$ v, and $\alpha_1 = 0.95$ for Type 201 transistors, used circuit constants (1% tolerance resistor values) with the following values (selected using equations 21 through 23):

$$\begin{aligned} R_3 &= 3,000 \text{ ohms} & E_2 &= 9.5 \text{ v} \\ R_1 = R_5 &= 12,000 \text{ ohms} & E_{CC} &= 20 \text{ v} \\ R_2 &= 2,200 \text{ ohms} \end{aligned}$$

The values chosen for E_{CC} and R_2 ensure operation of T_1 and T_2 in their active regions when they are turned on. For convenience, the voltages E_1 and E_2 were derived from a low-impedance bleeder connected between E_{CC} and ground. Bleeder resistance values of 2,400 ohms, 470 ohms, and 1,800 ohms (reading from E_{CC} to ground) were selected; they gave open-circuit values of $E_1 = 7.7$ v and $E_2 = 9.7$ v.

First, a value of $C_1 = 1.0$ μ f (measured value = 1.1 μ f) for low-frequency operation was chosen ($C_2 = 0$). The corresponding theoretical repetition period of the output is $2C_1(R_1 + R_5)$, equal to 0.053 sec. Measurements made with fourteen different transistors gave an output period of 0.033 sec \pm 11% and a duty cycle of 0.54 \pm 8%. When the impedance of the bleeder supplying the voltages E_1 and E_2 was reduced by a factor of 5, the ratio of measured to theoretical repetition periods was increased from the above value of 62% to a value of 72% without change in duty cycle.

Theoretically, repetition period and duty cycle of the output should be independent of supply voltage E_{CC} . Table 2 shows the measured performance for an average pair of transistors when E_{CC} is varied.

Table 2

E_{CC} in volts	Repetition Period in seconds	Duty Cycle
5	0.017	0.53
10	0.027	0.54
15	0.031	0.55
20	0.033	0.56
25	0.034	0.55

Next, a value of $C_1 = 0.001$ μ f (measured value = 0.00104 μ f) for higher-frequency operation was chosen. The value of C_2 for best shape of the output without affecting its period was determined experimentally to be about 500 μ f. (C_2 performs a function here corresponding to that of a small capacitor bypassing the cathode resistor of a

vacuum tube to improve its high-frequency response.) The corresponding theoretical repetition period of the output is 50 μ sec. Measurements made with the same fourteen transistors gave an output period of 40 μ sec \pm 20% and a duty cycle of 0.58 \pm 7%. In this case (compared with $C_1 = 1.0$ μ f), the low alpha cutoff frequency of some of the transistors* lengthened the rise and fall times and thus caused the average ratio of measured to theoretical periods of the output to be raised while increasing the variation of measured period.

The reasons cited under the one-shot multivibrator to explain discrepancies between theoretical and measured performance are also applicable here.

V. Conclusions

It has been demonstrated that some vacuum-tube triode-switching circuits of the multivibrator type can be transistorized with junction transistors without changing the circuit configuration. In the experimental transistor circuits cited, measured periods were about 70 to 80% of those calculated from the theory. Repeatability of results with different transistors was about \pm 10% of the center value.

The method of transistorization demonstrated is believed to be applicable to a great variety of vacuum-tube triode circuits of both switching and linear types.

VI. Acknowledgment

The above work was done under Bureau of Ships Contract N0bsr-63075.

The writer was helped by many people at Airborne Instruments Laboratory, including P. D. Strum, who supplied encouragement and comments, and M. Kellner, who did the experimental work described in Section III.

References

1. A. Eugene Anderson, "Transistors in Switching Circuits," Proc. I.R.E., vol 40, p 1541-1558, November 1952.
2. Richard F. Shea, "Transistor Operation: Stabilization of Operating Points," Proc. I.R.E., vol 40, p 1435-1437, November 1952.
3. Britton Chance et al, editors, "Waveforms," M.I.T. Radiation Laboratory Series, vol 19, p 192, McGraw-Hill Book Company, Inc., N.Y., 1949.

* Alpha cutoff frequency varied from about 0.2 to 1.4 Mc.

A SYNTHESIS PROCEDURE FOR LINEAR TRANSISTOR CIRCUITS

J. R. Burnett
Purdue University
Lafayette, Indiana

Summary

The design of transistor circuits is hindered by the loading effects demonstrated by transistors which is not usually present in vacuum tube circuits. A pure synthesis procedure takes into account these loadings. It is shown that a complete transistor can be removed at a time during a passive synthesis procedure, thus making possible an active synthesis procedure for transistors. The technique involves finding the relationship between the impedance parameters of a partitioned network and the original network. An example is shown to illustrate the technique.

Introduction

The primary purpose of looking into the synthesis of transistor circuits is to produce four terminal networks having prescribed frequency characteristics with gain. Secondary purposes of such a synthesis technique are to control impedance levels, element values, and even produce the frequency effects of RLC passive circuitry using only RC elements and transistors. Synthesis is needed since transistors are not amenable to the building block concept which has characterized the use of vacuum tubes in producing networks having prescribed frequency characteristics with gain. The disqualifying factor for transistors is that none of their admittance parameters are negligible, whereas vacuum tubes often have zero input admittance and zero transfer admittance from the output to the input terminals. This causes the loading effects of transistors.

The procedure described here makes full use of the existing passive network synthesis techniques and introduces a method of removing a transistor at a time from a ladder development.

Transistor Removal Equations

Linear transistors are characterized by four parameters which can be the impedance or admittance ones. Four are required because transistors are active networks, and none of these four can be neglected. The transistor removal equations are based upon the geometry in Fig. 1. The transistor parameters used here are based upon equations (1) and (2).

$$E_1 = I_1 z_{11} + I_2 z_{21} \quad (1)$$

$$E_2 = I_1 z_{12} + I_2 z_{22} \quad (2)$$

It is postulated that the network N which is characterized by the impedance parameters z_{11} ,

z_{12} , z_{21} , and z_{22} can be partitioned into two networks, N_a and N_b , which when connected as shown will produce the same overall impedance parameters as does N. Each of the networks N_a and N_b are also characterized by sets of impedance parameters with the appropriate subscripts. The parameters of network N may be found using the parameters of N_a and N_b .¹

$$z_{11} = \frac{z_{11a}(z_{11b} + z_{22a}) - z_{12a}z_{21a}}{z_{11b} + z_{22a}} \quad (3)$$

$$z_{22} = \frac{z_{22b}(z_{11b} + z_{22a}) - z_{12b}z_{21b}}{z_{11b} + z_{22a}} \quad (4)$$

$$z_{12} = \frac{z_{12a}z_{12b}}{z_{11b} + z_{22a}} \quad (5)$$

$$z_{21} = \frac{z_{21a}z_{21b}}{z_{11b} + z_{22a}} \quad (6)$$

In a synthesis procedure, the parameters of N would be known or assumed quantities, the parameters of N_a or N_b would be those of a transistor, and the other set of N_a or N_b would be the parameters of the network connected to the transistor. Equations (3), (4), (5), and (6) may be solved for the parameters of N_a or N_b .

$$z_{11a} = \frac{z_{11}z_{22b} - \Delta}{z_{22b} - z_{22}} \quad (7)$$

$$z_{22a} = \frac{z_{22}z_{11b} - \Delta_b}{z_{22b} - z_{22}} \quad (8)$$

$$z_{12a} = \frac{z_{12}z_{21b}}{z_{22b} - z_{22}} \quad (9)$$

$$z_{21a} = \frac{z_{21}z_{12b}}{z_{22b} - z_{22}} \quad (10)$$

$$z_{11b} = \frac{z_{11}z_{22a} - \Delta_a}{z_{11a} - z_{11}} \quad (11)$$

$$z_{22b} = \frac{z_{11a}z_{22} - \Delta}{z_{11a} - z_{11}} \quad (12)$$

$$z_{12b} = \frac{z_{12}z_{21a}}{z_{11a}-z_{11}} \quad (13)$$

$$z_{21b} = \frac{z_{21}z_{12a}}{z_{11a}-z_{11}} \quad (14)$$

Equations (7), (8), (9), and (10) yield the necessary relationships if the network N is to be realized by a passive network N_a followed by an active network N_b ; equations (11), (12), (13) and (14) yield the necessary relationships if N is to be realized by an active network N_a followed by a passive network N_b . These are not sufficient relationships that the passive network will be realizable itself. This matter is left to the ingenuity of the synthesizer in the selection of suitable transistors.

These equations or the geometry shown in Fig. 1 do not indicate that only one transistor can be removed to create a network N having a prescribed frequency characteristic with gain. Once the network N_a or N_b is removed in the form of a transistor, the parameters of the N_b or N_a can be further subdivided with due regard to realizability conditions. Ladder developments of four terminal networks are then possible as is the situation with passive network synthesis.

Example of the Synthesis Procedure

As an example to illustrate a few of the points concerned with this procedure, let the following open circuit transfer function be realized as a transistorized amplifier.

$$H(s) = K \frac{s}{(s+5)(s+0.1)} \quad (15)$$

An open circuit transfer function may be associated with the relationship

$$H(s) = \frac{z_{12}}{z_{11}} \quad (16)$$

The zeros of z_{11} may be associated with the poles of $H(s)$, and the zeros of z_{12} taken as the zeros of $H(s)$. There remains to select suitable poles of z_{11} .

The selection of these poles is also concerned with the impedance level of the N_b network if N_a is a transistor and with the realizability of the N_b network. The poles in this example are selected such that

$$z_{11} = 250 \frac{(s+0.1)(s+5)}{(s+0.09)(s+4.5)} \quad (17)$$

The selection of the multiplicative factor 250 is somewhat arbitrary, but large values quickly lead to unrealizable N_b . The synthesis will proceed upon the basis of removing transistors

by the removal equations and forming a ladder development of the passive parts which also realize the zero of the transfer impedance.

Realization

A 125 ohm resistor is first removed in series from z_{11} before a transistor is removed. This series resistor will reduce the gain but does provide for source resistance.

$$z_{11} = 125 + \frac{125s^2 + 698s + 74.4}{s^2 + 4.59s + 0.405} = 125 + z'_{11} \quad (18)$$

A transistor is next removed. The particular choice of transistor depends upon the z'_{11} . The value of z'_{11b} which must load this transistor is obtained by use of equation (11). Equation (11) can be rearranged to be as follows:

$$z'_{11b} = \frac{1}{g_{22t}} \left[\frac{-1 + g_{11t} z'_{11}}{1 - \frac{1}{r_{11t}} z'_{11}} \right] \quad (19)$$

As a necessary condition for z'_{11b} to be realizable the coefficients of the polynomials forming z'_{11b} must be all positive. Considering the numbers in z'_{11} and equation (17) and having available data of several transistor parameters, a suitable choice can be made. In this example

$$z'_{11b} = \left(\frac{1}{g_{22t}} \right) \frac{s^2(125g_{11t} + 1) + s(698g_{11t} - 4.59) + (74.4g_{11t} - 40)}{s^2(1 - \frac{1}{r_{11t}}) + s(4.59 - \frac{698}{r_{11t}}) + (0.405 - \frac{74.4}{r_{11t}})} \quad (20)$$

Clearly g_{11t} must be larger than $1/125$ mho and r_{11t} larger than 201 ohms. A large value of g_{22t} is desirable to maintain a low impedance level of z'_{11b} . An X22 grounded base which meets these specifications has the following parameters:

$$r_{11t} = 1770 \text{ ohms} \quad r_{21} = 1737 \text{ ohms}$$

$$r_{22t} = 604K \text{ ohms} \quad 1/g_{11t} = 118 \text{ ohms}$$

$$r_{12t} = 574K \text{ ohms} \quad 1/g_{22t} = 41500 \text{ ohms}$$

Using the appropriate parameter values in equation (20), there results

$$z'_{11b} = 25600 \frac{0.246s^2 + 5.93s + 1}{2.56s^2 + 11.5s + 1} \quad (21)$$

A shunt resistor must be removed from z'_{11b} to enable operating currents to reach the collector. This resistance should not be very large or else excessive power supply voltages will be required. In this example the resistance in shunt is 25,600 ohms minimum and is reasonable for a 0.5 ma operating point. The size of this resistance is governed by the $1/g_{22t}$, and the

closeness of the constant terms in the original z_{11} equation. The relative value of these constant terms is under the control of the designer through the selection of the poles of z_{11} and its multiplicative factor. The maximum resistance seen looking into the 1,1' terminals of a transistor is the r_{11} , and this results if the load is an open circuit. The minimum resistance seen looking into the 1,1' terminals is $1/g_{11}$, and this results if the load is a short circuit. Consequently, for the impedance level of the z_{11b} to be small, the impedance level of z_{11} should be about $1/g_{11t}$.

The shunt 25,600 ohm resistor is removed, followed by the removal of a series capacitor to realize the zero of z_{12} . The resulting impedance function is given by

$$z_{11b} = 25300 \frac{1 + 0.0478s}{1 + 0.414s} \quad (22)$$

At this point it is well to consider the characteristics of an impedance function which when placed in equation (11) can produce a realizable function. First and foremost, if transistors are used having positive parameters, the impedance function must have the same powers of s present in both the numerator and denominator. This restriction is evident in equation (20). If all these terms are not present, then some of the resultant coefficients will be negative. Transistors having negative parameters may be deliberately used to overcome this difficulty. This restriction is a reflection of the fact that a transistor represents a resistance to ground viewed from either set of terminals, and if it is to be removed from an impedance function, that impedance function must have a resistance to ground.

If another transistor is to be removed, it must also have a shunt resistance for furnishing bias current. If all the shunt resistance is removed from equation (22), then the resulting function will not yield a realizable loading impedance for the second transistor. Only part of it can be removed. If 50,600 ohms are removed in shunt, the resultant impedance function is

$$z_{11b} = 50600 \frac{1 + 0.0478s}{1 + 0.78s} \quad (23)$$

Again, using equation (19)

$$z_{11c} = \left(\frac{1}{s} \right) \frac{s(24209.1e^{-0.78}) + (506009.1e^{-1})}{s(0.78 \cdot \frac{2420}{n_{11t}}) + (1 - \frac{50600}{n_{11t}})} \quad (24)$$

For this transistor, g_{11t} must be larger than 3.59×10^{-5} mhos and r_{11t} must be larger than 50,600 ohms. These values suggest a grounded collector transistor. An ideal transformer could be removed at this point to lower the impedance level such that a grounded base transistor could be removed. If a TA 153 is operated grounded collector, its parameters are:

$$\begin{aligned} r_{11t} &= 800,336 \text{ ohms} & r_{21t} &= 18K \text{ ohms} \\ r_{22t} &= 18,056 \text{ ohms} & 1/g_{11t} &= 2720 \text{ ohms} \\ r_{12t} &= 800K \text{ ohms} & 1/g_{22t} &= 61.3 \text{ ohms} \end{aligned}$$

With these parameter values used in equation (24), there results

$$z_{11c} = 1150 \frac{1 + 0.00628s}{1 + 0.83s} \quad (25)$$

The impedance function z_{11c} admits of removal of a shunt resistance for bias purposes. The remainder of this impedance function may then be removed as a series resistance and finally a shunt capacitance. Figure 2 illustrates the complete circuit and the various impedances.

The TA153 operated as a grounded collector stage has a gain of slightly less than unity. This stage does serve to illustrate its use as an impedance transformer since the zero frequency resistance of z_{11b} is 50,600 ohms and the zero frequency resistance of z_{11c} is 1150 ohms. These stages allow a reduction in impedance level without appreciable loss in gain. Once the impedance level is low enough, a grounded base circuit may be used to produce gain.

Experimental Verification

The analytic solution of the circuit in Fig. 2 yields

$$z_{11} = 250 \frac{(s+0.1)(s+5)}{(s+0.09)(s+4.5)} \quad (26)$$

$$z_{12} = \frac{5.0 \times 10^4 s}{(s+0.09)(s+4.5)} \quad (27)$$

$$\frac{E_2}{E_1} = \frac{200}{(s+0.1)(s+5)} \quad (28)$$

Figure 3 is an experimental gain frequency curve of this amplifier. The frequency has been denormalized to facilitate testing. The break frequencies are 114 and 5710 cycles per second. The largest error is about 1/2 of a decibel. Discrepancies can be explained by element tolerances and frequency characteristics of the transistors which have been neglected.

Pole Zero Configurations

Equation (19) might be interpreted as computing the pole zero locations of a network to be placed in tandem with a transistor such that the loading effects of the transistor shift these poles and zeros to where the specifications demand they be. The pole zero configuration of the specifications, z_{11} , are shown in Fig. 4. The selection of the poles is almost arbitrary, but they were chosen close to the zeros to

produce two polynomials defining the impedance function without a large spread in coefficient values between corresponding powers of the numerator and denominator. The poles in this example were also chosen to produce an RC driving point impedance. The poles and zeros of the loading impedance for the first transistor are also shown in Fig. 4. Figure 4 may be interpreted as a root locus plot. The form of equation (19) suggests this concept. With the minus signs, the poles and zeros move in opposite directions to those encountered in servomechanisms. The gain factors are the g_{11t} and $1/r_{11t}$ for the zeros and poles respectively. The root locus of the effects of the second transistor can be shown similarly. This is illustrated in Fig. 5. Transistors having certain negative parameters may be used to cause the root locus plots to go in opposite directions. The example shown here has caused the poles to move to the right and the zeros to move to the left. With the opposite motion, RC networks can be transformed into RL networks, and vice versa.

Other Uses of Equations

The equations in the series (11), (12), (13), and (14) which were not used have little bearing upon this example. Equations (13) and (14) determine the transfer impedances of the network N_b . Equation (12) determines the z_{22} of N_b . The importance of these equations is that they show the poles of the impedance parameters include at least those in the transfer impedances. This is the justification of realizing a four terminal network by means of a ladder development of one of the driving point impedances.

With the specifications of a required transistor amplifier, the impedance parameters of the network may be determined by any of the methods used in passive network theory. As an example, if an amplifier is required which is terminated

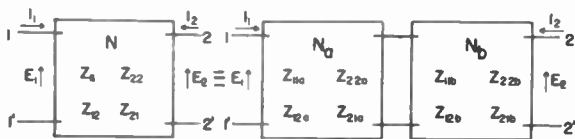


Fig. 1 - Block diagram of the partitioning.

in a resistance, the transfer function may be considered to be y_{12} divided by y_{22} plus the terminating resistance. The y_{22} thus obtained can then be developed in a ladder network in a manner similar to that shown in the numerical example.

Conclusion

A major difficulty associated with this procedure is the transistors. The parameters vary considerably with operating point and the characteristics of the network are likely to be sensitive to these parameter changes. Also, it is necessary to measure the parameters of the specific transistors contemplated for these circuits as the tolerances between transistors of the same type number are much too great. Effective transistors having parameters not available in single transistors can often be obtained by the connection of two or more transistors. In spite of these difficulties, the synthesis procedure has yielded a logical approach to the realization of transistorized amplifiers.

The material in this paper was carried out as part of contract No. D.A. 36-039 sc-15544 between the Signal Corps and Purdue Research Foundation.

References

1. "Transistor Circuitry Research Fourth Quarterly Progress Report," Purdue Research Foundation, April 1953, Lafayette, Indiana. Contract No. DA-36-039 sc-15544.
2. "The Synthesis of R-C Networks." E. A. Guillemin, Jour. Math. & Physics, April 1949.
3. "Control System Synthesis by Root Locus Method," W.R. Evans, Trans. AIEE 1950.

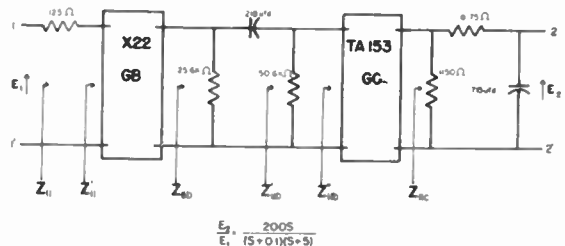


Fig. 2 - Synthesized amplifier.

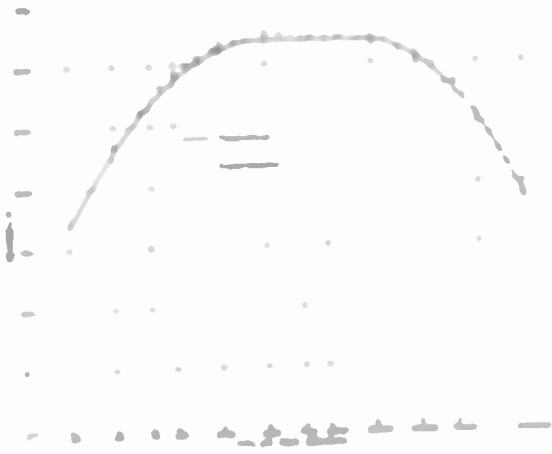


Fig. 3
Frequency response of the
synthesized amplifier.

Fig. 4
Root locus plot based by 127 parameter.



Fig. 5
Root locus plot based by TA151 translator.

NETWORK PARTITIONING TECHNIQUES APPLIED TO THE SYNTHESIS OF TRANSISTOR AMPLIFIERS

H. Markarian
School of Electrical Engineering
Purdue University
Lafayette, Indiana

Summary

Two methods for applying network partitioning techniques to the synthesis of transistor amplifiers are described. The syntheses of a differentiating and a Butterworth amplifier serve to illustrate and to bring out some of the advantages of these methods. In the case of the differentiating amplifier a passive network realizing a given transfer function is synthesized first. This is then partitioned, and each part resynthesized using transistors and passive elements. In the second example the partitioning is achieved by mathematical manipulation; and through a proper choice of transistor connections complex poles of the transfer function are realized without using inductors.

Introduction

When a network N_a with short circuit parameters y_{12a} and y_{22a} and a second network N_b with open circuit parameters z_{12b} and z_{11b} are connected in tandem as indicated in Fig. 1, the overall open circuit voltage transfer ratio is given by the following relation:

$$\left. \frac{E_2}{E_1} \right|_{I_2=0} = \frac{(-y_{12a}) K_b}{y_{22a} + \frac{1}{z_{11b}}} \quad (1)$$

where

$$K_b = \frac{z_{12b}}{z_{11b}} \quad (2)$$

This is merely an extension of the relation given by the Thevenin's Theorem and can be easily derived by taking the two-terminal pair network N_b to be the load of N_a . The convention concerning the positive direction of current at the terminals of a two-terminal pair network is responsible for the negative sign associated with y_{12a} in (1).

Application of Partitioning

The usefulness of the partitioning formula, as (1) may be so called, lies in the fact that it relates the overall voltage transfer ratio of a network to the indicated parameters of its two constituent two-terminal pair sections. For a given overall voltage transfer function it is thus possible to compile by the use of this relation a list of network pairs satisfying the specification. The most basic application of partitioning to the synthesis of transistor amplifiers consists of finding in such a list a network pair, one member of which can be identified as a transistor. The companion member in the pair can

then be realized using passive elements and put in tandem with the transistor to give the specified open circuit voltage transfer ratio within a multiplicative constant. Depending on whether the transistor takes the place of N_a or N_b , this procedure can be described as the removal of a transistor by partitioning starting from the sending or the receiving ends respectively. While the use of a list of acceptable network pairs helps to illustrate, its compilation is in practice unnecessary. With a knowledge of the parameters of the transistor to be used and some idea of the desired topology of the final network the designer can arrive at a satisfactory network pair after only few trials.

The merits of partitioning as a design tool become apparent when the transfer functions to be realized are complicated or when the amplifier is to have more than one transistor. This is obvious, for example, in the case of an amplifier with two transistors. Partitioning in such instances makes things easier by providing immediate access to the interstage coupling network.

Even though partitioning is generally carried out mathematically on the given open circuit voltage transfer function, it is found advantageous to follow a somewhat different course for its present application. The latter procedure involves the following steps.

1. A number of passive networks satisfying the given open circuit voltage transfer function are synthesized.
2. By inspection the one that looks the most favorable is chosen.
3. The chosen network is partitioned into N_a and N_b . Parameters y_{12a} , y_{22a} , z_{11b} and z_{12b} are determined.
4. N_a and N_b are resynthesized according to these parameters using transistors and passive elements.

An example that illustrates this procedure follows. In a second example, partitioning is done mathematically at the outset.

First Example

Let it be required to synthesize a differentiating transistor amplifier that realizes the following open circuit voltage transfer function:

$$\frac{E_2}{E_1} = \frac{As}{(s+2)(s+4)} \quad (3)$$

Multiplying the numerator and denominator of (3) by

$$\frac{1}{C(s+3)} \quad (4)$$

gives

$$\frac{E_2}{E_1} = \frac{\frac{As}{C(s+3)}}{\frac{(s+2)(s+4)}{C(s+3)}} \quad (5)$$

Considering that

$$\left. \frac{E_2}{E_1} \right|_{I_2=0} = \frac{z_{12}}{z_{11}} \quad (6)$$

$$= - \frac{y_{12}}{y_{22}} \quad (7)$$

the admittance functions specifying the passive network are

$$y_{12} = - \frac{As}{C(s+3)} \quad (8)$$

$$y_{22} = \frac{(s+2)(s+4)}{C(s+3)} \quad (9)$$

Figure 2 shows a passive network that realizes (8) and (9). Some experience at this point will help to see that it is possible starting at the dotted line to proceed forward and backward removing two transistors, one on each side. Therefore, this passive network can be considered to be satisfactory enough to serve as the framework for the amplifier design. The following parameters are determined for the partitioned network:

$$y_{12a} = - \frac{1/3 s}{C(s+3)} \quad (10)$$

$$y_{22a} = \frac{1/3 s}{C(s+3)} \quad (11)$$

$$z_{11b} = z_{12b} = \frac{C}{s + 8/3} \quad (12)$$

The correctness of (10), (11) and (12) can be verified by substituting them into (1), which gives

$$\frac{E_2}{E_1} = \frac{1/3 s}{(s+2)(s+4)} \quad (13)$$

It is to be pointed out that the impedance level of the final network as well as its topo-

logy depends on the choice of (4). Reverting momentarily to the idea of a list of acceptable network pairs, it is possible to think of (4) as some kind of an independent variable which the contents of the list are a function of. Thus, if after deciding upon a network pair the transistor amplifier does not turn out to be entirely satisfactory from practical considerations, it is necessary to try a different passive network pair corresponding to a different value of the independent variable (4). This may be accomplished by just changing the pole location, or it may require the introduction of additional factors in the numerator and denominator of (4). The latter artifice amounts to the introduction of redundant elements in the network, which may be required to provide otherwise lacking biasing paths.

To proceed with the synthesis, reference is made to Fig. 3, which gives not only the final result but also serves as a legend for the notation used in the calculations that follow. The section taken up first for transistorization is N_b , although it could just as well have been N_a . Recalling that

$$z_{11b} = \frac{C}{s + 8/3} \quad (12)$$

a shunt capacitor of $1/C$ farads is removed. This leaves to be realized

$$z'_{11b} = \frac{3C}{8} \quad (14)$$

For this section a grounded emitter connected junction type transistor, J #2, is chosen with the following open circuit parameters:

$$\begin{aligned} r_{11} &= 431.8 \ \Omega \\ r_{21} &= 31.8 \ \Omega \\ r_{12} &= -719.3 \ K \\ r_{22} &= 20.7 \ K \end{aligned} \quad (15)$$

If a load of $1000 \ \Omega$ is taken, substitution into (46) gives

$$z'_{11b} = 1486 \ \Omega \quad (16)$$

This is in parallel with the $1000 \ \Omega$ resistor providing the bias path to the input of the transistor. Now, invoking (14)

$$z'_{11b} = \frac{3C}{8} = \frac{1000 \times 1486}{1000 + 1486} \quad (17)$$

Therefore,

$$C = 1595 \quad (18)$$

The synthesis of N_b is thus completed. The value of C that has been established governs the synthesis of N_a . Substituting this value into (11) gives

$$y_{22a} = \frac{s}{3 \times 1595 (s+3)} \quad (19)$$

The removal of a series capacitor of a magnitude of $\frac{1}{9 \times 1595}$ F constitutes the first step, and leaves to be realized

$$y'_{22a} = \frac{1}{3 \times 1595} \quad (20)$$

The transistor chosen for N_a , J #1, is of the junction type with its emitter grounded, and has the following open circuit parameters:

$$\begin{aligned} r_{11} &= 428.7 \ \Omega \\ r_{21} &= 27.7 \ \Omega \\ r_{12} &= -840 \ \text{K} \\ r_{22} &= 10.2 \ \text{K} \end{aligned} \quad (21)$$

Making the source impedance equal to 600 ohms and substituting this value with (21) into (47) gives after inversion

$$y''_{22a} = \frac{1}{32800} \quad (22)$$

As this value is lower than what is required in (20), it is necessary to hang a resistor, R, across the output of the transistor J #1. The value of this is

$$R = \frac{1}{y'_{22a} - y''_{22a}} = 5600 \ \Omega \quad (23)$$

The synthesis of the differentiating transistor amplifier is completed; only the calculation of its gain remains. The simplest way of doing this is to determine y_{12a} and z_{12b} actually realized and to substitute these together with y_{22a} and z_{11b} into (1). These are computed to be

$$y_{12a} = \frac{s}{40.4(s+3)} \quad (24)$$

$$z_{12b} = -\frac{356000}{(s+2.567)} \quad (25)$$

and

$$\left. \frac{E_2}{E_1} \right|_{I_2=0} = \frac{882s}{(s+2)(s+4)} \quad (26)$$

Second Example

The synthesis of a Butterworth amplifier will be carried out using only resistors and capacitors. To do this, use is made of a quality of grounded emitter stages that exhibit a RL impedance at one end when a RC network is connected at the other end. Similarly, a RL network is transformed into a RC network. Using (46) and (47) it is very easy to prove this.

The open circuit system function of the amplifier to be synthesized is

$$\left. \frac{E_2}{E_1} \right|_{I_2=0} = \frac{A}{s^3 + 2s^2 + 2s + 1} \quad (27)$$

Multiplying the numerator and denominator by

$$\frac{1}{C(s+0.5)(s+1.05)} \quad (28)$$

gives

$$\frac{E_2}{E_1} = \frac{A}{\frac{C(s+0.5)(s+1.05)}{s^3 + 2s^2 + 2s + 1}}, \quad (29)$$

where the denominator now is

$$D(s) = \frac{s^3 + 2s^2 + 2s + 1}{C(s+0.5)(s+1.05)} \quad (30)$$

This is first expanded in a partial fraction and then its terms regrouped.

$$\begin{aligned} D(s) &= \frac{1}{C} \left(s + 0.45 + \frac{0.0957}{s+1.05} + \frac{0.6818}{s+0.5} \right) \\ &= \frac{1}{C} \left(s+0.05 + \frac{0.0957}{s+1.05} \right) + \frac{1}{C} \left(0.40 + \frac{0.6818}{s+0.5} \right) \\ &= \frac{1}{C} \left(s+0.05 \frac{s+2.964}{s+1.05} \right) + \frac{0.4}{C} \left(\frac{s+2.205}{s+0.5} \right) \end{aligned} \quad (31)$$

The expanded $D(s)$ is now substituted back into (29), obtaining

$$\frac{E_2}{E_1} = \frac{A}{\frac{C(s+0.5)(s+1.05)}{\frac{1}{C} \left(s+0.05 \frac{s+2.964}{s+1.05} \right) + \frac{0.4}{C} \left(\frac{s+2.205}{s+0.5} \right)}} \quad (32)$$

Comparing this with (1), the following term associations can be made.

$$\left. \begin{aligned} y_{12a} &= \frac{-A_1}{C(s+1.05)} \\ y_{22a} &= \frac{1}{C} \left(s+0.05 \frac{s+2.964}{s+1.05} \right) \\ K_b &= \frac{A_2}{s+0.5} \\ \frac{1}{z_{11b}} &= \frac{0.4}{C} \left(\frac{s+2.205}{s+0.5} \right) \end{aligned} \right\} \quad (33)$$

All the information needed for the synthesis of N_a and N_b is contained in (33) and (34). Even though not essential, it is helpful at this point to synthesize (33) and (34) into a passive net-

work. Reference to Fig. 4 showing the synthesized network clearly reveals the course to be followed.

Starting at the dotted line and proceeding right, a shunt resistor of 25C ohms magnitude is removed leaving

$$z'_{11b} = \frac{C}{0.36} \left(\frac{s+0.5}{s+2.394} \right). \quad (35)$$

To the left of the dotted line a shunt capacitor of $1/C$ farads is removed leaving

$$y'_{22a} = \frac{0.05}{C} \left(\frac{s+2.964}{s+1.05} \right). \quad (36)$$

These are both shown in Fig. 5. The syntheses of N_a and N_b have now progressed to points where the removal of a transistor from each is in order.

The transistor chosen for N_a , J #3, is of the junction type and is operated with its emitter grounded through a resistor of 200 ohms. The purpose of the resistor is to improve the stability of the circuit. The short circuit parameters of this stage are:

$$\begin{aligned} g_{11} &= 58.4 \times 10^{-6} \quad \Omega \\ g_{21} &= -3.92 \times 10^{-6} \quad \Omega \\ g_{12} &= 4.15 \times 10^{-3} \quad \Omega \\ g_{22} &= 14.5 \times 10^{-6} \quad \Omega \end{aligned} \quad (37)$$

The transistor in N_b , J #4, is also operated with its emitter grounded through a resistor of 200 ohms, and is again of the junction type. The open circuit parameters of this stage are:

$$\begin{aligned} r_{11} &= 422.4 \quad \Omega \\ r_{21} &= 218.4 \quad \Omega \\ r_{12} &= -552 \text{ K} \quad \Omega \\ r_{22} &= 8 \text{ K} \quad \Omega \end{aligned} \quad (38)$$

To remove the transistor J #3 from y'_{22a} , (36) and (37) are substituted into (52). The admittance remaining after the removal of the transistor simplifies to

$$y''_{22a} = 58.4 \times 10^{-6} \left[\frac{(50,000 - 293C)s + (148,000 - 307.2C)}{(14.5C - 50,000)s + (15.2C - 148,000)} \right], \quad (39)$$

which is physically realizable within the range,

$$485 \leq C \leq 3450. \quad (40)$$

On the other hand, the transistor J #4 can be removed from z'_{11b} by substituting (35) and (38) into (50), the remaining impedance simplifying to

$$z''_{11b} = 123.5 \left[\frac{(64.6C - 360,000)s + (32.3C - 865,000)}{(152 - C)s + (365 - 0.5C)} \right]. \quad (41)$$

This is physically realizable within the range

$$730 \leq C \leq 5570 \quad (42)$$

The ranges of the constant multiplier C for y''_{22a} and z''_{11b} respectively happen to be partially overlapping. Had this not been the case it would have been necessary to try different transistors or go back and change the partitioning.

Choosing a value of C common to both ranges,

$$C = 3450, \quad (43)$$

and substituting this into (39) and (41) gives

$$y''_{22a} = \frac{586}{10^6} s + \frac{1}{1790}, \quad (44)$$

and

$$z''_{11b} = 5115 + \frac{1}{\frac{38.5}{10^6} s + \frac{1}{62900}}. \quad (45)$$

This completes the synthesis of three pole Butterworth amplifier using resistors, capacitors and two transistors. The gain A in (27) is calculated to be 1550.

Conclusion

Partitioning has been described and two examples given illustrating its application to the design of transistor amplifiers. The representation has not aimed at comprehensiveness, its aim being to point at some of the possibilities of the method. The fact that C_c and the variation of α with frequency have been neglected should not be taken to demerit partitioning. It can just as effectively be used to provide compensation or to synthesize networks utilizing the transistor characteristics to realize specified system functions.

Acknowledgment

This paper is based on research done under the auspices of Purdue Research Foundation Project 820, sponsored by U.S. Army Signal Corps Engineering Laboratories, Fort Monmouth, New Jersey, through Contract DA 36-039 sc-15544, Project 132B.

Appendix

Following is a list of formulas referred to in the text. The order in which N_a and N_b are to be taken together with the conventional positive directions of voltages and currents are indicated in Fig. 1.

$$z_{11} = z_{11a} - \frac{z_{12a} z_{21a}}{z_{22a} + z_{11b}} \quad (46)$$

$$z_{22} = z_{22b} - \frac{z_{12b} z_{21b}}{z_{22a} + z_{11b}} \quad (47)$$

$$z_{12} = \frac{z_{12a} z_{12b}}{z_{22a} + z_{11b}} \quad (48)$$

$$z_{21} = \frac{z_{21a} z_{21b}}{z_{22a} + z_{11b}} \quad (49)$$

$$z_{11b} = -z_{22a} + \frac{z_{12a} z_{21a}}{z_{11a} - z_{11}} \quad (50)$$

$$z_{12b} = z_{12} \frac{z_{21a}}{z_{11a} - z_{11}} \quad (51)$$

$$y_{22a} = \frac{y_{12b} y_{21b}}{y_{22b} - y_{22}} - y_{11b} \quad (52)$$

$$y_{12a} = -y_{12} \frac{y_{21b}}{y_{22b} - y_{22}} \quad (53)$$

References

- L. Weinberg, "New Synthesis Procedures for Realizing Transfer Functions of RLC and RC Networks," Technical Report No. 201 at MIT Research Laboratory of Electronics, 1951.
- E.A. Guillemin, "Synthesis of RC-Networks," Journal of Mathematics and Physics, vol. 28, April 1949.
- J.R. Burnett, "A Synthesis Procedure for Linear Transistor Circuits," National Convention of the Institute of Radio Engineers, 1954.



Fig. 1

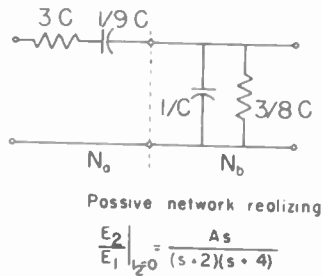


Fig. 2

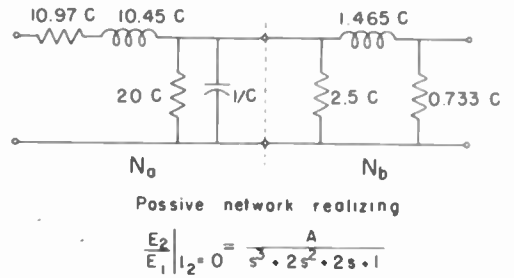


Fig. 4

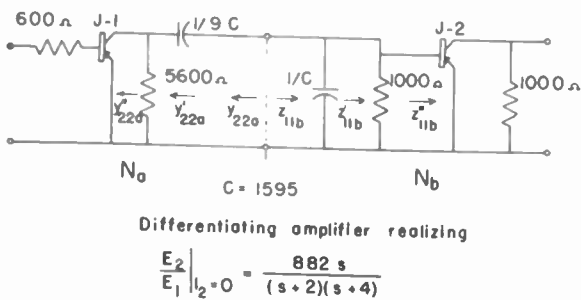


Fig. 3

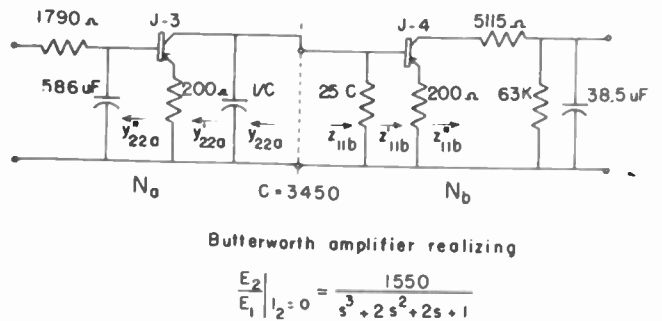


Fig. 5

A NEW EQUIVALENT CIRCUIT FOR JUNCTION TRANSISTORS

Ge Yao Chu
Sylvania Electric Products, Inc.
Electronics Division
Ipswich, Massachusetts

Abstract

Based upon the work of Shockley and Early on junction transistors, a new equivalent circuit consisting essentially of two transmission lines is presented. One line is used in the emitter circuit to represent Shockley's forward diffusion process of minority carriers in the base region and the other line used in the collector circuit represents the Early effect which feeds the collector signal back to the emitter circuit.

Approximate equivalent circuits for grounded base and for grounded emitter connections are also derived from this new circuit and are compared to conventionally established equivalent circuits. Thus a unified view is established.

Introduction

An equivalent circuit is useful to analyze the function of a device because (a) it allows a clearer insight into the operation of the device, and (b) it offers an analytical solution by well developed circuit theory. The first equivalent circuit for junction transistors was introduced by Shockley, et alii, for low frequency operation¹. Essentially, it consists of an equivalent current generator, known as the α generator, which responds to a signal current in the emitter (input) circuit and transfers this signal current into the collector (output) circuit. Since the impedance of the output circuit is very high and the input impedance is very low, the junction transistor will yield power gain as an amplifier. This type of operation is known as a grounded base application.

If the signal frequency is increased, the transit time of transferring charge carriers from the emitter to the collector in the base region becomes increasingly appreciable. The charge transfer efficiency will drop, thus the transfer current of the transistor will decrease. This charge transfer process takes place by diffusion, and is analogous to the charge flow on a transmission line. Therefore, equivalent circuits using a transmission line have been suggested. Yet, there is no equivalent circuit that gives a lucid analog to the established theory of the junction transistor. This paper attempts to fill this gap. The basic design theory for junction transistors has been published recently by Early³. This paper will use some of his results.

A "Two Transmission Line" Circuit

A new equivalent circuit for a one dimensional junction transistor is shown in Figure 1. The essential part of this circuit consists of two transmission lines shown by solid lines. One of the lines, connected primarily in the emitter circuit of the transistor and coupled unilaterally to the collector circuit, represents Shockley's forward diffusion process of the minority carriers in the base region^{1,2}. The other line, connected essentially in the collector circuit and coupled unilaterally back to the emitter circuit, represents the feedback effect due to space charge layer widening discovered by Early². Other behaviors of the transistors are represented by lumped circuit elements shown by the dotted lines. Among them, C_{Te} and C_{Tc} are the emitter and collector barrier capacitances respectively. G'_e is a lumped conductance representing an equivalent leakage in the emitter barrier due to imperfect injection efficiency of the emitter. Similarly G'_c is a conductance representing the leakage (e.g. due to the surface) between the collector and the base of a practical transistor. Between the internal base and the outer base terminal of the transistor is a lumped base resistance r'_b . Under normal conditions, G'_e and C_{Te} can generally be neglected and C_{Tc} , G'_c , and r'_b may not remain constant over wide frequency ranges. More exact representation of these parameters in an equivalent circuit is beyond the scope of this paper.

Diffusion Process

A one dimensional pnp junction transistor is schematically shown in Figure 2(a). Under a normal bias condition, the emitter injects holes into the base region. The excess hole density in the base region near the emitter end will cause diffusion of the holes toward the collector end. Holes diffused will be collected by the collector, since the latter is properly biased to receive holes. A small fraction of the holes will disappear during diffusion in the base region due to recombination. The diffusion and recombination processes can be described analytically by the continuity equation^{1,2} for a one dimensional case as follows:

$$\frac{\partial p'}{\partial t} = -\frac{p'}{\tau} + D \frac{\partial^2 p'}{\partial x^2} \quad (1)$$

where p' is the excess hole density of the n-type base region (which is equal to the difference between the hole density p and the thermal equilibrium hole density p_n); τ is the mean lifetime of holes; and D is the diffusion constant of

holes in the n region.

The boundary condition for p' at the emitter end ($x=0$) depends upon the injection level of the emitter. At the other end ($x=W_0$), because the electric field intensity in the collector-base barrier is high, any hole in that region will be accelerated toward the collector with high speed, and there will be no appreciable hole density in that region. Thus the boundary condition for p' at $x=W_0$ may be taken as zero. The solution of (1) with a normal d.c. bias condition will show a space distribution of p' such as the one shown in Figure 2(b). Because of this sloping characteristic, there is a hole density gradient in the x direction, and the diffusion of the holes in the forward direction takes place. The diffusion current I_p is related to the hole density gradient by:

$$I_p = -qD \frac{\partial p'}{\partial x} \quad (2)$$

Physical Analogy

These relations are analogous to the flow of charges in a transmission line. In fact, the differential equation to describe an RC transmission line has exactly the same form as (1). Transmission line theory⁵ gives the charge density on the line as

$$\frac{\partial Q}{\partial t} = -\frac{g}{c} Q + \frac{1}{r_0} \frac{\partial^2 Q}{\partial x^2} \quad (3)$$

where Q is the linear charge density of the line (i.e. charge per unit length), r is the series resistance per unit length of the line, c and g are the shunt capacitance and conductance per unit length of the line. By comparing (1) and (3), one obtains a complete analogy between the diffusion and recombination processes in the base region of the transistor and the RC transmission line as listed in Table 1.

The analogy shown in Table 1 suggests that the forward diffusion process of holes in the base region of the junction transistor can be represented by one RC transmission line (a) with the receiving end short circuited, but (b) with an output impedance corresponding to an open circuit. A short circuit transmission line followed by an equivalent current generator is, therefore, used for the forward transmission process as shown by the Shockley line in Figure 1.

Effect of Space Charge Layer Widening

As Early pointed out⁴, a variation of collector voltage V_0 will cause a variation of the collector-base barrier thickness x_m . This in turn will cause a variation of the base width W_0 . The variation of the base width will correspond to a short circuit position at the receiving end of the Shockley line, shifted back and forth according to the variation of V_0 . For a fixed input voltage on the line, this movement of the short circuit point will call for an in-

crease or decrease of the input current and thus a signal current corresponding to the variation of output voltage will appear at the input end. The dynamic characteristics of this feedback process is described by the same transmission line equation (Eq. (3)) in the case of a line and by the same continuity equation (Eq. (1)) in the case of a transistor. Therefore, this type of feedback process in the transistor due to the space charge layer widening can be represented by another RC transmission line which is coupled unilaterally from the collector side back to the emitter circuit as shown by the Early line in Figure 1. It will be shown later that the two lines are of the same length and the same propagation constant but different impedance levels.

Experimental Verification

Measured results of the forward current amplification factor α over a wide frequency range of junction transistors has verified the Shockley line in the equivalent circuit. Measured results of Early effect and variation of collector capacitance over wide ranges of frequencies and emitter bias currents can be explained by the Early line. These experimental results have been published elsewhere.

Analytical Proof

Following Shockley's diffusion theory and including the space charge layer widening effect, Early has obtained a set of theoretical expressions for the short circuit admittance of the above one dimensional junction transistors³. A slight alternative version of them including only the principal part of the transistor (called the elementary transistor) may be written as follows:

$$\begin{aligned} y_{ee} &= \frac{qI_0(1+j\omega\tau)^{\frac{1}{2}}}{kT \coth(W_0/L)} \coth \frac{W_0}{L}(1+j\omega\tau)^{\frac{1}{2}} \\ y_{ce} &= -\frac{qI_0(1+j\omega\tau)^{\frac{1}{2}}}{kT \coth(W_0/L)} \operatorname{csch} \frac{W_0}{L}(1+j\omega\tau)^{\frac{1}{2}} \\ y_{cc} &= \frac{I_c}{L} \left(-\frac{\partial W}{\partial V_c} \right) (1+j\omega\tau)^{\frac{1}{2}} \coth \frac{W_0}{L}(1+j\omega\tau)^{\frac{1}{2}} \\ y_{ec} &= -\frac{I_c}{L} \left(-\frac{\partial W}{\partial V_c} \right) (1+j\omega\tau)^{\frac{1}{2}} \operatorname{csch} \frac{W_0}{L}(1+j\omega\tau)^{\frac{1}{2}} \end{aligned} \quad (4)$$

where, y_{ee} , y_{ce} , y_{cc} , and y_{ec} are the four elements of the short circuit admittance matrix $[y]$ of the elementary transistor. The intrinsic current amplification factor α^* of the collector junction is assumed equal to unity.

In order to prove that the new equivalent circuit in Figure 1 is a correct representation of such an elementary transistor, it is now only necessary to show that a set of short circuit admittance matrix (called $[Y]$) of the solid line circuit in Figure 1 is equal to the $[y]$ matrix of Equation (4).

Applying transmission line theory, one can immediately find the elements of the $[Y]$ matrix as follows:

$$Y_{11} = \frac{i_1}{v_1} = \frac{i_1'}{v_1} = \frac{1}{Z_1} \coth \theta_1$$

$$Y_{21} = \frac{i_2}{v_1} = \frac{-i_2'}{v_1} = -\frac{1}{Z_1} \operatorname{csch} \theta_1 \quad (5)$$

$$Y_{22} = \frac{i_2}{v_2} = \frac{i_2''}{v_2} = \frac{1}{Z_2} \coth \theta_2$$

$$Y_{12} = \frac{i_1}{v_2} = \frac{-i_1''}{v_2} = -\frac{1}{Z_2} \operatorname{csch} \theta_2$$

Where $Z_1 = (z_1/y_1)^{1/2}$ and $Z_2 = (z_2/y_2)^{1/2}$ are the characteristic impedances of the Shockley and Early lines respectively, $\theta_1 = l_1(y_1 z_1)^{1/2}$ and $\theta_2 = l_2(y_2 z_2)^{1/2}$ are the overall propagation functions of the two lines. As,

$$y_1 = g_1 + j\omega c_1$$

$$y_2 = g_2 + j\omega c_2 \quad (6)$$

$$z_1 = r_1$$

$$z_2 = r_2$$

one obtains:

$$\theta_1 = l_1 [r_1 g_1 (1 + j\omega c_1 / g_1)]^{1/2}$$

$$\theta_2 = l_2 [r_2 g_2 (1 + j\omega c_2 / g_2)]^{1/2} \quad (7)$$

$$Z_1 = \frac{(r_1 / g_1)^{1/2}}{(1 + j\omega c_1 / g_1)^{1/2}}$$

$$Z_2 = \frac{(r_2 / g_2)^{1/2}}{(1 + j\omega c_2 / g_2)^{1/2}}$$

The coefficients of the [Y] matrix do conform to those of the [y] matrix. By identifying the corresponding terms, the relations in Table 2 are obtained.

Both lines are of the RC type with the same equivalent length with an equal propagation function, but with different characteristic impedances. The forward current amplification factor α of the transistor, which is particularly of general interest, equals the product of the injection efficiency γ and the transfer ratio β . Since only the latter is of concern here, one can find from Figure 1

$$\beta = \frac{Y_{21}}{Y_{11}} = \frac{-i_2'}{i_1} = -\operatorname{sech} \theta = -\operatorname{sech} \frac{\omega_0}{L} (1 + j\omega \tau)^{1/2} \quad (8)$$

Evidently β has also been correctly represented in this new circuit.

Approximate Equivalent Circuit for Grounded Base Connection

Replacing the two lines in the new equivalent

circuit with two appropriate RC filters, one obtains an approximate equivalent circuit for grounded base application. This circuit may be drawn in "T" form as shown in Figure 3.

In Figure 3, the distributed parameters of each line have been essentially lumped into a series resistance, a shunt capacitance and a shunt conductance. Besides, the latter shunt elements can be combined with C_{Te} , C_{To} , G_e' and G_o' to give C_e , C_o , G_e , and G_o as shown.

Again, the i_{1e} current generator in a conventional T circuit is of general interest. We can easily prove that the equivalent current generator i_2' shown in this T circuit is equal to the former as follows:

$$i_2' = \frac{i_1'}{1 + j\omega R_1 C_e} = \frac{i_1' / i_{1e}}{1 + j\omega R_1 C_e} i_{1e} = \frac{\alpha_o}{1 + j\omega R_1 C_e} i_{1e} \quad (9)$$

where $R_1 C_e = r_e C_e = 1 / (2\pi f_c)$ and when $v_2 = i_1'' = 0$.

The current generator i_1'' in the emitter circuit is an Early feedback generator. It may be converted into an equivalent voltage generator μV_2 if so desired. C_e and C_o may be identified as the total emitter and collector barrier capacitance. G_e and G_o may be taken as the total equivalent barrier leakage conductance. R_1 corresponds to the theoretical low frequency emitter resistance r_e , while R_2 is a similar equivalent resistance in the collector circuit due to the Early effect. These parameters are related to the physical constants of the transistor and to the circuit parameters of the conventional equivalent T circuit as shown in Table 2. Moreover, these parameters will have further significance in the grounded emitter circuit described below.

Approximate Equivalent Circuit for Grounded Emitter Connection

The T circuit is a basic equivalent circuit for junction transistors. Although it is most conveniently used for grounded base application, it can be used for grounded emitter application too. A π circuit in a ladder structure can be obtained which offers not only a simpler analysis in grounded emitter applications but also leads to some physical interpretations. This ladder circuit which has been known elsewhere is shown in Figure 4(a).

In Figure 4(a), all the elements have been expressed in terms of those parameters of the original T circuit. The form of this circuit suggests that grounded emitter junction transistors are analogous to vacuum tubes at high frequencies. Accordingly, a block diagram interpretation of this circuit is given in Figure 4(b).

This diagram illustrates that functionally the grounded emitter junction transistor operates as a system of three parts: a low pass RC filter, a wide band amplifier and an RC feedback network.

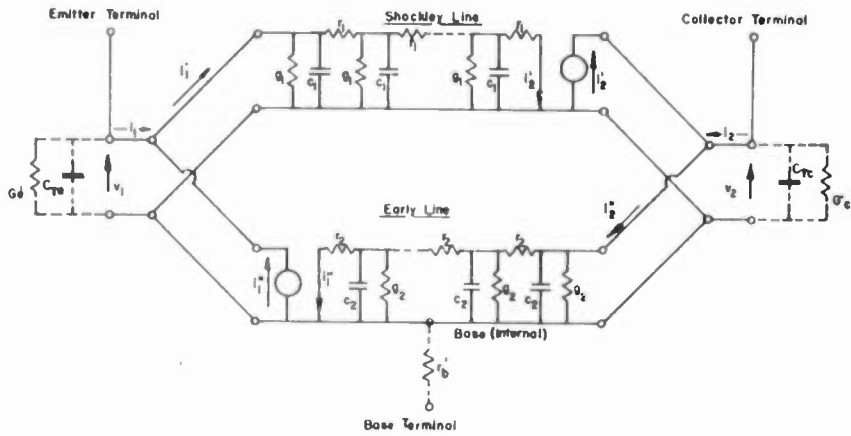


Fig. 1
A "two transmission line" equivalent circuit for junction transistors.

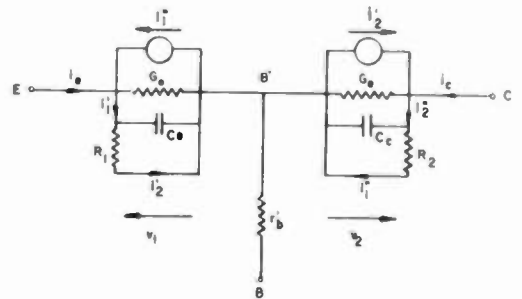


Fig. 3
An approximate equivalent circuit for grounded base applications.

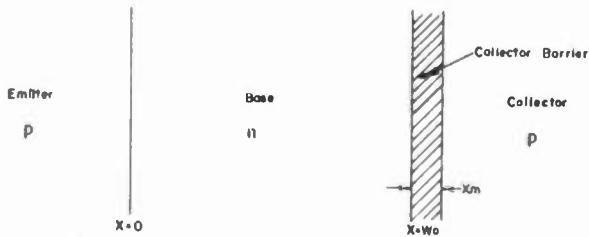


Fig. 2(a)
Schematic diagram of a p-n-p transistor.

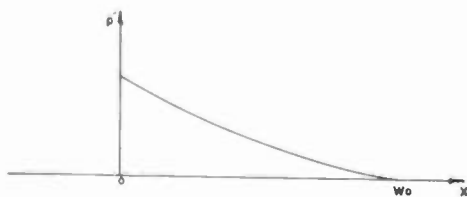


Fig. 2(b)
(Excess) hole density distribution along the base width under normal bias condition.

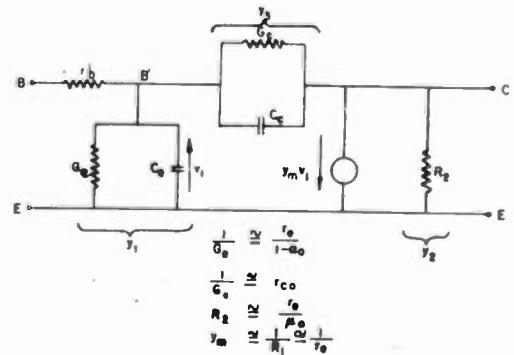


Fig. 4(a)
A ladder equivalent circuit for grounded emitter applications.

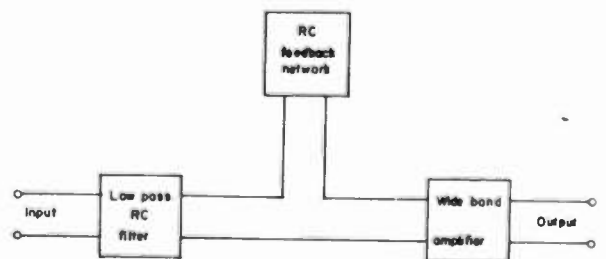


Fig. 4(b)
A block diagram interpretation of the grounded emitter junction transistor.

