PROCEEDINGS

of the

INSTITUTE ON

MODERN SOLID STATE CIRCUIT DESIGN

at

THE UNIVERSITY OF SANTA CLARA

SEPTEMBER 15-16, 1966

Co-sponsored by the

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

ELECTRICAL ENGINEERING DEPARTMENT
THE UNIVERSITY OF SANTA CLARA * CALIFORNIA

PREFACE

The analysis, design and synthesis of solid state circuits with the aid of computers is now practical. A two day institute was held at the University of Santa Clara on September 15-16, 1966 in order to present the approaches to modern solid-state circuit design. Topics which were discussed in the institute included the modelling and simulation of active devices, design implementation by computer, and the challenges of computer aided design. The institute provided a contempory understanding of the approaches to computer aided design as well as providing an opportunity for the attendees and the faculty to exchange ideas informally. The success of the first institute warranted the scheduling of a Second Institute on Modern Solid State Circuit Design which will be held on September 14-15, 1967 at the University of Santa Clara.

It is with great pleasure that I am able to acknowledge the cooperation of the National Aeronautics and Space Administration. We are particularly indebted to Dr. William Happ of the Electronics Research Center and Mr. Volta Torrey of the Technology Utilization Division of NASA Headquarters. Also, may I acknowledge the excellent assistance provided by the University and especially the encouragement provided by Dr. Robert J. Parden, Dean of the School of Engineering. Finally, I wish to acknowledge the unstinting devotion of the Secretary of the Department of Electrical Engineering, Mrs. M. Mahaffey.

Richard C. Dorf
Chairman, Electrical Engineering Department
University of Santa Clara
October 24, 1966

TABLE OF CONTENTS

PHYSICAL MODELS OF ACTIVE DEVICES

R. B. YARBROUGH
University of Santa Clara
Santa Clara, California

N67 13342

INTRODUCTION

The distinction between a physical model of an active device and a mathematical model or equivalent circuit of the same device may seem to be neither obvious nor important. While the physical model is established by combining insight into the physical phenomena and mathematics, it inevitably results in mathematical expressions which may be used to obtain an equivalent circuit. The end result may be essentially the same as an equivalent circuit obtained by measurements and the application of two-port network theory. Information about the device over its entire range of possible operation, however, may be measured only with difficulty. The physical model, even if only qualitatively correct, can provide understanding of the behavior of the device over a wide range of applications which can be made quite accurate with relatively few measurements.

The first model of the junction transistor which was applicable over a wide range of operating level was the large signal model proposed by Ebers and Moll [Ref. 1]. The form of the mathematical expression for the two-port equations is based on the static behavior of p-n junctions, and the model parameters are directly obtainable by the measurement of four short-circuit parameters of a two-port network. This model has good accuracy for static and low frequency operation but is lacking in dynamic information. A subsequent paper by Moll [ref 2] contains provision for the dynamics, particularly for switching, but is essentially non-physical in its approach. Linvill [Ref. 4] develops a model for the intrinsic transistor based on the continuity relations in the base region. This approach is physical and introduces a new set of parameters, based on the continuity equation, providing a linear physical model relating carrier densities and current densities. This model inherently contains the dynamic information of the intrinsic transistor and may be directly related to the Ebers and Moll equations.

This paper draws on the works of Ebers and Moll, Linvill and others to obtain a physical model of the junction transistor which may be used to derive any of the equivalent circuits of two-port network theory by applying the proper approximations. The only approximations other than the mathematical description of physical phenomena is the lumping of effects which are actually distributed.

THE JUNCTION TRANSISTOR

The junction transistor is a symmetrical device composed of two p-n junctions and an intervening base region. The physical model is developed by considering the junction characteristics and the base characteristics separately and then matching them together. The junction behavior as described by the Ebers and Moll equations has its basis in quantum mechanics, while the junction capacitance and surface leakage resistance is based primarily on classical electromagnetic theory. The active base region of the transistor is physically described in terms of classical physics while some of the parameters were obtained from statistical considerations. The model is completed by including ohmic losses in the body of the transistor as series elements.

For purposes of development the Ebers and Moll relations are considered first, then the junction capacitance and surface leakage and finally the active base. The combination of these three aspects then result in a general physical model.

STATIC JUNCTION RELATIONS

An intrinsic barrier voltage, ϕ , is developed when a junction exists between p-type and n-type semiconductors. This barrier potential is the difference in the fermi energy levels in the two materials which is in turn dependent on doping levels. Under thermal equilibrium an equal number of carriers cross this junction in both directions so that the net current is zero. The barrier potential is altered when a voltage is applied to the junction, and the resulting current is found to vary exponentially as qV/kT where q is the charge of an electron, V is the applied voltage, k is Boltzmann's constant and T is the temperature in degrees kelvin. When a second junction is placed nearby, the current at the first is affected by the voltage at the second in the same way but to a lesser degree. The form of the voltage-current relations for the junctions is then that given by Ebers and Moll:

$$\begin{aligned} I'_E &= a_{11}\left(e^{\frac{qV_E}{kT}} - 1\right) + a_{12}\left(e^{\frac{qV_C}{kT}} - 1\right) \\ I'_C &= a_{21}\left(e^{\frac{qV_E}{kT}} - 1\right) + a_{22}\left(e^{\frac{qV_C}{kT}} - 1\right) \end{aligned} \qquad (1)$$

where I_E' and V_E are the current and voltage at one junction (the emitter) and I_C' and V_C are those of the other junction (the collector). The junction voltages V_E and V_C are defined as drops from the p material to the n material, and the positive directions of currents are from p to n so that the expressions (1) are valid for either PNP or NPN transistors. The a_{ij} coefficients can theoretically be determined from the short circuit tests, however some inaccuracy arises in that the junction voltages are not directly measurable and the terminal currents differ from I_E' and I_C' by the leakage currents. This prevents the easy determination of these coefficients by simply making one voltage zero and the other a large negative value. The Ebers and Moll relations, while inaccurate for negative voltages due to leakage currents, are sufficiently accurate with positive voltages to allow the coefficients to be determined and yield the relations:

$$\begin{aligned} I_E' &= I_{EO}(e^{\frac{qV_E}{kT}} - 1) - \alpha_I I_{CO}(e^{\frac{qV_C}{kT}} - 1) \\ I_C' &= -\alpha_N I_{EO}(e^{\frac{qV_E}{kT}} - 1) + I_{CO}(e^{\frac{qV_C}{kT}} - 1) \end{aligned} \qquad (2)$$

where I_{EO} and I_{CO} are the reverse saturation currents of the junctions and α_N and α_I are the normal and inverse short circuit current gains.

OTHER JUNCTION EFFECTS

The potential barrier across the junction has an associated depletion layer, a region in which mobile charge carriers are absent except for those making up the currents through the junction. The depletion of the mobile charge carriers is instrumental in establishing the voltage, the bound charges remaining within the depletion region causing the electric field, which integrates to the barrier potential. The width of the depletion region varies with the barrier potential and thereby with the applied voltage. The depletion region stores charge in the bound charges, so that there is a capacitance associated with the junction. The capacitance is approximately inversely proportioned to the width, which in turn is proportional to a fractional power of the barrier potential or the junction voltage. The exact relation of width to the voltage depends on the

doping distribution in the depletion region, and the fractional power is between 2/3 and 1/2. This causes the junction capacitance for a one-dimensional junction theoretically to vary, as

$$C_j = \frac{C_o}{(1 - \frac{V_j}{\Phi_o})^{\frac{1}{n}}}, \qquad 2 < n < 3, \quad V_j < \Phi_o \,. \tag{3}$$

where Φ_o is the intrinsic barrier potential, dependent on doping levels, and C_o is the capacitance with no applied voltage. This is a lumped model of the junction capacitance, but the effect is actually distributed.

Because of various crystaline imperfections on the surfaces of the semiconductor there exists a path for current flow over the surface of the junction. The associated parameter is the surface leakage resistance. Such a resistance depends somewhat on the width of the junction, but is more or less constant, and is less of a distributed parameter than most others considered.

The width of the junctions also affect the so-called base spreading resistance. The effective ohmic impedance of the base region depends on the width of the base. As the base is sandwiched between the junctions, its width is also affected to some degree by the width of the depletion regions. As the emitter and collector regions are usually more highly doped than the base region, the depletion regions tend to be predominantly in the base. With a narrow base region, changes in the depletion widths can make significant changes in the base spreading resistance $r_{bb'}$. As the resistance is inversely proportional to the base width, the base spreading resistance for a one-dimensional transistor theoretically varies as

$$r_{bb'} = \frac{r_{bO}}{1 - \delta_E(1 - \frac{V_E}{\Phi_E})^{\frac{1}{n_e}} - \delta_C(1 - \frac{V_C}{\Phi_C})^{\frac{1}{n_c}}} \tag{4}$$

where δ_E is approximately the ratio of the unbiased width of the emitter junction to the unbiased base width and δ_C has the same relation referred to the collector junction. The term r_{bO} is the base resistance with no bias. This again is a lumped model of a

distributed phenomenon.

THE EXTRINSIC TRANSISTOR

It is now possible to begin the construction of the physical model by means of an equivalent circuit. Figure 1 shows such a circuit with bulk resistances $r_{ee'}$ and $r_{cc'}$ (usually small because of doping densities) surface leakage resistances r_{sc} and r_{se} base spreading resistance $r_{bb'}$ and the junction capacitances c_{je} and c_{jc}. The junction capacitances and base spreading resistance are indicated as distributed, but in usual practice the junction capacitances are assumed to be connected to the active base B'[1].

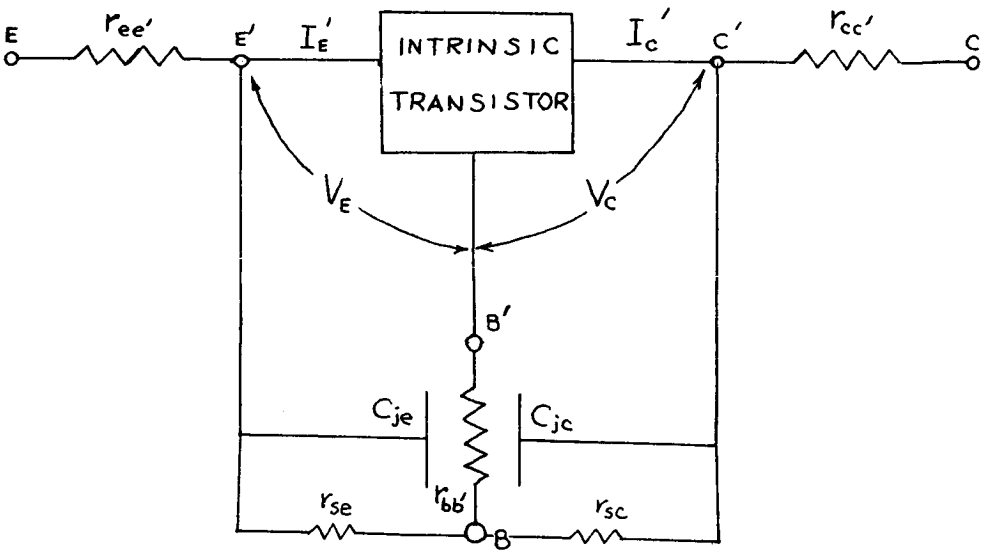


Figure 1. Transistor model showing extrinsic parameters.

---

[1] At the institute it was suggested that the junction capacitances be lumped by connecting 1/3 of each to B' and 2/3 to B.

The intrinsic transistor or active base region is shown as a three terminal black box with the variables of the Ebers and Moll equation indicated. No reference directions are given, as they depend on whether the transistor is PNP or NPN . For a PNP transistors the currents are into the box and the voltages V_C and V_E negative at B' . For the NPN transistor the directions are reversed.

INTRINSIC TRANSISTOR

Continuity Relations.

The continuity of charge density and of current within the base give rise to a partial differential equation [for instance see Refs. 3, 5, 6 or 7]. For a one-dimensional PNP transistor considering only minority carriers the equation becomes

$$\frac{\partial p}{\partial t} = \frac{p_o - p}{\tau_p} - \mu_p \frac{\partial}{\partial_p}(pE) + D_p \frac{\partial^2 p}{2x^2} \qquad (5)$$

where p is the hole density in the N-type base, p_o is the thermal equilibrium hole-density, τ_p is the recombination lifetime, μ_p the hole mobility and D_p is the diffusion constant. An analogous equation may be obtained from a section of a lossy transmission line as shown in Figure 2 if the current is assumed to be purely by diffusion ($E = 0$). If the charge per unit length q_ℓ is defined as

$$q_\ell = qpA, \qquad (6)$$

with q the magnitude of the charge on an electron and A the cross-section area of the base, the corresponding equation for charge per unit length is

$$\frac{\partial q_\ell}{\partial t} = \frac{qAp_o - q_\ell}{\tau_p} + D_p \frac{\partial^2 q_\ell}{\partial x^2}, \qquad (7)$$

and the equation associated with Fig. 2 is

$$\frac{dq_\ell}{dt} = GV_o - \frac{G}{C} q_\ell + \frac{1}{RC} \frac{d^2 q_\ell}{dx^2}. \qquad (8)$$

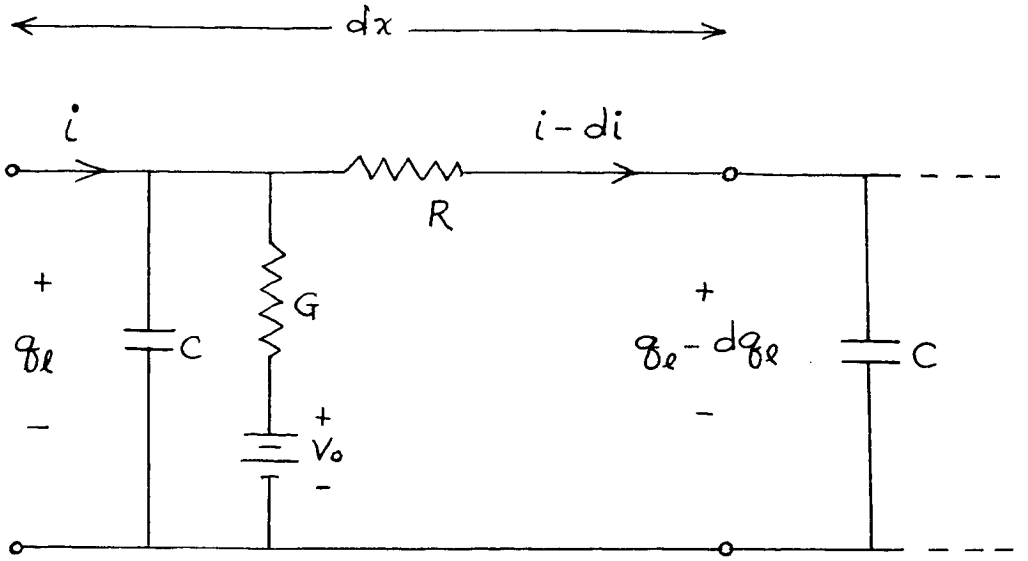


Figure 2. Transmission line analog of the continuity equation.

By matching corresponding coefficients of equations 7 and 8 it is seen that: $G\tau_p = C$, $1/RC = D_p$ and $V_o = qAp_o/C$.

The intrinsic transistor may be modelled by using the transmission line directly. A two section approximation leads to the hybrid p_i equivalent circuit. The principle difficulties with any circuit analogy are first the nonlinear voltage-current relations as in the Ebers and Moll equations (2). The second difficulty lies in the fact that the base width varies with the junction voltages, so that the length of the transmission line is somewhat uncertain. In any case it is necessary that the model of the intrinsic transistor must be mathematically equivalent to the Ebers and Moll equations in the static case. Any attempt to do this for the model of Eqn. (7) will require that $I_{EO} = I_{CO}$ and that $\alpha_I = \alpha_N$ because only the minority carrier densities were considered in the establishment of the equation. The difference in collector and emitter doping densities account for both the difference in reverse saturation currents and for differences in current gains (α) .

The saturation current for a junction in a one-dimensional diode is

$$I_o = qA\left[\frac{D_p p_{no}}{L_p} + \frac{D_n n_{po}}{L_n}\right],$$

where p_{no} and n_{po} are the thermal equilibrium minority carrier densities in the N and P materials respectively. L_p and L_n are the diffusion lengths. For a transistor the diffusion length associated with the base is modified by the factor of tanh (W/L) where W is the base width. Then for a PNP transistor with uniform base doping:

$$I_{CO} = qA\left[\frac{D_p p_{no}}{L_p \tanh(W/L_p)} + \frac{D_n n_{po}(C)}{L_n}\right]$$

$$I_{EO} = qA\left[\frac{D_p p_{no}}{L_p \tanh(W/L_p)} + \frac{D_n n_{po}(E)}{L_n}\right] \qquad (9)$$

Here it is seen that the difference in the reverse saturation currents arises from the difference in doping levels in the collector $[n_{po}(C)]$ and the emitter $[n_{po}(E)]$.

The difference in the α's is similarly accounted for in the difference of emitter and collector efficiencies. The currents accounted for in Eqn. 7 is the minority carrier current in the base. In addition to this, both the emitter and collector currents contain components of base majority carrier current. The efficiency of an emitter is determined by the ratio of base minority current injected from the emitter to the total emitter current. A collector efficiency could be similarly defined for those cases where the collector acts as an emitter, but here the term used is the collector's emitter efficiency. The emitter efficiency again depends on the doping densities of the p and N materials. The emitter efficiency and the collector's emitter efficiency respectively are:

$$\gamma_E = \frac{1}{1 + \frac{D_n}{L_n}\frac{L_p}{D_p}\frac{n_p(E)}{p_{no}} \tanh W/L_p} \tag{10}$$

$$\gamma_C = \frac{1}{1 + \frac{D_n}{L_n}\frac{L_p}{D_p}\frac{n_p(C)}{p_{no}} \tanh \frac{W}{L_p}} .$$

The current gains are proportional to the emitter efficiencies, and so are seen to be dependent on the doping levels.[1]

TRANSMISSION LINE ANALOG

In using the transmission line as a model, it is necessary to keep in mind that the relations of equations 7 and 8 are valid only for current-charge relations. This requires that a two-section transmission line approximation as shown in Fig. 3 is essentially a charge control model, that circuit relations are only valid between the charge on the capacitors and the currents. The terminal currents are related to the terminal voltages by the Ebers and Moll equations (2). The charges on the capacitors are related to the terminal voltages in an exponential form:

$$Q_E = Q_o e^{\frac{qV_E}{kT}} \quad ; \quad Q_C = Q_o e^{\frac{qV_C}{kT}} \tag{11}$$

---

[1] Note that $\gamma_C I_{CO} = \gamma_E I_{EO}$.

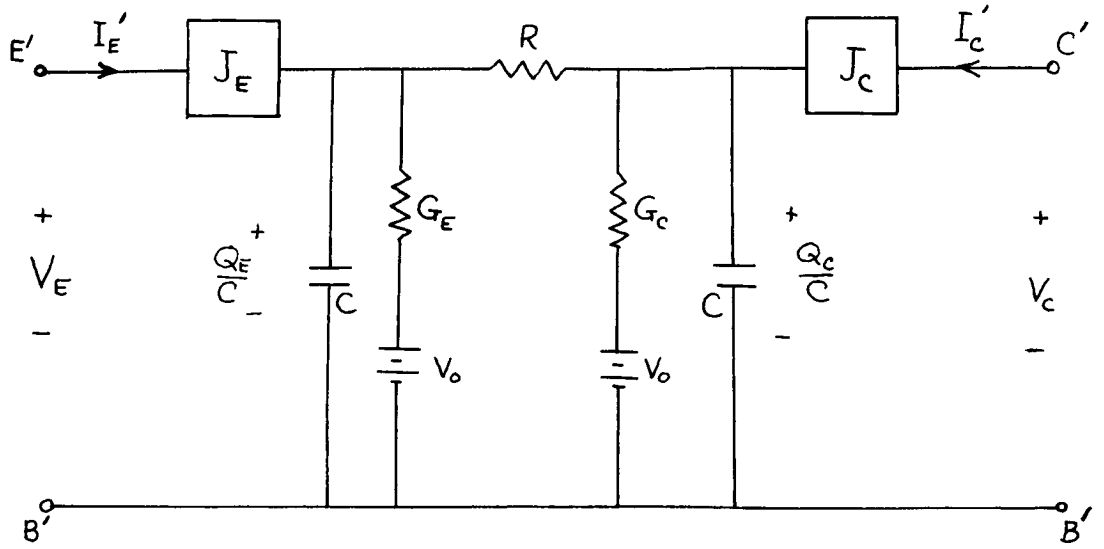


Figure 3: Two section transmission line model of a PNP intrinsic transistor.

The transmission line relations are written in terms of the capacitor voltages:

$$I_E = \frac{Q_E}{C}\left[G_E + \frac{1}{R}\right] - \frac{Q_C}{C}\left[\frac{1}{R}\right] - G_E V_o$$

$$I_C = \frac{Q_E}{C}\left[\frac{1}{R}\right] + \frac{Q_C}{C}\left[G_C + \frac{1}{R}\right] - G_C V_o \qquad (12)$$

With Q_E and Q_C in the form given in equations (11), a matching of coefficients with the Ebers and Moll equations (2) show that

$$RG_E = \frac{1 - \alpha_N}{\alpha_N}$$

$$RG_C = \frac{1 - \alpha_I}{\alpha_I}$$

$$V_o = \alpha_N R I_{EO} = \alpha_I R I_{CO} \quad .$$

It has been shown [Ref. 1] that $\alpha_I I_{CO} = \alpha_N I_{EO}$, and for the two section transmission analog it is necessary that the voltage source connected to G_E must be identical with that connected to G_C so

that both have the value V_o . From further consideration of open and short circuit conditions it may also be shown that $CV_o = Q_o$.

By measuring the cutoff frequency of the short circuit current gains, α_N or α_I, it is further possible to determine that

$$RC = \frac{1}{\alpha_N \omega_E} = \frac{1}{\alpha_I \omega_C} ,$$

which then determines that

$$Q_o = \frac{I_{EO}}{\omega_N} = \frac{I_{CO}}{\omega_I} ,$$

where ω_N is the cutoff frequency of α_N and ω_I is that of α_I .

It is to be noted that the parameters of the transmission line model can not be uniquely determined, but that they must be expressed in terms of one unknown. In the model shown in Figure 4 the diffusion resistance R is chosen for the unknown parameter. For normal biasing, with $V_E > 0$ and $V_C << -\frac{kT}{q}$, R is approximated by $r_\varepsilon/\alpha_N = kT/qI_E\alpha_N$, so that the emitter depletion layer block of Fig. 4 is replaced by a short circuit. The collector depletion layer is an open circuit with a voltage generator across the collector capacitance to account for

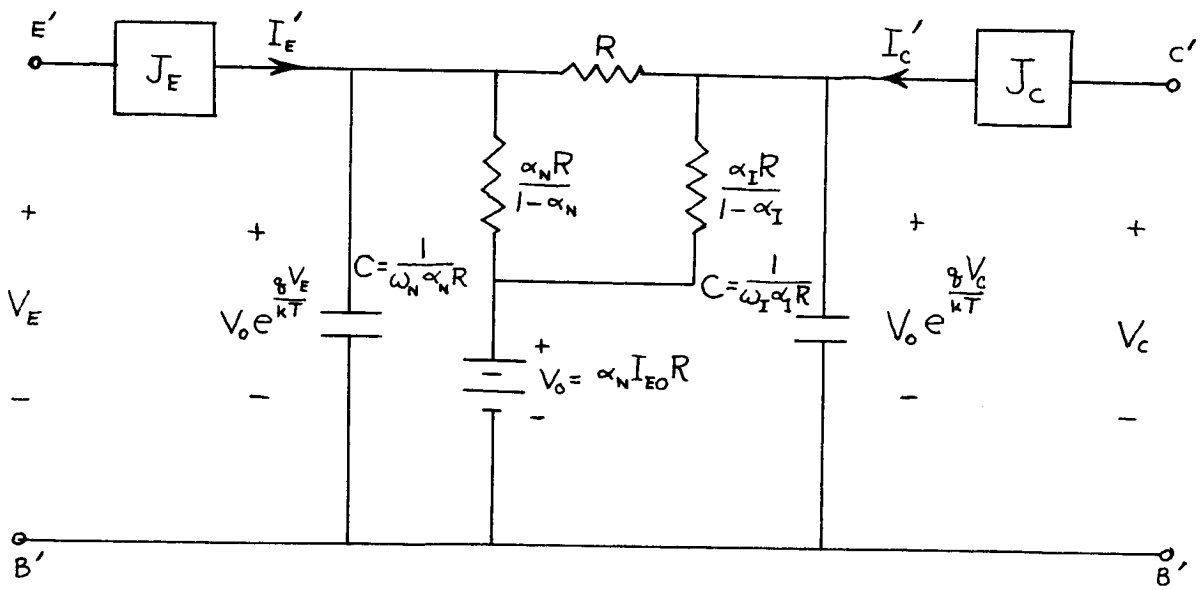


Figure 4: Two section model with parameters related to the Ebers and Moll equations.

modulation effects of the collector voltage on the base width, and the collector current then being the current through that generator as indicated in Fig. 5.

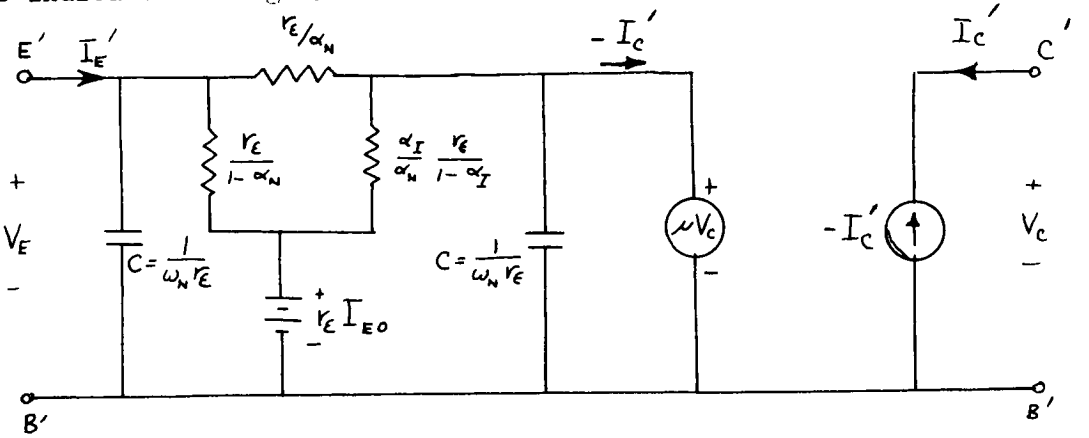


Figure 5: Transmission line model of a normally biased PNP transistor.

THE COMPLETE PHYSICAL MODEL

The complete model is obtained by imbedding the intrinsic transistor model of the previous section (Fig. 4) in the circuit of Fig. 1. This results in a PNP model as shown in Fig. 6.

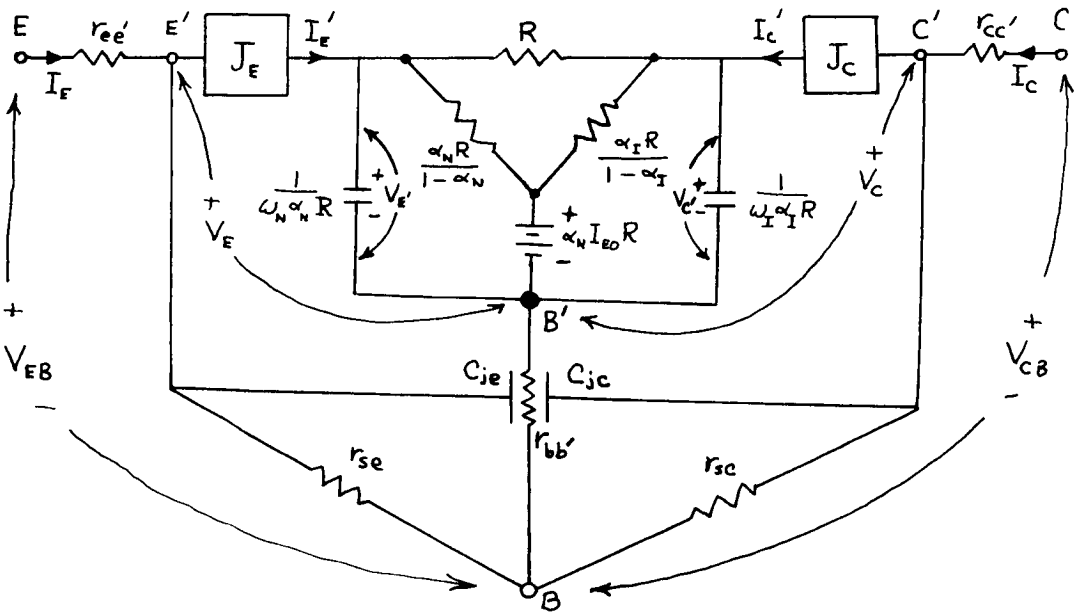


Figure 6: Physical model of a PNP transistor.

The model of Fig. 6 is a large signal model which is dependent on the relations:

$$V_{E'} = \alpha_N I_{EO} \, Re^{\frac{qV_E}{kT}}$$
$$V_{C'} = \alpha_I I_{CO} \, Re^{\frac{qV_C}{kT}} \qquad (13)$$

In reducing this model to an equivalent small signal equivalent circuit by the use of differentials it is necessary to relate both the reverse saturation currents (I_{CO}, I_{EO}) and the short circuit current gain (α_N, α_I) to the junction voltages. It is to be recalled that these parameters have relations to the base width of the forms given in Equations 9 and 10, and the base width in turn depends on the widths of the depletion layers. The width of a depletion layer is a fractional power function of the applied junction voltage, so that the chain rule must be applied to determine the small signal parameter (such as μ in Fig. 5) relating to base width modulation by the junction voltage.

The variations of the junction (depletion layer) capacitances with their respective voltages was discussed in the section on other junction effects.

FIELD EFFECT TRANSISTORS

The monopolar devices called field effect transistors (FET's) are different from the junction transistors (bipolar devices) in that the junction transistor action is dependent on minority carriers and majority carriers, while the FET is essentially a majority carrier device. A qualitative physical model of an FET may be obtained from physical reasoning based on the understanding of junction behavior and surface phenomenon [for instance see Ref. 7 or 8]. However the development of physical models for these devices have not to this author's knowledge been developed to the extent of those for the junction transistor. The model for the pre-pinchoff region was essentially established by Shockley [Ref. 9], but little has been done in the pinched off region of operation.

CONCLUSIONS

A large signal junction transistor equivalent circuit may be developed from a knowledge of the physical behavior of semiconductor

materials and junctions. The parameters of such a model may be determined by measurements of the Ebers and Moll short circuit parameters and the cutoff frequencies. The value of the large signal model based on physical phenomemon is that operation of the device over its entire range of operation (excluding avalanche effects which were not treated) may be predicted from a few static measurements and a few frequency determinations.

REFERENCES

1. J. J. Ebers and J. L. Moll,"Large Signal Behavior of Junction Transistors", Proc. IRE, V. 42, No. 12, Dec. 1954, pp. 1761-1772.

2. J. L. Moll, "Large-Signal Transient Behavior of Junction Transistors," Ibid, pp. 1773-1784.

3. J. L. Moll, Physics of Semiconductors, McGraw-Hill, 1964.

4. J. G. Linvill, "Lumped Models of Transistors and Diodes," Proc. IRE, V. 46, No. 6, June 1958, pp. 1141-1152.

5. J. G. Linvill, J. F. Gibbons, Transistors and Active Circuits, McGraw-Hill, 1961.

6. D. LeCroisette, Transistors, Prentice-Hall, 1963.

7. J. Lindmayer and C. Y. Wrigley, Fundamentals of Semiconductor Devices, D. Van Nostrand, 1965.

8. J. T. Wallmark and H. Johnson, Field-Effect Transistors, Prentice-Hall, 1966.

9. W. Shockley, "A Unipolar "Field Effect" Transistor, Proc. IRE, V. 40, pp. 1365-1367, Nov. 1952.

N67 13343

APPLICATIONS OF MODELS TO CIRCUITS

W. J. KERWIN
Ames Research Center, NASA
Moffett Field, California

INTRODUCTION

The various active device models for linear circuit application are considered from the standpoint of simplicity of use, and examples of their application to the design of an operational amplifier and an emitter-peaked voltage amplifier stage are given.

MODELS FOR LOW AND MEDIUM FREQUENCY USE

For our purpose here we will consider only the common-emitter models. Figure 1 shows the T model, the h parameter model, and the y parameter model. The values of the resistors are given in ohms in all cases. The data normally specified by the manufacturer include the h parameters (either common-base or common-emitter) at a specific voltage, current, and frequency (usually V_{ce} or V_{cb} = 5 V, I_c = 1 mA, and f = 1 Kcps), and occasionally the y parameters. These are usually given as a function of frequency at I_c = 20 mA and V_{ce} = 30 V. The T parameters are usually not available from the manufacturer's data sheet; however, as can be seen in Figure 1, the T model is the simplest of the three models and the one most frequently used.

The relationship of the T parameters to the h or y parameters is shown in Figure 2. These relations are all we need except for the case of mixed parameters. It is frequently true that the data specified include the common-base h parameters, except for h_{fb}, and here h_{fe} is usually given. The equation for converting to h_{fb} from h_{fe} is also given in Figure 2.

The operating conditions of the transistor are not usually V_{ce} = 5 V and I_c = 1 mA. To convert to the actual operating conditions we use the plots of the variation of the h parameters with collector current and with collector-to-emitter voltage that are normally supplied by the manufacturer. An example of the calculation of the operating h parameters of a 2N1132 for I_c = 0.75 mA and V_{ce} = 15 V is given in Figure 3, and the T parameters are calculated from the operating condition h parameters as shown. The approximation that $h_{fe} \gg 1$ has been made in deriving the equations shown in Figure 3.

In some cases only approximate values of the T parameters are needed. For this purpose the approximations shown below may be adequate. A constant λ is dependent on the particular device construction used with a usual value between 2 and 5. Here, the assumed value is $\lambda = 3$. In addition we need the emitter base diode resistance $r_\epsilon = (q/kTI_e) = (0.025/I_e)$ at 300° K. Substitution of these values into the equations below gives the results shown. For many cases it is adequate to assume that $r_d = \infty$.

$$r_e \cong \frac{\lambda - 1}{\lambda} r_\epsilon \cong \frac{0.0167}{I_e}$$

$$r_b \cong \frac{h_{fe} r_\epsilon}{\lambda} \cong \frac{0.00833 h_{fe}}{I_e}$$

ANALYSIS OF AN OPERATIONAL AMPLIFIER USING THE T MODEL

As an example of the use of the T model we will analyze the operational amplifier shown in Figure 4. Nominal values of the circuit parameters are included as a starting point. These may be changed after analysis, of course, but it is important that they be included at this point to give an idea of the magnitudes of the various terms in the analysis. The T model parameters are then calculated for each transistor at its respective operating point, resulting in the equivalent circuit shown in Figure 5. (All values of r_d have been assumed infinite for this example.) The voltages and currents specified on Figure 5 are used to develop the flow graph shown in Figure 6 which shows the interrelations of all voltages and currents in this amplifier circuit.

The flow graph is much too involved to work with at this point, and illustrates a general principle, namely, perform simple combinations and node eliminations before applying the general flow graph reduction formula. In this case, we will assume that the transfer voltage ratio e_{out}/e_{in} is the quantity of interest and we will simplify the graph with that in mind. After a few simple reduction steps the graph shown in Figure 7 is obtained. We now have only eight loops, but it can be seen that even here the use of the general reduction formula would be quite unwieldy. It is at this point, however, that we use the nominal values of all parameters to simplify the determination of the voltage ratio. If we look at the magnitude of the forward path terms, and of the loop transmissions (Fig. 8) we see that L_1, L_2, L_3, L_4, and L_7 can be neglected with an error of less than 1 part in 10,000 due to each term. We then have only three loops left and we have no multiple products, so that application of the general reduction formula is very simple. This flow graph simplification technique is extremely important as it makes the flow graph method much more practicable, and it is the flow graph technique which allows this method of simplification to be used. Note that in this case all of the forward path terms are retained.

We now insert the literal values for these retained forward path and loop terms realizing that even fairly large changes in the nominal values will still keep an accuracy of the order of 1 part in 1,000 since we neglected only terms less than 1 part in 10,000. This results in the transfer function shown, and when the nominal values are inserted (with $R_s = 0$) we obtain a voltage gain of 2.0042. A comparison of this calculated gain with the measured gains of six of these amplifiers, using the same feedback resistors in each case, was excellent as indicated in Figure 9. This amplifier has been constructed in hybrid integrated circuit form and is shown in Figure 10.

HYBRID-Π MODEL FOR HIGH FREQUENCY USE

At higher frequencies the internal capacities of the transistor can no longer be neglected. If we modify the T model to include the higher frequency effects, we obtain a model which is not as convenient to use as the hybrid-Π model. Our discussion here will therefore use a simplified form of the hybrid-Π model (Fig. 11). The values of C_{ob} and $\beta_o = h_{fe}$ are normally given by the manufacturer as well as a high frequency value of h_{fe} from which ω_T can be calculated. One parameter, r_{bb}', is usually missing, however. In some cases, y_{ie} is plotted versus frequency and an extrapolated value as $\omega \to \infty$ gives the reciprocal of r_{bb}'.

ANALYSIS OF AN EMITTER-PEAKED STAGE

As an example of the use of the hybrid-Π model we will analyze an emitter-peaked stage. The circuit schematic and the hybrid-Π equivalent circuit are shown in Figure 12 with a few defined relations. This circuit is analyzed by flow graph means in Figure 13. If we now put in nominal values (Fig. 14) and eliminate terms less than 0.1%, we obtain, with literal values, the transfer function shown in Figure 14. Then, substituting values for all except ω_e, we obtain a 1-zero, 2-pole function which can be modified by a change of C_e only while all other components are kept at their nominal value.

Figure 15 compares the bandwidth resulting for various values of C_e and therefore ω_e. It can be seen that the maximally flat 2-pole, 1-zero case shows a great improvement over the other possibilities, with a predicted value of C_e = 320 pF. In an experimental test a value of 330 pF gave maximally flat response, and a bandwidth of 5 Mcps. However, the model assumed a value of ω_T based on minimum specifications for the 2N2484 and it would be reasonable to expect that most transistors would exhibit considerably better response, as proved to be the case experimentally.

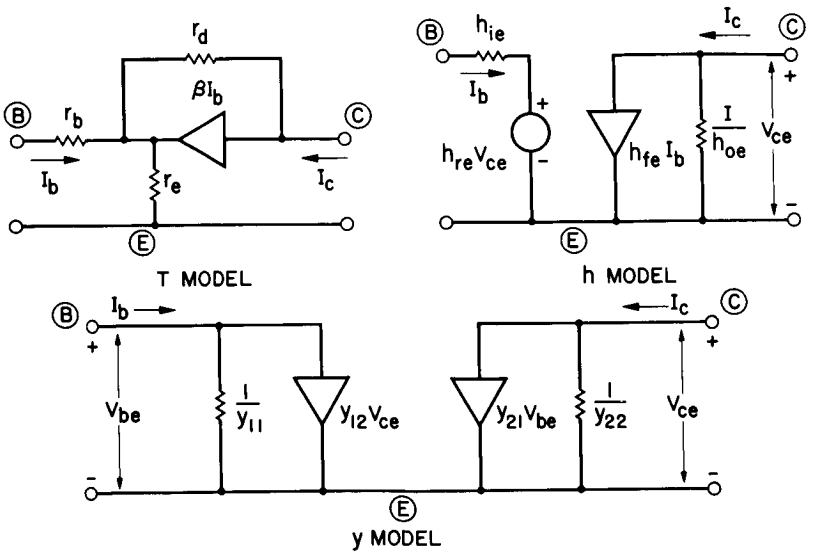
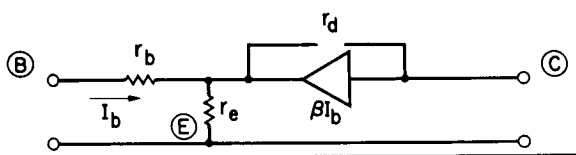


Figure 1.- Common emitter models.



T	COMMON EMITTER h	COMMON BASE h	y
r_b	$h_{ie} - \frac{(1+h_{fe})h_{re}}{h_{oe}}$	$\frac{h_{rb}}{h_{ob}}$	$\frac{1}{y_{11e}}\left(1+\frac{y_{21e}\,y_{12e}}{\Delta y}\right)$
r_e	$\frac{h_{re}}{h_{oe}}$	$h_{ib} - \frac{h_{rb}(1+h_{fb})}{h_{ob}}$	$-\frac{y_{12e}}{\Delta y}$
r_d	$\frac{1}{h_{oe}}$	$\frac{1+h_{fb}}{h_{ob}}$	$\frac{y_{11e}}{\Delta y}$
β	h_{fe}	$-\frac{h_{fb}}{1+h_{fb}}$	$\frac{y_{21e}}{y_{11e}}$

$$h_{fb} = -\frac{h_{fe}}{1+h_{fe}} \qquad \Delta y = y_{12e}y_{22e} - y_{12e}\,y_{21e}$$

Figure 2.- T parameters from h or y parameters.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA

2N1132

SPECIFICATIONS AT 1 mA, 5 V	SPECIFICATIONS AT .75 mA	SPECIFICATIONS AT 15 V	OPERATING CONDITION h PARAMETERS
$h_{fe} = 57$	×0.98	×1.1	61.4
$h_{ib} = 27\Omega$	×1.4	×1.01	38.2Ω
$h_{rb} = 2 \times 10^{-4}$	×1.0	×0.73	1.46×10^{-4}
$h_{ob} = .3 \times 10^{-6}\mho$	×0.9	×0.71	$0.19 \times 10^{-6}\mho$

$$r_e \cong h_{ib} - \frac{h_{rb}}{h_{fe} h_{ob}} = 25.7\,\Omega \qquad r_b = \frac{h_{rb}}{h_{ob}} = 768\,\Omega$$

$$r_d \cong \frac{1}{h_{ob} h_{fe}} = 85.7\,K\Omega \qquad \beta = h_{fe} = 61.4$$

Figure 3.- Typical conversion from h to T parameters.

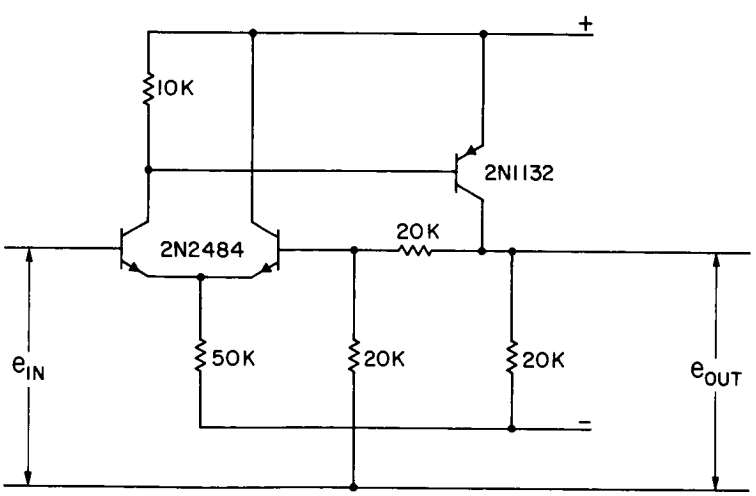


Figure 4.- Amplifier schematic.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA

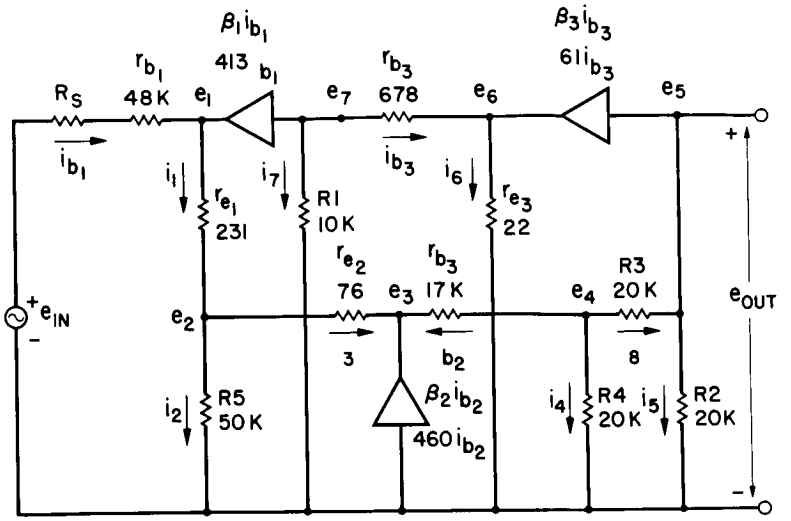


Figure 5.- Equivalent circuit for analysis.

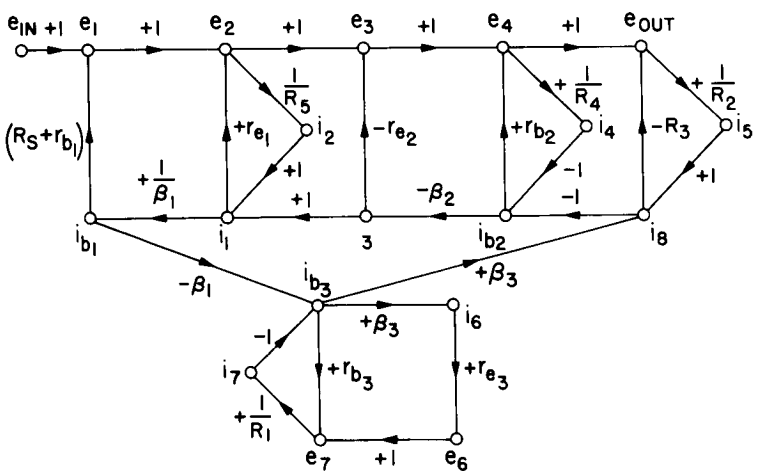


Figure 6.- Flow graph of the equivalent circuit.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA

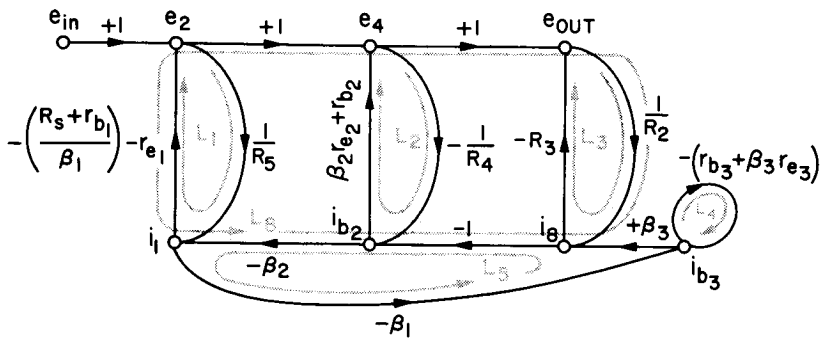


Figure 7.- Reduced flow graph.

$FP_1 = 28{,}060$ $\quad L_1 = -\left(\frac{R_s}{20.6M}\right)^*$

$FP_2 = 63$ $\quad L_2 = -2.6^*$

$FP_3 = 88$ $\quad L_3 = -1^*$

$FP_4 = 28{,}060$ $\quad L_4 = -.2^*$

$L_5 = -28{,}060$

$L_6 = -\left(\frac{R_s}{18K} + 8.0\right)$

$L_7 = -2.6^*$

$L_8 = -\left(\frac{R_s}{18K} + 8.0\right)$

$$\frac{e_{OUT}}{e_{IN}} \cong \frac{1+\frac{R_3}{R_4}+\frac{r_{b_2}+\beta_2 r_{e_2}}{\beta_2 R_5}\left(1+\frac{R_3}{R_4}\right)+\frac{R_3}{\beta_2 R_5}}{1+\frac{1}{\beta_1\beta_3}\left(R_S+r_{b_1}+\beta_1 r_{e_1}\right)\left[\frac{1}{R_4}+\frac{1}{R_2}\right]} = 2.0042 \quad (R_S=0)$$

* NEGLIGIBLE TERMS

Figure 8.- Transfer function simplification.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA

AMPLIFIER NUMBER	GAIN		Z_{IN}		Z_{OUT}	
	CALC.	MEAS.	CALC.	MEAS.	CALC.	MEAS.
1	2.0042	2.0042	116M	160M	20Ω	12.1Ω
2	2.0042	2.0046	116M	122M	20Ω	17.1Ω
3	2.0042	2.0049	116M	86M	20Ω	19.8Ω
4	2.0042	2.0053	116M	101M	20Ω	17.8Ω
5	2.0042	2.0030	116M	75M	20Ω	23.7Ω
6	2.0042	2.0028	116M	74M	20Ω	24.7Ω

Figure 9.- Amplifier performance.

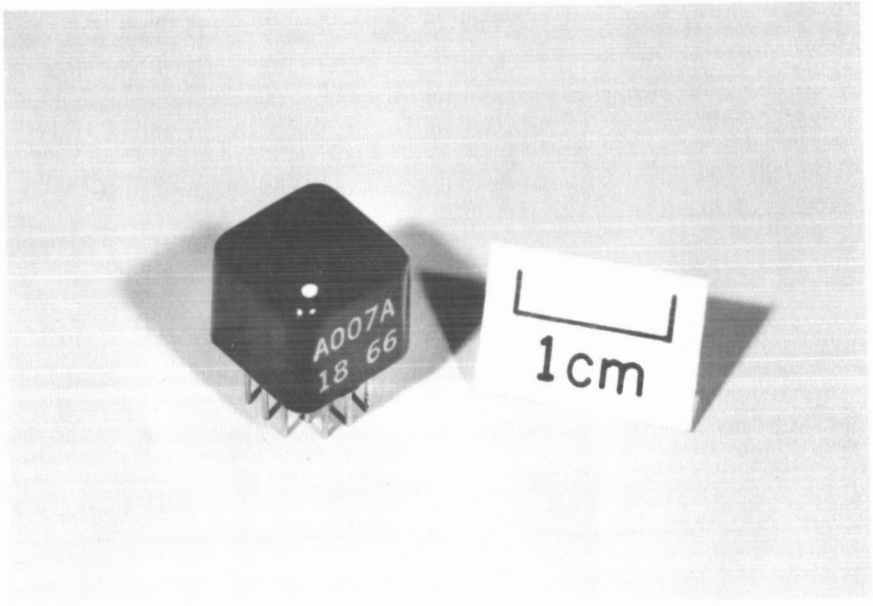
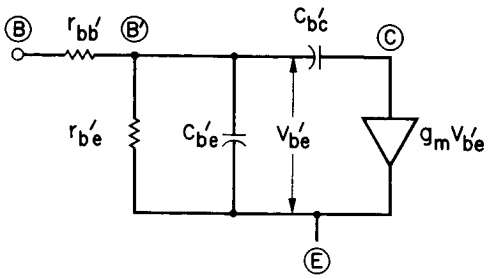


Figure 10.- Photograph of the amplifier.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA



APPROXIMATE RELATIONS

$1/r_{bb'} = Re(y_{ie})$ $\quad \omega \to \infty$

$C_{b'e} = \frac{1}{r_\epsilon \omega_T} - C_{ob}$

$r_\epsilon = \frac{.025}{I_e}$

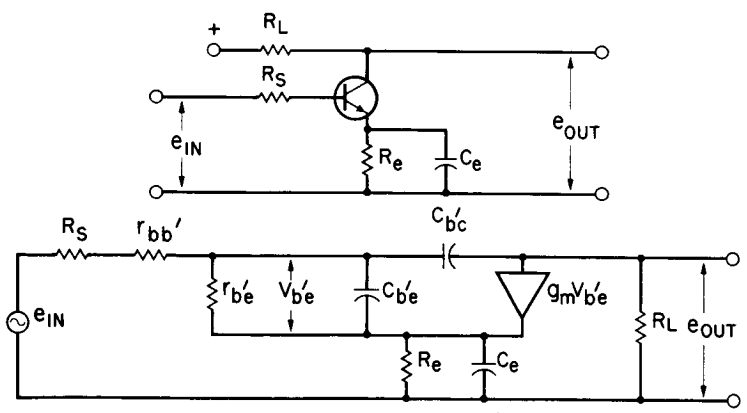
$r_{b'e} = \beta_o r_\epsilon$

$C_{b'c} = C_{ob}$

$\omega_T = 2\pi f |h_{fe}|_f$

$g_m = \frac{1}{r_\epsilon}$

Figure 11.- Hybrid-Π equivalent circuit.



$R'_S = R_S + r_{bb'}$

$Z = \frac{r_{b'e}\omega_\beta}{P+\omega_\beta}$

$Z_e = \frac{R_e\omega_e}{P+\omega_e}$

$\omega_\beta = \frac{1}{r_{b'e} C_{b'e}}$

$\omega_e = \frac{1}{R_e C_e}$

Figure 12.- Emitter peaked stage.

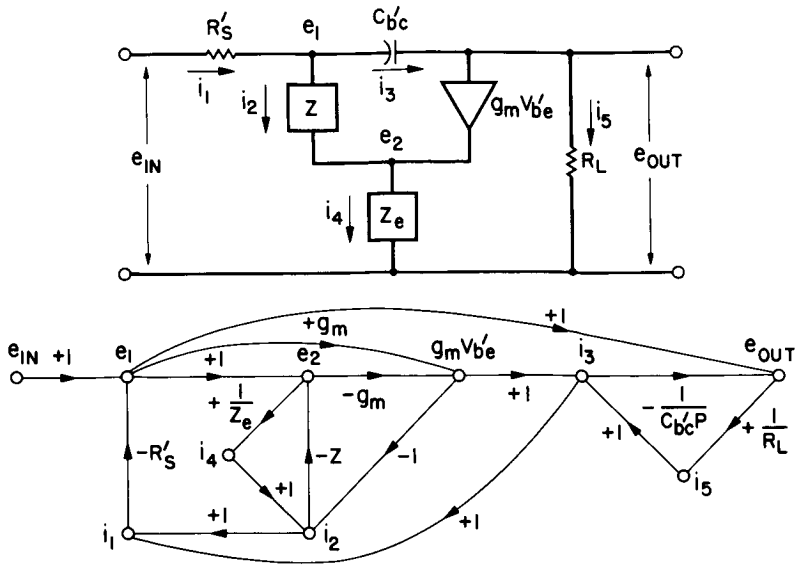


Figure 13.- Equivalent circuit and flow graph.

$R_L = 1K\Omega$ $\quad \beta = 400$

$R_S' = 1K\Omega$ $\quad r_{\epsilon} = 5\Omega$

$R_e = 100\Omega$ $\quad C_{be}' = 200$ pF

$C_e = 200$ pF $\quad r_{bc}' = 2K\Omega$

$\omega = 10^7$ rps $\quad C_{bc}' = 5$ pF

$\omega_e = 5 \cdot 10^7$ rps $\quad \omega_T = 800$ Mrps

$Z_e = \dfrac{100}{2 \cdot 10^{-8} p + 1}$ $\quad \omega_\beta = 2.5 \cdot 10^6$ rps

$Z = \dfrac{2000}{4 \cdot 10^{-7} p + 1}$

$$\frac{e_{OUT}}{e_{IN}} \cong \frac{-g_m R_L Z}{Z + Z_e + R_S' + g_m R_L R_S' Z C_{bc}' p + g_m Z Z_e \left[1 + C_{bc}' p (R_L + R_S')\right]}$$

$$\frac{e_{OUT}}{e_{IN}} \cong \frac{-1.67 \times 10^8 (p + \omega_e)}{p^2 + (1.18\omega_e + 1.25 \times 10^6) p + 1.8 \times 10^7 \omega_e}$$

Figure 14.- Transfer function and nominal values.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA

$$\frac{e_{OUT}}{e_{IN}} = \frac{-1.67\times10^{8}\,(p+\omega_e)}{p^2+(1.18\,\omega_e+1.25\times10^{6})\,p+1.8\times10^{7}\,\omega_e}$$

CASE 1	CASE 2	CASE 3
1 POLE ONLY	POLE-ZERO CANCELLATION	2 POLES - 1 ZERO
$\omega_e = \infty$	$\omega_e = 9.3\cdot10^{7}$ rps	$\omega_e = 3.1\cdot10^{7}$ rps
$C_e = 0$	$C_e = 108$ pF	$C_e = 320$ pF
$f_{-3dB} = 2.4$ Mcps	$f_{-3dB} = 2.9$ Mcps	$f_{-3dB} = 4.4$ Mcps

VOLTAGE GAIN = 9.3

Figure 15.- Choice of C_e for maximum bandwidth.

N67 13344

TOPOLOGICAL ANALYSIS OF PASSIVE NETWORKS

SHU-PARK CHAN
University of Santa Clara
Santa Clara, California

ABSTRACT

13344

In this paper we hope to give a brief introduction to the application of network topology to the analysis of passive networks. First, a minimum set of definitions is introduced. Some theorems in network topology are stated without proof* and topological formulas for computing various network functions are listed. Finally, a number of examples are given to illustrate the usefulness of these topological formulas in network analysis and to compare them to the conventional analysis methods.

Author

INTRODUCTION

Network topology (or network graph theory) may be described as a study of electrical networks in connection with their non-metric geometrical, namely, topological-properties through the investigation of their graphs. The application of graph theory to electrical networks can be dated back to the 19th century when, in 1947, Kirchhoff applied the topological concepts to formulate the loop-basis and node-basis network equations. The application of topological methods to the analysis and synthesis of electrical networks had not made much significant advancement until the last decade. During that period, network topology was gradually made known and recognized as the foundation for the modern network theory.

DEFINITIONS

Consider the passive network N (a bridged-T network with a load Z_5) with a voltage driver E as shown in Fig. 1. In this network, there are 4 nodes, (V_1, V_2, V_3 and V_4) and 5 branches (Z_1, Z_2, Z_3, Z_4, and Z_5). If each of the 5 branches is replaced by a line segment with the 4 nodes remaining unchanged, we obtain a graphical representation of N as shown in Fig. 2. Such a graphical representation is known as the linear graph (or simply graph) of the network N, and each of the 5 line segments is called an edge of the graph G, corresponding to a branch in N . The four nodes in G are called the vertices of G . Having shown how a graph is obtained from a given electrical network,

---

* The interested readers may find all the proofs and derivations in Ref. [1] or Ref. [2].

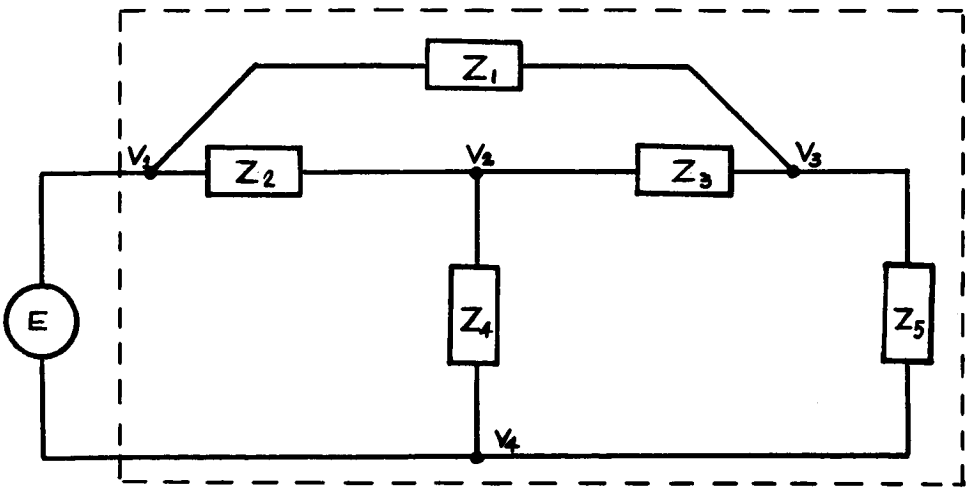


Fig. 1. A passive network N with a voltage driver E.

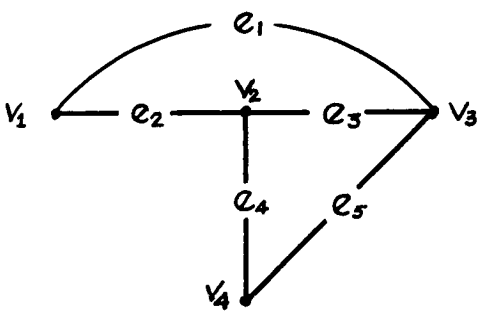


Fig. 2. The graph G of the network N of Fig 1.

we now define some of the topological terms as follows.

Definition 1. Linear Graph and Its Edges and Vertices.

A linear graph (or simply a graph) is a set of line segments called edges and points called vertices, which are the endpoints of the edges, interconnected in such a way that the edges are connected to (or incident with) the vertices.

Definition 2. Degree of a Vertex.

The degree of a vertex of a graph is the number of edges incident with that vertex.

Definition 3. Subgraph.

A subset G_i of the edges of a given graph G is called a subgraph of G. If G_i does not contain all of the edges of G, it is a proper subgraph of G.

Definition 4. Path and Path-set.

A path is a subgraph having all vertices of degree 2 except for the two endpoints which are of degree 1 and are called the terminals of the path. The set of all edges in a path constitute a path-set.

Definition 5. Circuit (loop) and Circuit-set (loop-set).

If the two terminals of a path coincide, the path is a closed path and is called a circuit (or loop). The set of all edges contained in a circuit is called a circuit-set (or loop-set).

Definition 6. Connectedness.

A graph or subgraph is said to be connected if there is at least one path between every pair of its vertices.

Definition 7. Tree.

A tree of a connected graph G is a connected subgraph which contains all the vertices of G but no circuits. The edges contained in a tree are called the branches of the tree.

Definition 8. 2-tree and k-tree.

A 2-tree of a connected graph G is a (proper) subgraph of G consisting of two unconnected circuitless subgraphs, each subgraph itself being connected, which together contain all the vertices of G. Similarly, a k-tree is a subgraph consisting of k unconnected circuitless subgraphs, each subgraph being connected; which together include all the vertices of G.

Definition 9. k-tree Admittance Product.

The k-tree admittance product of a k-tree is the product of the admittance of all the branches of the k-tree.

Example.

The graph G shown in Fig. 2 is the graph of the network N of Fig. 1. The edges of G are e_1, e_2, e_3, e_4, and e_5; the vertices of G are V_1, V_2, V_3, and V_4. A path of G is the subgraph G_1 consisting of edges e_1, e_2, and e_5 with vertices V_2 and V_4 as terminals. Thus, the set $\{e_1, e_2, e_5\}$ is a path-set. With edge e_4 added to G_1, we form another subgraph G_2 which is a circuit since as far as G_2 is concerned all of its vertices are of degree 2. Hence the set $\{e_1, e_2, e_4, e_5\}$ is a circuit-set. Obviously, G is a connected graph since there exists a path between every pair of vertices of G. A tree of G is the subgraph consisting of edges e_1, e_2, and e_4. Two other trees of G are $\{e_2, e_3, e_4\}$ and $\{e_1, e_3, e_5\}$. A 2-tree of G is $\{e_1, e_4\}$; another one is $\{e_2, e_5\}$; and still another one is $\{e_2, e_3\}$. Note that both $\{e_1, e_4\}$ and $\{e_2, e_5\}$ are subgraphs which obviously satisfy the definition of a 2-tree in the sense that each contains two

disjoint circuitless connected subgraphs, both of which together includes all the four vertices of G . On the other hand, $\{e_2, e_3\}$ is just one connected subgraph which contains only three vertices of G. Thus, $\{e_2, e_3\}$ does not seem to be a 2-tree. However, if we agree to consider $\{e_2, e_3\}$ as a subgraph which contains edges e_2 and e_3 plus the isolated vertex V_4, we see that $\{e_2, e_3\}$ will satisfy the definition of a 2-tree since it now has two circuitless connected subgraphs with e_2 and e_3 forming one of them and the vertex V_4 alone forming the other. Moreover, both subgraphs together indeed include all the four vertices of G . It is worth noting that a 2-tree is obtained from a tree by removing any one of the branches from the tree; and, in general, a k-tree is obtained from a (k-1)-tree by removing from it any one of its branches.

Finally, the tree admittance product of the tree $\{e_1, e_3, e_5\}$ is $y_1 y_3 y_5$; the 2-tree admittance product of the 2-tree $\{e_2, e_3\}$ is $y_2 y_3$ (with the admittance of a vertex defined to be 1).

NETWORK FUNCTIONS

Consider a network N with n independent nodes as shown in Fig. 3.

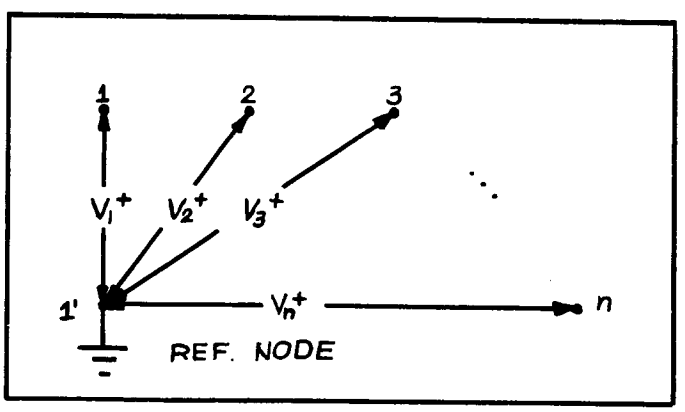


Fig. 3. A network N with n independent nodes

The node 1' is taken as the reference (datum) node. The voltages V_1, V_2, ... V_n (which are functions of s) are the transforms of the node-pair voltages (or simply node voltages) v_1, v_2, ..., v_n (which are functions of t) between the n nodes and the reference node 1' with the plus polarity marks at the n nodes. Using the node analysis method developed in Chapter 7, we obtain the n independent node equations:

$$
\begin{aligned}
y_{11}V_1 + y_{12}V_2 + \cdots + y_{1n}V_n &= I_1 \\
y_{21}V_1 + y_{22}V_2 + \cdots + y_{2n}V_n &= I_2 \\
&\cdots \\
y_{n1}V_1 + y_{n2}V_2 + \cdots + y_{nn}V_n &= I_n
\end{aligned}
\qquad (1)
$$

where for any j $(j = 1, 2, \ldots, n)$, y_{jj} is the sum of the admittances of all the branches connected to node j; and for any j and k $(j \neq k;\ j,k = 1,2,\ldots,n)$, y_{jk} is the negative of the sum of the admittances of all the branches connected between nodes j and k (thus all such branches are connected in parallel across the two nodes). Furthermore, I_j is the sum of the transforms of the known current sources at node j; and V_j is the transform of the voltage at node j with respect to the reference node '1'.

Written in matrix form, (1) becomes

$$
\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ & \cdots & & \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ \\ V_n \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \\ \\ I_n \end{bmatrix} \qquad (2)
$$

or, in abbreviated matrix notation,

$$
Y_n V_n = I_n \qquad (3)
$$

where Y_n is the node admittance matrix, V_n the n-vector* of the node voltage transforms, and I_n the n-vector of the transforms of the known current sources as expressed in (2).

Solving for the node voltages upon the application of Cramer's rule, we find

$$
V_j = \frac{\Delta_{1j}}{\Delta} I_1 + \frac{\Delta_{2j}}{\Delta} I_2 + \cdots + \frac{\Delta_{nj}}{\Delta} I_n \qquad (j = 1,2, \ldots, n) \qquad (4)
$$

---

* An n-vector is defined as a column matrix of order $n \times 1$.

where Δ is the determinant of the node admittance matrix Y_n, and Δ_{jk} is the (j,k)- cofactor of Δ.

Next, assume that the network N is a relaxed passive one-port (so that the initial conditions are all zero) and drive it with a single current source I_1 at nodes 1 and 1' as depicted in Figure 4.

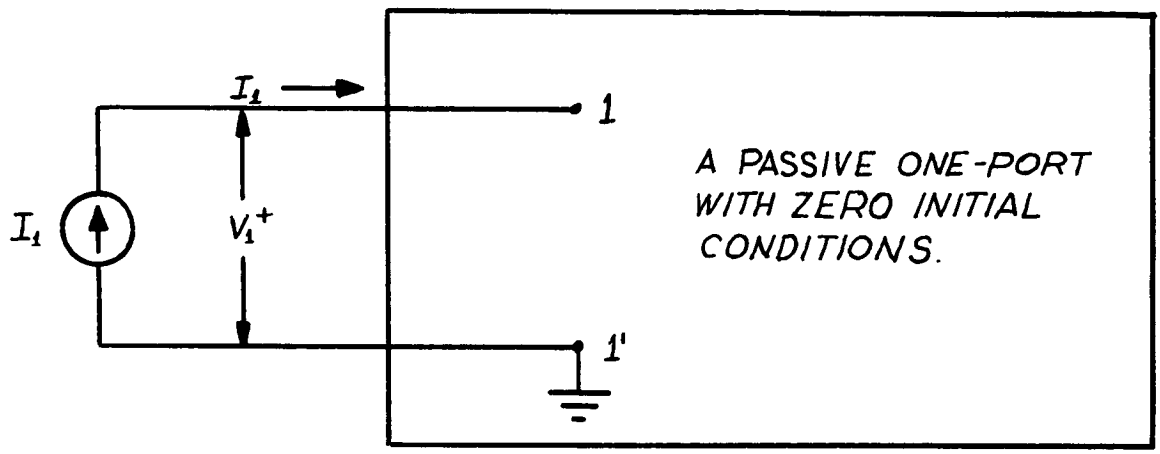


Fig. 4. The network N driven by a single current source

Thus, with the conditions

$$I_1 = I_{g_1}, \quad \text{and} \quad I_2 = I_3 = \cdots = I_n = 0 ,$$

Equation (4) becomes

$$V_j = \frac{\Delta_{1j}}{\Delta} I_1 \qquad (j = 1, 2, \ldots, n) \tag{5}$$

from which, by letting $j = 1$, we obtain

$$V_1 = \frac{\Delta_{11}}{\Delta} I_1 \quad . \tag{6}$$

Hence, the driving-point impedance function $Z_d(s)$ and its reciprocal, namely driving-point admittance function $Y_d(s)$ are readily obtained from (6) as

$$Z_d(s) \triangleq \frac{V_1}{I_1} = \frac{\Delta_{11}}{\Delta} \tag{7}$$

and

$$Y_d(s) = \frac{1}{Z_d(s)} = \frac{\Delta}{\Delta_{11}} \tag{8}$$

respectively.

We see from (7) and (8) that both $Z_d(s)$ and $Y_d(s)$ can be expressed in terms of the determinant of Y_n, Δ, and its (1,1) cofactor, Δ_{11}.

Next, consider a passive two-port driven by two current generators I_{g_1} and I_{g_2} at ports 1 and 2 respectively, as shown in Fig. 5.

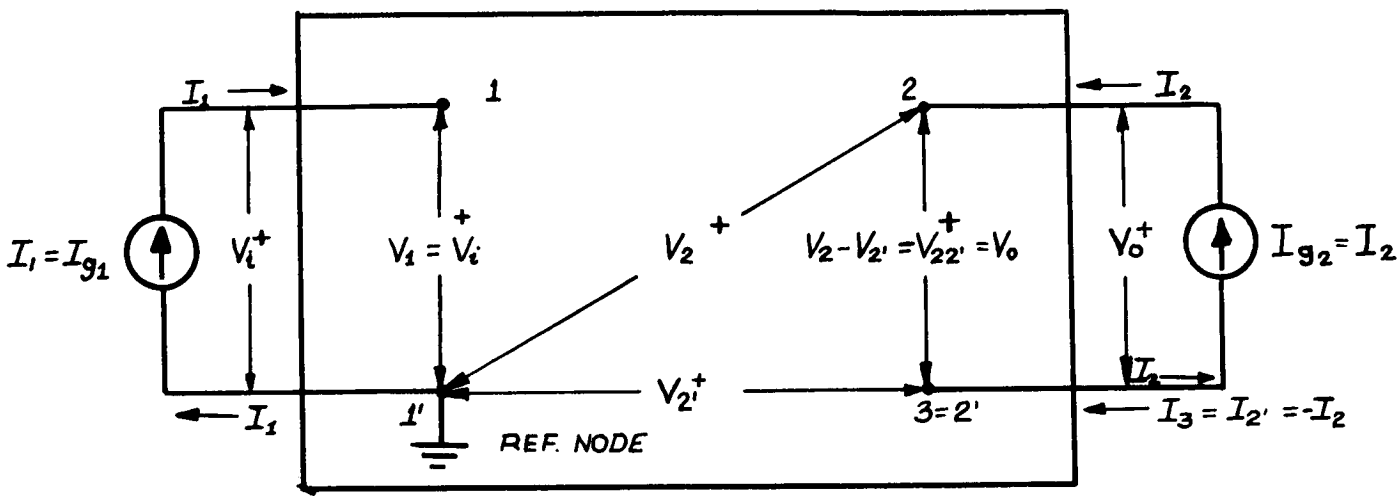


Fig. 5. A passive two-port driven by two current generators

The following observations are noted:

(1) Since there are only two sources, namely I_{g_1} and I_{g_2}, we have

$$I_1 = I_{g_1}$$

$$I_2 = I_{g_2}$$

$$I_3 = I_2' = -I_2$$

$$I_4 = I_5 = \ldots = I_n = 0$$

(2) The four external nodes are 1, 2, 2' and 1' with 1' chosen as reference node.

(3) With the notation used in the figure, we write

$$V_i = V_1 \quad \text{and} \quad V_o = V_{22'} = V_2 - V_{2'} = V_2 - V_3 \qquad \text{(since node 3 = node 2')} \tag{10}$$

Thus, from (4) (by letting $j = 1, 2,$ and 3) and with the conditions listed in (9) and (10), the following results are obtained:

$$V_i = \left(\frac{\Delta_{11}}{\Delta}\right)I_1 + \left(\frac{\Delta_{21} - \Delta_{2'1}}{\Delta}\right)I_2 = z_{11}I_1 + z_{12}I_2$$

$$V_o = \left(\frac{\Delta_{12} - \Delta_{12'}}{\Delta}\right)I_1 + \left(\frac{\Delta_{22} + \Delta_{2'2'} - \Delta_{22'} - \Delta_{2'2}}{\Delta}\right)I_2 = z_{21}I_1 + z_{22}I_2 \tag{11}$$

where

$$z_{11} = \left.\frac{V_i}{I_1}\right|_{I_2 = 0} = \frac{\Delta_{11}}{\Delta} \quad ; \qquad \text{(a)}$$

$$z_{21} = \left.\frac{V_o}{I_1}\right|_{I_2 = 0} = \frac{\Delta_{12} - \Delta_{12'}}{\Delta} \quad ; \qquad \text{(b)} \tag{12}$$

$$z_{12} = \left.\frac{V_i}{I_2}\right|_{I_1 = 0} = \frac{\Delta_{21} - \Delta_{2'1}}{\Delta} \quad ; \qquad \text{(c)}$$

and

$$z_{22} = \left.\frac{V_o}{I_2}\right|_{I_1 = 0} = \frac{\Delta_{22} + \Delta_{2'2'} - \Delta_{2'2} - \Delta_{22'}}{\Delta} \quad ; \qquad \text{(d)}$$

and are called the open-circuit impedances of the two-port.

If we solve (ii) for I_1 and I_2 using Cramer's rule, we find

$$I_1 = \frac{z_{22}}{\Delta_z} V_i + \frac{-z_{12}}{\Delta_2} V_o = y_{11}V_i + y_{12}V_o$$

$$I_2 = \frac{-z_{21}}{\Delta_z} V_i + \frac{z_{11}}{\Delta_z} V_o = y_{21}V_i + y_{22}V_o \qquad (13)$$

where the y's are the short-circuit admittances of the two-port:

$$y_{11} = \left.\frac{I_1}{V_i}\right|_{V_o = 0} = \frac{z_{22}}{\Delta_z} \; ; \qquad (a)$$

$$y_{21} = \left.\frac{I_2}{V_i}\right|_{V_o = 0} = \frac{-z_{21}}{\Delta_z} \; ; \qquad (b)$$

$$y_{12} = \left.\frac{I_1}{V_o}\right|_{V_i = 0} = \frac{-z_{12}}{\Delta_z} \; ; \qquad (c) \qquad (14)$$

$$y_{22} = \left.\frac{I_2}{V_o}\right|_{V_i = 0} = \frac{z_{11}}{\Delta_z} \; ; \qquad (d)$$

and

$$\Delta_z = \begin{vmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{vmatrix} = z_{11}z_{22} - z_{12}z_{21} \qquad (15)$$

Substituting (12) into (15) for a reciprocal two-port (so that $z_{12} = z_{21}$ and $y_{12} = y_{21}$) and after regrouping the terms we get

$$\Delta_z = \frac{1}{\Delta^2}(\Delta_{11}\Delta_{22} - \Delta_{12}^2) + (\Delta_{11}\Delta_{2'2'} - \Delta_{12'}^2) - 2(\Delta_{11}\Delta_{22'} - \Delta_{12'}\Delta_{21}) \qquad (16)$$

Next, recall the following identity for determinants*

$$\Delta_{ab}\Delta_{cd} - \Delta_{ad}\Delta_{cb} = \Delta_{abcd} \tag{17}$$

where a, b, c, d are positive integers less than n (n being the order of the determinant Δ) and Δ_{abcd} is the determinant obtained from by deleting rows a and c and columns c and d from Δ. Upon the application of (17), we can rewrite (16) as

$$\Delta_z = \frac{1}{2}\left(\Delta\Delta_{1122} + \Delta\Delta_{112'2'} - 2\Delta\Delta_{1122'}\right) \tag{18}$$

Substituting (12) and (18) into (14), and simplifying, we find the desired expressions for y_{11}, $y_{12}(=y_{21})$ and y_{22}:

$$y_{11} = \frac{\Delta_{22} + \Delta_{2'2'} - 2\Delta_{22'}}{\Delta_{1122} + \Delta_{112'2'} - 2\Delta_{1122'}} \quad \text{(a)}$$

$$y_{12} = y_{21} = \frac{\Delta_{12'} - \Delta_{12}}{\Delta_{1122} + \Delta_{112'2'} - 2\Delta_{1122'}} \quad \text{(b)} \tag{19}$$

$$y_{22} = \frac{\Delta_{11}}{\Delta_{1122} + \Delta_{112'2'} - 2\Delta_{1122'}} \quad \text{(c)}$$

Expressions in terms of network determinants and cofactors for other network transfer functions may be derived in a similar manner. Such expressions are given here also:

$$z_{12} \triangleq \frac{V_o}{I_1} = \frac{\Delta_{12} - \Delta_{12'}}{\Delta} \quad \text{(transfer impedance function) (a)}$$

$$G_{12} \triangleq \frac{V_o}{V_i} = \frac{\Delta_{12} - \Delta_{12'}}{\Delta_{11}} \quad \text{(voltage-ratio transfer function) (b)}$$

$$Y_{12} = Y_L G_{12} = Y_L\left(\frac{\Delta_{12} - \Delta_{12'}}{\Delta_{11}}\right) \quad \text{(transfer admittance function)(c)} \tag{20}$$

$$\alpha_{12} = Y_L Z_{12} = Y_L\left(\frac{\Delta_{12} - \Delta_{12'}}{\Delta}\right) \quad \text{(current-ratio transfer function) (d)}$$

---

* See, for example, Determinants and Matrices, by A. C. Aitken, 8th Ed., published by Interscience, 1954.

THEOREMS AND TOPOLOGICAL FORMULAS

The topological formulas for the various network functions of a passive one-port or two-port are derived from the following theorems:

Theorem 1*. Let N be a passive network without mutual inductances. The determinant Δ of the node admittance matrix Y_n is equal to the sum of all the tree-admittances of N, where a tree-admittance product $T^{(i)}(y)$ is defined to be the product of the admittance of all the branches of the tree $T^{(i)}$. That is,

$$\Delta = \det Y_n = \sum_i T^{(i)}(y) \quad . \tag{21}$$

Theorem 2*. Let Δ be the determinant of the node admittance matrix Y_n of a passive network N with $n+1$ nodes and without mutual inductances. Also let the reference node be denoted by $1'$. Then the (j,j) cofactor Δ_{jj} of Δ is equal to the sum of all the 2-tree-admittance products $T_{2_{j,1'}}(y)$ of N, each of which contains node j in one part and node $1'$ in the other. That is,

$$\Delta_{jj} = \sum_k T_{2_{j,1'}}^{(k)}(y) \tag{22}$$

where the summation is taken over all the 2-tree-admittance products of the form $T_{2_{j,1'}}(y)$.

Theorem 3*. The (i,j) cofactor Δ_{ij} of Δ of a relaxed passive network N with n independent nodes (with node $1'$ as the reference node) and without mutual inductances is given by

$$\Delta_{ij} = \sum_k T_{2_{ij,1'}}^{(k)}(y) \tag{23}$$

where the summation is taken over all the 2-trees-admittance products of the form $T_{2_{ij,1'}}(y)$ with each containing nodes i and j in one connected port and the reference node $1'$ in the other.

---

* For a complete proof of each of these theorems the interested reader may refer to either Ref. [1] or Ref. [3].

For example, the topological formulas for the driving-point function of a passive one-port can be readily obtained from (21) and (22) in Theorems (1) and (2) as stated in the next theorem.

Theorem 4. With the same notation as in Theorems (1) and (2), the driving-point admittance $Y_d(s)$ and the driving-point impedance $Z_d(s)$ of a passive one-port containing no mutual inductances at terminals 1 and 1' are given by

$$Y_d(s) = \frac{\Delta}{\Delta_{11}} = \frac{\sum_i T^{(i)}(y)}{\sum_k T_{2_{1,1'}}^{(k)}(y)} \quad \text{and} \quad Z_d(s) = \frac{\Delta_{11}}{\Delta} = \frac{\sum_k T_{2_{1,1'}}^{(k)}(y)}{\sum_i T^{(i)}(y)} \tag{23}$$

respectively.

For convenience, we define the following shorthand notation:

(a) $V(Y) \triangleq \sum_i T^{(i)}(y)$ = sum of all tree-admittance products (24)

and

(b) $W_{j,r}(Y) \triangleq \sum_k T_{2_{j,r}}(y)$ = sum of all 2-tree-admittance products with node j and the reference node r contained in different parts

Thus, (23) may be rewritten as

$$Y_d(s) = \frac{V(Y)}{W_{1,1'}(Y)} \quad \text{and} \quad Z_d(s) = \frac{W_{1,1'}(Y)}{V(Y)} \tag{25}$$

In a two-port network N, there are four nodes to be specified, namely nodes 1 and 1' at the input port (1,1'), and nodes 2 and 2' at the output port (2,2'), as illustrated in Fig. 6. But, for a 2-tree of the type $T_{2_{ij,1'}}$, only three nodes have been used, thus leaving the fourth one unidentified.

With very little effort, it can be shown that, in general, the following relationship holds:

$$W_{ij,1'}(Y) = W_{ijk,1'}(Y) + W_{ij,k1'}(Y)$$

or, simply,

$$W_{ij,1'} = W_{ij,1'} + W_{ij,k1'} \tag{26}$$

where i, j, k, and 1' are the four terminals of N with 1' denoting the datum node. The symbol $W_{ijk,1'}$ denotes the sum of all the 2-tree-admittance products, each containing nodes i, j, and k in one connected part and nodes k and i in the other.

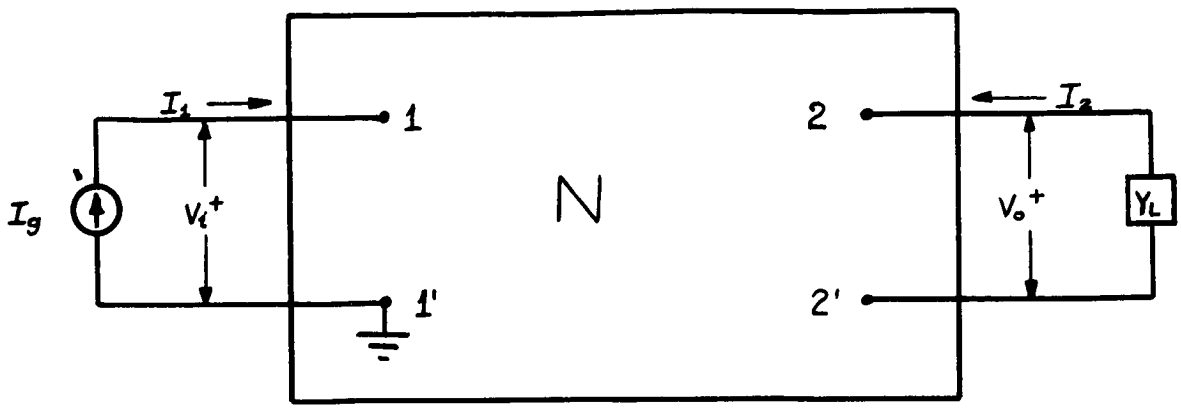


Fig. 6. A loaded passive two-port N .

We now state the next theorem.

Theorem 5. With the same hypothesis and notation as stated earlier in this paper,

$$\Delta_{12} - \Delta_{12'} = W_{12,12'}(Y) - W_{12',1'2}(Y) \tag{27}$$

It is interesting to note that (27) is stated by Percival* in the following descriptive fashion:

$$\Delta_{12} - \Delta_{12'} = W_{12',1'2'} - W_{12',1'2} = \left(\begin{matrix}1 & 2\\ 1' & 2'\end{matrix}\right) - \left(\begin{matrix}1 & 2\\ 1' & 2'\end{matrix}\right)$$

which illustrates the two types of 2-trees involved in the formula.

Hence, we state the topological formulas for a_{11}, z_{12}, and z_{22} in the following theorem:

Theorem 6. With the same hypothesis and notation as stated earlier in this Chapter,

$$z_{11} = \frac{W_{1,1'}(Y)}{V(Y)} \quad (a) \quad ; \quad z_{12}=z_{21} = \frac{W_{12,1'2'}(Y)-W_{12',1'2}(Y)}{V(Y)} \quad (b) \tag{28}$$

and

$$z_{22} = \frac{W_{2,2'}(Y)}{V(Y)} \quad (c)$$

* See Ref. [4] for a detailed discussion.

We shall now develop the topological expressions for the short-circuit admittance functions. Let us denote by $U_{a,b,c}(Y)$ the sum of all 3-tree-admittance products of the form $T_{3_{a,b,c}}(y)$ with identical subscripts in both symbols to represent the same specified distribution of vertices, Then, following arguments similar to those of Theorem 5, we readily see that

$$\Delta_{1122} = \sum_i T^{(i)}_{3_{1,2,1'}}(y) \triangleq U_{1,2,1'}(y) \qquad (a)$$

$$\Delta_{112'2'} = \sum_j T^{(j)}_{3_{1,2',1'}}(y) \triangleq U_{1,2',1'}(Y) \qquad (b) \qquad (29)$$

$$\Delta_{1122'} = \sum_k T^{(k)}_{3_{1,22',1'}}(y) \triangleq U_{1,22',1'}(Y) \qquad (c)$$

where 1, 1', 2, 2' are the four terminals of the two-port with 1' denoting the reference node (Fig. 6). However, we note that, in (29 (a)) and (29(b)) only three of the four terminals have been specified. We can therefore further expand $U_{1,2,1'}$ and $U_{1,2;1'}$ to obtain the following :

$$\Delta_{1122} + \Delta_{112'2'} - 2\Delta_{1122'} = U_{12',2,1'} + U_{1,2,1'2'} + U_{12,2',1'} + U_{1,2',1'2} \qquad (30')$$

For convenience, we shall use the short-hand notation ΣU to denote the sum on the right of (30'). Thus, we define:

$$\Sigma U \triangleq U_{12',2,1'} + U_{1,2,1'2'} + U_{12,2',1'} + U_{1,2',1'2} \qquad (30)$$

Hence, we obtain the topological formulas for the short-circuit admittances as stated in the followingtheorem:

Theorem 7. The short-circuit admittance functions y_{11}, y_{12}, and y_{22} of a passive two-port network with no mutual inductances are given by

$$y_{11} = \frac{W_{2,2'}}{\Sigma U} \quad (a) \; ; \quad y_{12} = y_{21} = \frac{W_{12',1'2} - W_{12,1'2'}}{\Sigma U} \quad (b) \; ; \qquad (31)$$

$$y_{22} = \frac{W_{1,1'}}{\Sigma U} \quad (c)$$

where ΣU is defined in (30) above.

Finally, following similar developments, other network functions are stated in Theorem 8.

Theorem 8. With the same notation as before,

$$Z_{12}(s) = \frac{W_{12,1'2'} - W_{12',1'2}}{V} \quad (a)$$

$$Y_{12}(s) = Y_L \left(\frac{W_{12,1'2'} - W_{12',1'2}}{W_{1,1'}}\right) \quad (b)$$

$$G_{12}(s) = \frac{W_{12,1'2'} - W_{12',1'2}}{W_{1,1'}} \quad (c) \qquad (32)$$

$$\alpha_{12}(s) = Y_L \left(\frac{W_{12,1'2'} - W_{12',1'2}}{V}\right) \quad (d)$$

EXAMPLES

Example 1. Consider the two-port N as shown in Fig. 7(a). The graph G of N is shown in Fig. 7(b).

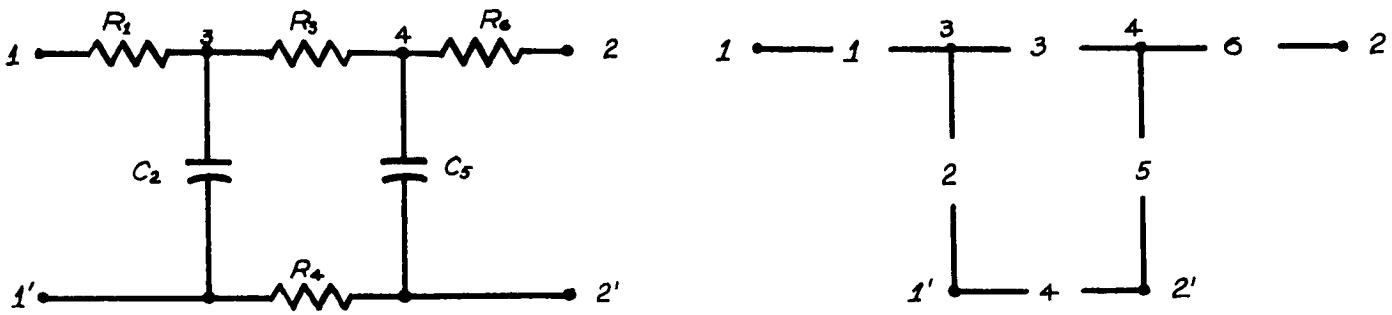


Fig. 7. (a) A passive two-port N, (b) The graph G of N

The admittances of the 6 branches of N are

$$y_1 = \frac{1}{R_1} \equiv G_1 \;;\; y_2 = sC_2 \;;\; y_3 = \frac{1}{R_3} \equiv G_3 \;;$$

$$y_4 = \frac{1}{R_4} \equiv G_4 \;;\; y_5 = sC_5 \;;\; y_6 = \frac{1}{R_6} \equiv G_6$$

The determinant Δ of the node-admittance matrix Y_n of N, taking 1' as the reference node, is given by

$$\Delta = \begin{array}{c|ccccc} & 1 & 2 & 2' & 3 & 4 \\ 1 & y_1 & 0 & 0 & -y_1 & 0 \\ 2 & 0 & y_6 & 0 & 0 & y_6 \\ 2' & 0 & 0 & (y_4+y_5) & 0 & -y_5 \\ 3 & -y_1 & 0 & 0 & (y_1+y_2+y_3) & -y_3 \\ 4 & 0 & -y_6 & -y_5 & -y_3 & (y_3+y_5+y_6) \end{array}$$

which, when expanded, yields

$$\begin{aligned}\Delta = {} & y_1^2y_3y_4y_6 + y_1^2y_4y_5y_6 + y_1^2y_4y_6^2 + y_1y_2y_3y_4y_6 + y_1y_2y_4y_5y_6 \\ & + y_1y_2y_4y_6^2 + y_1y_3^2y_4y_6 + y_1y_3y_4y_5y_6 + y_1y_3y_4y_6^2 + y_1^2y_3y_5y_6 \\ & + y_1^2y_5^2y_6 + y_1^2y_5y_6^2 + y_1y_2y_3y_5y_6 + y_1y_2y_5^2y_6 + y_1y_2y_5y_6^2 \\ & + y_1y_3^2y_5y_6 + y_1y_3y_5^2y_6 + y_1y_3y_5y_6^2 + y_1^2y_5^2y_6 + y_1^2y_4y_6^2 \\ & + y_1^2y_5y_6^2 - y_1y_3^2y_4y_6 - y_1y_3^2y_5y_6 - y_1^2y_5^2y_6 - y_1y_2y_5^2y_6 \\ & - y_1y_3y_5^2y_6 - y_1^2y_4y_6^2 - y_1y_2y_4y_6^2 - y_1y_3y_4y_6^2 - y_1^2y_5y_6^2 \\ & - y_1y_2y_5y_6^2 - y_1y_3y_5y_6^2 - y_1^2y_3y_4y_6 - y_1^2y_4y_5y_6 - y_1^2y_4y_6^2 \\ & - y_1^2y_3y_5y_6 - y_1^2y_5^2y_6 - y_1^2y_5y_6^2\end{aligned}$$

After cancellation of terms, Δ reduces to

$$\Delta = y_1y_2y_3y_4y_6 + y_1y_2y_3y_5y_6 + y_1y_2y_4y_5y_6 + y_1y_3y_4y_5y_6$$

Similarly, after expansion and cancellation of terms in each cofactor, we get

$$\Delta_{11} = y_1y_3y_4y_6 + y_1y_3y_5y_6 + y_1y_4y_5y_6 + y_2y_3y_4y_6 + y_2y_3y_5y_6 + y_2y_4y_5y_6 + y_3y_4y_5y_6$$

$$\Delta_{12} = y_1y_3y_4y_6 + y_1y_3y_5y_6 \quad ; \quad \Delta_{12'} = y_1y_3y_5y_6 \quad ;$$

$$\Delta_{22} = y_1y_2y_3y_4 + y_1y_2y_4y_5 + y_1y_2y_4y_6 + y_1y_3y_4y_5 + y_1y_3y_4y_6 + y_1y_2y_3y_5 + y_1y_2y_5y_6 + y_1y_3y_5y_6 \; ;$$

$$\Delta_{2'2'} = y_1y_2y_3y_6 + y_1y_2y_5y_6 + y_1y_3y_5y_6 \; ;$$

$$\Delta_{22'} = - y_1y_2y_5y_6 - y_1y_3y_5y_6$$

Hence,

$$z_{11} = \frac{\Delta_{11}}{\Delta} = \frac{y_1y_3y_4y_6+y_1y_3y_5y_6+y_1y_4y_5y_6+y_2y_3y_4y_6+y_2y_3y_5y_6+y_2y_4y_5y_6+y_3y_4y_5y_6}{y_1y_2y_3y_4y_6 + y_1y_2y_3y_5y_6 + y_1y_2y_4y_5y_6 + y_1y_3y_4y_5y_6} \quad (a)$$

$$z_{12} = z_{21} = \frac{\Delta_{12} - \Delta_{12'}}{\Delta} = \frac{y_1y_3y_4y_6}{y_1y_2y_3y_4y_6 + y_1y_2y_3y_5y_6 + y_1y_2y_4y_5y_6 + y_1y_3y_4y_5y_6} \quad (b)$$

$$z_{22} = \frac{\Delta_{22} + \Delta_{2'2'} - 2\Delta_{22'}}{\Delta} \quad (33)$$

$$= \frac{y_1y_2y_3y_4+y_1y_2y_3y_5+y_1y_2y_3y_6+y_1y_2y_4y_5+y_1y_2y_4y_6+y_1y_3y_4y_5+y_1y_3y_5y_6}{y_1y_2y_3y_4y_6+y_1y_2y_3y_5y_6+y_1y_2y_4y_5y_6+y_1y_3y_4y_5y_6} \quad (c)$$

Finally, substituting the admittance values into these expressions, we obtain the following results:

$$z_{11} = \frac{(C_2G_3C_5G_6 + C_2G_4C_5G_6)s^2 + (G_1G_3C_5G_6+G_1G_4C_5G_6+C_2G_3G_4G_6+G_3G_4C_5G_6)s + G_1G_3G_4G_6}{(G_1C_2G_3C_5G_6 + G_1C_2G_4C_5G_6)s^2 + (G_1C_2G_3G_4G_6 + G_1G_3G_4C_5G_6)s} \quad (a)$$

$$z_{12} = z_{21} = \frac{G_3G_4G_6}{(C_2G_3C_5G_6+C_2G_4C_5G_6)s^2 + (C_2G_3G_4G_6+G_3G_4C_5G_6)s} \quad (b) \quad (34)$$

$$z_{22} = \frac{(C_2G_3C_5 + C_2G_4C_5)s^2 + (C_2G_3G_4 + C_2G_3G_6 + C_2G_4G_6 + G_3G_4C_5)s + G_3G_4G_6}{(C_2G_3C_5G_6 + C_2G_4C_5G_6)s^2 + (C_2G_3G_4G_6 + G_3G_4C_5G_6)s} \quad (c)$$

Next, we calculate the short-circuit admittance functions. We see that

$$\begin{vmatrix} (y_4 + y_5) & 0 & -y_5 \\ 0 & (y_1 + y_2 + y_3) & -y_3 \\ -y_5 & -y_3 & (y_3 + y_5 + y_6) \end{vmatrix}$$

$$\Delta_{112'2'} = \begin{vmatrix} y_6 & 0 & -y_6 \\ 0 & (y_1 + y_2 + y_3) & -y_3 \\ -y_6 & -y_3 & (y_3 + y_5 + y_6) \end{vmatrix}$$

$$\Delta_{1122'} = \begin{vmatrix} 0 & 0 & -y_5 \\ 0 & (y_1 + y_2 + y_3) & -y_3 \\ -y_6 & -y_3 & (y_3 + y_5 + y_6 \end{vmatrix}$$

After expansion and cancellation of terms, we get

$$\Delta_{1122} = y_1y_3y_4 + y_1y_3y_5 + y_2y_3y_4 + y_2y_3y_5 + y_1y_4y_5 + y_2y_4y_5 + y_1y_4y_6 + y_1y_5y_6 + y_2y_4y_6 + y_2y_5y_6 + y_3y_4y_5 + y_3y_4y_6 + y_3y_5y_6$$

$$\Delta_{112'2'} = y_1y_3y_6 + y_2y_3y_6 + y_1y_5y_6 + y_2y_5y_6 + y_3y_5y_6$$

$$\Delta_{1122'} = y_1y_5y_6 + y_2y_5y_6 + y_3y_5y_6$$

$$\Delta_{1122} + \Delta_{112'2'} - 2\Delta_{1122'} = y_1y_3y_4 + y_1y_3y_5 + y_1y_3y_6 + y_1y_4y_5 + y_1y_4y_6 + y_2y_3y_4 + y_2y_3y_5 + y_2y_3y_6 + y_2y_4y_5 + y_2y_4y_6 + y_3y_4y_5 + y_3y_4y_6$$

Thus,

$$y_{11} = \frac{\Delta_{22} + \Delta_{2'2'} - 2\Delta_{22'}}{\Delta_{1122} + \Delta_{112'2'} - 2\Delta_{1122'}}$$

$$= \frac{y_1y_2y_3y_4+y_1y_2y_3y_5+y_1y_2y_3y_6+y_1y_2y_4y_5+y_1y_2y_4y_6+y_1y_3y_4y_5+y_1y_3y_4y_6}{\{y_1y_3y_4+y_1y_3y_5+y_1y_3y_6+y_1y_4y_5+y_1y_4y_6+y_2y_3y_4+y_2y_3y_5+y_2y_3y_6+y_2y_4y_5+y_2y_4y_6 + y_3y_4y_5 + y_3y_4y_6\}} \quad (a)$$

$$y_{12} = y_{21} = \frac{\Delta_{12'} - \Delta_{12}}{\Delta_{1122} + \Delta_{112'2'} - 2\Delta_{1122'}}$$

$$= \frac{-y_1y_3y_4y_6}{\{y_1y_3y_4+y_1y_3y_5+y_1y_3y_6+y_1y_4y_5+y_1y_4y_6+y_2y_3y_4+y_2y_3y_5 + y_2y_3y_6+y_2y_4y_5 + y_2y_4y_6+y_3y_4y_5+y_3y_4y_6\}} \quad (b)$$

$$y_{22} = \frac{\Delta_{11}}{\Delta_{1122} + \Delta_{112'2'} - 2\Delta_{1122'}} \quad (35)$$

$$= \frac{y_1y_3y_4y_6+y_1y_3y_5y_6+y_1y_4y_5y_6+y_2y_3y_4y_6+y_2y_3y_5y_6+y_2y_4y_5y_6+y_3y_4y_5y_6}{\{y_1y_3y_4+y_1y_3y_5+y_1y_3y_6+y_1y_4y_5+y_1y_4y_6+y_2y_3y_4 + y_2y_3y_5+y_2y_3y_6+y_2y_4y_5+y_2y_4y_6+y_3y_4y_5+y_3y_4y_6\}} \quad (c)$$

With the admittance values substituted into these expressions, we find

$$y_{11} = \frac{(G_1C_2G_3C_5+G_1C_2G_4C_5)s^2+(G_1C_2G_3G_4+G_1C_2G_3G_6+G_1G_2G_4G_6+G_1G_3G_4C_5)s+G_1G_3G_4G_6}{\{(C_2G_3C_5+C_2G_4C_5)s^2 + (G_1G_3C_5+G_1G_4C_5+C_2G_3G_4+C_2G_3G_6+C_2G_4G_6+G_3G_4C_5)s + (G_1G_3G_4+G_1G_3G_6 + G_1G_4G_6+G_3G_4G_6)\}} \quad (a)$$

$$y_{12} = y_{21} = \frac{}{\{(C_2G_3C_5+C_2G_4C_5)s^2+(G_1G_3C_5+G_1G_4C_5+C_2G_3G_4+ C_2G_3G_6+C_2G_4G_6+G_3G_4C_5)s + (G_1G_3G_4+G_1G_3G_6+G_1G_4G_6+G_3G_4G_6)} \quad (b)$$

and

$$y_{22} = \frac{(C_2G_3C_5G_6+C_2G_4C_5G_6)s^2+(G_1G_3C_5G_6+G_1G_4C_5G_6+C_2G_3G_4G_6+G_3G_4C_5G_6)s+G_1G_3G_4G_6}{\{(C_2G_3C_5+C_2G_4C_5)s^2+(G_1G_3C_5+G_1G_4C_5+C_2G_3G_4+C_2G_3G_6+C_2G_4G_6+G_3G_4C_5)s + (G_1G_3G_4+G_1G_3G_6 + G_1G_4G_6 + G_3G_4G_6)\}}$$

Example 2. In this example, we shall calculate the open-circuit and the short-circuit network functions of the two-port shown in Fig. 7(a). (Example 1). From the graph G of the network (Fig. 7(b)), we obtain the set of all trees of G as shown in Fig. 8.

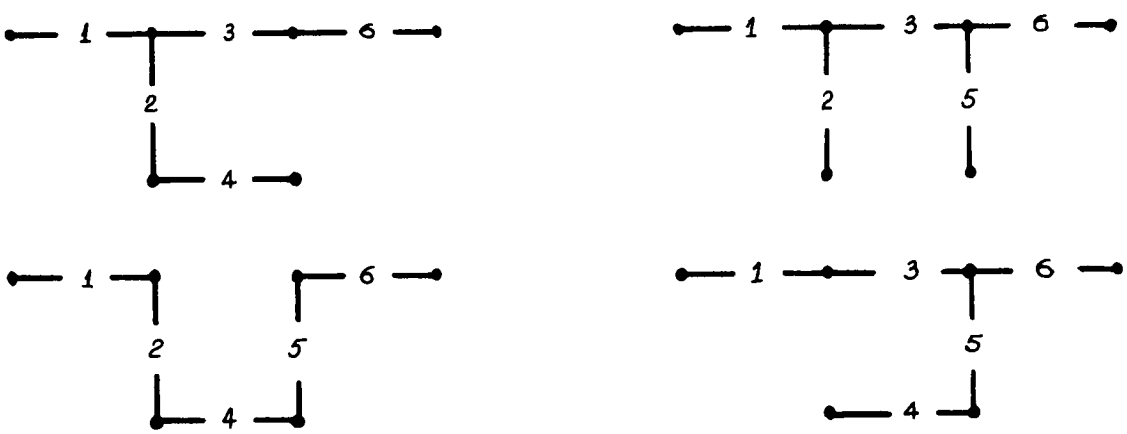


Fig. 8. The set of all trees of G

Thus, from Fig. 8, we get

$$V(Y) = y_1y_2y_3y_4y_6 + y_1y_2y_3y_5y_6 + y_1y_2y_4y_5y_6 + y_1y_3y_4y_5y_6$$

Next, we short vertex 1 to vertex 1' to obtain $G_{11'}$, which is shown in Fig. 9.

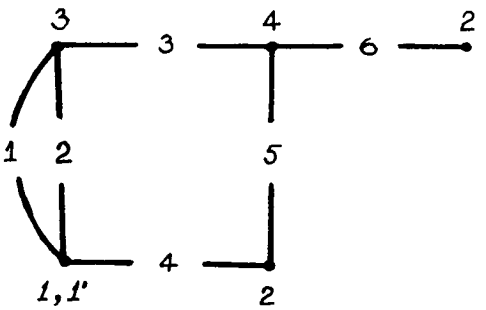


Fig. 9. The graph $G_{11'}$ for $W_{11'}$.

The trees of $G_{11'}$ are shown in Fig. 10.

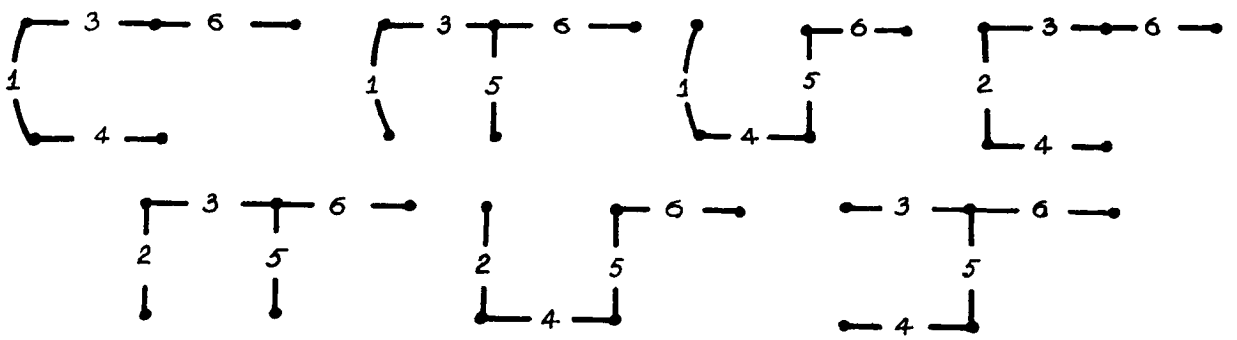


Fig. 10. The set of all trees of $G_{11'}$.

Hence, we get

$$W_{1,1'} = y_1y_3y_4y_6 + y_1y_3y_5y_6 + y_1y_4y_5y_6 + y_2y_3y_4y_6 + y_2y_3y_5y_6 + y_2y_4y_5y_6 + y_3y_4y_5y_6$$

To obtain $W_{2,2'}$, we short-circuit vertex 2 to vertex 2' to obtain $G_{22'}$, as shown in Fig. 11.

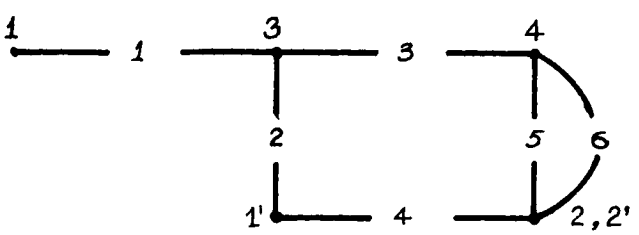


Fig. 11. The graph $G_{22'}$ for $W_{2,2'}$

The trees of $G_{22'}$ (which are 2-trees $T_{2_{2,2'}}$ of G) are shown in Fig. 12.

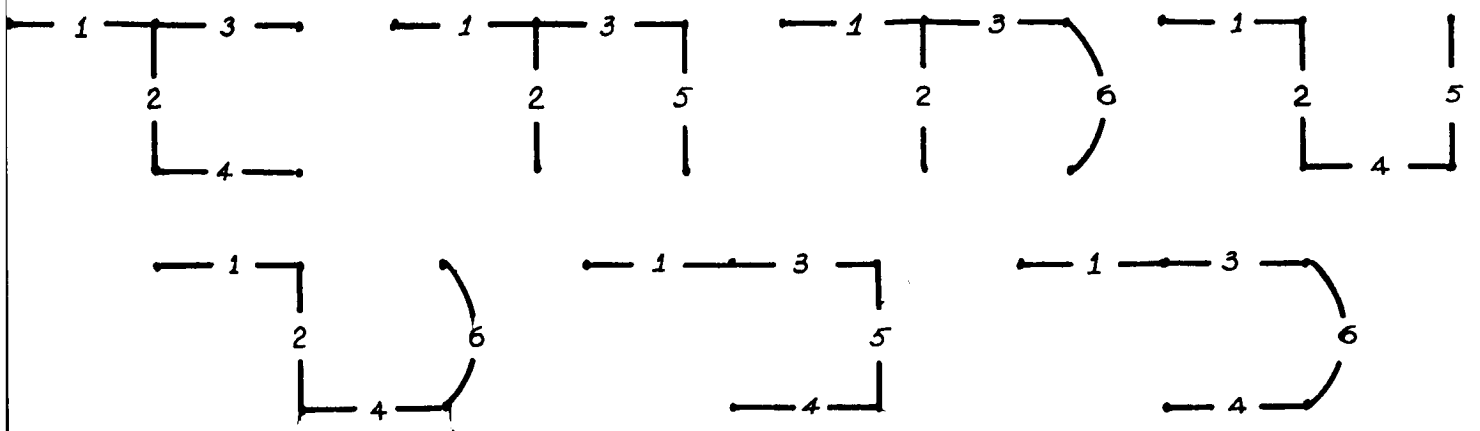
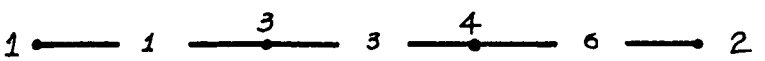


Fig. 12. The set of all trees of $G_{22'}$.

Hence, we find

$$W_{2,2'}(Y) = y_1y_2y_3y_4+y_1y_2y_3y_5+y_1y_2y_3y_6+y_1y_2y_4y_5+y_1y_2y_4y_6+y_1y_3y_4y_5+y_1y_3y_4y_6$$

Next, we find the set of all 2-trees $T_{2_{12,1'2'}}$ of G as shown in Fig. 13. The set of all 2-trees $T_{2_{12',1'2}}$ for the graph G in this example happens to be an empty set.



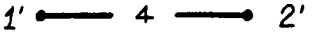


Fig. 13. The set of all 2-trees $T_{2_{12,1'2'}}$ of G.

Thus,

$$W_{12,1'2'} = y_1y_3y_4y_6 \quad \text{and} \quad W_{12',1'2} = 0$$

and

$$z_{11} = \frac{W_{1,1'}}{V}$$

$$= \frac{y_1y_3y_4y_6+y_1y_3y_5y_6+y_1y_4y_5y_6+y_2y_3y_4y_6+y_2y_3y_5y_6+y_2y_4y_5y_6+y_3y_4y_5y_6}{y_1y_2y_3y_4y_6+y_1y_2y_3y_5y_6+y_1y_2y_4y_5y_6+y_1y_3y_4y_5y_6} \qquad (a)$$

$$z_{12} = z_{21} = \frac{W_{12,1'2'} - W_{12',1'2}}{V}$$

$$= \frac{y_1y_3y_4y_6}{y_1y_2y_3y_4y_6+y_1y_2y_3y_5y_6+y_1y_2y_4y_5y_6+y_1y_3y_4y_5y_6} \qquad (b) \qquad (37)$$

$$z_{22} = \frac{W_{2,2'}}{V}$$

$$\frac{y_1y_2y_3y_4+y_1y_2y_3y_5+y_1y_2y_3y_6+y_1y_2y_4y_5+y_1y_2y_4y_6+y_1y_3y_4y_5+y_1y_3y_4y_6}{y_1y_2y_3y_4y_6+y_1y_2y_3y_5y_6+y_1y_2y_4y_5y_6+y_1y_3y_4y_5y_6} \qquad (c)$$

Finally, we shall obtain the short-circuit parameters topologically. Since a 3-tree of G contains 3 edges, we find, examining G in Fig. 11, one 3-tree $T_{3_{12',2,1'}}$, as shown in Fig. 14(a), nine 3-trees $T_{3_{1,2,1'2'}}$, as shown in Fig. 14(b); one 3-tree $T_{3_{12,2',1'}}$, as shown in Fig. 14(c); and one 3-tree $T_{3_{1,2',1'2}}$, as shown in Fig. 14(d).

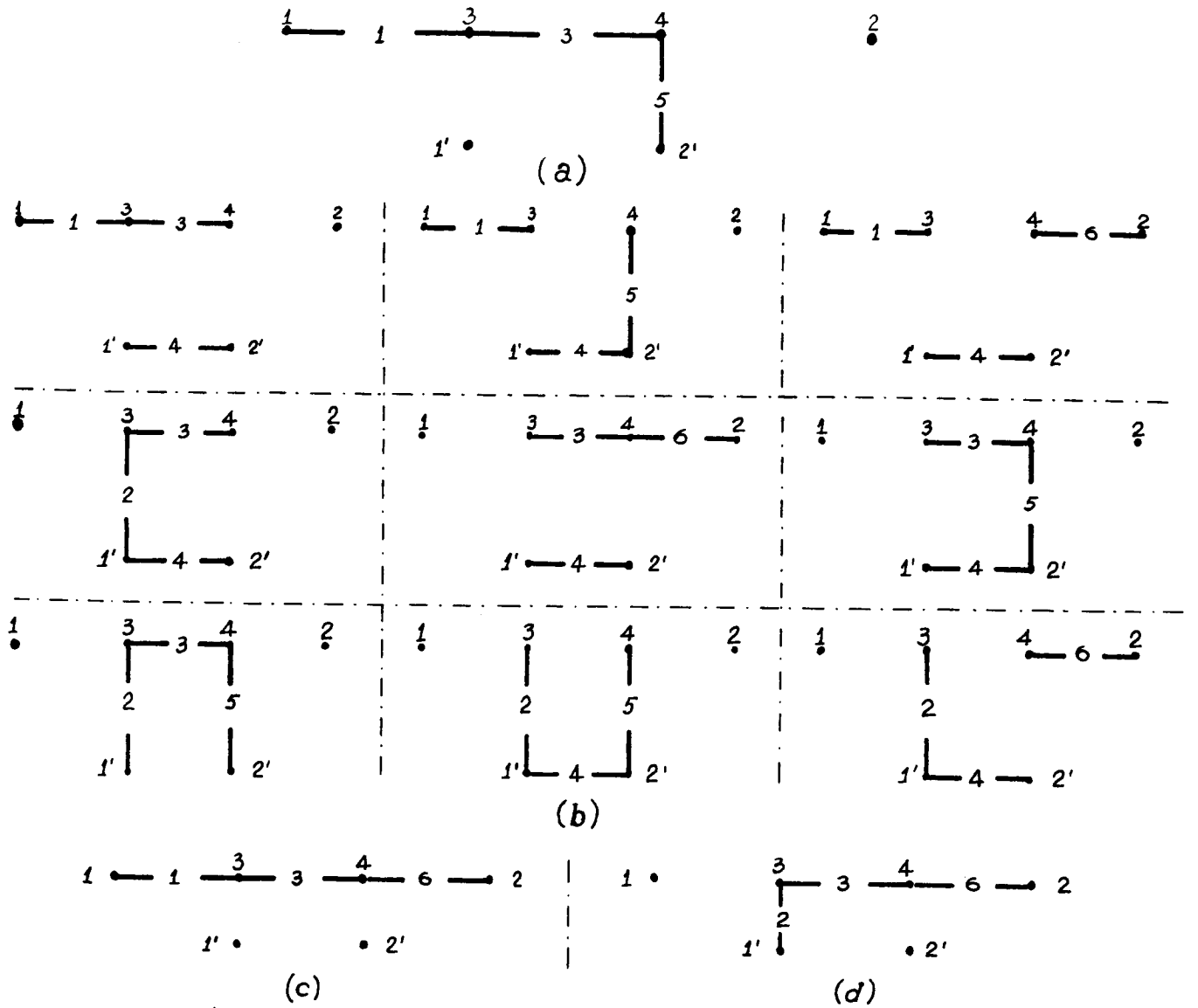


Fig. 14. The 3-trees of G: (a) $T_{3_{12',2,1'}}$, (b) $T_{3_{1,2,1'2'}}$, (c) $T_{3_{12,2',1'}}$ (d) $T_{3_{1,2',1'2}}$

Thus, from (30) and Fig. 14 we get:

$$\Sigma U \triangleq U_{12',2,1'} + U_{1,2,1'2'} + U_{12,2',1'} + U_{1,2',1'2}$$

$$= y_1y_3y_4+y_1y_3y_5+y_1y_3y_6+y_1y_4y_5+y_1y_4y_6+y_2y_3y_4+y_2y_3y_5+y_2y_3y_6+y_2y_4y_5$$

$$+ y_2y_4y_6+y_3y_4y_5+y_3y_4y_6$$

Also, earlier in this example, we have already obtained the expressions for $W_{1,1'}$, $W_{2,2'}$, $W_{12,1'2'}$, and $W_{12',1'2}$. Therefore, we get

$$y_{11} = \frac{W_{2,2'}}{\Sigma U}$$

$$= \frac{y_1y_2y_3y_4+y_1y_2y_3y_5+y_1y_2y_3y_6+y_1y_2y_4y_5+y_1y_2y_4y_6+y_1y_3y_4y_5+y_1y_3y_4y_6}{\{y_1y_3y_4+y_1y_3y_5+y_1y_3y_6+y_1y_4y_5+y_1y_4y_6+y_2y_3y_4+y_2y_3y_5+y_2y_3y_6+y_2y_4y_5+y_2y_4y_6 + y_3y_4y_5 + y_3y_4y_6\}} \quad (a)$$

$$y_{12} = \frac{W_{12',1'2} - W_{12,1'2'}}{\Sigma U} = y_{21}$$

$$= \frac{-y_1y_3y_4y_6}{\{y_1y_3y_4+y_1y_3y_5+y_1y_3y_6y_1y_4y_5+y_1y_4y_6+y_2y_3y_4+y_2y_3y_5+y_2y_3y_6+y_2y_4y_5+y_2y_4y_6 +y_3y_4y_5+y_3y_4y_6\}} \quad (b)$$

$$y_{22} = \frac{W_{1,1'}}{\Sigma U}$$

$$= \frac{y_1y_3y_4y_6+y_1y_3y_5y_6+y_1y_4y_5y_6+y_2y_3y_4y_6+y_2y_3y_5y_6+y_2y_4y_5y_6+y_3y_4y_5y_6}{\{y_1y_3y_4+y_1y_3y_5+y_1y_3y_6+y_1y_4y_5+y_1y_4y_6+y_2y_3y_4+y_2y_3y_5+y_2y_3y_6+y_2y_4y_5+y_2y_4y_6 +y_3y_4y_5+y_3y_4y_6\}} \quad (c)$$

Comparing (33) with (37), and (35) with (38), we see that the topological results are identical to those obtained by direct determinant expansion.

CONCLUSIONS

The main purpose of this paper has been to introduce to the reader the topological approach to the analysis of passive networks. We hope that the above discussions have accomplished this.

One of the obvious advantages of topological formulas over the conventional methods for the evaluation of determinants and cofactors in obtaining the driving-point immittance (i.e., impedance or admittance) functions is that the former avoids the usual cancellations inherent in the expansion of determinants and cofactors in the latter. Furthermore, topological analysis may be considered as a shortcut in evaluating network determinants and cofactors because there is no need to obtain the determinants and cofactors when topological formulas are used. The terms may be obtained directly from the network by inspection. This point has been adequately illustrated by the examples in the preceding section. Thus, topological methods also provide better physical insight to the problem since each term in the expression corresponds to a part of the system. (In the case of a passive one-port, each term in $Y_d(s)$ or $Z_d(s)$ either corresponds to a tree or to a 2-tree.)

Still another advantage of a topological analysis is that the analysis can easily be done by an automatic digital computer. [Ref. 5-10]

There are other advantages as well as disadvantages of the topological methods over the conventional analysis technique that have not been discussed here.[1,11] It is hoped, however, that our brief discussion on network topology in this work will give some inspiration to the interested reader so that he will pursue the subject further and that fruitful results will be derived from it.

REFERENCES

1. S.P. Chan, Introductory Topological Analysis of Electrical Networks to be published by Holt, Rinehart and Winston in 1967.
2. S. Seshu and M.B. Reed, Linear Graphs and Electrical Networks, Addison & Wesley, 1961.
3. W. Mayeda and S. Seshu, "Topological Formulas for Network Functions", University of Illinois Bulletin, vol. 55, No. 23, Nov. 1957.
4. R.H. Percival, "Solution of Passive Electrical Networks by Means of Mathematical Trees", J. Inst. Electrical Engineers (London), vol. 100 (Part III) pp. 143-150, May 1953.
5. W. Mayeda, "Digital Determination of Topological Quantities and Network Functions," Int. Tech. Report No. 6, Contract No. DA-11-022-ORD-1983, Elec. Engr. Research Lab., Univ. of Illinois, January 15, 1957.
6. F.J. MacWilliams, "Topological Network Analysis as a Computer Program," IRE Trans. on Circuit Theory, vol. CT-5, pp. 228-229; September 1958.
7. W. Mayeda and M.E. Valkenburg, "Network Analysis and Synthesis by Digital Computers", 1957 WESCON Convention Record, pt. 2, pp. 137-144.
8. W. Mayeda and M.E. Van Valkenburg, "Analysis of Nonreciprocal Networks by Digital Computer," 1958 IRE National Convention Record, pt. 2, pp. 70-75.
9. W. Mayeda, "Reducing Computer Time in the Analysis of Networks by Digital Computer", IRE Trans. on Circuit Theory, vol. CT-6, No. 1, March 1959.
10. H. Watanabe, "A Computational Method for Network Topology," IRE Trans. on Circuit Theory, vol. CT-7, No. 3, pp. 296-302, Sept. 1960.
11. S.P. Chan,"A Brief Introduction to Topological Techniques in Network Analysis", Electronics, Nov. 14, 1966.

N67 13345

TOPOLOGICAL ANALYSIS OF ACTIVE NETWORKS

R.M. CARPENTER
NASA Electronics Research Center
Cambridge, Massachusetts

ABSTRACT

A compatible series of computer programs is developed for circuits with large numbers of components, typically 30-100. The programs utilize the dichotomy entailed in the flowgraph associated with any active network. Algorithms are established (i) to dichotomize the network into voltage and current generators, (ii) to examine separately each set and the associated system of controls interrelating both sets, (iii) to establish a flowgraph in terms of a unique dichotomous representation for a given equivalent circuit. While the algorithms were developed for linear networks, the dichotomous procedures provide an approach which is not so restricted.

1. PROBLEM FORMULATION

1.1 Need for Dichotomous Techniques

The evaluation of network functions by matrix operations such as mesh or nodal analysis has inherent limitations associated with inversion or partial inversion of matrices. Typical limitations are

(a) Necessity for substantial memory
(b) Inaccuracies due to round-off error
(c) Large running time

The complexity of a matrix-based evaluation routine increases with the number of trees inherent in the network topology, which in turn appears to increase factorially with the number of network elements. As a consequence networks with 20 to 50 components are now considered beyond the capabilities of small computers such as the IBM 1620. To meet this need several alternative avenues are being explored:

(a) Increases in computer memory from 10^7 bits to 10^{10} bits.
(b) Decreases in computer operation times from 10^{-5} to 10^{-10} seconds.
(c) Utilization of relaxation techniques for matrix operations.
(d) Utilization of dichotomous procedures in place of matrix procedures.

Dichotomy is derived from Greek dicha (in two) and temnein (to cut) and refers to separating the system by appropriate cuts into two or more subsystems, followed by an analysis of each subsystem and of the relationships between them. Networks appear to be ideally suited for dichotomy into separate subsystems, which

appear more amenable to computer-oriented methods of analysis:

(a) A subsystem relating through variables, Kirchhoff's current law.

(b) A subsystem relating across variables, Kirchhoff's voltage law.

(c) Interrelations between the above subsystems, Ohm's law.

The dichotomous representation of lumped parameter systems is based on functional dependence between two variables associated with each parameter, and forms the foundations of flowgraph and similar network descriptions (Ref. 1-4). This dichotomy is not related to dicoptics employed by Kron and Happ (Ref. 5).

In Kron's approach dicoptics or tearing of systems into subsystems is essentially a search for judiciously selected subsystems to minimize interconnections and maximize intraconnections. The purpose of dicoptics is to reduce significantly the number of variables to be evaluated, while in the dichotomy entailed in flowgraphs the number of variables is doubled.

The dichotomy here proposed therefore increases the size of the associated matrix by analyzing two properties of each parameter, and aims at a reduction in the complexity of algorithms, not at a reduction in the size of the matrix associated with the system. Therefore, it is essential to distinguish between the concept of subsystem based on dichotomy defined above as separately analyzing two or more properties of each parameter of the system and the concept of subsystem resulting from dicoptics or tearing the system into parts analyzed by properties identical to those used before tearing.

1.2 Equivalent Circuit Description

A network with active and passive elements which may be nonlinear can be represented graphically by an equivalent circuit as shown in Fig. 1. The equivalent circuit describes:

- A set of vertices; each vertex is defined as an electrically distinct point.
- A set of elements; each element is defined by the direction from A, the origin vertex to B, the target vertex.
- The network topology, which is defined by associating with each element a set of vertices A and B. Similarly other network properties are defined by associating with each element other properties described by the symbols C,D,E,F,G,H.
- With each element E is associated a control function C, a control element D, and a generator function G. In controlled sources $D \neq E$ and in passive elements $D = E$.
- For independent sources D is absent; this case does not occur in computer-oriented calculations and is of academic interest only.
- The dichotomy of a network is accomplished by requiring that the properties D and E be split into two mutual categories, denoted by the binary symbols 1 and 0. The dichotomy in control function D is described by property C, and the

dichotomy in generator element E is described by property G.

The equivalent circuit, Fig. 2, contains the information needed to code the circuit in Fig. 1.

1.3 Review of Dichotomous Techniques

The dichotomy is based on constraints imposed by laws of physics and implemented by assigning each function a binary code 0 or 1.

- Across variables or voltages are coded by "0".
- Through variables or currents are coded by "1".
- No two voltage generators may be in parallel.
- No two current generators may be in series.
- The topology of voltage generators forms a tree or Z structure.
- The topology of current generators forms a link or Y structure.

Guidelines, algorithms and resulting programs for selection of an equivalent circuit from the schematic circuit diagram are excluded from the scope of this investigation. It is here assumed that sufficient information to construct the equivalent circuit in Fig. 2 is given. The objective of this investigation is to construct a dichotomous presentation of the network. A flowgraph is defined here as a network description in which an element is identified by:

(a) Two variables, current and voltage.
(b) A functional relation specifying the direction of functional dependence between the variables.
(c) A symbol or numerical relationship denoting this functional relationship.

The flowgraph is, therefore, a dichotomous network description, its specific form - matrix, directed graph or code - is, of course, immaterial and irrelevant.

Since the usage of the terms "block-diagram" and "flowgraph" as synonyms has led to much confusion, it appears appropriate to define block-diagram by the following properties:

(a) An oriented weighted graph with two distinct types of nodes: Contributive or summing point, Distributive or sampling point.
(b) A functional relation between nodes specifying the direction of functional dependence between the variables.
(c) A description of the functional relationship between variables.

The flowgraph has, therefore, distinct features which distinguish it from the block-diagram in several aspects:

(a) A flowgraph is a unique description of a network while a block-diagram description entails arbitrary

and subjective decisions.

(b) The functional dependence describing any network element expresses always a distributive node as a function of a contributive node, never otherwise.

(c) The functional relations governing interconnections between elements are dependent only on the network topology and are restricted to the three values -1, 0, +1. The functional dependence expresses always a contributive node in terms of distributive nodes, never otherwise.

From a draftsman's vantage point a flowgraph as defined here may, of course, be drawn as a block-diagram without violating the above definitions of either flowgraph or block-diagram.

1.4 The Closed System Formulation

The H tag of the computer input relates the quantities to be determined by the equivalent circuit and thus defines the problem uniquely, which, in turn, is essential for the construction of a unique dichotomous solution.

For a desired circuit response, such as a transfer function, current gain or output impedance, all entries except one will be coded "0" for a real network element, with an unknown "virtual" element coded a "1". To convert the open network to a closed network by inserting a fictitious element the system is viewed as a "black box" with input and output terminals and an unknown element connecting the two terminals. The element to be determined is usually coded E = 1. The artifice of setting $E(1) = 0 + j1$ is often used for computation of the unknown response.

For a closed system a constraint always exists which relates the network parameters: H = 0. If the unknown is E(1) = j, then H can be expanded as

$$H(j) = H(\bar{j}) + jH(j')$$

where $H(\bar{j}) = Re(H)$ and $H(j') = Im(H)$

In practice H is computed as a complex number, j is redefined as an unknown and H(j) = 0 is solved for the unknown $j = -H(\bar{j})/H(j')$.

1.5 Illustrative Example of Network Code

The voltage regulator, Fig. 1, is coded in Table 1. The controlled source, E = 3, specifies the direction of positive current flow from node 3 to node 1. A frequency independent current source is controlled by the current through element E = 6, and thus is encoded by C = 1, D = 6, F = 0, G = 1. The transmittance E = 8 with value E(8) = j1 is the unknown.

The problem solution is presented in coded form in Table 2. Thus algorithms are to be developed to produce computer print-out from the coded network. The computer readout, Table 2, is equivalent to the flowgraph in Fig. 4.

2. ALGORITHMS FOR FLOWGRAPH CONSTRUCTION

2.1 Flowgraph Construction

The dichotomous presentation consists of three basic flowgraph regions: V,W, for voltage and current, these are strictly algebraic relations; the third region T contains differential equations or subroutines, which are referred to as transmittances. To derive these relationships from the problem statement by the algorithms proceed as follows:

(1) The V matrix contains voltage relationships; the tree branches or G = 1 elements are determined in terms of the link branches or G = 0 elements.

(2) The W matrix contains current relationships of the link branches or G = 0 elements considering the tree branches or G = 1 elements as knowns.

(3) The T matrix contains the transmittances relating W and V matrices; the known and unknown variables are determined in code G in the problem statement.

(4) The assignment of known and unknown variables is opposite in the V and W subsystems from that in the T subsystem.

The V,W, and T matrices then form a closed flowgraph.

2.2 Coding Convention for Flowgraph

The V and W matrices denote interconnections between elements or Kirchhoff constraints and the T matrix describes the intra-connection or transmittance within each element. The flowgraph is interpreted in terms of signal flow as follows:

(a) The T,V, and W matrices denote signal flow from column to row entry.

(b) Signal flow in interconnections occurs from distributive to contributive nodes.

(c) Signal flow in intraconnections occurs from contributive to distributive nodes.

(d) Entries are finite; infinity is not permissible.

(e) Entries "+" or "-" imply +1 or -1; blanks imply zero.

Table 2 and Fig. 4 illustrate equivalent descriptions of a flowgraph.

2.3 Properties of V Matrix

The number of columns with entries equals the number of G = 1 elements in the equivalent circuit. Each G = 1 element is associated with one or more G = 0 elements to form a tie set, or closed sequence of adjoining elements. The algorithm must, therefore, examine the vertex origins and terminations of successive elements specified in inputs A and B, to determine which G = 1 elements must be selected to form a tie set with given G = 0 elements. From the network topology and the directions specified by A and B in the problem statement, the entries of V shown in Table 2 result.

2.4 Construction of V Matrix

Necessary data for construction are

(a) Entries A,B and G in the problem statement, Table 1.

(b) The definition of a tie associated with each G = 0 element.

(c) A definition for the sign associated with each G = 1 element in a given tie.

Establish a set of known voltages V consisting of elements E(i) for which G = 0 as:

$$V = (i/G(i) = 0)$$

and a conjugate set

$$\overline{V} = (i/G(i) = 1)$$

The set $\overline{V}$ contains only elements which are current generators. while the set V contains only voltage generators. The algorithm expresses the unknown voltages in terms of the known voltages.

$$\overline{V} = f(V)$$

There will be an entry +1 or -1 in the V(j,k) position of the V matrix if $j \epsilon \overline{V}$ and $k \epsilon V$. The entry will be ± 1 if E(j) occurs in a tie containing E(k). The entry will be -1 if E(j) and E(k) are in equal directions in the tie and +1 if in the opposite direction.

Select a tie of N elements with one element in $\overline{V}(j)$ and (N-1) elements in V(k) for each element in $\overline{V}(j)$. To illustrate in Fig. 1 for the tie E(4), E(5), E(6), clearly:

$$\overline{V}(j) = \overline{V}(5)$$
$$\underline{V}(k) = (00010100)$$
$$\overline{V}(5) = V(4) + V(6)$$

This may be written as a code:

$$\overline{V}(5) = (000+0+00)$$

This result can usually be read off immediately from the equivalent circuit and is then entered as column 5 of the V matrix, or can be expressed above as an algorithm in terms of inputs A,B and G.

2.5 Properties of W Matrix

The number of columns with entries equals the number of G = 0 elements in the equivalent circuit. Necessary data for construction are

(a) Entries A,B and G in the problem statement, Table 1.

(b) The definition of a cut-set as a single unknown G = 0 element which can be expressed in terms of one or more G = 1 elements on the basis of flow conservation through an imaginary closed surface.

(c) As assignment of sign based on the direction of flow.

Analagous to the V matrix two complementary sets exist:

Unknown currents: $\overline{W} = (i/G(i) = 0)$

Known currents: $W = (i/G(i) = 1)$

The set W will contain only current generators while the set $\overline{W}$ contains only voltage generators. A set of binary coded vectors is required to express the unknown currents in terms of only the knowns.

$$\overline{W} = f(W)$$

To illustrate, consider column W6 of Table 2 which yields W(k) = (01101011) and $\overline{W}$(j) = (10010100) and the entries $\underline{W}$ matrix $\underline{W}$(j,k) = (W(k), $\overline{W}$(j)) as listed.

2.6 Sign Determinator in W(j,k)

Let A($\overline{W}$) be the set of vertex origins of the G = 0 elements, and B($\overline{W}$) be the set of vertex terminations of these elements. In coded form from Table 1 A($\overline{W}$) = (20030300) and B($\overline{W}$) = (10040100) where a 0 indicates a G = 0 element. To reduce the length of the code rewrite A($\overline{W}$) = N(0110)and B(W) = N(1001) where N(0110) is understood to be a set containing only vertices 2 and 3 of the four vertices in the equivalent circuit.

To determine the sign in W(6) separate the vertices of the equivalent circuit, Fig. 3A, into two complexes. Define complexes A(C) and B(C) as sets of vertices connected by $\overline{W}$ elements, such that E(6) is the only $\overline{W}$ element connecting the two complexes. From the problem statement Table 1 A(6) = N(0010) and B(6) = N(1000). Algorithms are formulated to identify all G = 0 elements which are connected to A(6). These are A(C) = N(0011). Similarly B(C) = N(1100).

To find all G = 1 elements connecting A(C) and B(C), connect vertices 3 or 4 in A(C) to vertices 1 or 2 in B(C). The elements E(2), E(3), E(5), E(7) and E(8) meet this requirement and are listed in Table 4 with data on A and B from Table 1. It is then determined if A(i) is the A(C) for each E(i) and B(i) is in B(C) as shown in Table 4. The result in code for E(6) is W(6) = (0+-0-0--) as entered in the W(6) position in the $\underline{W}$ matrix of Table 2, which can be interpreted as I(6) = I(2) - I(3) - I(5) - I(7) - I(8).

The signs are then interpreted as the positive direction of current into A(C), the reference direction for the unknown current I(6) being specified from A(C) to B(C) or vertex A(6) = 3 to vertex B(6) = 1.

Similarly for W(4) from Table 1:

$$A(4) = N(0010) \text{ and } B(4) = N(0001)$$

Thus

$$A(C) = N(1110) \text{ and } B(C) = N(0001)$$

To find all G = 1 elements connecting vertices 1, 2 or 3 in A(C) to vertex 4 in B(C), note that E(5) and E(7) meet this requirement; hence W(4) = (0000+0+0).

2.7 Description of T Matrix

The $\underline{T}$ matrix describes the transmittances between the dichoto-

mous variables, current and voltage, as shown in Table 2.

If $D \neq E$, then E is an independent source. If $D = E$, then E is a passive element. Four special cases occur:

(a) $C = 1, G = 0$ - the transmittance is an impedance Z or a current controlled voltage generator.

(b) $C = 0, G = 1$ - the transmittance is an admittance Y or a voltage controlled current generator.

(c) $C = 0, G = 0$ - the transmittance is a dimensionless quantity relating the control and generator, or a voltage controlled voltage generator.

(d) $C = 1, G = 1$ - the transmittance is a dimensionless quantity relating the control and generator, or a current controlled current generator.

The case $D = 0$ denotes that E is an independent source and is not considered here.

3. APPLICATIONS

3.1 Closure of Open Graphs

An open system, Fig. 8, contains strictly dependent and strictly independent variables.

The flowgraph relates the variables X and Y by $Y = GX$, where G is the equivalent transmittance of the system. Closing the system by a "dummy" transmittance T, the "independent" variable is made $G = 1/T$, a function of the dependent variable $X = TY$. Assuming that any closed system is governed by a constraint $H = 0$, which is dependent on the network topology and is referred to as the topology equation, H can be expanded in terms of any parameter

$$H(T) = H(\bar{T}) + TH(T')$$

where $H(T)$ is the part of H devoid of T and $H(T')$ is the part of H which contains T with T factored out. We then solve this equation to obtain

$$T = -H(\bar{T})/H(T') = 1/G$$

The variables X and Y are judiciously chosen; for example, to obtain current gain $I(out)/I(in)$, choose $X = I(out)$ and $Y = I(in)$, so that $I(in) = G\ I(out)$; thus $1/G$ represents the current gain.

Since present techniques of flowgraph evaluation are based on closed systems, it is necessary to specify the problem in terms of of an unknown tagged parameter, the H input included in the flowgraph. No special algorithms are needed for processing it, provided the unknown is always treated as a transmittance associated with an element, not as a transmittance associated with a constraint due to interconnections.

3.2 Further Example of Computer Run

A second example is presented to illustrate

(a) The problem statement, including tagging of the unknown parameter.

(b) Construction of the flowgraph, including the closing parameter.

(c) Numerical values in proper form for future processing.

Fig. 5 shows the equivalent circuit of a transistorized band-pass amplifier (Ref. 6) coded for 1/G = V(out)/I(in), with Fig. 6 the coded equivalent circuit. Table 5 gives the problem statement and Table 6 the computer print-out from which the flowgraph in Fig. 7 is constructed.

Copies of this computer program have been qualified and are available at a nominal charge through project COSMIC, University of Georgia (Ref. 7).

3.3 Unique Features of Dichotomous Approach

The dichotomous approach here presented entails several original and unique features of primary interest to the systems analyst:

- A systematic procedure of problem formulation in terms of a closed system. This implies that the unknown to be evaluated forms an integral part in setting up the problem.
- An approach to circuit evaluation and design based on algorithms. This implies that the engineer formulate his problem to be understandable to a computer programmer who is not expected to acquire an understanding of the engineering aspects of the problem.
- A splitting-up of a larger system of equations into several "simpler" systems which can be analyzed with greater facility. The functional relationships describing the overall system are separated into two sets of algebraic relationships (Kirchhoff's voltage and current laws), and a set of first order differential equations relating them, if the system is replaced by an appropriate subroutine.

3.4 Extensions of the Method

The program described has been extended to include the evaluation of the flowgraph. The algorithms upon which this program is based will be the subject of a forthcoming report, (NASA/ERC/CQ 66-676). In addition, subroutines have been formulated to include sensitivity and reliability analyses, frequency response and transient response, as well as reduction techniques to obtain simplified flowgraph models.

REFERENCES

1. W. W. Happ and D. E. Moody, "Topological Techniques for Sensitivity Analysis", IEEE Transactions PG-ANE-11, No. 4, pp. 228-254, December, 1964.

2. W. W. Happ and J. B. Compton, "Flowgraph Models of Integrated Circuits and Film-Type Systems, IEEE Transactions on Aerospace, AS-2, pp. 259-271. 1964.

3. E. R. Robbins, D. E. Moody and W. W. Happ,"Modelling Piece-Wise Linear Systems by Topological Techniques", IEEE Transactions on Aerospace, pp. 252-264, June 1966.

4. W. W. Happ, "Flowgraphs as a Teaching Aid", IEEE Transactions on Education, pp. 69-80, July 1966.

5. W. W. Happ, "Special Cases of Orthogonal Networks", IEEE Transactions on Power Apparatus and Systems, Vol PAS-85, No. 8, August 1966.

6. D. A. Calahan, "Computer Design of Linear Frequency Selective Networks", Proceedings of the IEEE, Vol. 53, No. 11, November 1965.

7. "A Listing of Computer Programs Available from NASA Project COSMIC", Computer Center, University of Georgia, Athens, Georgia, August 1966.

8. R. M. Carpenter and W. W. Happ, "Algorithms for the Dichotomous Representation of Macrocircuits", Proceedings of the Fourth Annual Allerton Conference on Circuit and System Theory, October, 1966.

This lecture is based on a paper to be published as Ref. 8.

Table 1. Problem Statement for Voltage Regulator (Fig. 1)

A	B	C	D	E	F	G	H	Numerical
2	1	0	0	1	0	0	0	.500 E+2
2	3	0	2	2	0	1	0	.400 E-1
3	1	1	6	3	0	1	0	.100 E+3
3	4	0	0	4	0	0	0	.250 E+2
4	1	0	5	5	0	1	0	.100 E-1
3	1	1	6	6	0	0	0	.100 E+6
4	1	0	7	7	0	1	0	.200 E-1
3	1	1	8	8	0	1	1	.100 E+1

Table 2. Problem Solution - Flowgraph Construction

V	1	2	3	4	5	6	7	8	W	1	2	3	4	5	6	7	8	T	1	2	3	4	5	6	7	8
1		+							1									1	E							
2									2	-					+			2		Y						
3									3						-			3								
4					+		+		4									4				E				
5									5			I	+		-			5			I		Y			
6		-	-		+		+	+	6									6						Z		
7									7				+		-			7							Y	
8									8						-			8								Z

Table 3. Criteria for Dichotomous Assignment of C and G

G or C	Variable	Example	G	No Elements
0	Across	Voltage	0	In parallel
1	Through	Current	1	In series

	Function	Node			
			Y	Admittance	C = 0, G = 1
C	Control	Contributive	Z	Impedance	C = 1, G = 0
			W	Amplification	C = 1, G = 1
G	Generator	Distributive	V	Amplification	C = 0, G = 0

Table 4. Choice of Entries for W Matrix

E(i)	A(i)	B(i)	Is A in A(C)?	Is B in B(C)?	±
2	2	3	No	No	+
3	3	1	Yes	Yes	-
5	4	1	Yes	Yes	-
7	4	1	Yes	Yes	-
8	3	1	Yes	Yes	-

Table 5. Problem Statement for Band-Pass Amplifier (Fig. 5)

A	B	C	D	E	F	G	H	Numerical
2	1	1	1	1	0	0	0	.500 E+2
2	1	0	2	2	1	1	0	.400 E-5
2	1	0	3	3	2	1	0	.250 E-4
2	3	0	4	4	0	1	0	.500 E+2
3	1	1	5	5	0	0	0	.500 F+8
3	1	0	6	6	1	1	0	.200 E-8
3	4	0	7	7	1	1	0	.500 E-11
4	1	0	5	8	0	1	0	.500 E+2
4	1	0	9	9	0	1	0	.400 E+4
4	1	0	10	10	1	1	0	.100 E-8
4	1	1	11	11	2	0	0	.100 E-2
1	2	0	11	12	0	1	1	.100 E+1

Table 6. Computer Output for Flowgraph Construction

V	1	2	3	4	5	6	7	8	9	10	11	12
1		+	+	+								+
2												
3												
4												
5			-			+	+					
6												
7												
8												
9												
10												
11							-	-	+	+		
12												

W	1	2	3	4	5	6	7	8	9	10	11	12
1												
2	-											
3	-											
4	-				+							
5												
6					-							
7					-							+
8												-
9												-
10												-
11												
12	+											

T	1	2	3	4	5	6	7	8	9	10	11	12
1	Z											
2		Y										
3			Y									
4				Y								
5					Z							
6						Y						
7							Y					
8												
9									Y			
10										Y		
11											Z	Y
12												

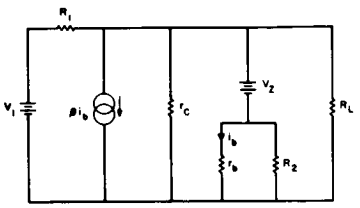


FIG 1 Equivalent circuit for voltage regulator

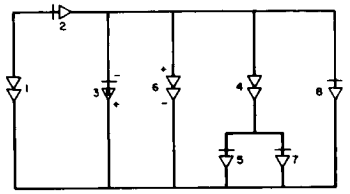


FIG. 2 Coded equivalent circuit, exhibiting dichotomy

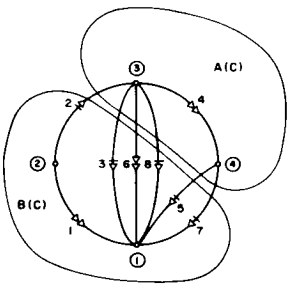


FIG. 3A Separation of equivalent circuit into cut-set for the evaluation of I (6)

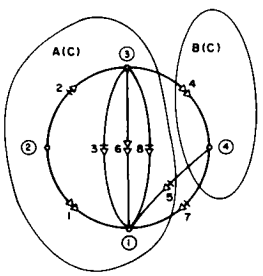


FIG. 3B Separation of equivalent circuit into cut-set for the evaluation of I (4)

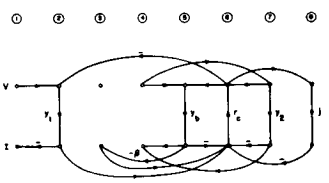


FIG 4 Flowgraph constructed from V, W, and T matrices

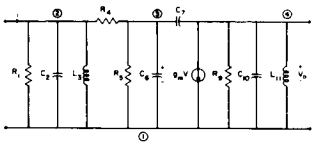


FIG 5 Band-pass amplifier equivalent circuit

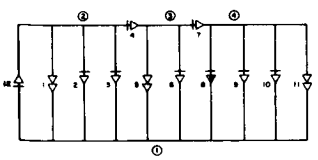


FIG 6 Coded equivalent circuit showing dichotomy

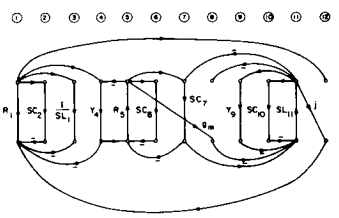


FIG 7 Flowgraph constructed from computer output V, W, T matrices

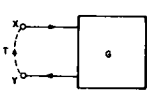


FIG 8. Closed System Representation, Y=GX, X=TY

N67 13346

MATRICES AND STATE VARIABLES*

JOHN STAUDHAMMER
Professor of Engineering
Arizona State University
Tempe, Arizona

ABSTRACT

Basic algorithms for matrix manipulations by computer are presented. Matrix formulation and solution of the network equations for a linear lumped parameter network are given. An example for these is shown. Advantages and disadvantages of this procedure are briefly discussed. References are made to pertinent literature.

INTRODUCTION

The greatest advantage of matrices in the theory of networks is in the systematic manner in which analysis may be carried out.

The intent of this paper is to introduce some computational algorithms for basic matrix operations, hoping that a short introduction to them will encourage engineers not already employing computers to avail themselves of this tool. Networks formulated as matrix equations will be shown to be readily analyzed by computer techniques; such as analysis is basic to any design.

The viewpoint adopted here is that the effect of changes made by the designer in the circuit may be observed in the results obtained by matrix analysis. Fundamentally, matrix methods are analysis tools; their usefulness to a designer of circuits lies in yielding readily to computation by standard computer programs found in every computing center.

The example given here is but an insignificantly small sample of applications of matrices to circuits. The vast bulk of literature readily and customarily accessed by circuit designers contains articles dealing with matrix applications to circuits.

One of the drawbacks of matrix methods will be found to be the accumulation of round-off errors. The minimization of such errors is of paramount importance in all matrix calculations. It is not the intent of this discussion to delve into numerically accurate procedures, but the user of any computer center should ascertain the accuracy of the programs he customarily uses. In general it will be found that iterative methods of solution are inherently more accurate.

*This paper is based on material contained in the forthcoming book "Computer-Aided Circuit Design" by W. W. Happ and J. Staudhammer (McGraw-Hill, 1967).

Recently a very good book [14] became available describing various large programs for use with large-scale computers. Unfortunately most of the programs described there can not be utilized without extensive reprogramming; also large computer installations have a relatively large cadre of sophisticated users well able to utilize their particular programs for their particular problems. This paper is intended for those who yet have to employ computers in circuit analysis to any great extent.

MATRIX ALGEBRA

Matrices are well known to electrical engineers and a working knowledge of them is assumed here. Interested readers will find ample references ranging from the practical [1] to the complete [2] exposition of the theory of matrices. However, algorithms for machine computation are summarized here as an aid for mechanization of some calculations.

A matrix is an array of elements, numbers of expressions, arranged in a rectangular pattern. The whole array is referred to by a single designator, say T; any element in the array is designated by a pair of subscripts. It is also customary to use an upper case letter for the array and the corresponding lower case letter for the element. Thus t_{12} would refer to the second element of the first row of the array T. When the whole array is referred to by a single letter, we speak of the whole matrix notation; when the matrix is described as a collection of elements, the term kernel index notation is used:

$$T = \left[t_{ij} \right]$$

Note that a matrix is a collection of numbers and a rule of ordering them in a prescribed pattern of rows and columns. Each row has the same number of elements and all columns must contain equal number of elements too. When the row dimension is one, the matrix is a row vector; a matrix consisting of a single column is called a column vector, or more simply a vector.

Matrix Addition

Two matrices A and B, having elements a_{ij} and b_{ij}, may be added together to give as the result the matrix C by the use of the formula

$$c_{ij} = a_{ij} + b_{ij} \tag{1}$$

provided the subscript ranges on A, B, and C are the same. Thus the matrices must have the same sizes: i.e., must possess the same number of rows and the same number of columns.

Equation 1 gives the algorithm for matrix addition; the equivalent FØRTRAN statements would be

```
      DØ  50   I = 1, N
      DØ  50   J = 1, M                          (2)
50    C(I,J) = A(I,J) +  B(I,J)
```

where the matrices A, B, and C were assumed to have been defined in a DIMENSIØN statement to possess at least N rows and M columns.

Matrix subtractions

$$D = A - B \tag{3}$$

merely introduces a minus in place of the plus in expression (2).

Matrix Multiplication

Two matrices P and Q may be multiplied together to yield the matrix R by the following formula

$$r_{ij} = \sum_{k=1}^{N} p_{ik} \quad q_{kj} \qquad \begin{matrix}(i = 1, \ldots.L)\\(j = 1, \ldots.M)\end{matrix} \tag{4}$$

From the subscript ranges it may be inferred that P has L rows and N columns; Q has N rows and M columns; and R has L rows and M columns. In whole matrix notation the above becomes

$$R = P \cdot Q \tag{5}$$

In order to obtain a product it is necessary that the row dimension of P equal the column dimension of Q; such matrices are said to be conformable in the order PQ. The product QP will not exist unless L = M. Since in general L ≠ N the products PQ and QP are not even of the same dimensions; thus in general

$$PQ \neq QP \tag{6}$$

This result is true even if L = M = N.

The FØRTRAN statements corresponding to the algorithm given in expression 4 are

```
      DØ  100  I = 1, L
      DØ  100  J = 1, M
      R(I,J) = 0.                                   (7)
      DØ  100  K = 1, N
100   R(I,J) = R(I,J) + P(I,K) * Q(K,J)
```

Simultaneous Equations

A set of linear algebraic equations

$$\begin{aligned}
a_{11} x_1 + a_{12} x_2 + \ldots + a_{1N} x_N &= b_1 \\
a_{21} x_1 + a_{22} x_2 + \ldots + a_{2N} x_N &= b_2 \\
&\vdots \\
a_{N1} x_1 + \ldots. + a_{NN} x_N &= b_N
\end{aligned} \tag{8}$$

may be rewritten as the matrix

$$A\,X = B \tag{9}$$

with

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ \vdots & & & \vdots \\ a_{N1} & \cdots & \cdots & a_{NN} \end{bmatrix}$$

$$X = \begin{Bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_N \end{Bmatrix} \qquad \text{and } B = \begin{Bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_N \end{Bmatrix}$$

The solution of such a set of equations is formally given by Cramer's Rule which computationally is nearly impossible to evaluate for large sets of equations since it requires (N-1)N! multiplications. On a large scale digital computer which performs each multiplication in 10 μs this amounts to 14 1/2 hours for a 12 x 12 matrix.

There exist a number of practical numerical methods for the solution of linear equations[3]. The simplest of these (although not "best") is the Gauss-Jordan procedure which requires about 1/2 N^3 multiplications and divisions for the solution[4]. The process consists of dividing the k-th equation by the coefficient a_{kk} and subtracting this newly formed equation from all other equations. These steps are done for k = 1, ... N. The method is usually applied to the augmented matrix G consisting of N rows and (N+1) columns by appending to A the vector B:

$$G = \left[A \vdots B\right] \tag{10}$$

In the augmented matrix the Gauss-Jordan algorithm is

$$g'_{kj} = g_{kj} / g_{kk} \qquad j = k+1, \ldots N+1$$

$$g'ij = g_{ij} - g_{ik}\, g_{kj} \qquad i = i \neq k \ldots\ldots N \tag{11}$$

The corresponding FØRTRAN statements are

```
      N1 = N+1
      DØ  90 K = 1,N
      K1 = K+1
      DØ 10 J = K1, N1
   10 G(K,J) = G(K,J) / G(K,K)
      DØ  30  I = 1,N
      IF (I - K)  20,  30,  20
```

```
20    DØ  25  J = K1, N1
25    G (I,J) = G(I,J) - G(I,K) * G(K,J)
30    CØNTINUE
```

Note that the solution vector X will appear in the (N+1)th column of the array G.

The above method will work as long as A(K,K) is not zero. Practical numerical roundoff error dictates the need of pivoting (i.e., having g_{kk} be the largest element in the subarray formed of the k to N rows and columns). Useful library programs should include such procedures; a further development is found in [4].

Matrix Inversion

The inverse B of a matrix A is defined by the relations

$$B \cdot A = A \cdot B = I \tag{12}$$

where I is a unit matrix (i.e., has zeros everywhere except that the main diagonal is all ones.)

Since the inverse may be post multiplying or premultiplying the original matrix and result in the same unit matrix, the matrix A must be square; i.e., it must have as many rows as columns. Non-square matrices possess no inverses.

The simplest (but not best) way of calculating the inverse matrix uses the Gauss-Jordan procedure mentioned above. The augmenting matrix in this case is a unit matrix I:

$$G = \left[A \;\vdots\; I \right] \tag{13}$$

G now has N rows and 2N columns; the formulas of expression 11 apply with j = K + 1, ... 2N.

The FØRTRAN statements for the creation of the matrix G and the Gauss-Jordan inversion are given below:

```
      N1 = N+1
      N2 = 2 * N
      DØ  8   I = 1, N
      DØ  5   J = 1, N
 5    G(I,J) = A(I,J)
      DØ  6   J = N1, N2
 6    G(I,J) = 0.
 8    G(I, N+I) = 1.
      DØ  90  K = 1, N
      K1 = K + 1
      DØ  10   J = K1, N2
10    G(K,J) = G(K,J) / G(K,K)
      DØ  30   I = 1,N
      IF (I-K) 20, 30, 20
```

(14)

```
20    DØ 25 J = K1, N2
25    G(I,J) = G(I,J) - G(I,K) * G(K,J)
30    CØNTINUE
90    CØNTINUE
```

The inverse of A will appear in columns N+1 to 2N of the array G.

Again control of numerical round-off dictates the use of pivoting; and economy in machine utilization requires that the matrix A be inverted in place. The above program may be modified to meet these requirements; such a version is discussed in [4].

Diagonalization

A given matrix Q may be decomposed into the product

$$Q = V\ P\ V^{-1} \tag{15}$$

where

$$P = \begin{bmatrix} P_1 & & & 0 \\ & P_2 & & \\ & & \ddots & \\ 0 & & & P_N \end{bmatrix}$$

and

$$V = \left[v_1 \vdots v_2 \vdots \cdots \vdots v_N \right]$$

provided the P_i are distinct.

The values $P_1, \ldots P_N$ are the eigenvalues, and the vectors $v_1, \ldots v_N$ are the eigenvectors of A.

Due to the special nature of P the matrix Q is the sum

$$Q = \sum_{k=1}^{N} P_k\ v_k\ w_k \tag{16}$$

where w_k is the k-th row of V^{-1}. (Expression 16 may be verified by direct multiplication of equation 15.)

The calculation of eigenvectors and eigenvalues is a major task for arbitrarily given Q; for a readable treatment of several useful methods the reader is referred to [3]. Here a method will be given which is useful in medium-sized problems; a theoretical justification is given in [5].

Starting with an arbitrary non-zero vector x_o the following sequence of vectors is calculated:

$$\begin{aligned} A\,x_o &= x_1 \\ A\,x_1 &= A^2\,x_o = x_2 \\ A\,x_2 &= A^3\,x_o = x_3 \\ &\vdots \\ A\,x_{(s-1)} &= x_{(s)} \end{aligned} \tag{17}$$

then for a sufficiently high s

$$x_{(s)} \cong P_1{}^s v_1$$

where P_1 is the largest eigenvalue in A.

The value of s is ascertained from the elements of the vector $x_{(s)}$:

$$\frac{x_{(s),i}}{x_{(s-1),i}} = P_1 \qquad i = 1, \ldots . N \tag{18}$$

i.e., when the ratios of two successive iterations of x approach a constant, this constant is P_1. In practice one element of the vector $x_{(s-1)}$ is made 1 by factoring out its value and discarding it. The reduced vector is then v_1. When v_1 does not change upon iteration, the process converged.

The iterative process described in 17 will converge if the P's are distinct and real. For complex P's a modification of the above, also described in [5], may be used.

The row inverse eigenvector w_1 is obtained by starting with an arbitrary non-zero row and premultiplying A analogously to equation 17. The resultant sequence of rows will converge to w_1 for distinct real eigenvalues.

Having found the largest eigenvalue and the corresponding v and w, the product $P_1\ v_1\ w_1$ is formed and subtracted from the original Q; the result is Q*:

$$Q^* = Q - P_1\ v_1\ w_1 \tag{19}$$

The matrix Q^* does not contain the eigenvalue P_1 any more, hence Q^* may be iterated for the next highest eigenvalue.

The above process is repeated until all eigenvalues and vectors are found.

In the case of complex P_i it is possible to start with an arbitrary nonzero complex vector (must contain at least one real and one imaginary part) and obtain convergence by the criteria described above; all comparisons will have to be made on complex quantities, which is not too difficult using FØRTRAN IV.

Matrix Differential Equations

Linear network equations can be written in the form of a set of first order differential equations:

$$A\,x = \frac{dx}{dt} \qquad x\Big|_{t=o} = x_o \tag{20}$$

where A is a square matrix and x is a vector of voltages and/or currents.

The solution of this state-variable equation is simply [7]

$$x = e^{At}\, x_o \tag{21}$$

where $e^{At} = I + At + \frac{1}{2!} A^2t^2 + \ldots$

By the use of equation 15 a reasonably simple solution is obtained:

$$x = \left[V P^o V^{-1} + V P V^{-1} \cdot t + \frac{1}{2!} V P^2 V^{-1} \cdot t^2 \ldots \right] x_o$$

$$= V \left[P^o + P \cdot t + \frac{1}{2!} P^2 t^2 + \ldots \right] V^{-1}\, x_o \tag{22}$$

However due to the fact that P is a diagonal matrix

$$P^m \cdot t^m = \begin{bmatrix} P_1{}^m t^m & & & 0 \\ & P_2{}^m t^m & & \\ & & \ddots & \\ 0 & & & P_N{}^m t^m \end{bmatrix} \tag{23}$$

and the bracket in equation 22 becomes

$$\left[\ \right] = \begin{bmatrix} e^{P_1 t} & & & 0 \\ & e^{P_2 t} & & \\ & & \ddots & \\ 0 & & & e^{P_N t} \end{bmatrix} \tag{24}$$

Thus by the use of equation 24 the solution vector for the circuit voltages/currents becomes

$$x(t) = V\left[\ \right]V^{-1} \quad x_o \tag{25}$$

Hence the time behavior of the circuit voltages and currents can be calculated simply from a knowledge of the eigenvalues and eigenvectors of the circuit matrix A.

For complex P's the exponentials go over into sine - cosine combinations and the arrays V and V^{-1} will contain complex coefficients.

COMPUTER PROGRAMS FOR MATRIX CALCULATIONS

Every large-scale computer installation has some capability for matrix manipulation. The installation's capability may range from a collection of matrix programs, such as the ones distributed by SHARE, to sophisticated systems which enable the user to express his matrix equations virtually in pure matrix notation. An example of such a system is Lockheed's system FAMAS, but virtually all large computer centers have their own brand of matrix compilers. Usually the sophisticated systems will compile a FØRTRAN IV program or else call on a library of FØRTRAN IV programs.

Here we shall discuss a minimal set of computer programs intended for unsophisticated users on a medium-scale computer, typically an IBM 1620.

Matrix addition and subtraction is usually conveniently handled within the program by insertion of the two nested DØ loops given in equation 2. Matrix multiplication is usually accomplished by the three nested DØ loops given in equation 7. Occasionally it is desirable from clarity standpoint to write these loops as separate subroutines. The subroutine arguments for addition are (A, B, C, M, N). A separate subroutine is usually written for subtraction, with the same argument list. No special provisions need to be made if A = B = C. Matrix multiplication subroutines are more common. The argument list would be (P, Q, R, L, M, N). Precaution must be exercised to have R ≠ P or Q.

A routine for the solution of simultaneous equations is usually provided by the computer manufacturer as part of the software package supplied with the computer. The routines supplied are usually some variants of the Gauss Method, although sometimes the Gauss-Seidel iteration method is utilized. In the former set of routines the determinant of the coefficient matrix is calculated as a by-product. Vanishing determinants mean that no unique solution exists; but unfortunately round-off errors in the calculations often result in small but non-zero determinants even though the exact value of the determinant is zero. Thus it becomes necessary to be able to identify true zero determinants.

The determinant of coefficient matrix will be zero or finite. The magnitude of the smallest finite value of an N by N determinant can be no less than the product of the N smallest elements in the coefficient matrix. Thus any determinant value falling into that range of values must be considered with due suspicion.

The chance of build-up of round-off errors is minimized by the use of iterative procedures, if they converge. In direct numerical procedures, the use of pivoting is indicated. In addition all matrix programs should be readily available in a form that does not require any changes in the compiled programs; specifically the DIMENSIØN statement should not be variable. In FØRTRAN IV this is easily accomplished by including in the calling sequence a parameter, say NA, which contains the numerical value of the array dimension. This number then becomes the variable DIMENSIØN information in the subroutine.

In FØRTRAN II recourse must be taken to relative indexing, where every element is referred to by a single subscript, the location of that element relative to the beginning of the array. Examples of such programs are discussed in (4).

Matrix inversion is another standard routine supplied by the computer manufacturer. Usually an in-place inversion routine is available; i.e., a routine that does not require additional space above the matrix itself.

A package of subroutines for eigenvalue-eigenvector calculation of matrices is available at most large computer installations. Centers using smaller computers usually have access to such programs but often only symmetric matrix routines can be found. The difficulty with unsymmetric matrices is that in general their eigenvalues and vectors are complex.

Accuracy considerations pervade the whole of matrix manipulations by computers. Detailed discussion of the computational advantages and disadvantages of various methods are found in [9]. Basic rules for the minimization of the effects of round-off are the following:

1. Avoid in subsequent calculations the use of the result of the subtraction of two nearly equal numbers.
2. Minimize the total number of arithmetic operations.

NETWORK STATE EQUATIONS

The usual procedure in analyzing an electrical network is to establish the Kirchhoff current law (KCL), the Kirchhoff voltage law (KVL), and the voltage-current relationships for the elements and thus obtain the loop equations and/or mesh equation for a given network.

The equations are then transformed to give the transformed network equations. Such analysis is thoroughly discussed in [10], but a simpler procedure is to utilize the state-variable approach. A thorough discussion of a computer program using this approach is in [11] where general RLC active networks are analyzed. Here an example is shown using only an RC active network.

Suppose that the differential equations (obtained by applying the KCL conditions at each node) for a network are

$$C \frac{dv}{dt} + GV + K = 0 \qquad (26)$$

The computational solution of equation 31 can be carried out by means of equation 25 ("direct computation") or by a faster method based upon direct expansion of the exponential as a power series [12]. The latter procedure is usually much preferred due to its controlled errors and its speed advantage.

TRANSISTOR CIRCUIT EXAMPLE

In order to demonstrate the equations derived above, the circuit given in Figure 1 was analyzed [6]. The equivalent circuit of the transistor is shown in Figure 2. This circuit model allows the rough calculation of DC and AC behavior provided the frequencies of interest are not appreciably higher than f_α. The application here does fulfill that requirement.

The equivalent circuit of the amplifier is shown in Figure 3 with the values indicated in Figures 1 and 2. The circuit differential equations may be written down by inspection.

In this example there are 15 voltages (E_{in}, V_1 ... V_{14}) to be considered; the network equations make up 15 simultaneous linear differential equations. Note that current balance expressions (KCL) for nodes E_{in}, V_4, V_8, and V_{12} do not contain terms involving capacitors. Thus the rank (maximum size of non-zero sub-determinant) of [C] is at most 11. The rank could be less if some of the circuit elements were of particular values. In this case they are not, and the eigenvalues P_1 P_2 ... P_{11} are found as described in equations 15 through 19. These values represent the poles of the system matrix. The dominant roots were found to be

$$P_1 = 556 \text{ and } P_2 = 5.33 \times 10^5$$

corresponding to a low frequency half-power point of 90 cps and an upper half-power frequency of 89 kc.

The upper half power frequency is suspiciously high since the design called for a 20 kc value. Actual measured response of the amplifier was about 40 kc, due to a miscalculation in the value of the 5600 μμf. The half pwr. freq. is calculated from terms which are the sum of the 15 μf coupling capacitor and the 5600 μμf capacitor. Thus the accuracy required in the calculations involving this capacitor is about 1 part in 10^4, a requirement not met by the routines used in the calculations.

The above demonstrates vividly the need of working with a separate high frequency circuit where the interstage coupling capacitors are shorted. Matrix methods will work here also, although some automatic circuit analysis programs will obviate the need to derive different equations valid for given frequency regions. For electronic amplifiers this is standard procedure discussed in any textbook. Note however that the motivation here is a desire for higher accuracy in the calculations rather than the need for easier calculations. A procedure for automating the derivation of approximate circuit models is described in [13].

with C = capacitance matrix; G = conductance matrix; K = DC sources vector; V = voltage vector ("state vector").

It is desired to bring this set of equations into the "normal form"

$$\frac{dV}{dt} = AV + B \tag{27}$$

Formally this may be accomplished by premultiplying eq.26 with C^{-1} and putting every term not containing the derivative on the right side of the equation: this process can be carried out provided that C^{-1} exists. Generally, det C = 0 and hence C^{-1} does not exist. In these cases the set of equations in 26 is reduced to yield a diagonal form for C, say by the Gauss-Jordan procedure outlined earlier. Whenever a whole column of zeros is produced, the column is switched with the first following column having a non-zero element on or below the main diagonal. In this manner the original set (equation 26) is reduced to the following

$$\begin{bmatrix} C_{11} & \vdots & C_{12} \\ \cdots & \cdots & \cdots \\ C_{21} & \vdots & C_{22} \end{bmatrix} \frac{d}{dt} \begin{bmatrix} V_1 \\ \cdots \\ V_2 \end{bmatrix} + \begin{bmatrix} G_{11} & \vdots & G_{12} \\ \cdots & \cdots & \cdots \\ G_{21} & \vdots & G_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ \cdots \\ V_2 \end{bmatrix} + \begin{bmatrix} K_1 \\ \cdots \\ K_2 \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix} \tag{28}$$

with $C_{21} = 0$ and $C_{22} = 0$. In general $C_{12} \neq 0$ and the above sets of linear equations are solved:

$$V_2 = -G_{22}^{-1} \left(K_2 + G_{21} V_1\right) \tag{29}$$

$$\frac{dV_1}{dt} = -\left[C_{11} - C_{12} G_{22}^{-1} G_{21}\right]^{-1} \left[G_{11} - G_{12} G_{22}^{-1} G_{21}\right] V_1 - \left[C_{11} - C_{12} G_{22}^{-1} G_{21}\right]^{-1} \left[K_1 - G_{12} G_{22}^{-1} K_2\right] \tag{30}$$

This is now of the form of equation 27.

Provided that the inversions indicated can be carried out, equation 30 is solved, subject to the initial conditions, in the form

$$V_1(t) = e^{At} V_1(0) - A^{-1} B \tag{31}$$

For the existence of this solution the following must hold: det $G_{22} \neq 0$; det $(C_{11} - C_{12} G_{22}^{-1} G_{12}) \neq 0$; det $(G_{11} - G_{12} G_{22}^{-1} G_{21}) \neq 0$. If either of these determinants vanish (because of particular circuit parameter values) degeneracy exists in the equations and some of the variables in V_1 may be expressed in terms of others in that set. Computer programs exist for automatically performing the elimination of such surplus variables one at a time until the minimum number of variables is found [11].

SUMMARY AND EXTENSIONS

As was demonstrated above matrix manipulations are very easily programmable in a higher order computer language such as FORTRAN. However, efficient programs are much more complicated than the ones indicated here; also, programming for minimum round-off error is a very challenging and exacting task.

Once a library of matrix programs is established it is a simple matter to analyze given sets of circuit equations. After a time the computer user demands more accurate and faster routines which can handle much wider classes of problems. What usually goes unnoticed is that the user becomes accustomed to dealing with problems which he could not do without computer aid. Thus the advantages of matrix procedures include simplicity, modularity, and expandability. Against these must be counted the relative cumbersomeness of inputting, accuracy and timing difficulties (unless specific precautions are taken, at the expense of simplicity) and large computer memory requirements. This latter is so because for an N by N matrix a storage of N^2 locations must be reserved even though most of these locations will be filled with zeros.

The input difficulties are partially eased by some special programs for general circuit analysis, such as IBM's ECAP [16]. Developments in this field are fast, and are directed mostly toward the writing of fast, reliable programs for rather general non-linear circuit analysis programs.

The development of these general circuit analysis computer programs for direct design use is the next major step to be completed. One can incorporate these programs into a semi-automatic design loop consisting of a feedback path which will adjust the circuit parameters and the circuit topology such as to meet some prescribed performance criteria.

On the other hand it is possible to apply methods similar to Mitrovics' method: the circuit equations, containing some literal parameters, can be solved for in terms of those parameters using a symbol manipulation language, such as IBM's FORMAC [8]. In this procedure some performance criteria for the network can be established directly in literal form and values for these parameters can be obtained by suitable iteration. With the wider availability of compilers of this kind this latter procedure should become useful in automated circuit design.

BIBLIOGRAPHY

1. Pipes, L. A., "Matrix Methods in Engineering", Prentice-Hall, New York, 1962.

2. Gantmacher, F. R., "The Theory of Matrices", Chelsea, New York, 1960.

3. Faddeev, D. K. and V. N. Faddeeva, "Computational Methods in Linear Algebra", Freeman, San Francisco, 1963.

4. Staudhammer, J., "Matrix Manipulation by Computer: Direct Numerical Procedures", TM 2860/000, System Development Corporation, April 1966.

5. Frazer, R. A., W. J. Duncan, and A. R. Collar, "Elementary Matrices", Princeton U. Press, 1958.

6. Soule, D. F., "Conventional Design Procedure" in W. W. Happ and J. Staudhammer, "Computer Aided Circuit Design", McGraw-Hill, 1967.

7. Gupta, S. C., "Transform and State Variable Methods in Linear Systems", Wiley, New York, 1966.

8. Staudhammer, J., "Summary of FØRMAC" TM-2640, System Development Corporation, Santa Monica, California, September 1965.

9. Wilkinson, J., "Rounding Errors in Algebraic Processes", Prentice-Hall, New York.

10. Seshu, S. and N. Balabanian, "Linear Network Analysis", Wiley, New York, 1963.

11. Pottle, C. "Comprehensive Active Network Analysis by Digital Computer - A State-Space Approach", Proceedings of the Third Allerton Conference, October 1965.

12. Liou, M. L., "A Novel Method for Evaluating Transient Responses" Proceedings IEEE, Vol. 54, No. 1, January 1966.

13. Happ, W. W. and J. Staudhammer, "Topological Techniques for the Derivation of Models with Pre-assigned Accuracy", Archiv der elektrischen Ubertragung, June 1966.

14. Kuo, F. F. and J. F. Kaiser, "System Analysis by Digital Computer" Wiley, New York, 1966.

15. Gentleman, W. M. and G. Sande, "Fast Fourier Transforms - For Fun and Profit", Proceedings, 1966 Fall Joint Computer Conference, San Francisco, November 1966.

16. IBM Corporation, "1620 Electronic Circuit Analysis Program - Users Manual", Publication H20-0170-1, White Plains, New York, 1966.

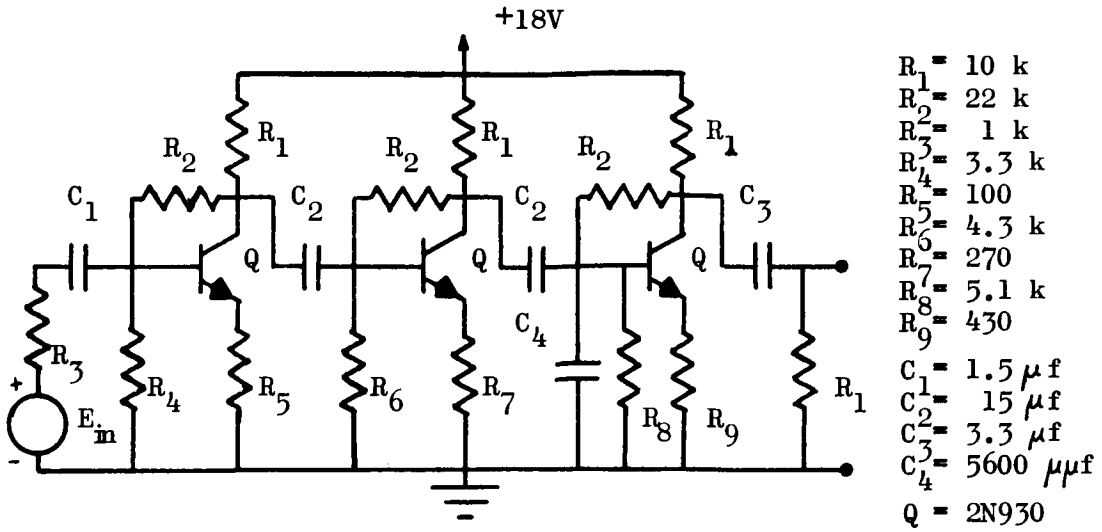


Figure 1

Three Stage RC Coupled Transistor Amplifier

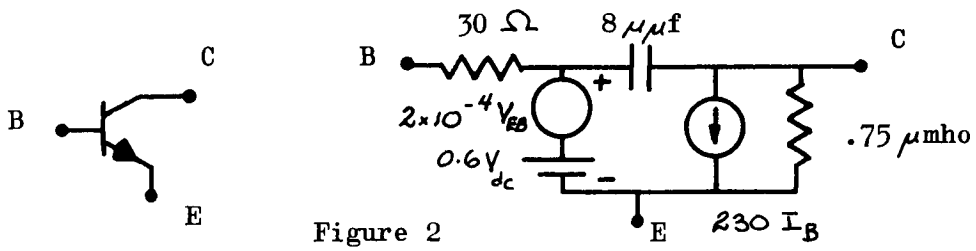


Figure 2

Equivalent Circuit for the Transistors

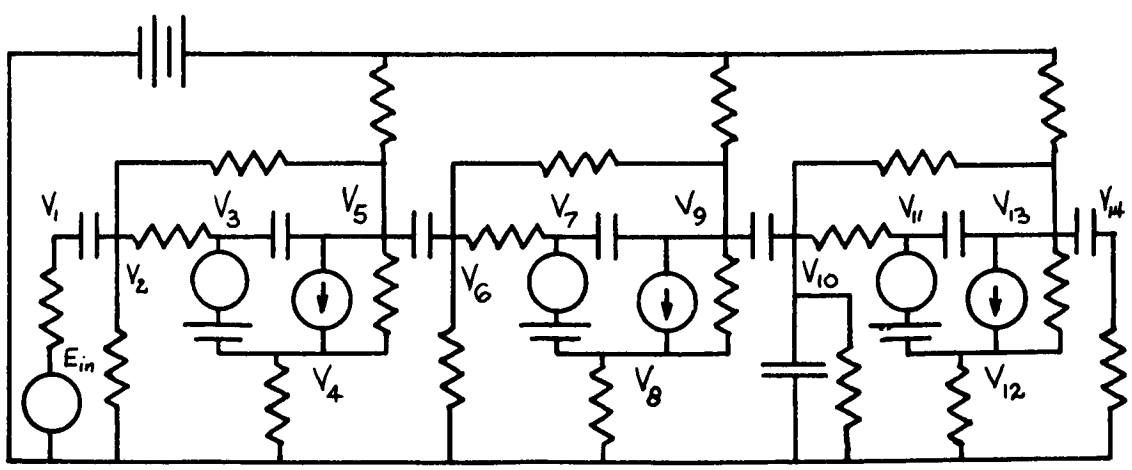


Figure 3

AC-DC Equivalent Circuit of the Three Stage Amplifier

N67 13347

RC ACTIVE NETWORKS

W. J. KERWIN
Ames Research Center, NASA
Moffett Field, California

INTRODUCTION

We are concerned here with the techniques available for the realization of various driving point or transfer functions without the use of inductors or transformers. We would like to be able to realize all the possible types of functions available with the use of inductive components, possibly even do more. Our purpose is to reduce the overall weight and size, and to eliminate all magnetic materials, as is necessary in probes for deep space magnetic field measurements. Another goal is cost reduction since, for example, the elimination of inductors allows circuit integration which can lead to lower cost and greater reliability in quantity production.

We will define an RC active network as one containing only resistors, capacitors, transistors, and diodes.

A COMPARISON OF RC PASSIVE, RLC, AND RC ACTIVE NETWORKS

In the passive RC network the available functions are extremely limited as indicated in Figure 1. The passive RLC network has a greatly increased freedom both for driving point and transfer functions (Fig. 2). If we now consider the RC network in which the R can be either positive or negative, our freedom is increased (Fig. 3); however, the poles are still severely restricted and this is not a significant improvement. If we also have negative capacity available to us (Fig. 4), all of the RLC pole, zero positions are possible as well as driving point zeros anywhere in the right-half plane. The possibility of right-half plane poles was not considered, as we are concerned here only with stable functions.

THE NEGATIVE IMPEDANCE CONVERTER

One of the early methods of RC active synthesis used the negative impedance converter (NIC) to produce -R or -C and thereby allow the design of ±R, ±C networks. The NIC can best be described by the hybrid g parameters (Fig. 5) in which the input impedance (Z) of an arbitrary 2-port is indicated as a function of the g parameters and the load impedance (Z_L). When $g_{11} = g_{22} = 0$ and $g_{12}g_{21} = 1$, the input $Z = -Z_L$.

A typical synthesis procedure using the NIC, the Linvill method, is shown in Figure 6. The poles of Z_{21} are determined by the subtraction of two driving point functions and it is this subtraction which allows complex poles. An example is given in Figure 7, in which

any positive value of K in the second-order system may be obtained by changing R and C appropriately. (VNIC refers to a voltage negative-impedance converter.)

THE OPERATIONAL AMPLIFIER

A high gain ($G \to \infty$) amplifier can be used as the active element and, if an input RC network is combined with a feedback RC network (Fig. 8), we find that the transfer voltage ratio has a transfer admittance in the denominator, that is, the overall poles are determined by the product of the poles of y_{21}^{A} and the zeros of y_{12}^{B}. Since the zeros of a passive RC transfer function are nearly unrestricted, we can realize most of the transfer functions we wish to obtain. An example is shown in Figure 9 in which a second-order bandpass function is obtained. Other methods are available using operational amplifiers, in particular the analog computer simulation method described in Reference 1. Although this method uses a large number of active elements and resistors, only the minimum possible number of capacitors is required. If each realization is restricted to a low-order system, the sensitivity to element change and amplifier gain change is far superior to that of most other methods (comparable to that of gyrator systems). This method is therefore suited to the realization of poles of very high Q.

THE GYRATOR

The use of the gyrator in RC active circuits is undergoing very extensive investigation at the present time. The gyrator can be described quite simply by the use of the y parameters (Fig. 10). The input admittance is written as a function of the load admittance (Y_L) on a 2-port network and the y parameters of the network. The gyrator property of impedance inversion occurs when $y_{11} = y_{22} = 0$ and $y_{12}y_{21} = -1$. We then have the very useful property that a capacitive load produces an inductive reactance at the input. This property allows the direct replacement of all inductors in an LC filter with gyrators and capacitors. The design effort is almost eliminated. One way of looking at the action of the gyrator is shown in Figure 11. The inductive input impedance results from the -R and -C which could be produced with two negative impedance converters. A gyrator could be built in this way; however, this is not the best approach from other standpoints, but it illustrates the principle. Figure 12 shows a gyrator used to synthesis a second-order bandpass filter. The 1 Ω above the gyrator symbol indicates the value of the gyration resistance (γ) and thus a 1 fd capacitor would be gyrated to a 1 Hy inductor. A more usual practical value would be a gyration resistance of 10^4 Ω and in this case a 1000 pF capacitor would be gyrated to 0.1 H since $L = \gamma^2 C$.

THE CONTROLLED SOURCE

This method uses low-gain voltage amplifiers in most cases, although a current amplifier could also be used. Only the voltage amplifier will be considered as the low output impedance and high input

impedance greatly simplify the design and allow cascading without additional isolation stages as would be required with the NIC and gyrator examples previously given (Ref. 2). Of course, in the case of direct replacement of the inductors in an RLC filter, no isolation stages are needed; but a new problem arises, that of the frequent ocurrence of inductors not having one terminal grounded. This generally will require two gyrators per inductor. Typical amplifier specifications for use as a controlled source are given in Figure 13. The gain stability requirement of ±0.04% is predicated on a pole Q no greater than 10 and a ±1% overall system stability. This type of sensitivity problem exists in many of the RC active synthesis methods, the overall gain sensitivity to amplifier gain change being of the order of $S = 2Q$. The gyrator does not exhibit this characteristic nor does the analog computer simulation method. Both are more complex procedures and would normally be used only where a pole $Q > 10$ is necessary.

Figure 14 illustrates the use of a controlled voltage source of gain K to produce a low pass and a high pass second-order voltage transfer function. The procedure is quite simple, a pole Q of 5 (coefficient of P is $1/Q = 0.2$) is obtained when $K = 2.80$. The element values shown are not optimum in minimizing gain, but they illustrate the principle. An additional function is needed to allow synthesis of most types of filters, and that is a means of realizing $j\omega$ axis zeros. If we sum the outputs of the high and low pass second-order networks (Fig. 14), we obtain the circuit shown in Figure 15 and the indicated 2-zero, 2-pole function. This gives continuously variable $j\omega$ axis zeros at the expense of considerable complexity. This network can be simplified as shown in Figure 16. The only restriction here is that the zeros be located at a greater distance from the origin than the associated poles. One arbitrary constant (k) exists and is chosen, on the basis of element spread or sensitivity requirements; a good average value is $k = 2$. The design is then very simple for any given transfer function. If the poles are at a greater distance than the zeros, then the network shown in Figure 17 is used; in this case, only three capacitors are required. Amplifier gains between 2 and 3 are usually sufficient.

DESIGN EXAMPLE USING A CONTROLLED SOURCE

The pole-zero positions of an elliptic function filter are shown in Figure 18. This filter has an 0.18 dB pass-band ripple and a 39.3 dB stop-band ripple, and is of the sixth order. Since four zeros are required on the $j\omega$ axis, two of the 2-pole, 2-zero networks are needed, and in addition a 2-pole only network is required to complete the 6-pole function. Figure 19 shows the resulting filter after scaling to a maximum resistance of 100 KΩ and to a cutoff frequency of 3160 cps (Ref. 3). The transfer function of each section of the overall transfer function is also given. Note that the highest Q required is approximately 9, and that the greatest amplifier gain is 2.762. In this case k was chosen as 2. The experimental performance is shown in Figure 20. A cutoff slope of over 200 dB/octave is obtained even though only moderately high pole Q's are required. The agreement with the calculated performance is excellent.

A variety of RC active synthesis procedures was considered which, in appropriate combinations, should be suited to nearly all applications. It is hoped that we have stimulated sufficient interest here that you will consider the subject in more detail as given in the references, and elsewhere. Reference 4 is recommended as an excellent brief treatment of the subject.

REFERENCES

1. Somerville, M. H.; and Tomlinson, G. H.: Filter Synthesis Using Active RC Networks. J. Electronics Control, vol. 12, no. 5, May 1962, pp. 401-420.

2. Sallen, R. P.; and Key, E. L.: A Practical Method of Designing RC Active Filters. IRE Transactions on Circuit Theory, March 1955, pp. 74-85.

3. Kerwin, William J.: An Active RC Elliptic Function Filter. 1966 IEEE Region Six Conference Record.

4. Huelsman, L. P.: Handbook of Operational Amplifier Active RC Networks. Burr-Brown Research Corp., 1966.

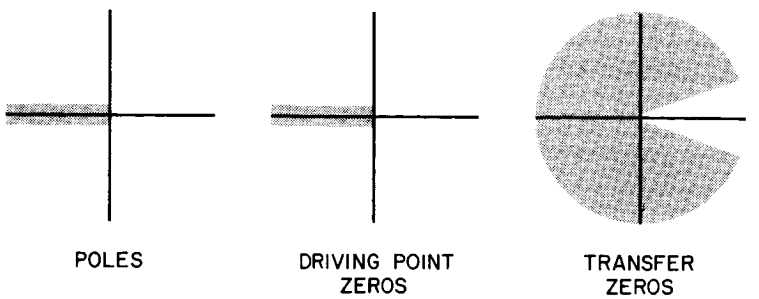


Figure 1.- Pole-zero positions for +R, +C networks.

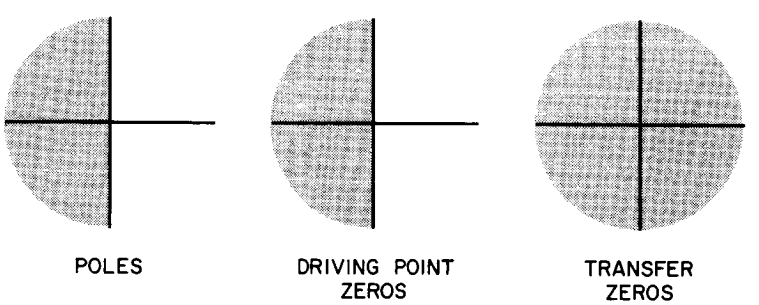


Figure 2.- Pole-zero positions for RLC networks.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA

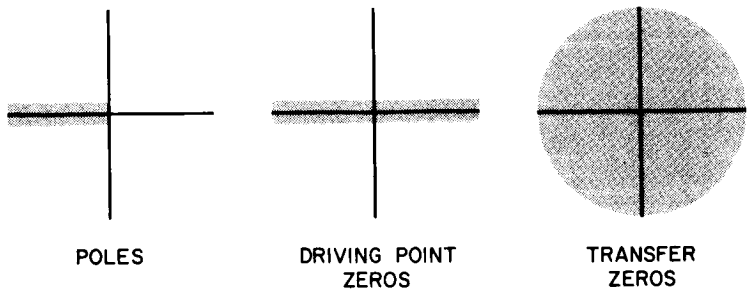


Figure 3.- Pole-zero positions for ±R, +C networks.

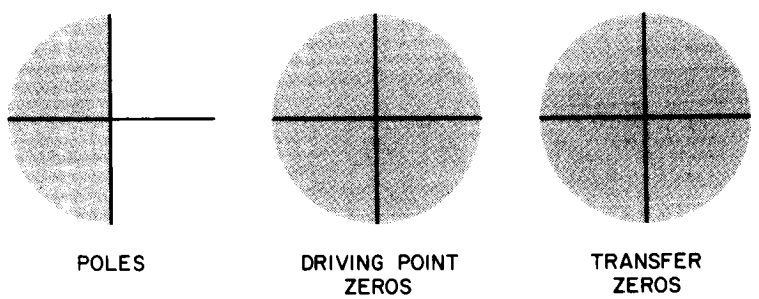


Figure 4.- Pole-zero positions for ±R, ±C networks.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA

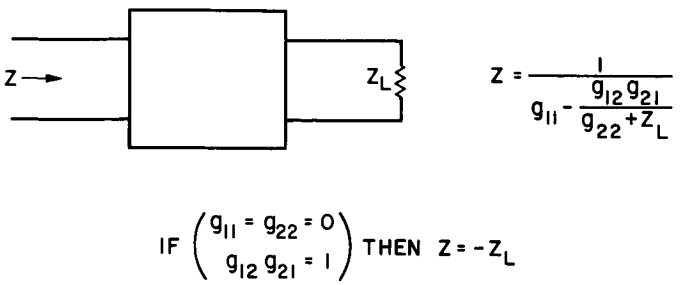


Figure 5.- The negative impedance converter.

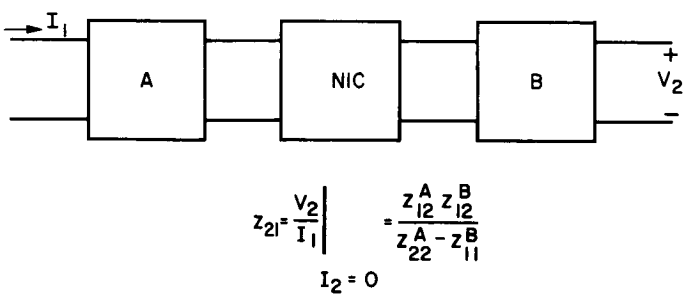
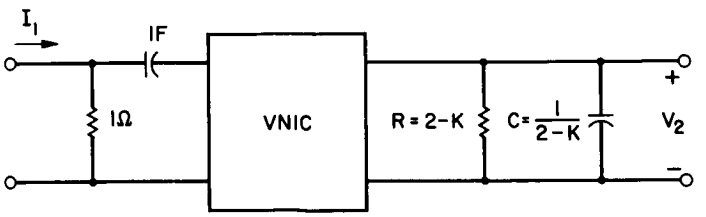


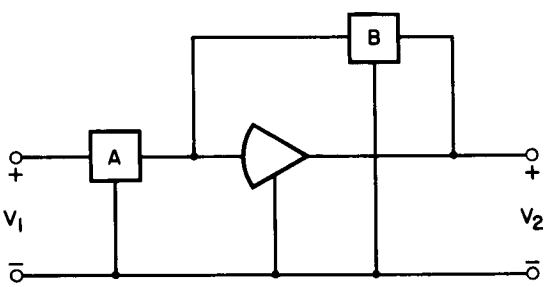
Figure 6.- The Linvill synthesis method.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA



$$Z_{21} = \frac{V_2}{I_1} = \frac{P}{p^2 + Kp + 1}$$

Figure 7.- An example of VNIC synthesis.



$$\frac{V_2}{V_1} = \frac{-y_{21}^{A}}{y_{12}^{B}}$$

Figure 8.- An operational amplifier synthesis (A and B are passive RC networks).

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA

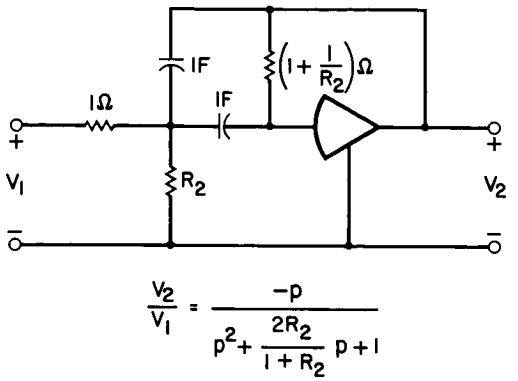
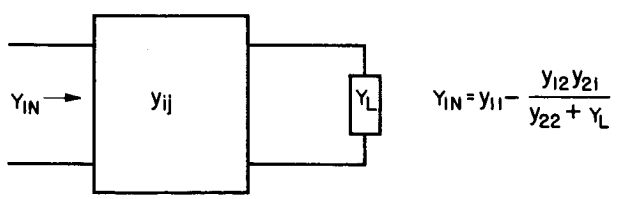


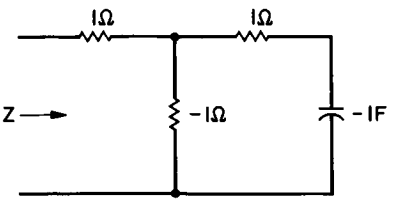
Figure 9.- An example of an operational amplifier synthesis.



$$\text{IF} \left\{ \begin{matrix} y_{11} = y_{22} = 0 \\ y_{12}\, y_{21} = -1 \end{matrix} \right\} \text{THEN} \quad Y_{IN} = \frac{1}{Y_L}$$

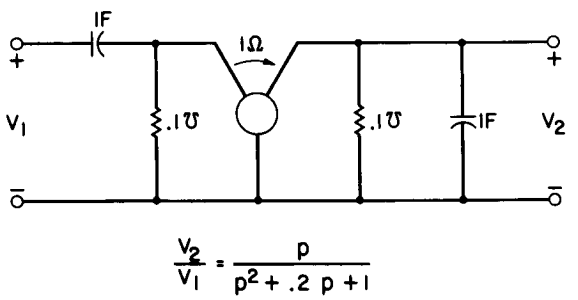
Figure 10.- The gyrator.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA



$$Z = 1 + \cfrac{1}{-1 + \cfrac{1}{1 - \cfrac{1}{p}}} = p$$

Figure 11.- A simple (?) inductor.



$$\frac{V_2}{V_1} = \frac{p}{p^2 + .2\,p + 1}$$

Figure 12.- An example of a gyrator synthesis.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA

1. NO DC POTENTIAL AT INPUT OR OUTPUT
2. NO CAPACITORS
3. GAIN STABILITY OF ±.04 PERCENT OVER 50°C
4. Z_{IN} > 50 megohms
5. Z_{OUT} <25 ohms
6. LOW POWER
7. FREQUENCY RESPONSE – DC TO 5 megacycles

Figure 13.- Controlled source requirements.

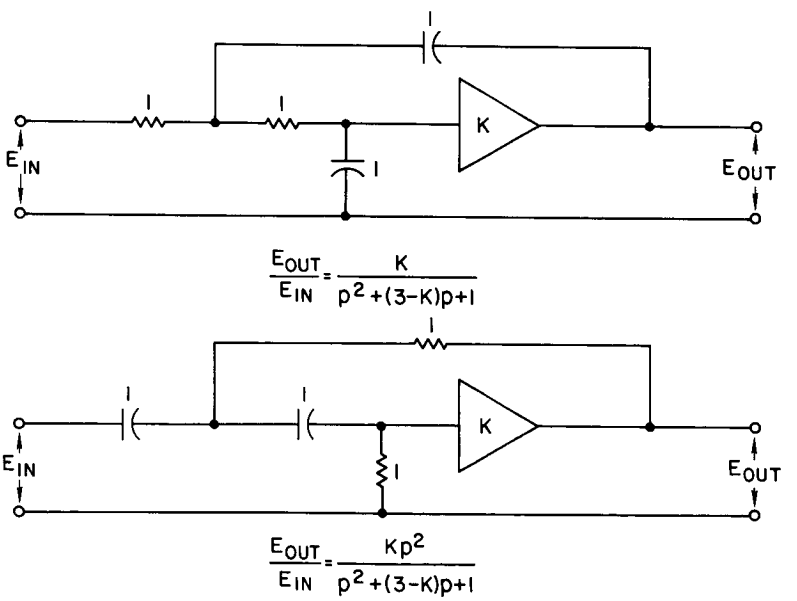
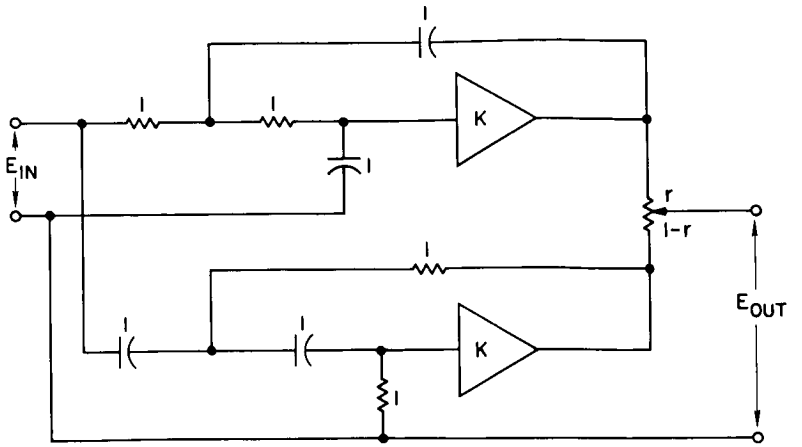
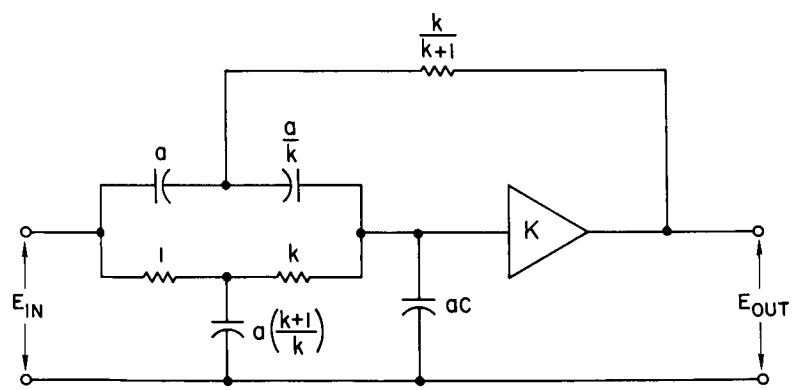


Figure 14.- Controlled source 2-pole realization.



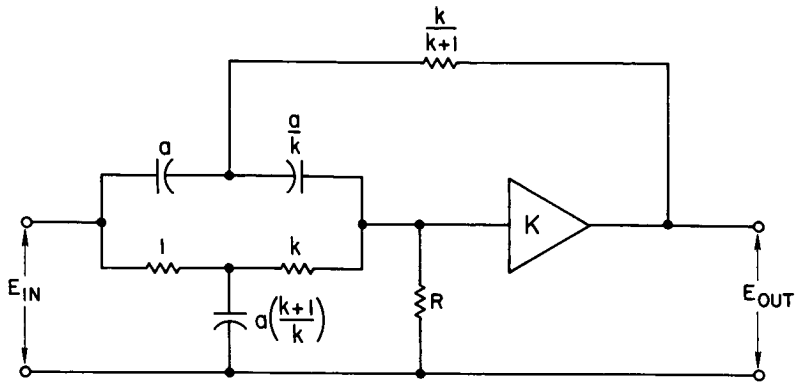
$$\frac{E_{OUT}}{E_{IN}} = \frac{Kr\left(p^2 + \frac{1-r}{r}\right)}{p^2+(3-K)p+1}$$

Figure 15.- Controlled source 2-pole, 2-jω axis zero realization (zeros controllable).



$$\frac{E_{OUT}}{E_{IN}} = \frac{\left(\frac{K}{(k+1)C+1}\right)\left(p^2+\frac{1}{a^2}\right)}{p^2+\frac{(k+1)\left(C+\frac{2-K}{k}\right)}{a\left[(k+1)C+1\right]}p+\frac{1}{a^2\left[(k+1)C+1\right]}}$$

Figure 16.- Controlled source 2-pole, 2-jω axis zero realization.



$$\frac{E_{OUT}}{E_{IN}} = \frac{K\left(p^2 + \frac{1}{a^2}\right)}{p^2 + \left(\frac{k+1}{a}\right)\left(\frac{1}{R} + \frac{2-K}{k}\right)p + \frac{1 + \frac{k+1}{R}}{a^2}}$$

Figure 17.- Controlled source 2-pole, 2-$j\omega$ axis zero realization.

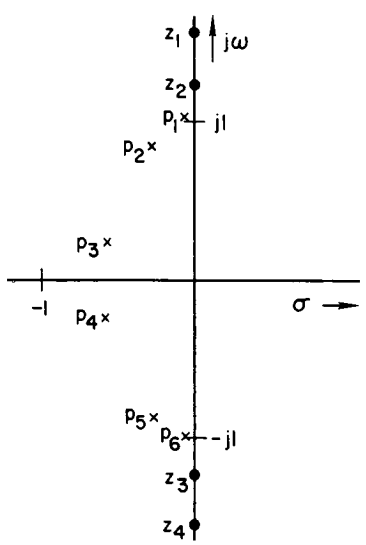


Figure 18.- Pole-zero positions for a sixth-order low-pass filter.

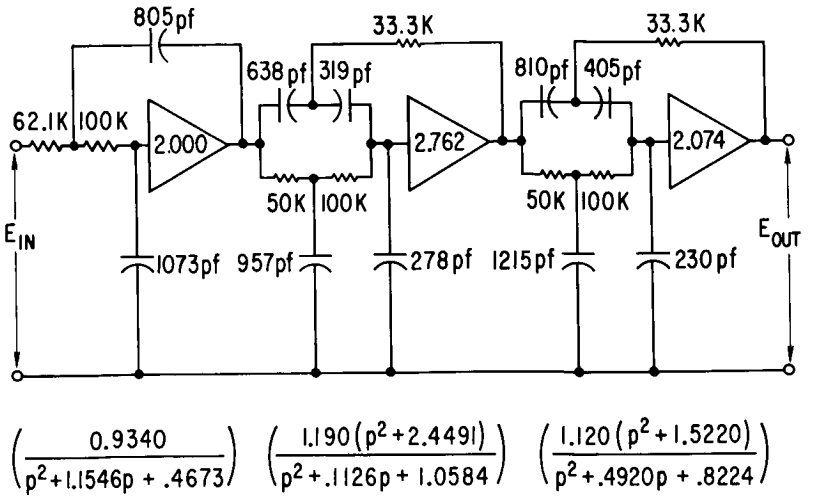


Figure 19.- Schematic of the low-pass filter.

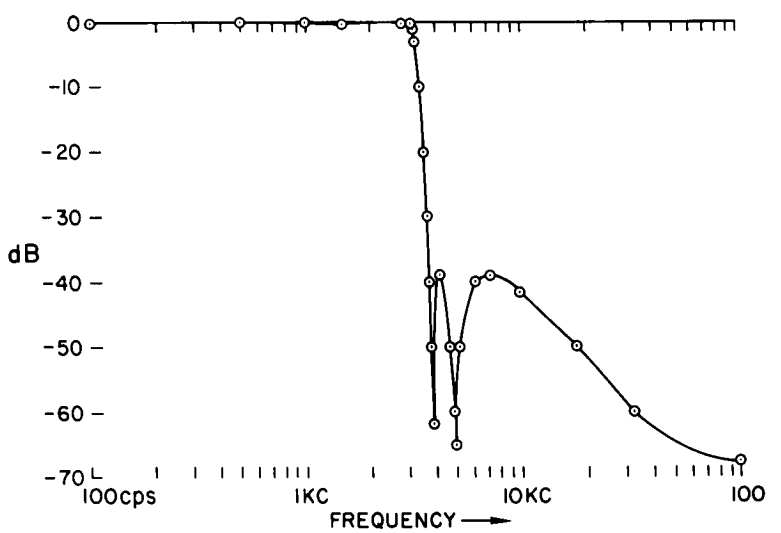


Figure 20.- Measured performance of the low-pass filter.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA

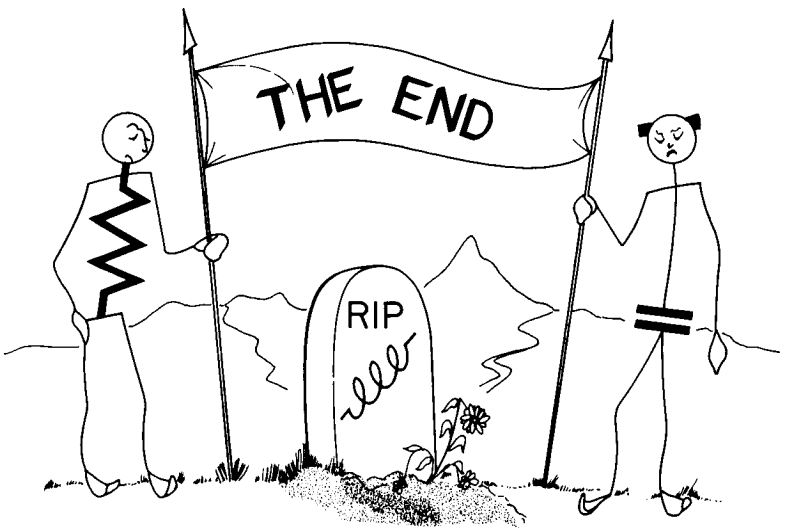


Figure 21

N67 13348

INDUCTORLESS FILTERS

H. J. Orchard
Lenkurt Electric Co., Inc.
San Carlos, California 94070

Attempts to produce inductorless filters with RC-active networks lead to circuits which are very much more sensitive to component tolerances than conventional LC filters. An alternative, and apparently optimum, solution is merely to replace each inductor in a conventional doubly loaded LC filter by a gyrator-capacitor combination.

Over the past 15 years there has been an intensive search for a practical method of making inductorless filters. Spurred on by the development of the negative-impedance converter[1] (NIC) and a subsequent paper by Linvill,[2] a tremendous effort has been applied to the synthesis of filterlike transfer functions using active elements together with resistors and capacitors. These active elements have included both NICs[3] and controlled sources.[4]

Frustrating all this work has been the bugbear of the high sensitivity of the resulting filter performance to tolerances on both the active and passive components. Everyone working in the field has acknowledged this problem, and no paper has been complete without a sensitivity analysis and an assurance that the advocated design method does, in fact, minimize the sensitivity. An excellent summary by Blecher[5] has shown that all design methods in this category have, at best, approximately the same order of sensitivity, and hence, from a practical point of view, one is free to choose, without penalty, whichever is the easiest to make. One is tempted to conclude that this high sensitivity is an inevitable part of the price which must be paid for avoiding inductors.

As an example of the order of magnitude of the problem, one may quote the RC-active bandpass filters described by Kinariwala[6] and Sipress.[7] These reproduce the performance of relatively simple multiplex-telephone channel filters in the ranges 12-16 kHz and 16-20 kHz, respectively. They are excellent designs of their class, and yet they require components stable to within about one part in 10^4 to achieve a passband constant to within 0.1 dB.

This paper was first published in Electronics Letters, vol. 2, p. 224, June 1966, and is reprinted here by permission of the Institution of Electrical Engineers.

Much of the enthusiasm for RC-active filters arises from a misunderstanding of the order of sensitivity to component tolerances which exists in conventional LC filters. If one designs a flat-passband reactance ladder filter to operate from a resistive source into a resistive load, and arranges that, at the frequencies of minimum loss over the passband, the source delivers its maximum available power into the load, one finds, to a first order of approximation, that, at every frequency in the passband and for every component, the sensitivity of the loss to component tolerances is zero. This is easily checked by noting that, when one has zero loss in a reactance network, a component change, either up or down, can only cause the loss to increase; in the neighborhood of the correct value, the curve relating loss to any component value must therefore be quadratic, and, consequently, d(loss)/d(component) must be zero.

Our ability to make high-quality filters meeting stringent specifications relies heavily on this desensitizing property which occurs at zero loss in a doubly loaded reactance network. It may be noted, in passing, that this property does not occur in either singly loaded or predistorted dissipation-compensated filters,[8] which accounts for why such networks are not widely used.

The sensitivity of an RC-active filter increases rapidly with the degree and sharpness of the filter characteristic, whereas, in an LC filter, it gets worse only by virtue of second-order effects becoming noticeable. In the LC equivalents to the bandpass filters described by Kinariwala and Sipress, for example, the component tolerances allowable for the same quality of passband are about 100 times greater. For difficult filters, this factor may easily exceed 10^3. In this respect, the LC filter possesses a most valuable property, which familiarity has tended to obscure. In attempting to make inductorless filters it would be wise to try, as far as possible, to retain this unique property; fortunately it can be done very easily.

The solution is to design a conventional doubly loaded LC ladder filter to meet the specification, and simply replace each inductor in the filter by a gyrator[9,14] terminated by a capacitor. This gyrator-capacitor network, like the LC network, is dissipationless and passive, and has exactly the same low sensitivity. The only remaining problem is a practical one of how to make a gyrator which is good enough.

The basic component available for eliminating inductors is the transistor, which is both an active and a nonreciprocal device. When used in making an NIC, for example, the nonreciprocal property is thrown away, and the active property is carefully retained. For a gyrator, we must do just the reverse--throw away the active property and retain the nonreciprocity.

Bearing in mind the need for using a common-earthed power supply, the most useful form in which to construct a gyrator is by direct simulation of the components of the admittance matrix, expressed as

$$\begin{bmatrix} 0 & -g \\ g & 0 \end{bmatrix} = \begin{bmatrix} 0 & -g \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ g & 0 \end{bmatrix}$$

The two component matrices represent voltage-controlled current sources of opposite polarity and direction, connected in parallel. By using a reasonable amount of negative feedback, the mutual conductances g of the two controlled sources can easily be made stable to within one or two parts in 10^3. The equivalent inductance of a capacitively terminated gyrator is $L = C/g^2$, and thus it is at least as stable as a "real" inductor.

The main difficulty lies in getting the terms in the principal diagonal close enough to zero. Assuming that both principal diagonal terms in the admittance matrix are equal at ϵ, and that one uses a perfect capacitor, the maximum Q factor obtainable from the equivalent inductance is

$$Q_{max} = g/2\epsilon.$$

A Q factor of 500 thus demands $g = 1000\,\epsilon$.

Many published gyrator circuits[10,11,12,13] have obtained low values of ϵ by placing negative resistances in parallel (or series) with the ports of an imperfect gyrator, to cancel substantial amounts of residual conductance. For the present application this is unacceptable, because it would be equivalent to making inductors with a Q factor of perhaps 5, and multiplying this up to 500 by negative resistance. One would have to maintain a critical balance to better than 1% between positive and negative resistances, and this is just what we want to avoid. This kind of difficulty is always likely to arise if such handy elements as operational amplifiers are used as building blocks.

Instead, the low value of ϵ must be obtained by careful design of the gyrator as a whole, using negative feedback both to stabilize g and reduce ϵ. Our first experiments in this direction have been very successful, and a direct replacement of the inductors in a sharp seventh-degree filter has verified the expected low sensitivity. Work is proceeding and further results on gyrator design will be reported shortly.

References

1. Merrill, J. L., "Theory of the negative impedance converter," Bell Syst. Tech. J., 1951, vol. 30, pp. 88-109.
2. Linvill, J. G., "RC active filters," Proc. IRE, 1954, vol. 42, pp. 555-564.
3. Yanagisawa, T., "RC active networks using current inversion type negative impedance converters," IRE Trans., 1957, vol. CT-4, pp. 140-144.
4. Sallen, R. P., and Key, E. L., "A practical method of designing RC active filters," ibid, 1955, vol. CT-2, pp. 74-85.

5. Blecher, F. H., "Application of synthesis techniques to electronic circuit design," IRE Internat. Convention Record, 1960, vol. 8, pt. 2, pp. 210-217.
6. Kinariwala, B. K., "Synthesis of active RC networks," Bell Syst. Tech. J., 1959, vol. 38, pp. 1269-1316.
7. Sipress, J. M., "Synthesis and sensitivity of active RC networks," D.E.E. Dissertation, Polytechnic Institute of Brooklyn, New York, June 1960.
8. Darlington, S., "Synthesis of reactance 4-poles," J. Math. Phys., 1939, vol. 18, pp. 257-353.
9. Tellegen, B. D. H., "The gyrator, a new electric network element," Philips Res. Rep., 1948, vol. 3, pp. 81-101.
10. Ghausi, M. S., and McCarthy, F. D., "A realization of transistor gyrators," Semiconductor Prod. Solid State Technol., 1964, vol. 7, no. 10, pp. 13-17.
11. Morse, A. S., and Huelsman, L. P., "A gyrator realization using operational amplifiers," IEEE Trans., 1964, vol. CT-11, pp. 277-278.
12. Sharpe, G. E., "The pentode gyrator," IRE Trans., 1957, vol. CT-4, pp. 321-323.
13. Bogert, B. P., "Some gyrator and impedance inverter circuits," Proc. IRE, 1955, vol. 43, pp. 793-796.
14. Holt, A. G. J., and Taylor, J., "Method of replacing ungrounded inductors by grounded gyrators," Electronics Letters, 1965, vol. 1, p. 105.

ITERATIVE OPTIMIZATION TECHNIQUES FOR CIRCUIT DESIGN

N67 13349

G. C. TEMES
Ampex Corporation, Redwood City, California

ABSTRACT

Iterative optimization procedures are used if no classical synthesis technique is available for the solution of the design problem. Such is the case if the configuration and/or some of the element values are constrained, if unavoidable parasitics or prespecified active devices must be incorporated in the circuit, etc. This paper gives a survey of optimization methods which were usefully employed in such situations.

INTRODUCTION

The "insertion loss" synthesis methods developed by Darlington, Cauer and Piloty, as well as other classical network design techniques provide powerful tools for circuit design. However, if practical constraints are placed on the configuration and on the element values, or if available devices (active elements, piezoelectric resonators, etc.) must be accommodated, the fast and direct methods of classical circuit synthesis break down. To find out what the best circuit is, under the given restrictions, iterative design techniques must be utilized.

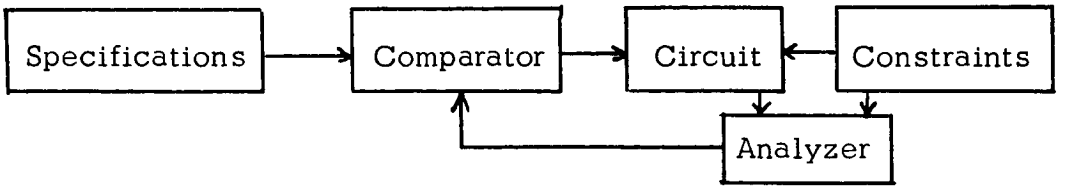
The price paid for the increased generality and flexibility of these iterative methods is heavy; on comparable problems, the iterative technique typically requires about 10 times the computer time needed for direct synthesis. Hence, they complement, rather than replace, the classical methods.

A (somewhat arbitrary) classification of iterative techniques can be made according to their function. Thus, we shall distinguish between direct design techniques, which start out from the specified performance and result in the final circuit; approximation techniques, which supply the transfer function satisfying the specifications; and, finally, realization techniques, which produce the circuit from a given transfer function. The schematic block diagrams for the three methods are shown in Fig. 1.

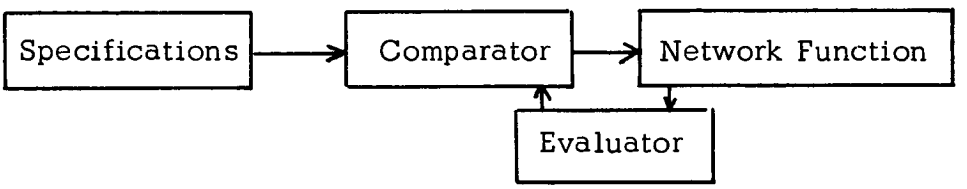
Any iterative optimization methods should, ideally, satisfy the following requirements:

1) Speed of execution

1. Direct design techniques:



2. Approximation techniques:



3. Realization techniques:

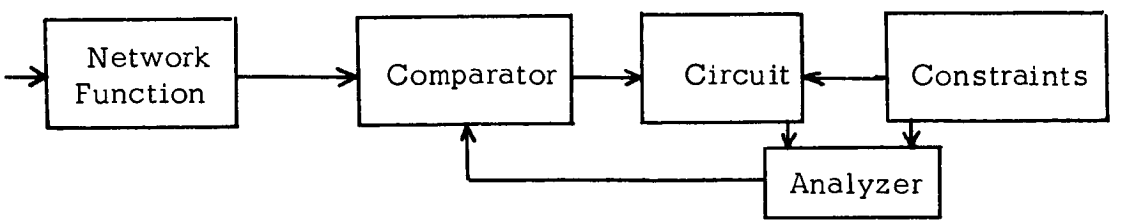


Figure 1

2) Ease of programming
3) Ease of feeding in input data, configuration
4) Readily interpreted output
5) Assured convergence to global optimum
6) Flexibility in accepting a wide variety of circuits and constraints
7) Accuracy of output
8) Insensitivity to round-off errors, ill-conditioned inputs.

These properties, unfortunately, are contradictory. A reasonable compromise, based on the application and anticipated frequency of use, must be found. For a program that is widely and often used, for example, requirements 1), 3) and 4) outweigh requirement 2), and vice versa.

DEFINITIONS AND NOTATIONS

The frequency or time response of the circuit is usually the property that must be optimized. Accordingly, the independent variable x used will normally represent either frequency or time. The range of approximation will be denoted by $a \leq x \leq b$, or, briefly, $[a, b]$. Let the desired response be $F(x)$; the actual response after the jth iteration, $F_j(x)$. The difference of F and F_j gives the error; using a positive weight function $w(x)$, the error can be weighted differently at various points of $[a, b]$; the weighted error is $w(x)\,[F(x) - F_j(x)]$.

The optimization is performed by minimizing some quantity ϵ related to the weighted error. The quantity ϵ will be called the error-criterion. The three most frequently used choices for the error-criterion are listed below:

a. Least pth error. The quantity to be minimized is chosen as

$$\epsilon = \int_a^b \left\{ w(x) \left| F(x) - F_j(x) \right| \right\}^p dx \,. \qquad (1)$$

Frequently used is $p = 2$; then ϵ is called the mean squared error. Alternatively, $p = 1$ can be selected; then ϵ is the mean absolute error.

b. Minimax error. Here, the minimized quantity is the maximum value of the weighted error in the approximation range:

$$\epsilon = \max_{[a, b]} \left\{ w(x) \left| F(x) - F_j(x) \right| \right\} \,, \qquad (2)$$

The minimization of the ϵ of (2) is equivalent to the minimization of the ϵ of (1), for $p \to \infty$.

The approximation obtained by choosing the minimax error criterion is also often called Chebyshev-approximation.

c. Maximally flat error. The error and a maximum possible number of its derivatives vanish at some inband value of x, x_o:

$$\begin{aligned} & F(x_o) - F_j(x_o) = 0 \\ & \frac{d}{dx}\left\{F(x) - F_j(x)\right\}_{x = x_o} = 0 \\ & \cdots\cdots\cdots \\ & \frac{d^n}{dx^n}\left\{F(x) - F_j(x)\right\}_{x = x_o} = 0 \quad , \end{aligned} \tag{3}$$

where

$$a \leq x_o \leq b \, . \tag{4}$$

This approximation can be obtained from the previously discussed types by using $a \to b \to x_o$. It is used if the range $[a,b]$ is narrow, or if the neighborhood of x_o is predominantly important in the approximation.

The iteration proceeds by changing some parameters a_i contained in F_j. For direct design or realization, the a_i are usually the element values; for approximation they can be the coefficients of the transfer function or the critical frequencies of the circuit.

Naturally, in actual computations the calculations are carried out only at a number of sample points, typically $10 \sim 100$. Hence, in all preceding (and following) equations, the integrals should be replaced by sums and the derivatives by difference quotients, in practical applications. Note, however, that in some cases the differentiation can be carried out analytically and yields manageable results (see, e.g., Ref. [18]).

OPTIMIZATION TECHNIQUES

Linear Programming[1]

In some cases, the error function ϵ depends linearly on the circuit parameters a_i. If there are no constraints on the a_i, or if all constraints on these are also linear in all a_i, then the methods of linear programming -- specifically, the simplex method[1] -- can be applied to the (constrained) minimization of ϵ.

Linearization[2]

The range of applicability of linear programming and other linear methods can be extended by utilizing truncated MacLaurin series. E.g., the relation between the $(j+1)$th and jth approximants can be expressed approximately in terms of Δa_i, the increments of the circuit parameters as follows

$$F_{j+1}(x, a_i + \Delta a_i) \cong F_j(x, a_i) + \sum_{i\omega} \frac{\partial F_j(x, a_i)}{\partial a_i} \Delta a_i. \qquad (5)$$

(The derivatives can be evaluated numerically.) Using equations similar to (5), all nonlinear $\epsilon(a_i)$ and constraint equations can be transformed into expressions which are linear in the Δa_i. Then the Δa_i, rather than the a_i, are found, using linear methods.

Linearization is used in several of the optimization methods described below. It has also been successfully utilized in other iterative circuit design procedures.[7, 9, 13, 14]

Steepest Descent[2, 3]

This method involves changing all a_i in such a direction (in the $\underline{\epsilon}$, $\underline{a_i}$ space) that the rate of change of $\underline{\epsilon}$ is the fastest. This will be achieved if the change is along the negative gradient vector:

$$-\nabla\epsilon = \left(-\frac{\partial \epsilon}{\partial a_1}, -\frac{\partial \epsilon}{\partial a_2}, \ldots, \right). \qquad (6)$$

Equation (6) gives only the direction of the change in the a_i space, but not its size. Several procedures have been developed for finding this so-called step size. These will be discussed below. They are equivalent to a one-dimensional minimization.

Step Size Determination

A very simple method for finding the step size is illustrated in Fig. 2, for a two parameter circuit. Proceeding from an initial approximation represented by point P_0, the next point P_1 is obtained by a multidimensional Newton-Raphson process.[2, 3] From Fig. 2, after some calculation,[3]

$$\Delta a_j = -\frac{\epsilon(a_i)\,\frac{\partial \epsilon}{\partial a_j}}{\sum_{(i)} \left(\frac{\partial \epsilon}{\partial a_i}\right)^2} \qquad (7)$$

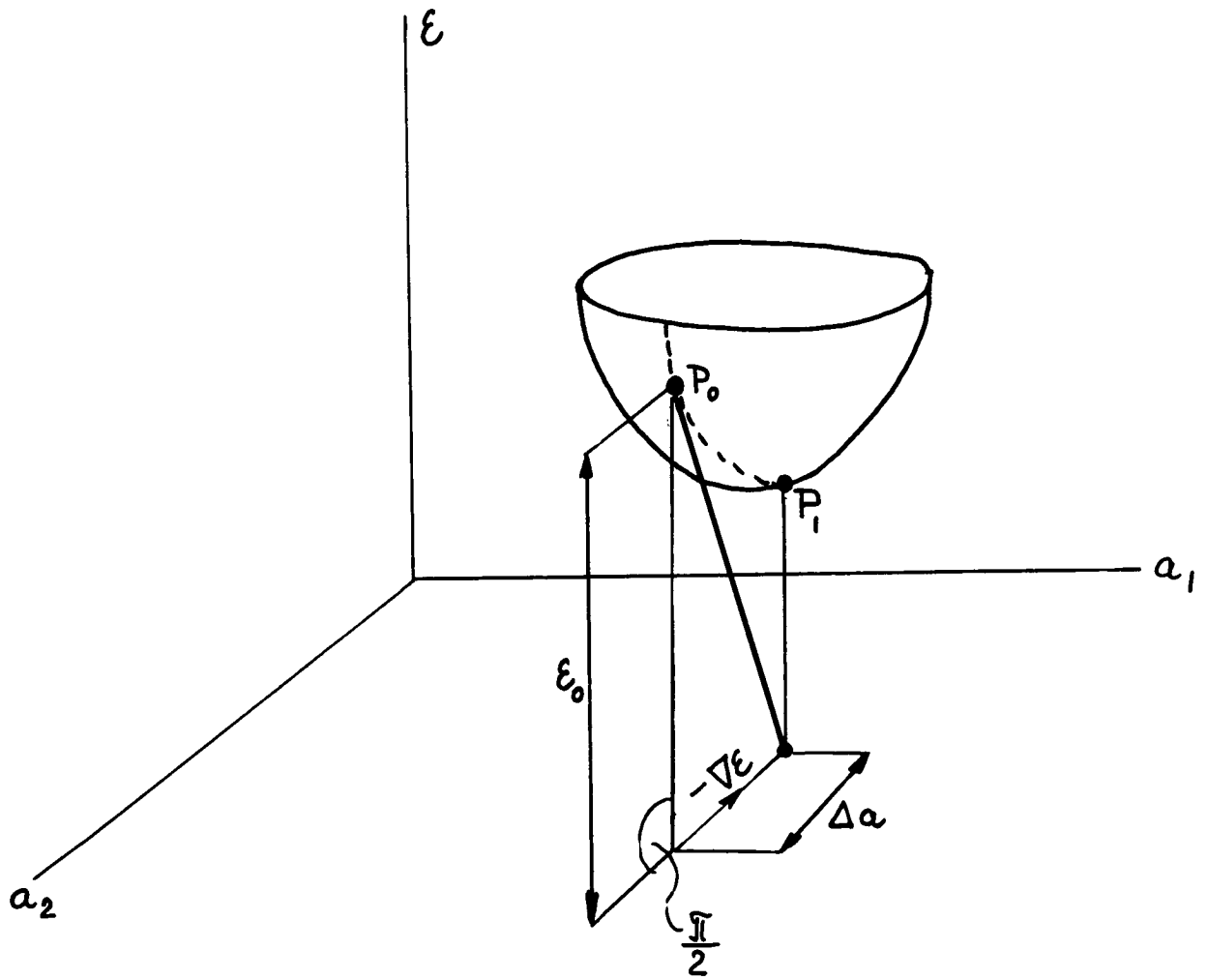


Figure 2

results for the change in the a_i. ϵ must be evaluated after each iteration and if it does not decrease any more, the step size must be cut (e.g., by a factor of 10).

A better way to obtain the reduced step size is to use quadratic interpolation. This is done by interpolating the $\epsilon(a_i)$ surface at three equally spaced points on the curve which is its intersection with the (ϵ, $\nabla\epsilon$) plane (Figs. 2 and 3), and finding the minimum of the resulting quadratic parabola. The resulting step size is given in Fig. 3.

An effective, but very tedious, method for finding the step size is to approximate the (ϵ, a_i) surface by a quadratic surface and find its minimum. The resulting equation [3], however, contains all partial derivatives of the form

$$\frac{\partial^2 \epsilon}{\partial a_i \partial a_j}$$

and hence its evaluation is too laborious.

Third-order polynomial interpolation has also been used (Refs. [5], [6], [18]) to find one-dimensional minimum.

A possible method for finding the step size is to use search techniques, described below.

Search techniques

A conceptually simple method for minimizing ϵ in one variable (step size determination) or in many variables (direct minimization) is to evaluate it for a large number of parameter and independent variable values and compare the results. The parameter values may be picked randomly or systematically.

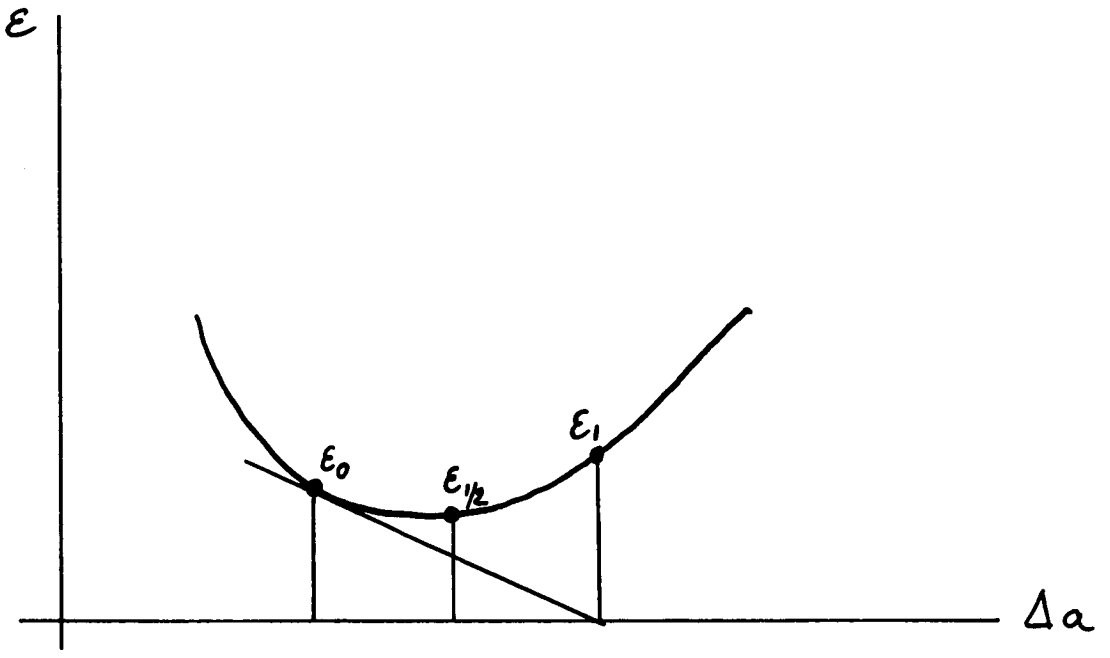
For one-dimensional systematic search, the use of the Fibonacci numbers ($1,1,2,3,5,8,\ldots, y_k, y_{k+1}, y_{k+2}=y_k+y_{k+1}, y_{k+3}=y_{k+1}+y_{k+2}, \ldots$) provides a highly efficient algorithm [16]. Let

$$\epsilon_{j+1}(a + \bar{\theta}\,\Delta a) = \epsilon_j(a)\,. \tag{8}$$

Here, ϵ_j (ϵ_{j+1}) is the error after the jth (j + 1th) iteration; Δ a represents the known direction of the parameter increment; finally, scalar $\bar{\theta}$ is the unknown optimum step size, minimizing ϵ_{j+1}. Now $\bar{\theta}$ can be found in the following steps:

1) Find an upper bound θ_u by evaluating ϵ_{j+1} for

$$\theta = 1,\ 1+k,\ 1+k+k^2,\ \ldots,\ \sum_{i=0}^{i} k^i = \theta_u \tag{9}$$



$$(\Delta a_j)_0 = - \frac{\varepsilon(a_j) \frac{\partial \varepsilon}{\partial a_j}}{\sum_{(i)} \left(\frac{\partial \varepsilon}{\partial a_i} \right)^2}$$

$$\Delta a_j = (\Delta a_j)_0 \frac{1}{4} \frac{\varepsilon_1 - 4\,\varepsilon_{1/2} + 3\,\varepsilon_0}{\varepsilon_1 - 2\,\varepsilon_{1/2} + \varepsilon_0}$$

Figure 3

where $\bar{i}$ is the first number such that $\epsilon_{j+1} \geq \epsilon_j$ and where

$$k = \frac{1+\sqrt{5}}{2} = \lim_{k\to\infty} \frac{y_{k+1}}{y_k} \cong 1.618 \qquad (10)$$

is the limiting Fibonacci ratio. Choose

$$\theta_L = 0 . \qquad (11)$$

Now $\bar{\theta}$ is bracketed between θ_L and θ_u.

Next these bounds are moved closer, using a subdivision of the range $[\theta_L , \theta_u]$:

2) Calculate

$$\begin{aligned} \theta_a &= \theta_L + (2-k)(\theta_u - \theta_L) \\ \theta_b &= \theta_L + (k-1)(\theta_u - \theta_L) \end{aligned} \qquad (12)$$

and

$$\begin{aligned} \epsilon_a &= \epsilon(a + \theta_a \Delta a) \\ \epsilon_b &= \epsilon(a + \theta_b \Delta a) . \end{aligned} \qquad (13)$$

3) Compare ϵ_a and ϵ_b :

α. If $\epsilon_a < \epsilon_b$, then (for a convex $\epsilon(a)$ function) $\theta_L \leq \bar{\theta} \leq \theta_b$. Hence, replace θ_u by θ_b , θ_b by θ_a . Calculate a new

$$\theta'_a = \theta_L + (2-k)(\theta_b - \theta_L). \qquad (14)$$

Now the process can be repeated.

β. If $\epsilon_a > \epsilon_b$, then $\theta_a \leq \bar{\theta} \leq \theta_u$. Hence, by choosing

$$\begin{aligned} \theta'_L &= \theta_a \\ \theta'_b &= \theta_b \end{aligned} \qquad (15)$$

and

$$\theta'_b = \theta_a + (k-1)(\theta_u - \theta_a) , \qquad (16)$$

the process can be repeated.

γ. If $\epsilon_a = \epsilon_b$, replace (θ_u, θ_L) by (θ_a, θ_b) and return to 2.

4) If $(\theta_u - \theta_L)$ became less than a specified tolerance, choose

$$\bar{\theta} = \frac{\theta_u + \theta_L}{2} .$$

Alternative (but less efficient) one-dimensional search algorithms are described in Refs. [7], [16] and [18].

A multivariable search technique is given by Murata[4] . The method uses a sequence of linear searches, optimizing each parameter individually, then starting again with the first parameter. The linear search described is again less effective than the Fibonacci-search.

Least Squares

Particularly simple formulae result from choosing a least squares error criterion and using the linearization techniques described above[2]. For a set of sample points $[x_k]$, combining eqs. (1) and (5) gives, to a first approximation

$$\epsilon \cong \sum_{(k)} \left\{ w \left[F - F_j - \sum_{(i)} \frac{\partial F_j}{\partial a_i} \Delta a_i \right] \right\}^2_{x = x_k} . \qquad (17)$$

This must be minimized with respect to the Δa_i. From

$$\frac{\partial \epsilon}{\partial a_i} = 0 \qquad i = 1, 2, \ldots n \qquad (18)$$

a system of n linear equations is obtained for the n unknown Δa_i. Having found the Δa_i,

$$a_i \rightarrow a_i + \Delta a_i \qquad i = 1, 2, \ldots n \qquad (19)$$

and the process is repeated.

A Hybrid Method

A combination of linearization, linear programming and linear search has been described by Ishizaki et al[7]. The method uses the Chebyshev error-criterion of eq. (2). It proceeds in the following steps:

1) Using eq. (2),

$$| w (x_k) [F (x_k) - F_{j+1}(x_k)] | \leq \epsilon ; a \leq x_k \leq b. \qquad (20)$$

Also, by eqs. (5) and (20), approximately

$$| w [\Gamma - F_j] - w \sum_{(i)} \frac{\partial F_j}{\partial a_i} \Delta a_i |_{x = x_k} \leq \epsilon ; a \leq x_k \leq b. \qquad (21)$$

Hence, introducing

$$\xi_i \stackrel{\Delta}{=} \frac{\Delta a_i}{a_i} \qquad i = 1, 2, \ldots n$$

$$\xi_{n+1} \stackrel{\Delta}{=} \epsilon , \qquad (22)$$

the linear inequalities

$$\pm w \sum_{(i)} \frac{\partial F_j}{\partial a_i} a_i \xi_i - \xi_{n+1} \pm w [F_j - F] \leq 0 \qquad (23)$$

are obtained at all sample points x_k. Eq. (23) can be supplemented by the linearized forms of any constraint equations. E.g., positive a_i values can be maintained throughout the iteration by specifying

$$- \xi_i \leq 1 . \qquad (24)$$

2) Equation (23) and the constraints are now linear in the ξ_i; hence, $\xi_{n+1} \equiv \epsilon$ can be minimized, subject to these inequalities, using the simplex method[1].

3) Let the solution vector obtained from step 2) be $\bar{\xi}$. Then, due to the first-order approximation used throughout, only the direction, but not the magnitude, of $\bar{\xi}$ is acceptable. Hence, a step size determination must be performed, using linear search. This yields the necessary changes in the parameter-values.

4) The process is repeated until the decrease in ϵ is less than a predetermined value.

5) In order to save time in the calculation of the partial difference-quotients, a linear transformation of the unknowns can be used[7]. This makes it possible to use the results of previous linear searches in finding the approximate values of the partial derivatives.

Generalized Remez-Method

Minimax error can also be achieved using a generalized Remez-technique[8, 9] This method is based on the fact that (under nondegenerate conditions) ϵ as given by eq. (2) is minimized if $w\,[F - F_j]$ has $n+1$ extrema, with alternating signs but equal magnitudes, in $[a, b]$. Hence, by (5), to a first-order approximation,

$$w(x_k)\,\left[\, F(x_k) - F_j(x_k) - \sum_{(i)} \left.\frac{\partial F_j}{\partial a_i}\right|_{x_k} \Delta a_i \,\right] = (-1)^k E$$

$$k = 0, 1, \ldots n \qquad (25)$$

$$x_k \in [\, a, b \,].$$

The iteration proceeds in the following steps:

1) An initial approximation F_o is found. This must have the property that $w\,[F - F_o]$ has $n+1$ alternating (but in general unequal) extrema. Such F_o may be found, e.g., by matching it to $F(x)$ at n internal points of (a, b).

2) The $n+1$ largest alternating extrema of $w\,[F - F_o]$ are located, using, e.g., linear search techniques. Let these be $[x_k^o]$.

3) Equations (25) are used, with $j = o$ and $x_k = x_k^o$, to find the Δa_i and E. This involves solving a system of $n+1$ simultaneous linear equations for the n unknown Δa_i and for E.

4) F_1 is found by changing all a_i to $a_i + \Delta a_i$. Then, steps 2), 3) and 4) are repeated, until the Δa_i and/or the change in E are less than their prespecified tolerances.

A variation of this procedure uses a prescribed error-ripple E. This decreases the number of unknowns by one and thus simplifies somewhat the calculations. The price paid is a larger error-amplitude.

The generalized Remez-method has been used successfully in both time-domain[12] and frequency-domain[9] circuit design.

Zero-Shifting Techniques

Chebyshev-approximation can also be achieved by the zero-shifting technique[10]. This involves the following steps:

1) F_o is again found by matching it to $F(x)$ at n internal points of (a,b). Let these zeros of $F-F_o$ be $[z_\ell]$.

2) The $n+1$ extrema of $w[F-F_o]$ are located. Let these be $[x_k]$.

3) Approximately

$$\epsilon(x_k) \cong G(x_k) \prod_{\ell=1}^{n} (x_k - z_\ell); \quad k = 0, 1, \ldots n \tag{26}$$

and hence, if the zeros are shifted by $[\Delta z_k]$, the approximate relation

$$\Delta(\log_e |\epsilon(x_k)|) \cong -\sum_{\ell=1}^{n} \frac{\Delta z_\ell}{x_k - z_\ell}; \quad k = 0, 1, \ldots, n \tag{27}$$

is valid.

Therefore, in order to change $|\epsilon|$ at all $[x_k]$ values to a new equal magnitude E

$$\log_e |\epsilon(x_k)| + \Delta(\log_e |\epsilon(x_k)|) = \log E \tag{28}$$

or

$$\log_e |\epsilon(x_k)| - \sum_{\ell=1}^{n} \frac{\Delta z_\ell}{x_k - z_\ell} = \log E \qquad k = 0, 1, 2, \ldots, n. \tag{29}$$

Eqs. (29) give $n+1$ simultaneous linear equations for $[\Delta z_\ell]$ and $\log E$.

A more heuristic version of this technique merely matches F_o to F at n points z_k, finds the error-extrema and evaluates their

average value. Next, any two adjacent z_k are moved slightly closer to (farther from) each other if the absolute value of the extremum between them is larger (smaller) than the average. This adjustment process is continued until all error-extrema are of the same size.

Coefficient Matching

All previously discussed methods are equally useful in direct design or in approximation. The technique described below[11] is usable only for realization. Hence, it can only be used if an explicit algebraic expression is known for the desired transfer function.

The method involves matching the coefficients c_i' of the desired rational transfer function F to the coefficients of F_j obtained from the actual network after the jth iteration. Let these "actual" coefficients be c_i. The error in the ith coefficient is defined by

$$\epsilon_i = K\, c_i' - c_i. \tag{30}$$

Here, K is an arbitrary constant, independent of i. Now if the variable admittances in the circuit are y_k $(k = 1, 2, \ldots, n-1)$, then using a linearization technique, similar to that applied in eq. (5), the change in ϵ_i can be expressed in terms of the corrections in the y_k and in K:

$$\Delta\epsilon_i = -\sum_{k=1}^{n-1} \frac{\partial c_i}{\partial y_k}\, \Delta y_k + c_i'\, \Delta K. \tag{31}$$

For the first-order correction of all remaining coefficient errors,

$$\Delta\epsilon_i = -\epsilon_i \qquad i = 1, 2, \ldots, n. \tag{32}$$

Equations (31) - (32) give n simultaneous linear equations in the required modifications of the admittances (Δy_k) and in ΔK. Note that the partial derivatives in (31) can be found using simple topological methods[11].

Nonlinear Programming Methods

More sophisticated methods of nonlinear programming[5, 6] utilize not only the current but also the previous value of the gradient vector. These methods require less iteration cycles for convergence than either the steepest descent or the linearized least squares method. However, the computing time for each cycle is greater and so is the necessary programming effort.

These techniques have been used successfully in circuit design[17, 18]. They are described in detail elsewhere (by Dr. Huber) in these Proceedings.

Constraint Types

Constraints can be placed on the configuration of the network, as well as some (or all) of its element values. If the element values is exactly prespecified, it cannot be used as a variable parameter in the optimization. Frequently, however, only bounds are given on the variations of some element values a_i. These are normally in the form

$$\ell_i \leq a_i \leq u_i \tag{33}$$

(where usually $\ell_i \geq 0$) or equivalently,

$$(a_i - \ell_i)(u_i - a_i) \geq 0 . \tag{34}$$

Sometimes bounds are also placed on the ratios of element values (due, e.g., to practical limitations on resonator Q's or capacitance-ratios). These are usually of the form

$$\ell_{ij} \leq \frac{a_i}{a_j} \leq u_{ij} , \tag{35}$$

or, for positive a_i, a_j:

$$(a_i - \ell_{ij} a_j)(u_{ij} a_j - a_i) \geq 0 . \tag{36}$$

More complicated (e.g., frequency or time dependent) constraints are also possible, but seldom used.

Incorporation of Constraints in the Optimization

If the constraints are linear, or have been linearized using truncated MacLaurin series, they can be incorporated into linear programming procedures[11] without difficulty. Another straightforward method for handling constraints is to use them only as a check in the course of the iteration, on the successive designs. If a parameter is found to violate a bound in any iteration cycle, it is made equal to the extreme value permitted by that bound. This parameter will not thereafter be changed in the subsequent iteration. After the iteration has converged, a one-dimensional optimization can be performed in the range of this parameter, to improve the final design.

The Carroll-Fiacco-McCormick Technique

A useful technique[15-18] for handling constraints is to incorporate them directly into the error-function. Considering, e.g., the constraints given in eq. (33), an additional term of the form

$$d_k = r_k \sum_{(i)} \left\{ \frac{1}{u_i - a_i} + \frac{1}{a_i - \ell_i} \right\} \tag{37}$$

is added to the error-function. If an initial approximation satisfying all conditions is chosen and r_k assigned an arbitrary value, the presence of d_k will prevent the optimization process from violating the boundary. Having found this initial optimum, r_k is decreased (e.g., by a factor of 10) and the optimization repeated. As $r_k \to 0$, the sequence of approximants converge to the constrained optimum. Combination of this method with the Fletcher-Powell optimization technique[6] has been used successfully in circuit design, using either the least squares[17] or the minimax[18] error criterion.

CONVERGENCE CONSIDERATIONS

All optimization procedures described in this paper place certain conditions on the $\epsilon(a_i)$ function. Even the most powerful methods require $\epsilon(a_i)$ to be at least convex in the approximation range, to assure that the iteration does not end up in a local minimum.

Unfortunately, the functions commonly encountered in circuit optimization do not satisfy even this modest requirement[18]. E.g., one classic problem in this field is the iterative design of lossy filters, using the lossless design as an initial approximation. One specific circuit has been used as an example by two authors for illustration purposes in articles describing novel optimizing methods. They ended up with two different circuits, neither of which represented the true optimum, illustrating that the error surface had at least three minima in the parameter-space.

For this reason, the choice of proper initial approximation is of considerable importance. Since some methods (zero-shifting, least squares) are less sensitive to the closeness of the initial approximation than others (e.g., the Remez-method), the formers may be used as a preamble to faster but more sensitive techniques. Search may precede the optimization, in order to find a good initial approximation; it may also follow it, to assure that the optimum found is really the global one and not just a local minimum.

In practical applications, a large problem can often be partitioned into a sequence of smaller ones, if it can be established that some a_i are substantially effective in changing the response only in certain parts of the interval $[a,b]$. This circumstance can be found out either by physical considerations or by calculating the partial derivatives of the initial approximation with respect to all a_i at various points of the $[a,b]$ range. Once the crucial a_i and their effective subregions have been identified, the optimization can be performed in each subregion separately. The resulting circuit can then further be improved, by conventional techniques; however, it is usually extremely close to the absolute optimum.

REFERENCES

[1] S. I. Gass, "Linear Programming, Methods and Applications," McGraw-Hill, New York, 1958.

[2] M. R. Aaron, "Use of least squares in system design," IRE Trans. on C. T., Dec. 1956.

[3] A. D. Booth, "An Application...", Quart. J. Mech. and Appl. Math., Dec. 1949.

[4] T. Murata, "The use of adaptive constrained descent in system design," 1964 IEEE Conv. Rec., Pt.1.

[5] R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization," Computer J., June 1963.

[6] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," Computer J., July 1964.

[7] Y. Ishizaki, H. Wakazuki and H. Watanabe, "Some considerations on min-max approximation problem for network design," Proc. 1965 Allerton Conference; also, Nippon Electric Co., Monos. DEB 451 & 485.

[8] E. Remez, "Sur le calcul effectif des polynomes d'approximation de Tchebychef," C. R. Acad. Sci. Paris, 1934.

[9] G. C. Temes and J. A. C. Bingham, "Iterative Chebyshev approximation technique for equalizer synthesis," Proc. Allerton Conference, 1965.

[10] H. J. Maehly, "Methods for fitting rational approximations," J. of ACM, 1960.

[11] D. A. Calahan, "Computer design of linear frequency selective networks," Proc. IEEE, Nov. 1965.

[12] R. Fischl, "Optimal Chebyshev approximation at discrete points in the time domain," Proc. Midwest C.T. Conference, 1964.

[13] C. A. Desoer, "Network design by first-order predistortion technique," IRE Trans. on C.T., Sept. 1957. This is not an iterative method, but can easily be expanded into one.

[14] J. D. MacDonald and G. C. Temes, "A simple method for the predistortion of filter transfer functions," IEEE Trans. on C.T., Sept. 1963.

[15] C. W. Carroll, "The created-response-surface technique for optimizing nonlinear restrained systems," Op. Res., 1961.

[16] A. V. Fiacco, G. P. McCormick, "Computational Algorithm...", Management Sci., No. 4, 1964 and "The Sequential...", ibid., No. 2, 1960.

[17] P. O. Scheibe and E. A. Huber, "The Application of Carroll's optimization technique to network synthesis," Proc. Allerton Conference, 1965.

[18] A. D. Waren and L. S. Lasdon, "Practical filter design using mathematical optimization," ibid.

N67 13350

NETWORK DESIGN VIA NONLINEAR PROGRAMMING

E. A. HUBER
Sylvania Electronic Systems-West, Mountain View, Calif.

ABSTRACT

The network design problem is considered a minimization problem and iterative gradient techniques are used to automatically adjust the network element values to approximate a desired network response with minimum error. Constraint equations are used to insure physical realizability. With this formulation the design problem becomes a nonlinear programming problem. The result is an automatic procedure which uses the circuit designer's experience in setting up the problem but requires no tedious labor on his part.

INTRODUCTION

Suppose we have a general computer program to analyze linear networks. It may perform a straightforward nodal or mesh analysis or may use more complicated topological methods. In any event, we assume that it computes specified network responses from a description of the structure or graph of the network. Suppose that we also have a synthesis requirement in terms of a desired network response. We can then consider a brute force synthesis procedure that will force the actual response to approximate the desired response by automatic adjustment of the element values and repeated use of the analysis program.

Consider the network shown in Figure 1; it is desired to make this a first-order Butterworth filter by adjusting the element C1.

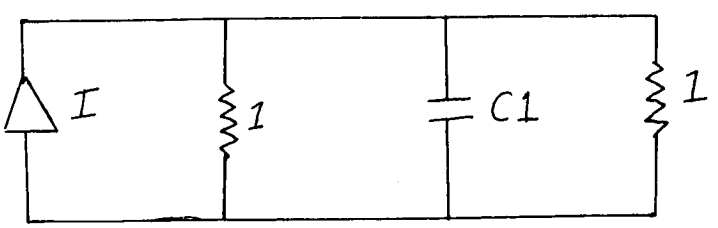


Figure 1 First Order Butterworth Filter

Denote C1 by the variable x and assume that the frequency function of interest is $\mathrm{Re}\left\{Z_{in}(\omega)\right\}$. A straightforward nodal analysis of the network would give

$$\hat{H}(\omega;x) = \mathrm{Re}\left\{Z_{in}(\omega;x)\right\} = \frac{1}{1+\omega^2 C_1^2} = \frac{1}{1+\omega^2 x^2} \quad (1)$$

The desired response obtained from a first-order Butterworth filter is

$$H(\omega) = \mathrm{Re}\left\{Z_{in}(\omega)\right\} = (1+\omega^2)^{-1} \quad (2)$$

Before we may proceed we must assign some measure of error between $\hat{H}(\omega;x)$ produced by the network and the specified $H(\omega)$. If we choose a least squares error criterion, we can establish the error function

$$E(x) = \int_{-\infty}^{+\infty} \left(H(\omega) - \hat{H}(\omega;x)\right)^2 d\omega = \int_{-\infty}^{+\infty} \frac{\omega^4(x^2-1)^2}{(1+\omega^2)^2\,(1+\omega^2x^2)^2}\, d\omega . \quad (3)$$

Note that even for a simple problem, this straightforward approach has resulted in an error function which is very nonlinear in the parameter x. In general, there is no assurance that E(x) will be convex.

The proposed approach does not include an analytical expression for $\hat{H}(\omega;x)$ as given in (1); instead, $\hat{H}(\omega;x)$ will be computed at a number of discrete frequency points using the analysis program. Thus, numerical integration is implied in (3), so the limits of the integration and the frequency increments to be used are an important part of the specification of the error criterion. Iterative gradient techniques like those discussed by Dr. Temes will be used to minimize E(x), but there is no assurance that these techniques will produce a solution which is physically realizable. Thus, there is a need to constrain the region of x over which a solution will be acceptable. It is clear, for example, that we require $x \geq 0$. In general, nonlinear constraint inequalities such as

$$\frac{\omega_0 L_1}{R_1} \leq 100$$

will be required to insure practical solutions. In addition, the inclusion of constraints gives the designer additional leverage in controlling the solutions obtained by the automated procedures.

THE NONLINEAR PROGRAMMING PROBLEM

If we consider several adjustable parameters and denote them by the n-component vector

$$\bar{X} = (x_1, x_2, \ldots x_n)^T,$$

we can state the nonlinear programming problem as follows:

Determine a vector $\bar{C}$ which minimizes the nonlinear function $E(\bar{X})$ where $\bar{X}$ is an n-component vector subject to nonlinear constraint inequalities of the form

$$q_i(\bar{X}) \geq 0 \quad i, = 1, 2, \ldots m. \tag{4}$$

If either $E(\bar{X})$ or any $q_i(\bar{X})$ is nonlinear, the problem can be considered a nonlinear programming problem. For the network design problem we assume that the structure of the network, the desired response, the measure of error and initial estimates of the element values, $\bar{X}^0$, have been given.

CARROLL'S RESPONSE-SURFACE TECHNIQUES

Carroll's optimization technique as developed by Fiacco and McCormick[1,2] transfers the problem of minimization of a nonlinear function with nonlinear constraints to an unconstrained minimization problem by forming a new function

$$\varphi(\bar{X};r) = E(\bar{X}) + r \sum_{i=1}^{m} \frac{1}{q_i(\bar{X})} \tag{5}$$

where r is a real positive parameter. The first term in equation (5) is the function to be minimized, the second term is the penality for adding the constraints. An iterative procedure is used to minimize equation (5) for a strictly monotonic decreasing sequence of r values, $\{r_j\}$, to obtain a sequence of points $\{\bar{C}(r_j)\}$ that respectively minimize $\varphi(\bar{X}; r_j)$. In the limit $C(r_k) \to \bar{C}$ as $k \to \infty$.

Computer programs are written to calculate the sequence of minima $\{\bar{C}(r_j)\}$. An initial estimate $\bar{X}^0$ is given by the designer as input. The initial perturbation parameter r_0, which depends on $\bar{X}^0$ is then calculated and iterative gradient procedures[3] are used to determine $\bar{C}(r_0)$ by minimizing $\varphi(\bar{X};r_0)$. The perturbation parameter is reduced and the minimization repeated so that at the jth stage we

minimize $\phi(\bar{X};r_j)$ using $\bar{C}(r_{j-1})$ as the starting point. If $E(\bar{X})$ and $-q_i(\bar{X})$ are convex, proofs[4] of the convergence of $\bar{C}(r_j)$ can be given. In practice, it is doubtful that any assurance can be given concerning the convexity of $E(\bar{X})$ using a general analysis program of the standard variety. Thus, some art is left in the design problem in that convergence will depend on the initial estimate $\bar{X}^0$. This approach has been used successfully for a non-trivial design problem[5] in which a nodal analysis routine was used to adjust eight parameters of a two-stage band pass amplifier.

If one is serious about using the computer for design, perhaps the general analysis program should be abandoned. The desirability of using analysis methods which result in convex error functions should be obvious. Also, since the bulk of the computing time is spent in minimization using gradient techniques, analysis methods which allow exact partial derivations as opposed to approximations by ratios of differentials will greatly increase the efficiency of the minimization procedure[3,6]. These objectives become feasible if we restrict the structure of the networks being designed. Lasdon and Waren[7], for example, obtain exact partial derivatives but not convex error functions by restricting the networks to doubly-terminated filter structures and using ABCD parameters to compute insertion loss. They use an error criterion which either maximizes the amount by which the insertion loss exceeds the specifications or minimizes the amount by which the insertion loss fails to meet the specifications. This type of error criterion is much more desirable than that of least squares for filter design problems.

LINEAR PROGRAMMING SEQUENCE

Another approach to the solution of the nonlinear programming problem is to transfer it to a sequence of linear programming problems by using the linear terms of a power series expansion about the initial estimate $\bar{X}^0$ and subsequent estimates $\bar{X}^j$ of the parameters. Watanabe[8,9] et al formulate the problem as follows.

The error criterion is chosen as a nonuniform Tchebycheff approximation of the form:

$$E(\bar{X}) = \max_{\omega} \; W(\omega) \left| \hat{H}(\omega;\bar{X}) - H(\omega) \right| \tag{6}$$

where $W(\omega)$ is a positive scalar weighting function. The constant inequalities are of the form

$$q_k(\omega_i;\overline{X}) \le 0 \qquad \begin{matrix} k = 1, \ldots \ell \\ i = 1, \ldots m_k \end{matrix} \tag{7}$$

Consider a set of m frequency points for which equation (6) is to be applied. The problem described by equations (6) and (7) may then be reformulated so that we minimize ϵ subject to constraints of the form

$$\begin{aligned} &\hat{H}(\omega_i;\overline{X}) - H(\omega_i) - \frac{\epsilon}{W(\omega_i)} \le 0 \qquad i = 1, \ldots m \\ &-\hat{H}(\omega_i;\overline{X}) + H(\omega_i) - \frac{\epsilon}{W(\omega_i)} \le 0 \qquad i = 1, \ldots m \\ &q_k(\omega_i;\overline{X}) \le 0 \quad i = 1, \ldots m_k \quad k = 1, \ldots \ell \end{aligned} \tag{8}$$

The problem formulated in (8) is still nonlinear because both $\hat{H}(\omega_i;\overline{X})$ and $q_k(\omega_i;\overline{X})$ are nonlinear in the parameters $\overline{X}$. This problem may be reformulated into a linear problem by approximating $\hat{H}(\omega_i;\overline{X})$ near $\overline{X}^r$ by $\sum_{j=1}^{n} \frac{\partial \hat{H}(\omega_i;\overline{X}^r)}{\partial x_j} \Delta x_j + \hat{H}(\omega_i;\overline{X}^r)$ and $q_k(\omega_i;\overline{X})$ by $\sum_{j=1}^{n} \frac{\partial q_k(\omega_i;\overline{X}^r)}{\partial x_j} \Delta x_j + q_k(\omega_i;\overline{X}^r)$. Thus for the first inequality in (8) we obtain,

$$\sum_{j=1}^{n} \frac{\partial \hat{H}(\omega_i \overline{X}^r)}{\partial x_j} \Delta x_j - \frac{\epsilon}{W(\omega_i)} + \hat{H}(\omega_i;\overline{X}^r) - H(\omega_i) \le 0 \tag{9}$$

$$i = 1, \ldots m$$

The other inequalities in (8) are expanded in like fashion. Assuming an initial estimate $\overline{X}^0$ and an ϵ which satisfy (8), we compute a sequence of points $\overline{X}^1, \overline{X}^2, \ldots$. The point $\overline{X}^{r+1}$ is determined from $\overline{X}^r$ using the first order approximation in (9) to determine Δx_j. Each iteration is a linear programming problem (linear in Δx_j) which can be solved using the SIMPLEX method.[4]

CONCLUSIONS

Both examples of nonlinear programming techniques for solving network design problems involve a sequence of minimizations. If the constraint inequalities are ignored, only a single minimization is required but the results may not be acceptable. The inclusion of constraint inequalities, however, gives a great deal of leverage in controlling the design and is felt to be worth the additional complexity. In general, even the nonlinear programming techniques may require more than one attempt for an acceptable solution. Nevertheless, a number of designers have been successful in the design of crystal filters, group delay equalizers and matching networks using these techniques.

REFERENCES

1. A. V. Fiacco and G. P McCormick, "Programming under Nonlinear Constraints by Unconstrained Minimization: A Primal -- Dual Method, " Research Analysis Corporation Technical Paper RAC-TP-96; Bethesda, Md; (September, 1963).

2. A. V. Fiacco and G. P. McCormick, "The Sequential Unconstrained Minimization Technique for Nonlinear Programming. Algorithim II, Optimum Gradients by Fibonacci Search, " Research Analysis Corporation Technical Paper TAC-TP-123, McLean, Virginia, (June, 1964).

3. R. Fletcher and M.J.D. Powell, "A Rapidly Convergent Descent Method for Minimization", Computer Journal Vol. 6 (1963), pp. 163-168.

4. T. L. Saaty and J. Bram, Nonlinear Mathematics, McGraw-Hill, (1964); pp. 141-144, 106-112.

5. P. Scheibe and E. Huber, "The Application of Carroll's Optimization Technique to Network Synthesis, " Proceedings of Third Annual Allerton Conference on Circuit and System Theory, University of Illinois; (1965) pp. 182-191.

6. R. Fletcher and C. M. Reeves, "Function Minimization by Conjugate Gradients, "Computer Journal; (July, 1964); pp. 149-154.

7. L. S. Lasdon and A. D. Waren, "Optimal Design of Filters with Bounded, Lossy Elements", IEEE Transactions on Circuit Theory; (June, 1966); pp. 175-187.

8. Y. Ishizaki, H. Wakazuki and H. Watanabe, "Some Considerations on Min-Max Approximation Problems for Network Design, " Proceedings of Third Annual Allerton Conference on Circuit and System Theory, University of Illinois, (1965); pp. 786-795.

9. Y. Ishizaki and H. Watanabe, "An Iterative Method for Network Design as a Nonlinear Programming Problem, " Nippon Electric Ltd., Monograph DEB, 485, Kawaski, Japan, 1964.

THE SENSITIVITY OF ACTIVE CIRCUITS

N67 13351

R. C. DORF
University of Santa Clara
Santa Clara, California

ABSTRACT

The sensitivity of an active circuit to parameter variations is an important consideration in the analysis and design of modern solid-state circuits. In this paper, several useful sensitivity indices will be discussed and illustrated. It will be shown that the root sensitivity which is calculated on the basis of the state vector differential equation is particularly suitable for digital computer calculations.

INTRODUCTION

An active circuit, whatever its nature, is subject to a changing environment, aging, ignorance of the exact values of the circuit parameters, and other natural factors which affect a circuit. The variation of the parameters of a circuit will often have an important effect on the performance of the circuit. Thus, the sensitivity of the performance of a circuit to parameter variations is of prime importance and is the subject of this paper.

The sensitivity of a circuit is defined as the effect of parameter variations on the performance of the circuit. Thus, in general, the sensitivity of a circuit is defined as

$$S_{p_i}^{I_j} = \frac{\Delta I_j}{\Delta p_i} \tag{1}$$

where I_j is the jth performance index and p_i is the ith parameter. For example, if we are concerned with the node voltage e_1 and the effect of the resistance R_1, we have

$$S_{R_1}^{e_1} = \frac{\Delta e_1}{\Delta R_1} \tag{2}$$

where ΔR_1 is the variation in the resistance R_1. Alternatively, one may utilize the normalized sensitivity measure which is defined as

$$S_{p_i}^{I_j} = \frac{\Delta I_j / I_j}{\Delta p_i / p_i} \tag{3}$$

In active circuit analysis, the performance indices of interest are, among others: node voltages, transient response, ac response, gain, and the circuit transfer function. In order to accomplish computer-aided circuit analysis and design, one often selects sensitivity measures which are readily calculated using a digital computer. As we shall find in the ensuing paragraphs, several sensitivity indices are more readily utilized in computer-aided analysis.

In order to illustrate the effect of parameter variations let us consider the open-loop circuit of Fig. 1 and the feedback circuit of Fig. 2. The effect of a change in the circuit $A(s) + \Delta A(s)$ of

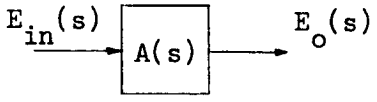


Figure 1.

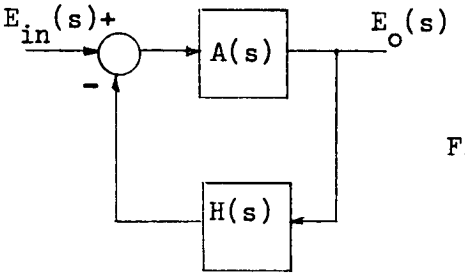


Figure 2.

the open-loop circuit results in the output

$$E_o(s) + \Delta E_o(s) = (A(s) + \Delta A(s))E_{in}(s)$$

or

$$\Delta E_o(s) = \Delta A(s) E_{in}(s) \tag{4}$$

Thus, the change in the circuit results in a proportional variation in the output. For the closed-loop feedback system of Fig. 2, we have

$$E_o(s) + \Delta E_o(s) = \frac{A(s) + \Delta A(s)}{1 + (A(s) + \Delta A(s))H(s)} E_{in}(s) \qquad (5)$$

Then, the change in the output is approximately

$$\Delta E_o(s) = \frac{\Delta A(s)}{1 + AH(s)} E_{in}(s) \qquad (6)$$

Comparing equations (4) and (6) we note that the change in the output has been reduced by the factor (1 + AH(s)) which is usually much greater than one over the range of frequencies of interest. Thus, we find that the sensitivity of an active circuit can be reduced by the introduction of feedback.

BODE SENSITIVITY

The definition of sensitivity attributed to Bode is stated in terms of the transfer function of a circuit, $T(s) = E_o(s)/E_{in}(s)$, as[1]

$$S^T_{\Delta A} = \frac{\Delta T(s)/T(s)}{\Delta A(s)/A(s)} \qquad (7)$$

In the limit, for small incremental parameter changes, we obtain

$$S^T_{dA} = \frac{dT/T(s)}{dA(s)/A(s)} \qquad (8)$$

Thus, we may show using Eqn. (8), that the sensitivity of the open-loop circuit of Fig. 1 is equal to one. The sensitivity of the feedback circuit of Fig. 2 is

$$S^T_{dA} = \frac{1}{1 + AH(s)} \qquad (9)$$

Similarly, the sensitivity of the transfer function to the feedback network is

$$S^{T(s)}_{dH(s)} = \frac{-AH(s)}{1 + AH(s)} \qquad (10)$$

Therefore, we may reduce the sensitivity of an active circuit S_A^T by establishing $AH(s) >> 1$. However, the sensitivity S_H^T of the circuit to the feedback network then becomes approximately 1 . Thus, the feedback circuit must be constructed of reliable components which will not vary with environmental changes.

In order to illustrate the effectiveness of feedback in reducing the sensitivity of an electronic circuit let us consider the simple circuit shown in Fig. 3. The signal flow graph of the amplifier is shown in Fig. 4.

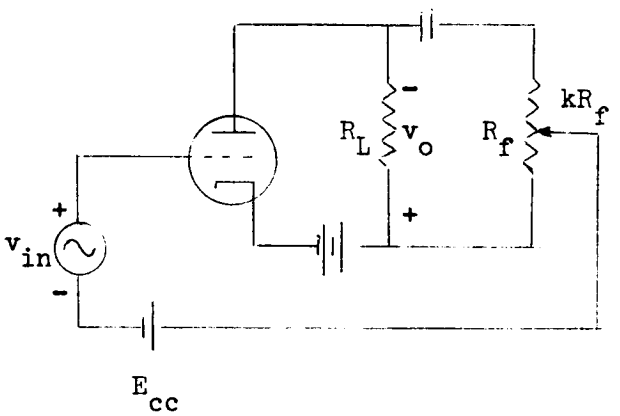


Figure 3.

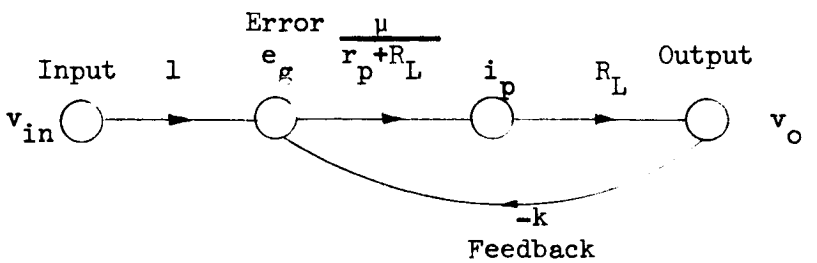


Figure 4.

Clearly, the gain without the feedback is $A = \dfrac{\mu R_L}{r_p + R_L}$. The sensitivity of the open loop system to changes in μ is equal to one. The sensitivity of the closed loop circuit is

$$S_{\mu}^{T} = \frac{\partial T/T}{\partial \mu/\mu} = \frac{1}{1 + Ak} \qquad (11)$$

where $T = \frac{A}{1+Ak}$. For a typical value of $A = 20$ and if $k = 0.5$, then $S_{\mu}^{T} = 1/11 = 0.091$.

ROOT SENSITIVITY

The sensitivity index utilized by Bode is useful for illustrating the concept of sensitivity and the value of the introduction of feedback in order to reduce the sensitivity of an electronic circuit. However, it is not a particularly useful index for computer analysis or design. Another, more potentially useful index is defined in terms of the characteristic roots of the circuits and is written as

$$S_{p_j}^{r_i} = \frac{\partial r_i}{\partial p_j / p_j} \tag{12}$$

In this definition, r_i = the ith characteristic root, p_j = jth parameter and the circuit transfer function is written as

$$T(s) = \frac{K \prod_{m=0}^{M} (s + z_m)}{\prod_{i=1}^{n} (s + r_i)} \tag{13}$$

The sensitivity index $S_{p}^{r_i}$ is defined as the root sensitivity of a circuit. The evaluation of the root sensitivity of a circuit may be obtained utilizing root locus methods.[2] As an example, let us consider the feedback circuit shown in Fig. 2, where

$$A(s) = \frac{K}{s(s+\beta)} \quad \text{and} \quad H(s) = 1$$

The characteristic equation of this circuit is $s^2 + \beta s + K = 0$ or written in root locus form we have

$$1 + \frac{K}{s(s + \beta)} = 0 \tag{14}$$

Considering the case where the nominal values of $K = 0.5$ and $\beta_o = 1.0$ the resulting characteristic roots are $r_1 = -0.5 + j0.5$ and $r_2 = r_1^*$. The locus of roots for this circuit as a function of the gain K is shown in Fig. 5. For a $\pm$ 20% change in K we evaluate the root locations by root locus methods as shown in Fig. 5. Thus the root sensitivity for r_1 is

$$S^{r_1}_{+\Delta K} = \frac{\Delta r_1}{\Delta K/K} = \frac{+j0.09}{+0.2} = 0.45 \underline{/ +90^\circ} \tag{15}$$

The pole β also may vary as a result of environmental changes, so that $\beta = \beta_o + \Delta\beta$. Then, the effect of $\Delta\beta$ is represented in the characteristic equation as

$$s^2 + s + \Delta\beta s + 0.5 = 0 \tag{16}$$

since the nominal value of β is $\beta_o = 1$ and the nominal value of gain is $K = 0.5$. Rewriting Eqn (16) in root locus form we obtain

$$1 + \frac{\Delta\beta s}{s^2 + s + 0.5} = 0 \tag{17}$$

We note that the denominator is equal to the unchanged characteristic equation when $\Delta\beta = 0$. The root locus for changes in $\Delta\beta$ is shown in Fig. 6. We also note that for small changes in β one may use the departure vector as an approximation to the locus of roots. Evaluating the root sensitivity from the root locus, we obtain

$$S^{r_1}_{+\Delta\beta} = \frac{0.16 \underline{/-131^\circ}}{0.20} = 0.80 \underline{/-131^\circ} \tag{18}$$

$$S^{r_1}_{-\Delta\beta} = \frac{0.125 \underline{/39^\circ}}{0.20} = 0.625 \underline{/39^\circ} \tag{19}$$

One observes that the angle of the root sensitivity is as important a factor as the magnitude since the direction of the movement of the root indicates the change in the relative stability of the circuit. Comparing the sensitivity of the root due to K and β we find that the sensitivity of the root due to the pole β is more important owing to the

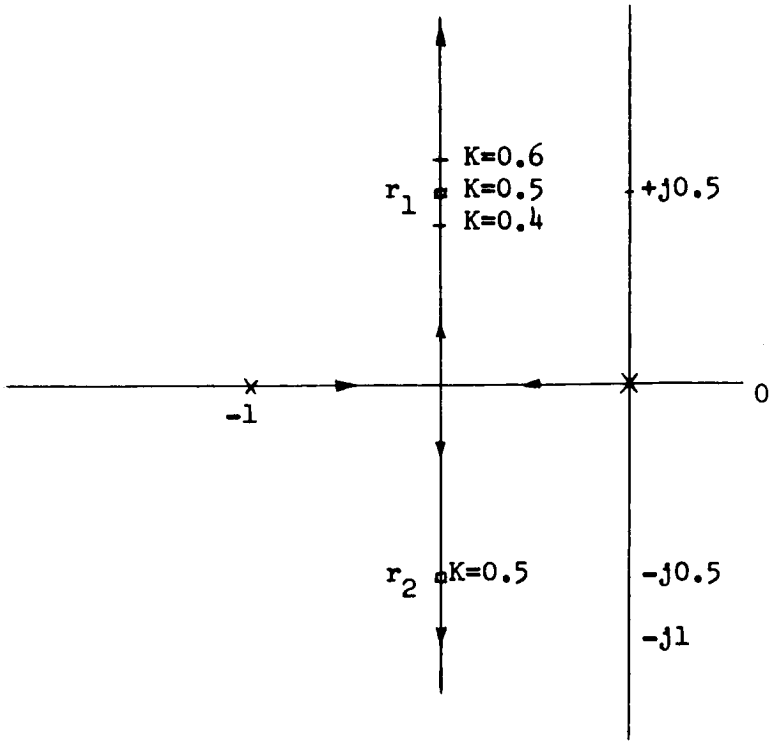


Fig. 5 - The root locus for K

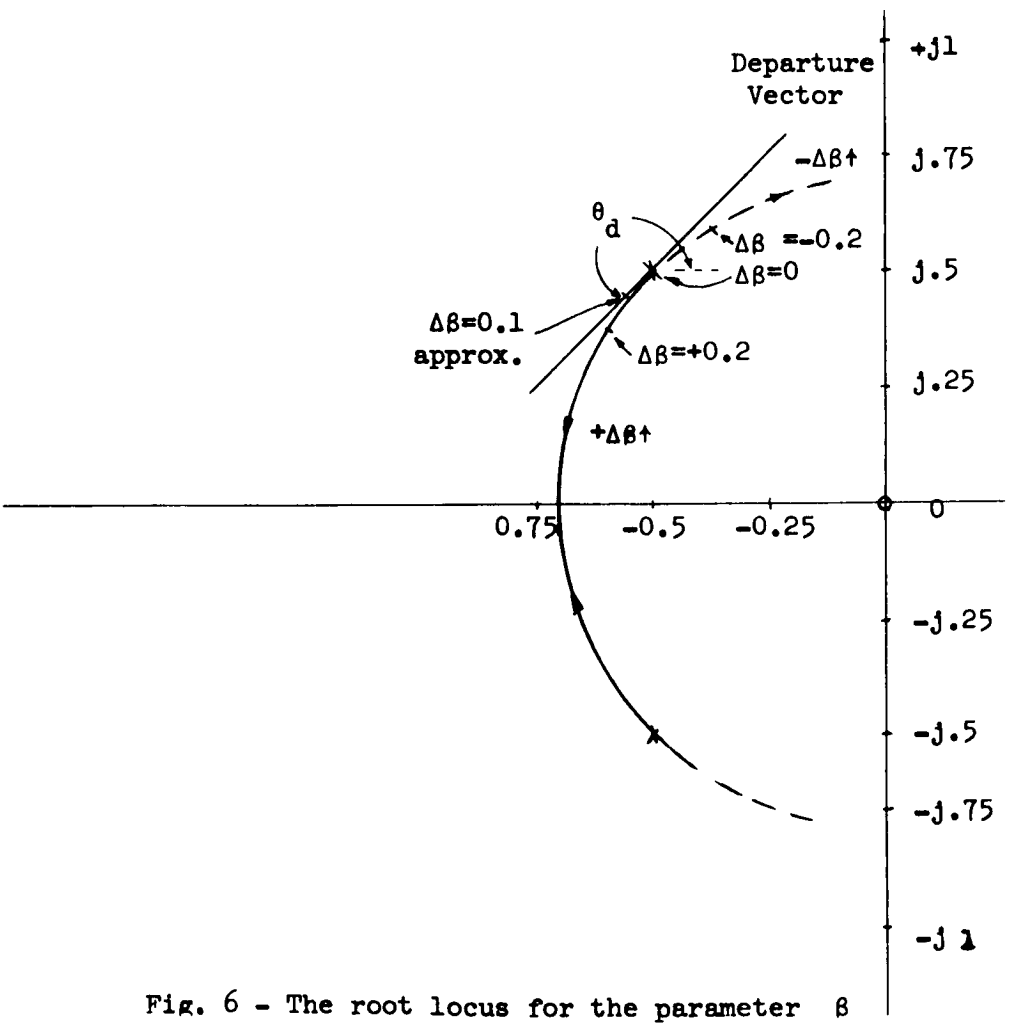


Fig. 6 - The root locus for the parameter β

larger magnitude of $S^{r_1}_{\Delta\beta}$ as well as the direction of the roots which is towards lower damping for a reduction in β .

SENSITIVITY IN THE TIME-DOMAIN

Several sensitivity measures may be developed which will be useful for computer evaluation. A sensitivity measure which is particularly useful for nonlinear circuits is defined in terms of the sensitivity coefficients. The sensitivity coefficients (or variables) are defined in the time-domain as

$$v_i(t) = \frac{\partial x_i(t)}{\partial p} \tag{20}$$

where $x_i(t)$ = the ith state variable and p is the parameter that is varying due to environmental changes. Therefore, for a set of n state variables we may define the sensitivity vector as

$$\underline{v}(t) = \frac{\partial \underline{x}}{\partial p} \tag{21}$$

where $\underline{x} = (x_1, x_2, \ldots, x_n)^T$ = state vector.[3] The state variables are commonly selected as the capacitor voltages and the inductor currents for an active circuit. The state vector differential equation for the circuit is written as

$$\dot{\underline{x}} = \underline{f}(\underline{x}, \underline{u}, t) \tag{22}$$

where $\underline{u}$ = the vector of input signals. Equation (22) may be written as

$$\underline{F}(\dot{\underline{x}}, \underline{x}, \underline{u}, t) = 0 \tag{23}$$

and one may obtain the derivative of $\underline{F}$ with respect to p as

$$\frac{\partial \underline{F}}{\partial p} = \frac{\partial \underline{F}}{\partial \dot{\underline{x}}} \frac{d\underline{x}}{dp} + \frac{\partial \underline{F}}{\partial \underline{x}} \frac{d\underline{x}}{dp} + \frac{\partial \underline{F}}{\partial p} \frac{dp}{dp} = 0 \tag{24}$$

Then, for example, for the linear system we obtain

$$\frac{\partial \underline{F}}{\partial \dot{\underline{x}}} = -I, \quad \frac{\partial \underline{F}}{\partial \underline{x}} = A, \quad \frac{\partial \underline{F}}{\partial p} = \frac{\partial \Delta}{\partial p}\,\underline{x} = D\underline{x}, \quad \text{and thus}$$

$$\frac{d\dot{\underline{x}}}{dp} = A\,\frac{d\underline{x}}{dp} + D\underline{x} \tag{25}$$

For the definition of the sensitivity coefficients $\underline{v} = \frac{\partial \underline{x}}{\partial p}$ equation 25 becomes

$$\dot{\underline{v}} = A\underline{v} + D\underline{x}(t) \tag{26}$$

The solution of this linear sensitivity equation may be obtained by computer methods.[4] The solution of equation (26) may also be written as

$$\underline{v}(t) = \phi(t)\,\underline{v}(o) + \int_o^t \phi(t-\tau)\,D\,\underline{x}(\tau)\,d\tau \tag{27}$$

where $\phi(t)$ is the transition matrix, e^{At}. For example, consider the second-order system described by the linear state vector differential equation

$$\dot{\underline{x}} = A\,\underline{x} + B\,\underline{u} \tag{28}$$

where $A = \begin{bmatrix} 0 & 1 \\ -p & -3 \end{bmatrix}$ and the nominal value of $p = -2$. Then, the matrix D is as follows:

$$D = \frac{dA}{dp} = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix} \tag{29}$$

and the sensitivity vector is

$$\underline{v}(t) = \phi(t)\,\underline{v}(o) + \int_o^t \begin{bmatrix} \phi_{12} & (t-\tau) \\ \phi_{22} & (t-\tau) \end{bmatrix} (-x_1(\tau))d\tau \qquad (30)$$

The typical transient response of the two sensitivity coefficients is shown in Fig. 7.

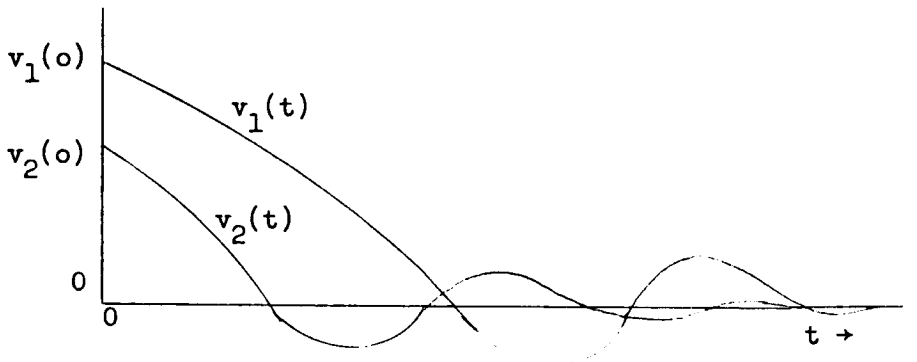


Figure 7.

In order to obtain a suitable sensitivity measure utilizing the sensitivity coefficients, one might use

$$S = \int_o^\infty (\underline{v}^T\underline{u})dt = \int_o^\infty (v_1^2(t) + v_2^2(t))dt \qquad (31)$$

Using the measure as represented by equation 31, a designer could include the effect of sensitivity in the computer-aided design procedure.

ROOT SENSITIVITY AND THE STATE VECTOR FORMULATION

The root sensitivity of a linear system represented by the time-domain vector differential equation

$$\dot{\underline{x}} = A\,\underline{x} + B\,\underline{u} \qquad (32)$$

may be determined by a digital computer program.[5] The change in any root r_i for a change in the parameters of A is

$$dr_i = \frac{R(s) * dA}{tr\ R(s)} \Bigg|_{s=r_i} \tag{33}$$

where R(s) is the adjoint matrix and dA is the differential change in A . The asterick indicates the inner product of two matrices, that is

$$A * B = a_1 b_1 + a_2 b_2 + \ldots \tag{34}$$

where a_i = ith row of A and b_i = ith column of B . The symbol tr denotes the trace of a matrix. Equation (33) is obtained by using an algorithm for the characteristic roots and the characteristic matrix which is particularly applicable to digital computer calculation.[6] Then, we have

$$R(s) = Is^{n-1} + R_1 s^{n-2} + R_2 s^{n-3} + \ldots \tag{35}$$

The algorithm for generating $R_k(s)$ is

$$R_k = AR_{k-1} - d_k I \tag{36}$$

where $d_k = \frac{1}{k} tr\ AR_{k-1}$, and $R_o = I$.

As an example of this method let us reconsider the feedback circuit of Fig. 2 with a transfer function (see Eqn. (14)

$$T(s) = \frac{0.5}{s^2 + \beta s + 0.5} \tag{37}$$

and the nominal value of β is 1.0 . The resulting characteristic root of interest is $r_1 = -0.5 + j0.5$ and we wish to determine the root sensitivity of r_1 due to a small change in β . The circuit with the transfer function of eqn. (37) may be represented by the state vector equation

$$\dot{\underline{x}} = \begin{bmatrix} 0 & 1 \\ -0.5 & -\beta \end{bmatrix} \underline{x} + \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix} e_{in}$$

$$= A\,\underline{x} + B\,u \tag{38}$$

Therefore, $dA = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} d\beta$, $R_o = I$, $d_1 = \text{tr } A = -\beta = -1$

$$R_1 = Ar_o - d_1 I = \begin{bmatrix} 1 & 1 \\ -0.5 & 0 \end{bmatrix} \tag{39}$$

Thus, equation (36) becomes

$$R(s) = Is + R_1 = \begin{bmatrix} (s+1) & 1 \\ -0.5 & s \end{bmatrix} \tag{40}$$

Using equation (33) to evaluate the root change we obtain

$$dr_1 = \left. \frac{R(s) * dA}{\text{tr } R(s)} \right|_{s=r_1}$$

$$= \left. \frac{\begin{bmatrix} (s+1) & 1 \\ -0.5 & s \end{bmatrix} * \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} d\beta}{(2s+1)} \right|_{s=r_1}$$

$$= \frac{-r_1\, d\beta}{(2r_1+1)} \tag{41}$$

Since $r_1 = -0.5 + j0.5$, we have a root sensitivity

$$S^{r_1}_{+\beta} = \frac{dr_1}{d\beta/\beta} = 0.5\ \underline{/-135^\circ} \tag{42}$$

This root sensitivity calculation may be compared with that obtained in Eqn. (18) for incremental changes in the parameter β. This method of calculating the root sensitivity of an active circuit which is based on the time-domain equations describing the circuit is exceedingly useful for digital computer calculation and a program is available at present.[5]

CONCLUSIONS

The sensitivity of an active circuit to parameter variations is an important consideration in the analysis and design of active circuits. Several useful sensitivity indices have been discussed and illustrated. The root sensitivity measure which is calculated on the basis of the state vector differential equation is particularly useful for digital computer calculations. Alternatively, the sensitivity coefficients are particularly useful for indicating the sensitivity if an analog or hybrid computer is available for computational purposes. In any case, the usefulness of sensitivity measures to indicate the variation of a circuit to parameter variations is of prime importance and should be considered in the design of modern solid-state circuits.

REFERENCES

1. H. W. Bode, Network Analysis and Feedback Amplifier Design, Von Nostrand, Princeton, N.J., 1945.
2. R. C. Dorf, Modern Control Systems, Addison-Wesley Publishing Co., Reading, Mass., March, 1967.
3. R. C. Dorf, Time Domain Analysis and Design of Control Systems, Addison-Wesley Publishing Company, Reading, Mass., 1965
4. R. C. Dorf, "System Sensitivity in the Time-Domain," Proceedings of the 3rd Allerton Conference on Circuit and System Theory, pp. 46-62, 1965.
5. B. S. Morgan, Jr., "Computational Procedures for Sensitivity Coefficients in Time-Invariant Multivariable Systems," Proceedings of the 3rd Allerton Conference on Circuit and System Theory, pp. 252-258, 1965.
6. D. K. Faddeev and V. N. Faddeeva, Computation Methods of Linear Algebra, Freeman and Co., San Francisco, 1963, pp. 229.

N67 13352

COMPUTER ASSISTANCE IN CIRCUIT ANALYSIS AND DESIGN

WILLIAM W. HAPP
NASA Electronics Research Center
Cambridge, Massachusetts

JOHN STAUDHAMMER
Arizona State University
Tempe, Arizona

ABSTRACT

Salient desirable features of computer programs for circuit analysis are discussed. Gross characteristics of some computer programs are given. A survey to determine the availability of computer routines for solid state circuits discusses: scope of program extent of present use, range of applications, availability, interchange of programs and feedback of information. The feasibility of a pool of programs for circuit designers is examined.

SCOPE OF CIRCUIT DESIGN PROGRAMS

Copeland (1) of Bell Telephone Laboratories recently discovered a new oscillation mode from a computer program of a modelled Gunn effect oscillator. This serves to illustrate that the scope of circuit analysis computer programs goes far beyond the mere corroboration of a designer's calculation.

Breadboarding and testing of electronic circuits prior to fabrication is increasingly replaced by computer programmed analysis. According to a recent cost analysis (2) this new method is more economical than conventional testing of a circuit, and further allows the compilation of statistical records for reliability and production yield, which by bench testing and conventional calculations is uneconomical and impractical.

A detailed survey of computer programs for circuits should include programs for production yield prediction, photomask generation and similar steps in the manufacture of integrated circuits. Likewise computer programs on circuits are intimately tied to topics ranging from devices to systems.

A great number of organizations have developed their own programs for the analysis of circuits. In fact very little original work is needed if one is satisfied to start with some matrix description of the linear network, as was pointed out elsewhere in this Proceedings. Difficulties arise when nonlinear networks are to be analyzed or when some topological description is used to enter the network into the computer. For nonlinear networks the existence of a solution is not always easy to assess (in essence every element must satisfy the Lipshitz conditions (20).) For general topological inputs an input language must be developed, as was described in Mr. Carpenter's paper, or as is implemented in ECAP. The simplest of such "home-brewed" programs may be a collection of subroutines to step through a set of network relations which are entered as data cards. In this case the program may be designed to automatically carry out complex number calculations without having to be concerned with the mechanics of implementation. One example of such a FORTRAN coded program, which produces a FORTRAN deck from input data cards, was described in (21).

The most complicated of the circuit analysis programs presently in wide use are the general circuit analysis programs. There are presently well over a dozen of these in general use and it is expected that in the near future a veritable avalanche of these might become available. An assessment of these programs for a given user will have to be made. The salient desirable features of these programs should include (6):

1. Simple Input. A clerk with no circuit analysis background and no knowledge of computing should be able to enter the required input data. Input should include
   a. Topology;
   b. Circuit Values;
   c. Excitations;
   d. Output Modes.

Each of the above categories should be separate without the need of re-inputting all data.

2. Variable Models. In active circuit analysis one of the chief considerations is the equivalent circuit used for the active components. There should be choice to utilize a variety of equivalent circuits as the accuracy of the desired analysis dictates.

3. Nonlinearities. A wide range of these must be considered. Typical ones include saturation and reverse voltage breakdown, but could include thermal considerations also.

4. Selective Outputs. Due to the low cost of computation, it is no trick at all to envelop the circuit designer in reams of output data. What is required is a set of options for significant outputs - and some automation in ignoring most numbers.

5. Automatic Parameter Modification. As pointed out in 1. changes in parameter values should not require complete re-inputting. Such modifications are necessary in tolerance analysis and in automated design procedures.

6. Error Checks. The reliability and accuracy of the answers provided should be easy to assess. Generally every answer, no matter how inaccurate, is printed with maximum precision in every program. No such automatic error checks are presently available in circuit analysis programs.

For automated design usually these programs are combined with some optimization procedures. Performance criteria are calculated and are compared against the desired values. Error measures are derived and are then used to adjust the parameters and/or the topology of the test circuit. For such use the above criteria for analysis programs must be extended to

7. Optimization. This must be done simultaneously for the various parameters in the network. Procedures for this task are not worked out in general; however, much effort is currently expended in this field (22,6.)

8. Flexible Objective Description. In every design effort there is some function, often only verbally circumscribed, which must be optimized. Unless the objective function can be described in quantitative terms to a computer program, no optimization can begin. Presently many objectives ("simplicity", "reliability", "ease of trouble shooting", etc.) can be expressed only incompletely and with great difficulty in numerical form. Further developments in this area are necessary either for writing compilers which "understand" more verbal descriptions, or in educating the designers to use less qualitative descriptors.

FEATURES OF SOME PROGRAMS

In light of the above desired properties of general network analysis programs four programs, available to qualified users, are compared here. This list is meant to be by no means exhaustive or even suggestive of programs available. Very limited descriptions are given here for ECAP, NET-1, PREDICT, and CIRCUS.

1. Electronic Circuit Analysis Program (ECAP). Probably the widest used program, developed originally by Norden Division of United Aircraft Corp. in cooperation with IBM. This program is a direct descendant of Branin's Transistor Analysis Program (TAP) (24) and is freely distributed by IBM (25). Versions exist for the IBM 1620 and the 7090/94 computers; users have modified this program for virtually all other computers. The program performs DC, AC, and Transient Analysis of piecewise linear networks. By means of program controlled switches it is possible to model most mildly nonlinear networks. The user must provide the equivalent circuits which are to be used for the active devices; in short, the program input is the equivalent network to be analyzed. The program performs stepwise integration. Solution pieces are fitted together at various boundaries defined by switch actions. The program consists of approximately 7500 FORTRAN statements and is quite easy to use. DC and AC analysis (9 separate frequencies) of the three stage amplifier given earlier in this Proceedings (see the paper "Matrices and State Variables") took about 20 minutes on an IBM 1620-II; it took less than a minute on a CDC 3400.

The program provides for automatic parameter variations, and is limited to 50 nodes and 200 branches in the 7094 version.

2. NET-1. Also a descendant of Branin's TAP, this program is designed to handle nonlinear networks with minimal modeling required. The program uses a tape of equivalent circuit parameters for given active network types. Thus all one needs to do is call out the transistors and diodes by type. Unfortunately the equivalent circuit used is fixed; it must accommodate every conceivable use of the active device. Consequently some 35 parameters are needed for the modified Ebers-Moll model. Normally this would not be objectionable, but the use of many parameters (whether needed or not) slows down the computations. Taking advantage of "operating range" simplifications could speed up this program; also the limitation to Ebers-Moll models precludes the accurate analysis of field effect transistor circuits.

For integration the Certaine-Adams method (27), a predictor-corrector procedure, is used. The program exists in two forms: a FAP program for the IBM 7040 and 7090 computers and a MAP program for the IBM 7044 and 7094 computers. Adaptation of the program to other machines has met with little success elsewhere. The program is available from Los Alamos Scientific Laboratory.

3. PREDICT. This program was developed at IBM, Owego, N.Y. for studies of radiation effects in circuits. The program is designed for the IBM 7094 computer. It uses the Beaufoy-Sparks change control model for active devices, but will handle other nonlinearities as a mathematical subroutine. The transistor equivalent circuit must be input to the program, thus simplified equivalent circuits may be used if their use leads to sufficient accuracy.

The program apparently sacrifices running time for accuracy; a fourth order Runge-Kutta integration routine is used for integrating the resultant nonlinear differential equations.

4. CIRCUS. This program, currently being documented at Boeing Corp. in Seattle, is also a radiation effect analyzer. It is, however, written almost entirely in FORTRAN IV with about 150 machine dependent instructions. Versions of this latter portion exist for the IBM 7094, CDC 6600, GE-635, and the Univac 1108 computers. Here again the Beaufoy-Sparks model is used with built-in device values. Integration is speeded up by the use of an "exponential integration routine". The program will be made conditionally available.

It is of interest to compare some of these programs as to their execution time for a given problem. The matter of convenience of input and output is rather subjective: programmers find that their particular programs are easiest to converse with. Thus no attempt was made to compare them for other than execution time. Three circuits, an inverter, a video amplifier, and a bistable multivibrator, were used (23). Results, in arbitrary time units, are given below.

	INVERTER	AMPLIFIER	MULTI.
PREDICT	15.24	15.18	1.5
NET-1	5.24	4.7	0.5
CIRCUS	4.18	2.72	0.15

SURVEYS ON CIRCUIT ANALYSIS AND DESIGN PROGRAMS

Several organizations and individuals have conducted surveys on computer programs not limited in scope to circuits, among these:

a. A Department of Defense (3) survey consists of single page abstracts submitted by interested individuals, and includes the name of the originator of the program.

b. The COSMIC project (4,5) administered by the Computer Center at the University of Georgia at Athens, under a NASA grant is establishing a library of verified programs. The grant is administered by NASA/ Technology Utilization, with the primary purpose of making programs generated by NASA available to industry for a nominal fee. Input data to COSMIC are at present primarily from NASA program libraries at Marshall Space Flight Center in Huntsville, Manned Space Center at Houston and the Electronic Research Center at Cambridge, Massachusetts.

c. The SHARE project is an information interchange restricted to IBM equipment users. Good sources of information are IBM users who are well aware of the scope and limitations of SHARE. Design automation workshops are held several times a year by SHARE participants.

d. The Journal of the Association for Computing Machinery (ACM) maintains a periodic listing of new programs.

Several surveys are directed specifically towards circuits; among these:

a. A survey by Kuo (6), assessing the scope of major network programs, such as ECAP, NET-1 and TAP. A critical evaluation of program languages and limitation imposed by the type of computer are examined.

b. A literature survey by Yang (7) at Villanova University with about 200 cross-referenced papers on Computer-Aided Circuit Design.

c. The forthcoming January 1967 issue of the International Journal of Electrical Engineering Education (8) published in Britain is devoted entirely to programs suitable for courses in electronics.

d. Surveys by several "controlled circulation" journals usually high-light one specific area in circuit design.

e. A survey (9) specializing in computer approaches to flowgraphs and dichotomous techniques emphasizes teaching aspects.

In addition to these surveys irregular publications on new programs appear in the Transactions of the IEEE.

SEMINARS AND CONFERENCES

An effective distribution of new programs occurs frequently at specialized seminars held under the sponsorship of universities or technical societies in a rather informal manner. The following seminars are primarily devoted to computer-aided circuit design,and informal lecture notes have been the source of distribution for many programs and reports, well in advance of their regular publication. Among such seminars are:

January 17-18, 1967	University of Wisconsin Solid-state Circuit Design Rolf Schuenzel, Director Engineering Extension
Jan. 19-20, 1967	University of Wisconsin at Milwaukee Reliability Aspects of Electronics
Jan. 31 - Feb. 2, 1967	New York University - A course for experienced circuit designers without prior knowledge of computers.
Feb. 28 - March 2, 1967	MIT Kresge Auditorium Computer Aided Circuit Design Sponsored by NASA/ERC
April 3-7, 1967	University of California, Los Angeles Automated Circuit Analysis Sam Houston, Head Engineering Extension

SEARCH AND QUALIFICATION PROCEDURES

It is conservatively estimated that at least 2000 programs are now in use in electronics and related fields. It is therefore desirable to establish search procedures to locate a desired program. It may also be appropriate to determine standardization and qualification procedures to assure the usefulness and reliability of a program.

A list of subject descriptors used at NASA-ERC in searching for routines is given in Table 1. Further refinements and additions of descriptors are planned. Every program or subprogram is coded with three subject identifiers, and pertinent information as to program language and accessories required. This information is processed on punched cards and interchanged among interested users. Interested parties are encouraged to have their names added to the information file.

Several hundred programs are so far coded on a provisional basis, and a detailed report is in preparation. Specific information on individual programs will be made available through COSMIC (5). A list of representative programs is given in Table 2. Extensive work is going on at NASA-ERC on the program "Network Analysis for System Applications" described by Mr. Carpenter earlier. Extensions to various fields are documented by NASA reports (10-19).

CONCLUSION

This paper pointed out some of the problem areas in large scale network analysis programs. The combination of these programs with optimization programs should be the next major step in automating circuit design. Once the circuit is known, automated procedures for photomask generation can take over. Such program systems are in experimental use at several major installations; simplified versions should become widely available in the next few years.

One question that naturally arises in this context is whether a special interest group has not de-facto formed in this field. Apart from circuit designers and not quite programmers, the engineers engaged in extensive computer usage are developing skills and expertise in a different field: computer using. It is therefore not too surprising that a movement exists for the establishment of a Computer Users group, preferrably within the professional group framework of IEEE. It is only a matter of time, we feel, before such a group will be formed to aid and assist its membership in this never-never-land between circuit hardware and "circuit software". Meanwhile the most feasible route for the exchange of automated circuit design information appears to be attendance at conferences such as this one, and the development of personal contacts in this field.

Table 1. Information File of Computer Programs for Electronics

Subjects Identifiers

100 System Oriented

110 Model Definition
120 Topology
130 Logic
140 Simulation
150 Strategy
160 Information Theory

200 Arithmetic Oriented

210 Matrix Techniques
220 Relaxation Techniques
230 State-Space Techniques
240 Non-Linear Techniques
250 Numerical Procedures
260 Mathematical Functions

300 Real-time Oriented

310 Frequency Domain
320 Transient Domain
330 Stability Analysis
340 Sampled Data
350 Control Techniques
360 Function Synthesis

400 Statistically Oriented

410 Data Acquisition
420 Tests and Measurements
430 Reliability
440 Stochastic Processes
450 Data Processing
460 Monte Carlo Techniques

500 Device Oriented

510 Process Control
520 Properties of Materials
530 Fabrication Technology
540 Device Characterization
550 Component Technology
560 Information Display

600 Design Oriented

610 Methods and Techniques
620 Cost Diagnostics
630 Linear Circuits
640 Digital Circuits
650 Integrated Circuits
660 Instrumentation

700 Environment Oriented

710 Thermal
720 Nuclear
730 Sensors
740 Weapons
750 Space
760 Life

800 Interface or Related Topics

Table 2 Computer Programs for Circuit Design

Programs	Originators or Users
D.C. Voltage Regulator Design	Moore School
Design of Sampled Data Systems by Linear and Quadratic Programming	Moore School
Network Synthesis	GE - Florida
Gate Assignment	
Load Factor Analysis	
Logic Diagram	
Timing Analysis	
Module Assignment	
Path Routing	
Steady State Analysis of Parametric Amplifiers	Purdue University
Transfer Function and Frequency Response	Bendix
Inverse LaPlace Transform and Time Response	
Filter Programs	
Monte Carlo Analysis	
Matrix Analysis of an Equivalent Circuit Frequency Response	
Worst Case Analysis of Digital Circuits	
Simulation of a Tunnel-diode delay-line Memory Cell	Oregon State Univ.
Tunnel Diode V-I Characteristics	
Statistical Analysis of Amplifier Parameters	
Transistor Amplifier Characteristics from given h parameters	
SLAM - Logic Simulation	TRW
NAP - AC Analysis Program	General Dynamics
LAPL - Analysis by La Place Methods	
PACER - Failure Rates from MIL-H-217	
FORMAC - Non-numeric Calculations	IBM - Cambridge
Linear Programming - Optimization with Linear Constraints	
STRESS - Monte Carlo	
POP - Non Linear Optimization	
LNAR - Linear Network Analysis and Realization	(Calahan)
STANPAC	GE - Phoenix
DEUCE	British Marconi
Worst Case Flip-Flop Design	Prof. Ley - N.Y.U.
Transistor Shunt-series Feedback Pair	
Transistor RL Feedback	
Transistor RC Feedback	
Transistor Shunt Peaked	
Vitro NAP - AC Circuit Analysis Program	Vitro - Silver Spring, Md.

Table 2 Computer Programs for Circuit Design (Cont'd)

Programs	Originators or Users
Transient Analyzer Generator	JPL
Design and Test of Digital Systems	Sperry
A.C. and D.C. Circuit Analysis Program	ARINC
CRAM Computerized Reliability Assessment Method	ARINC
Redundancy Techniques to the Reliable Design of Digital Computers	ARINC
Computer Programmed Diode Reliability	Hughes
Logic Circuit Evaluation	Hollander Associates
Evaluation Criteria for Associative Memories	
Criteria for Systems Trad-Offs	Sylvania & ARINC
Cost and Availability Program	ARINC
General Purpose System Simulator	IBM
General Operating System Simulation Program	GE - Tempo
Simple Digital Device Simulator	AMF
(BLODI) Block Diagram Compiler (B and C)	Bell Laboratories
(PATSI) Block Diagram Compiler	Lincoln Laboratories
Mathematical Automated Reliability and Safety Evaluation Program	Mathematica
System Reliability Prediction by Function	ARINC
System Reliability Prediction by Function	Federal Electric
Prediction of Circuit Drift Malfunctions of Satellite Systems	IBM
Mathematical Simulation for Reliability Prediction	Sylvania
General Effectiveness Methodology	Computer Applications
TOPIC: Design of Logic Circuits using Monte Carlo Methods	Washington University
FACTOR: Finds roots of polynomials with real coeff's	Univ. of Michigan
TCHDEL: finds roots of all-pole transfer functions	Univ. of Michigan
RATTCH: finds zeros of chebyshev rational functions	Univ. of Michigan
LAPLAC: calculates step and impulse re sponses of a specified rational function having simple poles	Univ. of Michigan

REFERENCES

1. Copeland, J. A., "Effect of External Circuitry on Gunn Diodes", Electron Device Research Conf., June 29 - July 1, 1966.

2. Kopp, C., "Review, Application and Evaluation of Computer Methods for Circuit Reliability Analysis", Contract N164-10516, U. S. Naval Ammunition Depot, Crane, Indiana.

3. Roach, J. M., "Survey of Studies and Computer Programming Efforts", Report DEM 1/AD 622676 Office of the Director, Research and Engineering Department of Defense, Washington, D.C.

4. "Quarterly Abstracts", NASA/MSFC/Code MS-T, Huntsville, Alabama 35812.

5. "A Listing of Computer Programs Available from NASA COSMIC", Computer Center, University of Georgia, Athens, Georgia 30601, August 1966.

6. Kuo, F. F., "Network Analysis by Digital Computer", Proc. IEEE, Vol. 54, No. 6 pp 820-829, June 1966, also F. F. Kuo and J. F. Kaiser (editors), "System Analysis by Digital Computer", Wiley 1966.

7. Yang, T., "A Bibliography on Computer-Aided Circuit" Analysis and Design, Villanova University NASA Grant NGR-39-023-004, May 1966.

8. Private Communication, Dr. M. Hartley, Editor, International Journal of Electrical Engineering Education, Manchester College of Science and Technology, Manchester, England.

9. Carpenter, R. M., and Happ, W. W., "Algorithms for the Dichotomous Evaluation of Macrocircuits", NASA/ERC-CQ Report # 66-634.

10. Happ, W. W., "Flowgraph Techniques for Closed Systems", IEEE Trans. AES No. 3, pp 252-264, May 1966.

11. Happ, W. W., "Flowgraphs as a Teaching Aid", IEEE Trans. on Education, Vol. E-9, No. 2, pp 69-79, June, 1966.

12. Happ, W. W., Robbins, R. E., and Moody, D. E., "Modeling Piece-Wise Linear Systems by Topological Techniques, IEEE Trans. AES, June 1966.

13. Hartman, W. F., and Carpenter, R. M., "Desensitizing of Microcircuits to Variations in Temperature and Production Spread, Phase II - Computer Aided Sensitivity Analysis, "NASA/ERC-CQ Report # 66-650.

14. Rombalski, W. F., and Carpenter, R. M., "Desensitizing of Microcircuits to Variations in Temperature and Production Spread Phase IV: Tolerance in Performance Criteria due to Production Spread of Parameters", NASA/ERC-CQ Report # 66-651.

15. Hasdorff, L., and Grosky, W. I., "Numerical Computation of the Time Responses of a System from its Transfer Function", NASA/ERC-CQ Report # 66-659.

16. Hasdorff, L., and Grosky, W. I., "A Computer Oriented Two-Parameter Stability Analysis for Electronic Circuits", NASA/ERC-CQ Report # 66-658.

17. Happ, W. W., and Staudhammer, J., "Topological Techniques for the Derivation of Models with Pre-assigned Accuracy", Archiv der elektrischen Ubertragung, vol. 6, 329-335, A.E.U. 20, 1966.

18. Hasdorff, L., and Harrington, E. V., "A Computer-Oriented Approach to Coding, Evaluating and Optimizing for Acoustic Systems", NASA/ERC-CQ Report # 66-652.

19. Hasdorff, L., and Harrington, E. V., "Computer-Oriented Approach for Coding, Evaluating and Optimizing Electrical-Mechanical-Hydraulic Systems" NASA/ERC-CQ Report # 66-657.

20. Desoer, C. A., and Katzenelson, J., "Nonlinear RLC Networks" BSTJ, Vol. 44, p 161, January 1965. See also Trans. IEEE on Circuit Theory, September 1966.

21. Staudhammer, J., "Network Analysis Programs for the GE-225", Paper presented at the GET Users Meeting, April 1965; Research Report, Arizona State University, Tempe, Arizona.

22. Wilde, D. J., "Optimum Seeking Methods", Prentice-Hall, 1964.

23. Kuo, F. F., Private Communication.

24. Branin, F H., Jr., "DC and Transient Analysis of Networks Using a Digital Computer", IRE National Convention, Record, Part 2, p 236, 1962.

25. 1620 Electronic Circuit Analysis Program (ECAP), (1620 - EE-02x) User's Manual, IBM Publication H20-0170-1, IBM Corporation, White Plains, N.Y.

26. Malmberg, A. F., Cornwell, F. L., and Hofer, F. N., "NET-1 Network Analysis Program", Report LA-3119, Los Alamos Scientific Laboratory, August 1964.

27. Ralston, A., "A First Course in Numerical Analysis", McGraw-Hill, 1965.

28. "Automated Digital Computer Program for Determining Responses of Electronic Systems to Transient Nuclear Radiation", IBM Corp., Owego, N.Y., Report 54-521-5, July 1964.

APPENDIX A

LIST OF PARTICIPANTS

BAUGH, RICHARD
Hewlett Packard
Palo Alto, California

COMSTOCK, JAMES
Friden, Inc.
San Leandro, California

DAVIDSON, RAYMOND A.
Chairman, Electrical & Electronics Engineering Department
San Fernando Valley State College
Northridge, California

DAWSON, DARROW F.
Lawrence Radiation Laboratory
Livermore, California

DEBOO, GORDON
NASA - Ames Research Center
Moffett Field, California

DECKERT, KENNETH
IBM Corporation
San Jose, California

DEROSE, CLIFFORD
Philco Corporation
Palo Alto, California

DICKIE, RONALD E.
Ryan Aeronautical Company
San Diego, California

EDENHOFER, J. R.
Autonetics
Anaheim, California

FAZARINC, ZVONKO A.
Hewlett Packard
Palo Alto, California

GRAHAM, J. F.
Collins Radio Company
Newport Beach, California

GRANT, TERRY L.
NASA - Ames Research Center
Moffett Field, California

HAINES, GEORGE W.
Sprague Electric Company
North Adams, Massachusetts

HANSEN, H. W.
Sylvania Electronic Systems
Mt. View, California

HARLEY, THEODORE P.
IBM Corporation
San Jose, California

HARRISON, DEAN
NASA - Ames Research Center
Moffett Field, California

HARTLEY, ROBERT L.
Integrated Circuit Engineering Corporation
Phoenix, Arizona

HESTER, H. G.
IBM Corporation
San Jose, California

HILL, CHARLES
Hewlett Packard
Palo Alto, California

HODSON, D.
Sylvania Electronic Systems
Mt. View, California

HOWARD, BILL
University of California
Berkeley, California

HULL, R.
Sylvania Electronic Systems
Mt. View, California

HYNES, DAVID
Sprague Electric Company
North Adams, Massachusetts

IRVING, DONALD W.
FMC Corporation
Santa Clara, California

IVY, R. M.

JOHNSON, A. K.
Lockheed Missiles & Space Co.
Sunnyvale, California

JOHNSON, E. T.
IBM Corporation
San Jose, California

JOHNSON, R. A.
Collins Radio Company
Newport Beach, California

JOHNSON, ROGER M.
Lockheed Missiles & Space Co.
Sunnyvale, California

JONES, WILLIAM L.
Electrical Engineering Department
Utah State University
Logan, Utah

KAUFMAN, MURLAN
Tektronix
Beaverton, Oregon

KAWABATA, FRED
Tektronix
Beaverton, Oregon

LANSDON, DAN
Hewlett Packard
Palo Alto, California

LAPATRA, JACK
Electrical Engineering Department
University of California
Davis, California

LEE, JOHN
General Electric Company
San Jose, California

LEE, RICHARD
Hewlett Packard
Palo Alto, California

LEE, ROBERT
NASA - Ames Research Center
Moffett Field, California

LINKWITZ, SIEGFRIED
Hewlett Packard
Palo Alto, California

LIU, CHUNG
IBM Corporation
San Jose, California

MAGNUSON, WALDO G.
Lawrence Radiation Laboratory
Livermore, California

MARTIN, LEROY C.
Tucson, Arizona

MCKENZIE, D. C.
Signetics Corporation
Sunnyvale, California

MILES, R. S.
Autonetics
Anaheim, California

MILLER, R. A.
Sylvania Electronic Systems
Mt. View, California

MITCHELL, DONALD
Fairchild Semiconductor
Mt. View, California

MUNOZ, ROBERT M.
NASA - Ames Research Center
Moffett Field, California

NEILSON, E. D.
Signetics Corporation
Sunnyvale, California

O'BRIEN, TOM
Hewlett Packard
Palo Alto, California

ONG, A. M.
U. S. Naval Electronics Laboratories
San Diego, California

PADDOCK, JOHN
NASA - Ames Research Center
Moffett Field, California

POTTER, RON
Hewlett Packard
Palo Alto, California

PRUCHA, M. JOHN
NASA - Ames Research Center
Moffett Field, California

QUEEN, GREG
Fairchild Semiconductor
Mountain View, California

REED, JOHN
Autonetics
Anaheim, California

ROE, JOHN
Electrical Engineering Department
University of Missouri
Columbia, Missouri

ROSE, ED
Advanced Technology Division
American Standard
Mountain View, California

ROTH, PETER
Hewlett Packard
Palo Alto, California

SHOU, SHAO-HAN
Union Carbide
Mountain View, California

SIEVERS, WILLIAM
Fairchild Semiconductor
Mountain View, California

SLOATE, H.
General Electric Company
Syracuse, New York

STEMPIN, CARL
Lockheed Missiles & Space Co.
Sunnyvale, California

STEWART, J. A.
Lenkurt Electric Co.
San Carlos, California

STOFT, PAUL E.
Hewlett Packard
Palo Alto, California

SYN, WAI MUN
IBM Corporation
San Jose, California

TANIKAWA, R. K.
Autonetics
Anaheim, California

THAYER, LOUIS
Friden, Inc.
San Leandro, California

THOMAS, LEE C.
Bell Telephone Laboratories
Holmdel, New Jersey

TURNER, D. P.
IBM Corporation
San Jose, California

VAN SANDWYK, J.
Sylvania Electronic System
Mountain View, California

VASQUEZ, RICHARD
Hughes Aircraft Co.
Culver City, California

VERDUYN, FRITZ M.
IBM Corporation
San Jose, California

WEAR, LARRY L.
Hewlett Packard
Datamec Division
Mountain View, California

WESTIN, PAUL E.
Dalmo Victor Company
Belmont, California

WILLIFORD, J. G.
Collins Radio Company
Newport Beach, California

WISEMAN, NEIL
Friden, Inc.
San Leandro, California

YEN, CHU-SUN
Hewlett Packard
Palo Alto, California