# THEORY AND APPLICATIONS OF FLOW GRAPHS

## CHARLES S. LORENS

## TECHNICAL REPORT 317

### JULY 15, 1956

## RESEARCH LABORATORY OF ELECTRONICS
### MASSACHUSETTS INSTITUTE OF TECHNOLOGY
### CAMBRIDGE, MASSACHUSETTS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 317                                      July 15, 1956

THEORY AND APPLICATIONS OF FLOW GRAPHS

Charles S. Lorens

This report is based on a thesis submitted to the Department of
Electrical Engineering, M.I.T., July 1956, in partial fulfill-
ment of the requirements for the degree of Doctor of Science.

Abstract

Our object is threefold: to develop a basic theory of flow graphs; to extend their
known application to electrical networks; and to formulate the natural relation between
flow graphs and the properties of discrete statistical systems.

Part I deals with the first two objectives. Much of the linear theory of flow graphs
parallels linear equation theory in such a manner that considerable meaning is given
to many classical operations of formulation, inversion, and reduction. Part I covers
additive functional systems; inversion; electrical network formulation; mathematical
determinants; partitions and expansions of a determinant; rank, positive definite,
eigenvalues, and eigenvectors; group multiplication; and transcendental equations.

Part II deals with the third objective. The pictorial representation of a discrete
statistical system specifies its properties. With flow graphs these properties are
conveniently calculated and interpreted directly from the representative graph. The
material of Part II covers recurrent events, transient systems, stationary state distri-
butions, sensitivity and variation, generating functions, channel capacity, and informa-
tion structure.

# Table of Contents

# I. THEORY OF FLOW GRAPHS WITH APPLICATIONS TO ELECTRICAL NETWORKS

## 1. INTRODUCTION

When a man goes out to have a house built he likes to see the drawings and sketches of the house before he puts too much money into the project. When he buys a new suit of clothes he likes to see how they look on himself before he buys them. It is the same way with an engineer. He likes to see how his ideas "look" before he commits himself to spending a great deal of time on them. In putting systems together, he likes to lay them out to get a notion of the over-all operation of the system and to see how much equipment will be involved. In this way he eliminates many mistakes at the beginning and obtains an idea of how to go about realizing his system. Having the system laid out in functional form, he is able to work on individual components and, if need be, fit the components together from time to time to check on the combined operation. This procedure allows him to modify his system as he learns more about it and the equipment necessary to make it operate.

1.1 Mathematical Structure and Intuition. In the past few years a new development in mathematics has allowed the applied mathematician to set up problems on a functional component basis and then work on the components individually, much as the engineer works on his components. The development has been pushed so far that the solutions are being obtained directly from the functional representation of the problem. The importance of this mathematical development is that there is a structure associated with the mathematical statements. This mathematical structure corresponds very closely to the original physical problem.

In visualizing a problem we have an intuitive idea of what the problem looks like. The method of setting down the problem should include this intuition. Having set the problem down along with our intuitive ideas, we should be able to solve the problem directly from this semi-intuitive representation. Just as the engineer builds components and tests them with others, so should the applied mathematician be able to obtain component solutions and fit them together to get the whole solution. With this procedure the mathematician gains considerable insight on the solution mechanism and its requirements.

1.2 Missing the Point. Some will strongly argue that algebraic representation and manipulation are sufficient for obtaining solutions, but I fear that they have missed the point. The point is this: an algebraic equation carries very little intuition about the physical problem. With algebraic manipulation we do not get a full insight of the solution mechanism and how the solution comes about.

1.3 Background. S. J. Mason (8, 9) recognized the importance of this structural representation and did a considerable amount of development work on it. He called the work "flow graphs," partly because it represents the flow of the solution process.

Much of his work was directed at the solution of linear electronic problems. In this field of mathematics, it was soon found that rather complicated problems could be put down by inspection. Through standard reduction methods the solution could be obtained directly from the formulation without the necessity of algebraic equations. Subsequent work (10) shows that it is possible to put the solution down directly from the formulation, thus eliminating the reduction procedure.

Previous to Mason's work, C. E. Shannon (11) worked out a number of the properties of what are now known as flow graphs. Unfortunately, the paper originally had a restricted classification and very few people had access to the material. A. Tustin (12) applied flow graphs to dc machinery; and P. C. Watson (14) originated the application of flow graphs to matrix equations.

More recent work (1-7) has been devoted to the application of flow graphs to electrical network and statistical systems.

1.4 More Elegant Mathematical Methods. This past development has given the outsider the feeling that flow graphs represent a "trick solution" that gives "quick results" that could have been obtained by other more elegant mathematical methods.

It is very true that if I can solve a problem by flow graphs, you can obtain the solution through mathematical manipulation. But it is also true that I will have a better insight into what the problem involves and how to modify it in order to obtain better and more desirable results.

Flow graphs are a precise mathematical tool representing a method of mathematical manipulation that parallels intuitive ideas about the mathematics.

1.5 Motivation. This discussion leads up to asking about the basic relation between algebra and flow graphs and, subsequently, the mathematical manipulation. The useful things that can be done in algebra must be useful in flow graphs. The algebraic formulations must be capable of being formulated with flow graphs. Specific manipulations must correspond to specific manipulations of the flow graphs and, thus, to manipulation of the physical problem.

The basic aim of Part I is to formulate flow graphs on a logical exact mathematical basis. Because of the ease with which formulations and solutions of linear problems are obtained, much of the initial material will be restricted to linear equations. Linear network analysis provides excellent examples for demonstrating the various properties of linear flow graphs.

1.6 Summary of Part I. For the benefit of those unfamiliar with flow-graph manipulation, Part I begins with a brief summary of the more important rules. A few examples

give you a chance to sharpen your skill and become acquainted firsthand with the objectives of flow-graph formulation and reduction.

In Section 3 the basic ideas of setting up and manipulating a general additive flow graph are introduced. Formulation is a process of making mathematical statements. Reduction is a process of substituting an equivalent statement for a group of statements.

In Section 4, the manipulation of flow graphs is extended by three methods of inversion – a means of inverting the dependencies of the variables. The meaning of inversion in standard mathematical terminology has never been clearly understood. However, the use of flow graphs puts quite a bit of meaning into the word. Two general methods of inversion that apply to additive functional relations have been formulated. One method is particularly applicable to changing the direction of dependency without appreciably changing the structure of the graph. The third method is applicable to the general theory of linear equations.

Formulation of flow graphs for linear electrical networks provides the topic of Sections 5 and 6. A number of interesting approaches have been evolved which facilitate rapid formulation so that the flow graph retains many of the intuitive properties of the network. Most electrical circuits are built in cascade for design considerations. Section 5 develops a method of cascading the representative flow graphs in the same manner as that in which the electrical circuit elements are cascaded. Section 6 deals with conventional network analysis formulated in terms of flow graphs.

One of the most important aspects of these two formulations for electrical networks is that customary circuit approximations can be made after the formulation instead of before.

Flow graphs represent a mathematical tool that displays the interdependency of the system of variables. The display allows us to use our intuition in the production of various properties of the system.

The general theory and methods of flow graphs are closely tied in with those of standard mathematical manipulations. With flow graphs some of the standard methods become simpler to understand and to apply.

Determinants are a good example of the simplification that flow graphs afford. In Section 7 a firm connection is established between the standard mathematical determinant, $D$, and the flow-graph determinant, $\Delta$, formulated by S. J. Mason (10),

$$\Delta = (-1)^n D \tag{1}$$

where $n$ is the number of dependent variables.

Cramer's rule for the solution of a set of linear equations as the ratio of two determinants follows with slight modification.

On the basis of further work with determinants in Section 7 it is easy to establish the ideas of partitions and expansions of a determinant. These ideas are particularly important because they readily allow the application of approximations to the solution of the flow graphs.

The partitioning of the graph also brings out a method of multiplying a group of variables by a constant or multiplying a number of branch coefficients by a constant without disturbing the external characteristics of the graph and equivalent dependencies. This modification provides a means of reduction without changing the structure of the graph. It also provides a means of changing an electrical network without changing its external characteristics. At the end of the section it is shown for a graph in which all branches have arbitrary values that the maximum possible number of branches that can be made unity without changing the graph's external properties is equal to the number of nodes in the graph, excluding those that are strictly independent or dependent.

A number of special properties of determinants, such as rank, positive definite, eigenvalues, and eigenvectors, are included in Section 8.

S. J. Mason worked out a method of solving matrix equations. This material is included in Section 9. The method is applicable for both right-hand and left-hand systems.

The concluding section takes up the properties of the group operation of multiplication and shows how it is applied to flow graphs involving transcendental functions.

## 2. FLOW GRAPHS

This section gives an introduction to the mechanics of linear flow graphs. The ideas of formulation and the rules for reduction are quite simple and can be easily understood by anyone who has a knowledge of algebra. Possible questions as to why
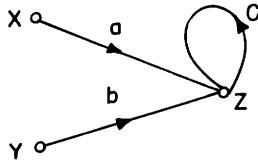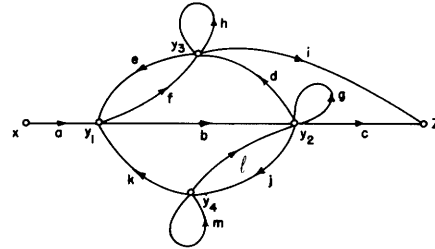


Fig. 1. Flow-graph equation.



Fig. 2. Set of flow-graph equations.

certain things are done will be answered in the sections that follow. Some of the end results are introduced here in order to give perspective to the general theory as presented in Section I.

2.1 **Drawing Out Equations.** A flow graph represents a system of equations. The graph is set up on the idea that a particular variable is equal to a number of other variables multiplied by different constants. Of course, the particular variable can also be multiplied by a constant and added in with the other variables. The equation

$$z = ax + by + cz$$

has the flow-graph representation of Fig. 1. Thus, the nodes represent the variables and the branches represent the multiplication operation on these variables. Converging branches represent a summation.

The system of equations

$$y_1 = ax + ey_3 + ky_4$$

$$y_2 = by_1 + gy_2 + \ell y_4$$

$$y_3 = fy_1 + dy_2 + hy_3$$

$$y_4 = jy_2 + my_4$$

$$z = cy_2 + iy_3$$

is represented by the flow graph of Fig. 2.

5

2.2 Equivalent Dependencies. Flow graphs provide a means of solving a set of equations without becoming entangled in algebraic manipulation. By solving a set of equations you are essentially eliminating unnecessary variables and finding equivalent dependencies. In Fig. 2 you may be interested in eliminating the variables $y_1$, $y_2$, $y_3$, and $y_4$ to find the equivalent dependency between x and z. One of the variables must be independent so that the other variable can be dependent upon it. In the example, x is independent, since all of its branches point away from the variable.

2.3 Loops and Paths. Equivalent dependencies in flow-graph notation are found by the loop rule. Here a loop is a closed sequence of variables, no variable appearing more than once. With each loop there is associated a loop product (sometimes also called the loop) equal to the product of the coefficients of the branches going in the direction of the loop. The loops and loop products of Fig. 2 are

$y_1 \, y_2 \, y_3$ : bde $\qquad$ $y_2$ : g

$y_1 \, y_2 \, y_4$ : bjk $\qquad$ $y_3$ : h

$y_1 \, y_3 \qquad$ : fe $\qquad$ $y_4$ : m

$y_2 \, y_4 \qquad$ : $j\ell$

The loops on the single variables are called self-loops. Disjoint loops are loops having disjoint sequences of variables as the loop $y_1 y_3$ and the loop $y_2 y_4$ are disjoint.

In like manner a path is a sequence of variables but it does not close on itself. There are three paths from x to z in Fig. 2. These paths and their path products are

$x \, y_1 \, y_2 \, z$ : abc

$x \, y_1 \, y_3 \, z$ : afi

$x \, y_1 \, y_2 \, y_3 \, z$ : abdi

2.4 Loop Rule. The loop rule states that the dependency is equal to the ratio of the sum of all paths between the variables multiplied by their cofactor and all divided by the determinant of the system, Eq. 2.

$$T_{jk} = \frac{\Sigma_r \, L_{jk_r} \cdot \Delta_{jk_r}}{\Delta} \qquad (2)$$

The cofactor of a path is the determinant of the system of variables not involved in the path.

The determinant of a system of variables is calculated on the basis of the system's loops as follows:

$$\Delta = 1 \cdot S_1 - \Sigma_r \, L_r^1 \cdot S_r + \Sigma_r \, L_r^2 \cdot S_r - \Sigma_r \, L_r^3 \cdot S_r + \ldots \qquad (3)$$

where $L_r^n$ is the loop product of n disjoint loops of two or more variables, and $S_r$ is the product of one minus the self-loops disjoint from the loops in $L_r^n$. For example, the determinant of the system in Fig. 2 is calculated in the following manner:

1. the product of one minus the self-loops

$$1 \cdot S_1 = (1-g)(1-h)(1-m)$$

2. the sum of the loops taken one at a time, each multiplied by any disjoint self-loops

$$-\Sigma_r \, L_r^1 \cdot S_r = -\{bde(1-m) + bjk(1-h) + ef(1-g)(1-m) + \ell j(1-h)\}$$

3. the sum of the product of the disjoint loops taken two at a time, each multiplied by any disjoint self-loops

$$+\Sigma_r \, L_r^2 \cdot S_r = ef\ell j$$

There are no combinations of three or more disjoint loops of two or more variables. Thus the determinant of the system is

$$\Delta = (1-g)(1-h)(1-m) - \{bde(1-m) + bjk(1-h) + ef(1-g)(1-m) + \ell j(1-h)\} + ef\ell j$$

The cofactor of the path $xy_1y_2z$ is computed on the basis of the disjoint variables $y_3$ and $y_4$ to be

$$\Delta_{abc} = (1-m)(1-h)$$

The cofactor of the other path, $xy_1y_3z$, is based on the variables $y_2$ and $y_4$.

$$\Delta_{afi} = (1-m)(1-g) - \ell j$$

Thus the dependency of z on x is

$$T_{xz} = \frac{abc(1-m)(1-h) + afi\{(1-m)(1-g) - \ell j\} + abdi(1-m)}{\Delta}$$

This dependence is the statement that

$$x \cdot T_{xz} = z$$

Of course, once you know the rule, this answer could be written down directly by inspection of the flow graph. By solving the original set of equations you will find that it is the exact solution. The following sections will show that the exact solution will always be obtained by this method. Other methods lead to approximate solutions.

From the flow graph of Fig. 2 the following calculations are made upon inspection of the loops and paths through the graph:

7

$$T_{xy_1} = \frac{a\{(1-h)(1-g)(1-m) - \ell j(1-h)\}}{\Delta}$$

$$T_{xy_2} = \frac{ab(1-h)(1-m)}{\Delta}$$

$$T_{xy_3} = \frac{af\{(1-g)(1-m) - \ell j\} + abd(1-m)}{\Delta}$$

$$T_{xy_4} = \frac{abj(1-h)}{\Delta}$$

The determinant is always the same for the same flow graph.

There is much more to flow graphs than solving equations. The interesting part is that they are quite useful in analyzing electrical networks and statistical systems.

This section has shown some of the mechanics of formulation and reduction. In the following sections details of the mechanics and their application to electrical networks are presented.

8

# 3. GENERAL ADDITIVE FORMULATION AND REDUCTION

In the formulation of problems, we are faced with the need to make statements. We can say that something is true, not true, or more or less true than something else.

At present, flow graphs are formulated only on positive statements, represented by the equality sign in algebra.

3.1 Positive Statements. Positive statements are represented by nodes (small round circles). The statement, "The quantity of electric current is x," is represented by a particular node with an x beside it. In this way a system of variables is set up by setting up a system of nodes.
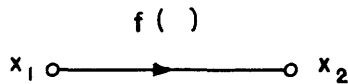


Fig. 3. Functional relation.

The functional operation between two variables is represented by a line (or branch) between the two nodes with an arrowhead pointing in the direction of the functional dependence. The function corresponding to the branch is placed directly on the branch. For example, the variables $x_1$ and $x_2$ related by the function

$$f(x_1) = x_2$$

are represented by the two nodes $x_1$ and $x_2$ with a branch between them in the direction of $x_2$, Fig. 3. The nose of the branch refers to the dependent end $x_2$; the tail of the branch refers to the independent end $x_1$. Multiplication by the factor "a" is indicated by placing an "a" on the branch.

The group operation of addition is represented by the convergence of branches from independent variables onto a single node, the dependent variable. Each of the branches is some function of the independent variable. For example, Fig. 4 represents the equation

$$h(y) + f(x_1) + ax_2 - x_3 = y$$

The divergence of branches is not given any mathematical meaning other than the dependency of the variable. Thus, where the group operation is limited to addition it becomes possible to represent sets of equations by graphical means. The set of equations

$$y_1 = a(x) + e(y_3)$$

$$y_2 = b(y_1) + f(y_2) + g(y_3)$$

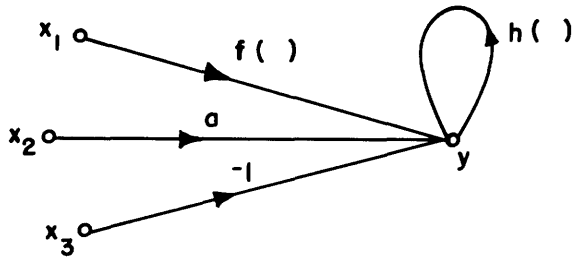$$y_3 = d(y_2)$$

$$z = c(y_2)$$
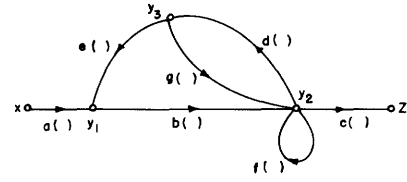
9

Fig. 4. Additive equation.
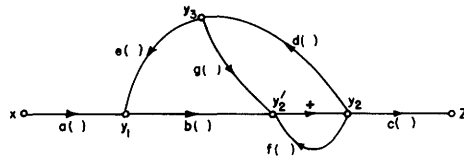


Fig. 5. Flow-graph representation.



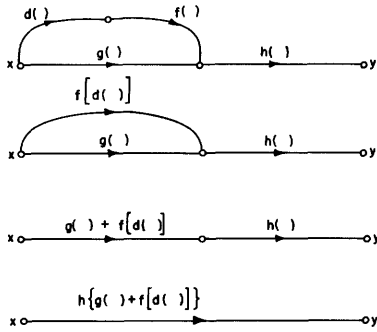Fig. 6. Splitting the node "$y_2$".

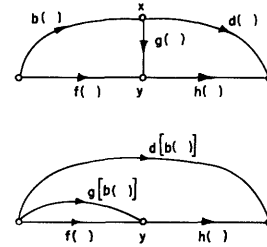

Fig. 7. Cascade and parallel reduction.



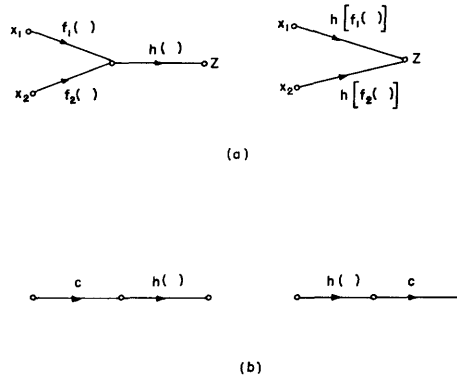Fig. 8. Elimination of the node "x".



(a)



(b)

Fig. 9. Linear conditions.

10

is represented by the flow graph of Fig. 5. A consistent graph is obtained when not more than one positive statement is made about each node.

3.2 Paths and Loops. The paths and loops are important properties of a flow graph. A path is a sequence of nodes in which each node appears no more than once. The path starts at a particular independent node and ends at a dependent node, so that the path has a direction as does the path $xy_1y_2z$ in Fig. 5. With each path is associated the functional dependence in the direction of the path. The important paths are those that have a functional dependence between each node of the path and in the direction of the path. Thus the important paths can usually be observed by inspection by moving through the graph in the direction of the arrows, going by each node no more than once. A loop is a path that starts and ends at the same node. The sequence of nodes $y_1y_2y_3$ of Fig. 5 forms a loop. For loops the starting node is immaterial. Only the direction and the functional dependence is important.

3.3 Node Splitting. Independent nodes are usually thought of as sources, in that the graph and the dependency usually flow from these nodes. In like manner, the dependent nodes are thought of as sinks. However, it is not possible to apply this notation too rigidly because some variables are sinks in the sense that they depend on other variables, and sources in the sense that they are the independent variables in other relations. Splitting up the source and sink aspect of a node is commonly known as node splitting.

Splitting the node $y_2$ of Fig. 5 into source $y_2$ and sink $y_2'$ produces the graph of Fig. 6, where the two variables $y_2$ and $y_2'$ are dependent, by the relation, $y_2' = y_2$. The plus sign is sometimes used to represent unity multiplication.

3.4 Strictly Independent and Dependent Nodes. In the reduction and inversion process it is convenient to have a variable strictly independent or strictly dependent. The term "strictly independent" refers to variables dependent upon no other variable. This situation can be recognized in flow-graph notation when all of the branches are divergent from the variable's node. "Strictly dependent" refers to a variable upon which no other variable is dependent. Such a variable is recognized by the fact that only branches converge on the variable's node. It is always possible to make the node $x$ strictly dependent by defining the new relation, $x = x'$, and placing this relation on the graph. This is essentially "drawing off" a variable in order to get a better look at it.

The process of making a variable embedded in the graph strictly independent is to first make it strictly dependent and then use the process of inversion which is described in a following section.

3.5 Reduction. One of the more important advantages of flow graphs comes from the ability to reduce them to simpler graphs without ever having to revert to the algebraic

equations that they represent. The reduction process can be carried out at least as fast as, and usually considerably faster than, it can be carried out with algebraic equations. Usually in the manipulation of algebraic equations, there are several ways of bringing about the reduction. Likewise, with flow graphs there are several ways. However, with flow graphs it is usually easier to see the possible variations in the reduction.

3.6 Dependency and Equivalent Relations. Reduction hinges entirely on the dependency of variables and equivalent relations. The final objective is usually to eliminate inter-mediate variables, substituting equivalent relations so that the dependency is directly between the desired variables in the desired form and direction.

Based on the representative mathematics, the equivalent branch of a number of parallel branches is the sum of the branch functions. The equivalent branch of an isolated chain of cascade branches is the multiple function of a function. The reduction of the flow graph in Fig. 7 is an example. Each reduction operation can be verified by writing out the equations and performing the operations algebraically. For linear graphs the multiple function is the product of the branch coefficients.

Further elimination is accomplished by substituting equivalent branches for the ones eliminated. The equivalence, of course, is determined from the dependence of the remaining nodes. The only dependence that is affected is the dependence that passes through the node that is to be eliminated.

In Fig. 8, the variable x is eliminated by substituting the dependency through x. Note that in general it would not have been possible to have eliminated the variable y.

The two conditions for a function h to be linear are these:

$$h[f_1(x_1) + f_2(x_2)] = h[f_1(x_1)] + h[f_2(x_2)] \tag{4a}$$

$$h(cx) = ch(x) \tag{4b}$$

These conditions are given in flow-graph notation in Figs. 9a and 9b, where the right and left figures are equivalent.

3.7 Specification of Properties. For the elimination of the most general node, where there is more than one branch coming in and going out, some of the properties of the functions must be specified. It should be recognized that the properties need not be specified completely for each function, and not at all for others. With flow-graph manipulation we have a better idea of the properties that each function should have in order to get a solution by a particular reduction procedure. Two specific types of functions are treated in the following sections: linear functions and transcendental functions.

The important point on reduction is that reduction of a flow graph is carried out as a process of eliminating variables by substituting equivalent relations. The graphical representation allows us to take advantage of the flexibility of the process to attain different forms of the solution.

12

## 4. INVERSION OF DEPENDENCY

Inversion plays an important part in making changes in the dependency to facilitate the reduction or formulation of a graph or to make the final dependency of the graph in the desired direction. This section presents three essentially different methods of inverting the dependency in a flow graph. The first two methods are general; the third is limited to linear graphs.

Inversion changes the independent variable to a dependent variable and vice versa. It is essentially a process of turning around and is so represented in manipulating the flow graph. Inversion has meaning only for two types of paths: a path from a strictly independent variable to a strictly dependent variable, and a path that forms a loop.

In the process of inverting a path, we must be careful to preserve the mathematical exactness of the formulation. At each node along the inversion path certain changes in the branch functions and structure must be made in order to preserve the mathematical exactness. The three following sections present different ways of preserving this exactness. A group of comparative examples is given in section 4.4.

4.1 First Inversion Method. Consider the general node $z$ shown in the flow graph of Fig. 10a, where we are interested in inverting the path from $y_1$ through $z$ to $x_1$. It is assumed that the inverse exists. If it did not, the inversion could not be performed. Figure 10a is the representation of the equations

$$f_1(y_1) + f_2(y_2) + h(z) = z$$

$$g_1(z) = x_1 \tag{5}$$

$$g_2(z) = x_2$$

One method of inverting Eqs. 5 gives Eqs. 6:

$$f_1^{-1}(u) = y_1$$

$$z - h(z) - f_2(z) = u$$

$$g_1^{-1}(x_1) = z \tag{6}$$

$$g_2(z) = x_2$$

This set is represented in Figs. 10b and 10c. Thus the general inversion rule is formulated:

1. Invert the direction of the path, splitting all nodes along the way and inserting the inverse function on the inverted branches.

2. At each split node attach all of the outgoing branches (except for the inverted branch) onto the first node representing the original variable and attach all the incoming
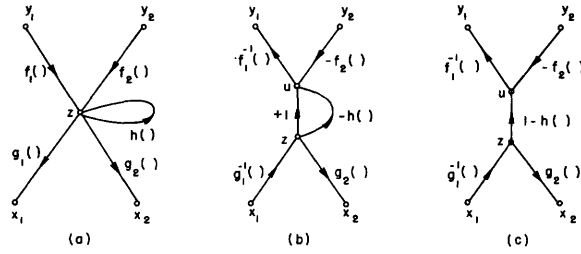
13

Fig. 10. (a) A general node "z". (b) First inversion method.
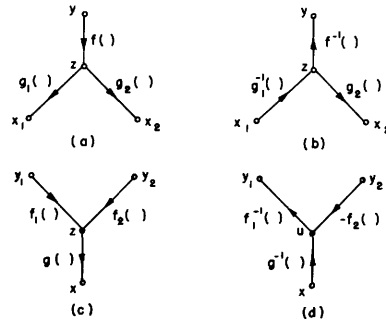(c) First inversion method.



Fig. 11. Degenerate cases of first inversion method. (a) First
degenerate condition; (b) inverse of Fig. 11a; (c) second
degenerate condition; (d) inverse of Fig. 11c.

branches (except for the inverted branch) to the second node, multiplying each function
by -1.

This rule simplifies for the special case of either one branch leaving a node or one
branch entering a node. These degenerate conditions with their inverses are pictured
in Fig. 11. Note that only in the case of one branch entering the node before or after
inversion does the inverted node still remain the former variable.

4.2 Second Inversion Method. The second general method involves another type of
inversion which superficially avoids the inverse functions of the branches.

The method is quite important in linear graphs, where it avoids the introduction of
fractions.

Consider again the set of Eqs. 5 and their graphical representation, Fig. 10a. An
inversion is indicated by Eqs. 7, which are represented in Fig. 12.

$$y_1 - f_1(y_1) - f_2(y_2) + z - h(z) = y_1$$

$$z - g_1(z) + x_1 = z \tag{7}$$
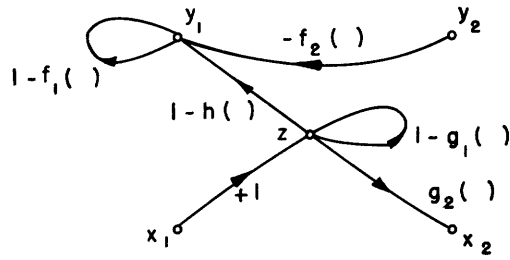
$$g_2(z) = x_2$$
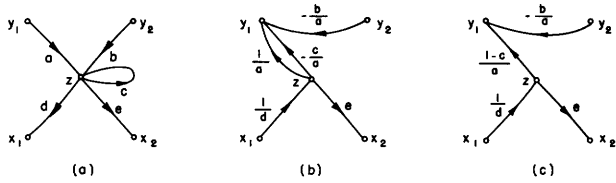
14

Fig. 12. Second inversion method.



Fig. 13. (a) A linear node "z". (b) Third inversion method. (c) Third inversion method.
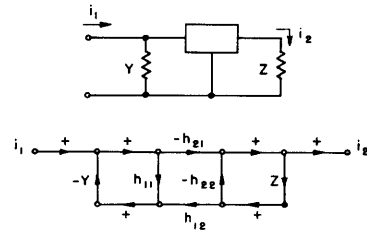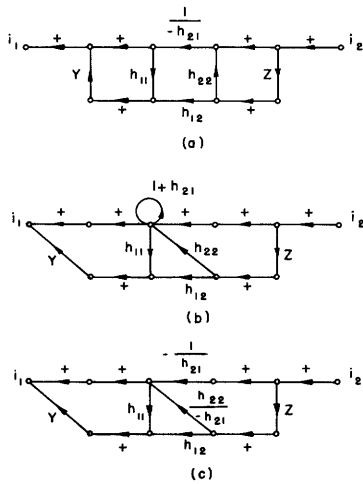


Fig. 14. Loaded transistor and flow graph.



Fig. 15. (a) First inversion method. (b) Second inversion method. (c) Third inversion method.
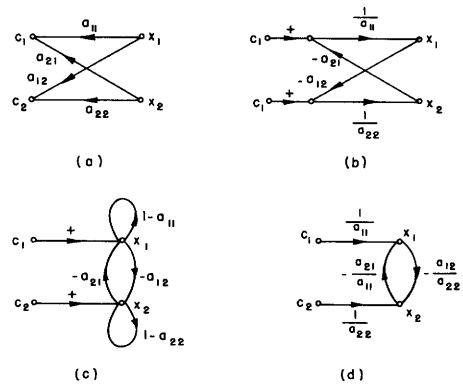


Fig. 16. (a) Second-order set of equations. (b) First inversion method. (c) Second inversion method. (d) Third inversion method.

From this type of inversion we are able to formulate the following general rule:

    1. Take the nose of each branch to be inverted and place it with its tail, changing the branch function to one minus the former branch function.

    2. Take the nose of all incoming branches along with the nose of the inverted branch, changing the sign of the incoming branch functions. The former self-loop function is changed to one minus the former function.

Inversion of a loop of a linear graph by this type of inversion corresponds to an interchange of the columns of the determinant of the representative linear set of equations.

There is another inversion method corresponding to this second method which differs only in an algebraic sign. However, it has the very poor characteristic of unnecessarily introducing a large number of self-loops.

4.3 Third Inversion Method. Restricting ourselves to linear portions of a graph, a third type of inversion is possible. We need the linear property so that superposition will be valid. This method was originally developed by S. J. Mason (9).

Consider the linear graph of Fig. 13a representing Eqs. 8.

$$ay_1 + by_2 + cz = z$$

$$dz = x_1 \tag{8}$$

$$ez = x_2$$

A method of inversion of the path $y_1$ to $x_1$ through $z$ is represented by Eqs. 9, which are represented in Figs. 13b and 13c.

$$\frac{1}{a}z - \frac{c}{a}z - \frac{b}{a}y_2 = y_1$$

$$\frac{1}{d}x_1 = z \tag{9}$$

$$ez = x_2$$

Thus we are led to the following rule for the inversion of linear graphs:

    1. Invert the path changing the branch function to its reciprocal.

    2. Move the nose of each incoming branch along with the nose of the inverted branch, multiplying the branch function by minus the value of the inverted branch function.

4.4 Comparison. As a means of comparison consider inverting the main forward path of the flow graph representing the loaded transistor, Fig. 14. Formulation of this graph is described in Section 5.

The three inversion methods are illustrated in Figs. 15a, 15b, and 15c. Another example of inversion is the inversion of the two main paths $x_1C_1$ and $x_2C_2$ for Fig. 16a, representing a second-order set of linear equations. For comparison, the three

16

methods are shown in Figs. 16b, 16c, and 16d. The first method has the advantage of leaving the form unchanged. The second method seems more complicated than the other two but it has the advantage of not being involved with inverse functions or fractions. The third method retains the original set of variables.

Inversion represents a convenient method of manipulating the dependency of a flow graph. Its two main uses are as an aid in reducing a graph to the desired relations and as an aid in setting up the graph.

## 5. FORMULATION OF CASCADE ELECTRICAL NETWORKS

Two major steps are involved in the working of a problem: formulation and solution. At the start, neither of these steps is completely obvious. With flow graphs, the same steps are necessary.

This section deals with the mathematical formulation in terms of the flow graphs of electrical elements in cascade. We present individual flow graphs for the different elements and then show how these graphs are cascaded in the same order in which the elements are cascaded in the physical circuit. A more general element, which allows the formulation of circuits involving feedback, is then introduced.

This type of formulation is important, since many electrical engineers find that building networks in cascade greatly simplifies the work of designing the individual circuits. The mathematics that represents these cascade systems should also look like a cascade system.

5.1 Two Terminal-Pair Networks. Cascade circuits are usually made by cascading simple two terminal-pair networks. There are two parallel formulations for these networks: one emphasizes the current gain; the other, the voltage gain. Either formulation, or a combination, can be used.

The three-terminal electrical network, Fig. 17a, is conveniently represented mathematically by Eqs. 10a, if we are interested in the current gain of the device, or by Eqs. 10b, if we are interested in the voltage gain of the device. The flow graphs for these two formulations are given in Figs. 17b and 17c.

$$e_1 = h_{11} i_1 + h_{12} e_2$$
$$i_2 = -h_{21} i_1 - h_{22} e_2$$

current gain equations (10a)

$$i_1 = g_{11} e_1 - g_{12} i_2$$
$$e_2 = g_{21} e_1 - g_{11} i_2$$

voltage gain equations (10b)

5.2 Cascading. The two terminal-pair flow-graph elements are cascaded in the same manner in which the physical elements are cascaded. In order to obtain a consistent flow graph, the output of a current-gain element drives the input of another current-gain element. The same is true for voltage-gain elements. The cascading of current-gain elements or voltage-gain elements shown in Fig. 18 is accomplished on the basis of Eqs. 11.

$$i_2 = i_3$$
$$e_2 = e_3$$

(11)

The procedure consists of connecting the flow-graph elements with lateral branches of
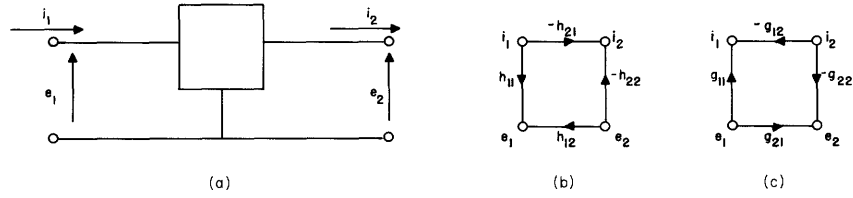
Fig. 17. (a) Two terminal-pair network. (b) Current gain. (c) Voltage gain.
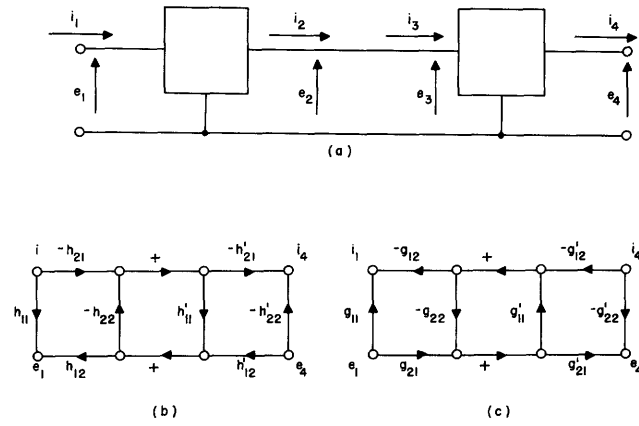


Fig. 18. (a) Cascading networks. (b) Current gain. (c) Voltage gain.
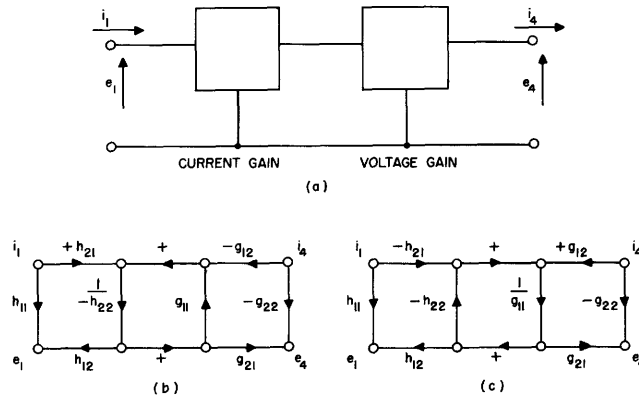


Fig. 19. (a) Mixed cascading. (b) Current inversion. (c) Voltage inversion.

(a)

(b)

(c)

(d)

(e)

(f)

20
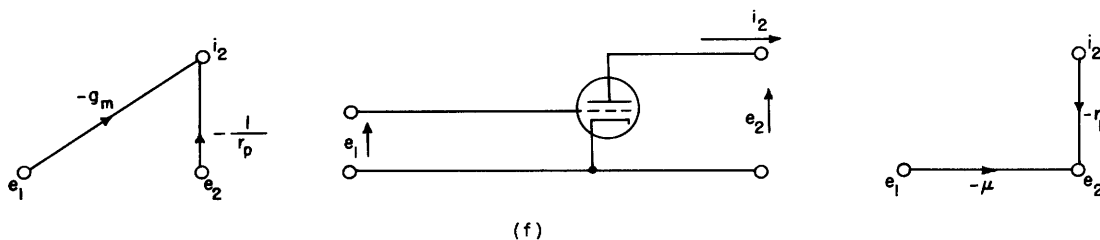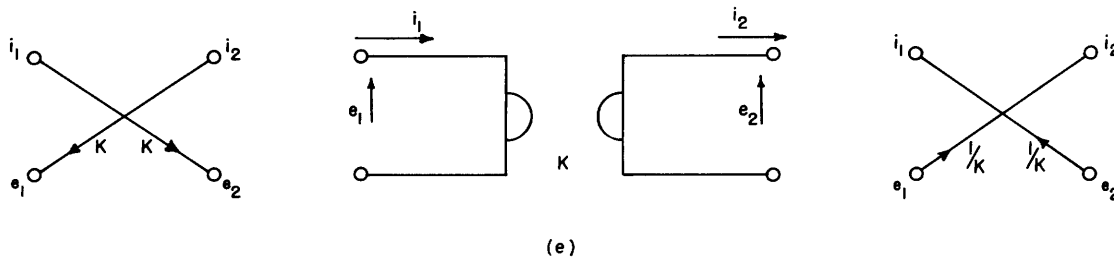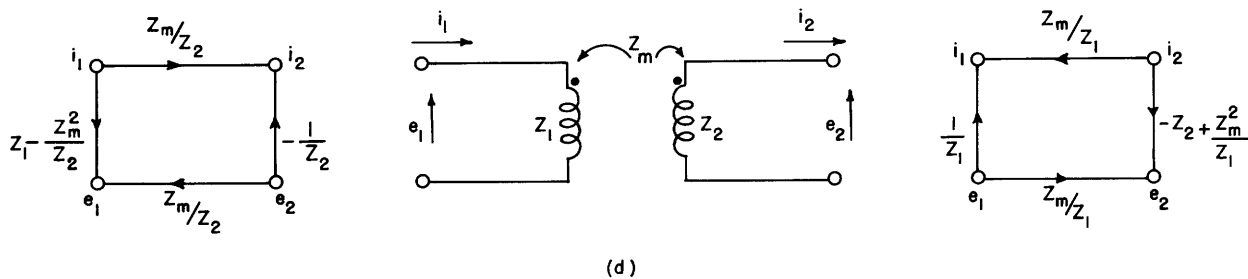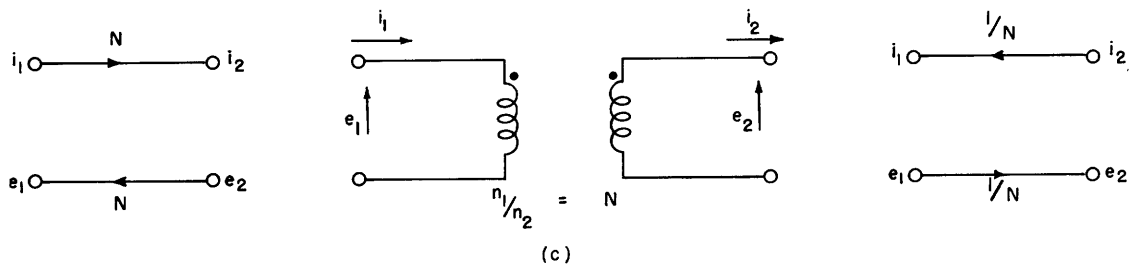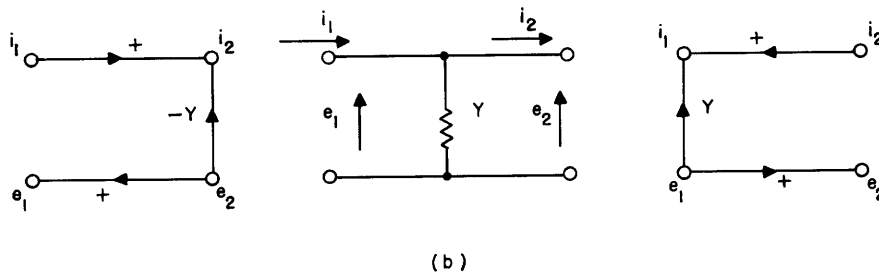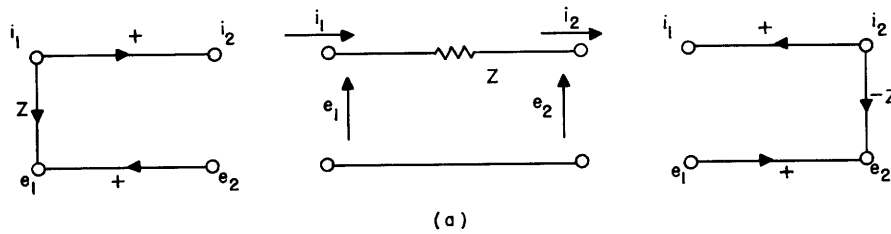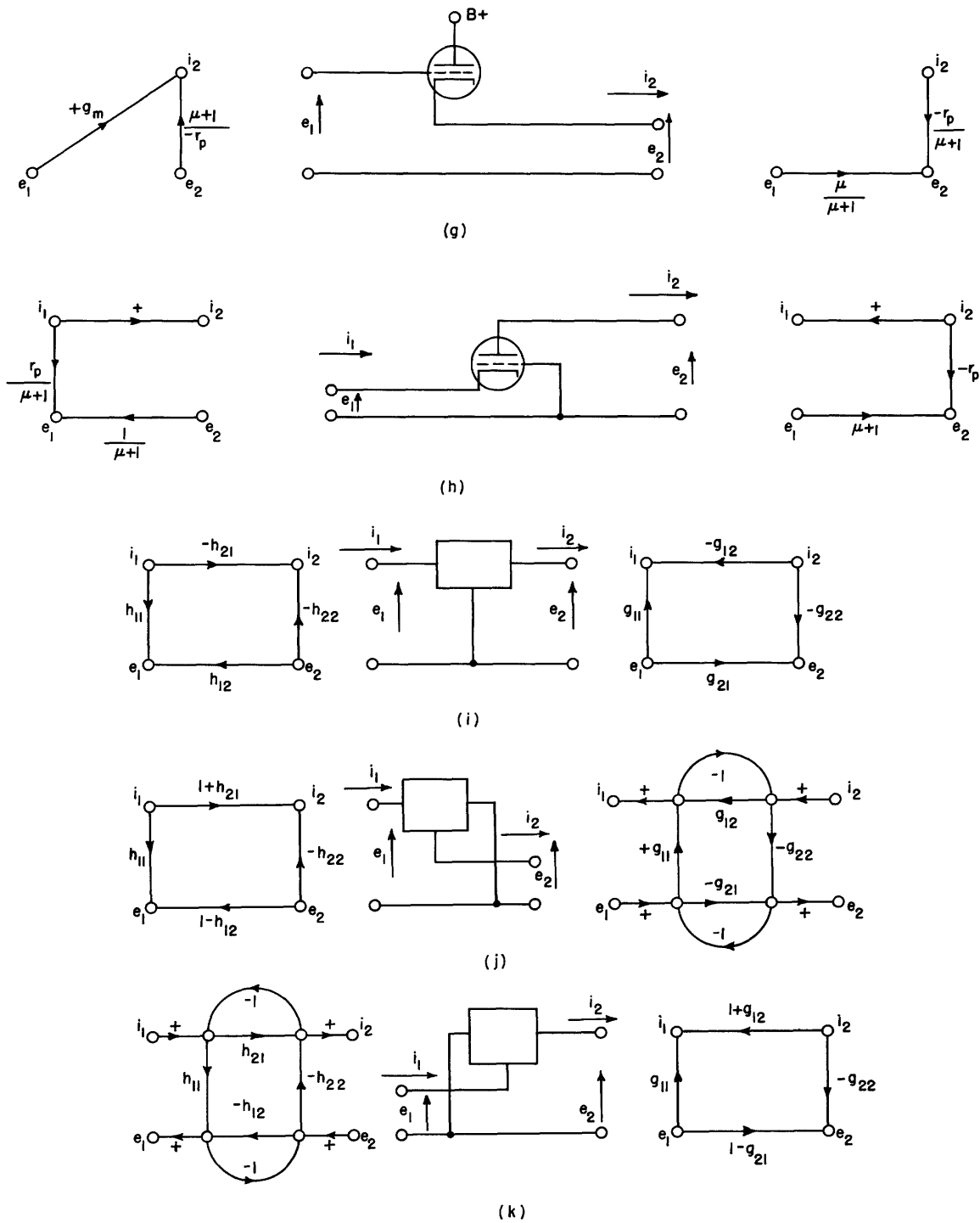
Fig. 20. Elements of cascade networks: (a) series impedance; (b) shunt admittances; (c) ideal transformer; (d) physical transformer; (e) ideal gyrator; (f) grounded-cathode amplifier; (g) cathode follower; (h) grounded-grid amplifier; (i) grounded transistor amplifier; (j) grounded-output transistor amplifier; (k) grounded-input transistor amplifier.

21

+1 that have the same direction as the branches in the elements. This type of cascading then satisfies the consistency requirement of making no more than one positive statement about each variable.

The advantage of this formulation comes from the fact that the flow-graph element has the same characteristics as the electrical element and is cascaded in the same physical relation. For example, the input impedance of the current-gain device is observed by looking into the device from the left side. In like manner, if we look into the representative flow graph from the left, we see the input impedance $h_{11}$. Another important characteristic of the current-gain device is the forward current gain, $h_{21}$. In the flow graph this gain is represented by the forward-pointing arrow between the input and output currents.

By setting up the flow graphs in this manner, all of the currents appear on the upper level, and all of the voltages are on the lower level. This is convenient from a dimensional standpoint because all of the branches that point downward are impedances, and all of the branches that point upward are admittances. Lateral branches are dimensionless, and in bilateral networks the lateral branches are equal.

5.3 Mixed Cascading. In order to have a current-gain device cascade with a voltage-gain device, we have to perform an inversion at the output of the first device or at the input of the second device. There are a number of inversion methods. One method involves inverting the path and its coefficients and changing the sign of any incoming branches. The theory of inversions is presented in the preceding section.

The flow graph of a current-gain device cascaded with a voltage-gain device is shown in Fig. 19b and Fig. 19c. In Fig. 19b, the output branch of the current-gain device is inverted; in Fig. 19c, the input branch of the voltage-gain device is inverted.

5.4 Elementary Components. It is interesting to see what specific electrical elements look like in this flow-graph formulation. Figure 20 shows the more common elements used in pure cascade networks. The current-gain representation is on the left; the voltage-gain representation, on the right. The gyrator is neither a current-gain nor a voltage-gain device but a crossover from one to the other. The grounded-cathode and cathode-follower vacuum-tube amplifiers are inherently voltage-gain devices, so that no current-gain equivalents actually exist. However, they can be represented as voltage-to-current-gain devices, as indicated in the diagrams.

In practice, the voltage-gain model for the transistor is not used. It is included here only for completeness. Both grounded-base and grounded-emitter constants are published for the transistor. The grounded-input transistor provides a method of going from one system to the other.

The most practical method of connecting devices in cascade is to have voltage circuits drive voltage circuits and current circuits drive current circuits. If the system contains both transistors and vacuum tubes the changeover is usually made in the
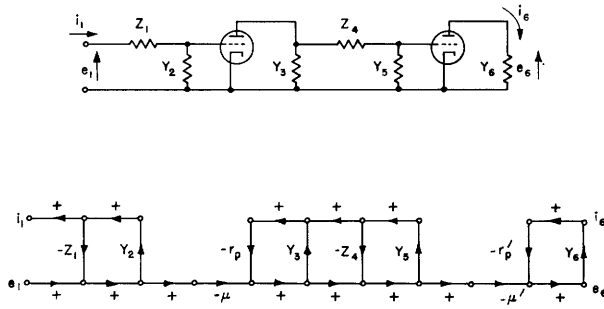
Fig. 21.  First example:  vacuum tubes
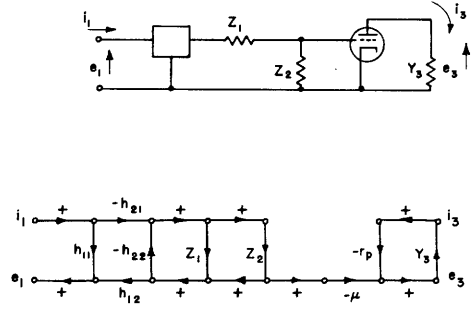in an electrical ladder.



Fig. 22.  Second example:  transistor
driving a vacuum tube.

vacuum-tube unit.  Otherwise an inversion is necessary, as described before.

Because of the consistency requirement, termination of a current system is accomplished with a series impedance; termination of a voltage system, with a shunt admittance.

The importance of this formulation stems from the fact that networks are usually built in cascade.  Being able to set up the mathematical formulation so that it carries along the intuitive feeling of the cascade structure helps us to understand the implications of the mathematics.

5.5 Examples of Simple Cascading.  In order to illustrate these points, a number of examples are included.  Note how the flow graph that represents the physical circuit "strings out" in the same manner as the circuit.

The first example (Fig. 21) shows two vacuum-tube amplifiers in an electrical ladder.  The flow graph is set up on a voltage-gain basis by noting the physical position of the elements in the circuit.  From the flow-graph diagram it is easy to see how isolation is obtained with the use of vacuum tubes.

The second example illustrates the joint use of current- and voltage-gain devices (Fig. 22).  The transistor network essentially terminates at the input to the vacuum tube so that the element $Z_2$ is considered to be in series.

The third example shows the use of a vacuum-tube device driving a current device (Fig. 23).

The equivalent circuit for a transistor in terms of resistances and a current gain is also a cascade circuit, as is indicated in the fourth example, Figs. 24a, 24b, and 24c.  More usable flow graphs for approximations are obtained by inverting the loop containing $r_b$ and $r_c$ in Fig. 24a and the loop containing $r_e$ and $r_c$ in Fig. 24b.

The fifth example demonstrates the use of an inversion (Fig. 25).  The inversion is necessary in order to counteract the effect of the changeover of the gyrator from current to voltage gain.  The input to the shunt-admittance model could not be inverted, since the input impedance is zero.

23

5.6 General Cascade Elements. Many important cascade circuits employ some type
of feedback. The feedback is usually introduced in the third arm or across the input
and output arms of the two terminal-pair network. The feedback may come from another
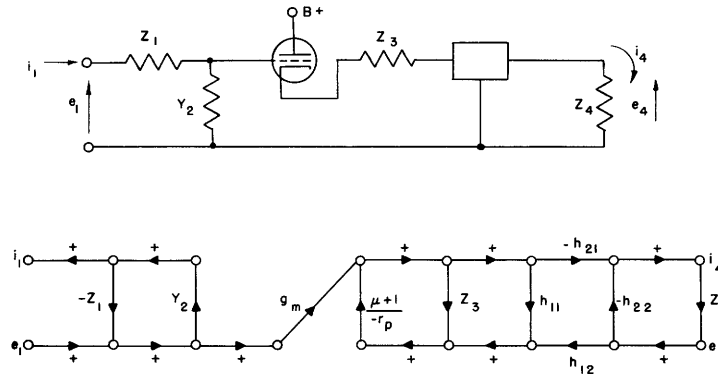


Fig. 23. Third example: cathode follower driving a transistor.

part of the circuit or from the circuit itself, as in the case of the vacuum-tube ampli-
fier with an unbipassed cathode resistor.

Naturally, it takes a more general circuit than the simple circuits shown in Fig. 20
to handle this type of situation. The more general circuits with their corresponding
flow graphs are shown in Fig. 26. Figures 26a and 26b are current-gain representa-
tions. Figures 26c and 26d are voltage-gain representations. Figures 27a and 27b
present the model for the gyrator which is neither a current- nor voltage-gain device
but a crossover from one to the other.

With the set of diagrams in Figs. 26 and 27, it is possible to obtain the different
configurations of the transistor and vacuum tube illustrated in Fig. 20 either directly
or with the use of an inversion. Different methods of inversion are discussed in
Section 4.

5.7 Single-Element Feedback. Where a single element is placed in the base lead or
in shunt with the device over the top, the points of the general flow graph are folded
in and connected by a branch representing the element. Figure 28 is an example. The
use of a single element either in shunt or in the base of the ideal transformer or gyrator
produces a simple flow graph upon reducing the graph to a standard two terminal-pair
flow graph. These elements with their flow graphs are shown in Fig. 29. Current-
gain graphs are on the left; voltage-gain graphs, on the right. In the case of the
gyrator it is interesting to see how the backward coupling can be made zero when
K is positive.

Several types of vacuum-tube amplifiers have relatively simple flow graphs. These
amplifiers, with their flow graphs, are shown in Fig. 30.

24
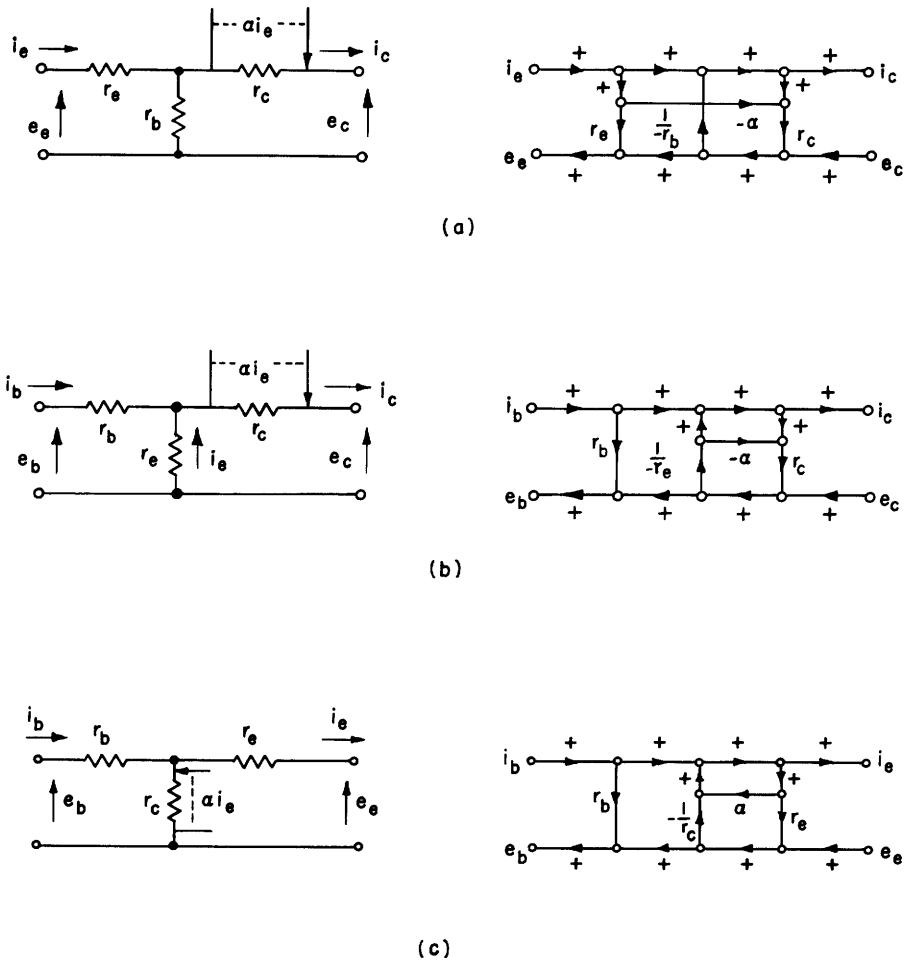
Fig. 24. Fourth example: (a) grounded-base transistor; (b) grounded-emitter transistor; (c) grounded-collector transistor.
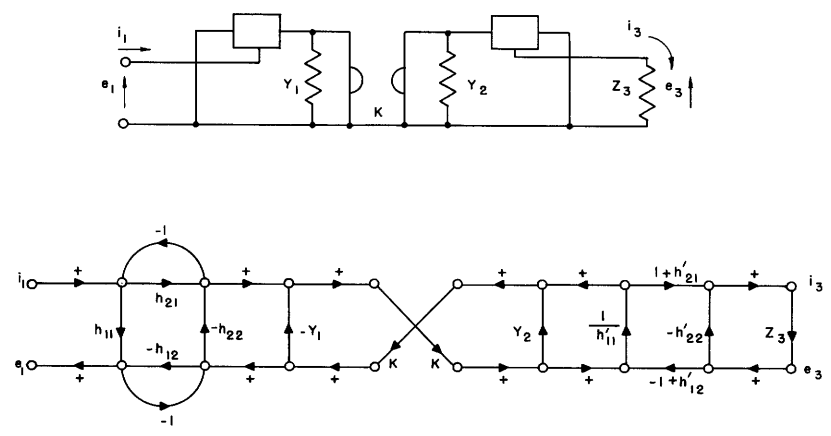


Fig. 25. Fifth example: a necessary inversion.

**5.8 Cascading General Elements.** Formulating the mathematics with flow graphs on an element basis permits the cascade of the flow-graph elements in the same order as that in which the physical elements are cascaded. The general procedure is to have current-gain elements drive current-gain elements and voltage-gain elements drive voltage-gain elements.

The general elements of Fig. 26 are divided into two groups – current-gain and voltage-gain. Each element has two inputs and one output or one input and two outputs.



(a)



(b)



(c)



(d)
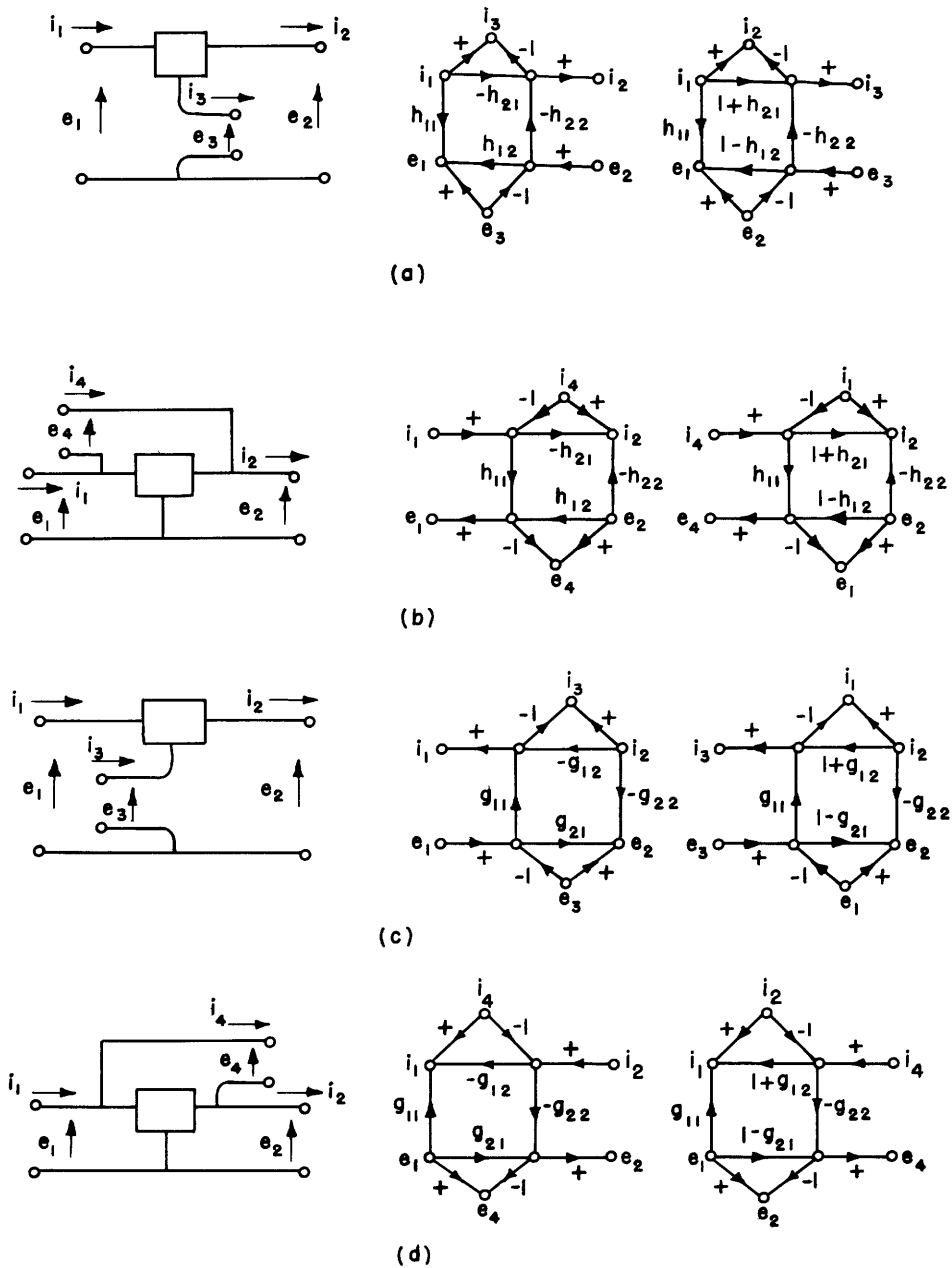
Fig. 26.  (a) Current gain, base input.  (b) Current gain, shunt input.
         (c) Voltage gain, base input.  (d) Voltage gain, shunt input.

26
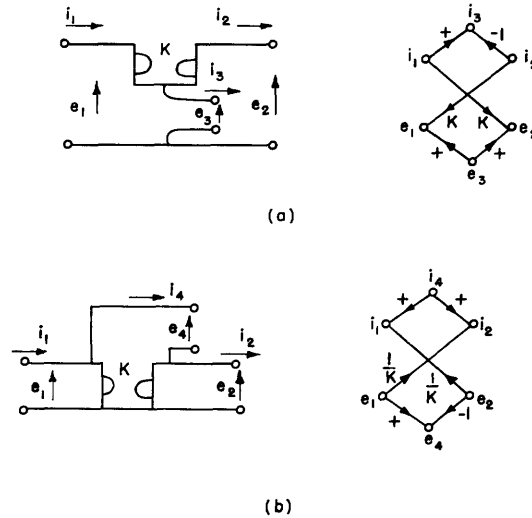
Fig. 27.   Gyrator model:  (a) current gain,  base output;
(b) voltage gain,  shunt output.

The difference between the input and output terminals is indicated by the direction of
the leads on the device.   The input is from the left; the output is to the right.   The con-
nection between the output and the input of the flow-graph elements is made on the basis
of Eqs. 12.

$$e_{out} = e_{in}$$

$$i_{out} = i_{in}$$

(12)

For example, consider the circuit in Fig. 31.   The voltage-gain flow graph of the
gyrator is used, since the output of the cathode follower is on the voltage basis.   The
usual procedure is to cascade the flow-graph elements in the same order as the physical
elements, making connections with lateral branches of +1.   The direction of the lateral
connection branches is the same as within the flow-graph elements.   By reversing the
sign of the current, a voltage-gain output becomes a current-gain input, and vice versa.
A good example is the vacuum tube that has both the cathode and plate as outputs.   With
a change in sign the cathode operates as a current-gain output instead of a voltage-gain
input.   This example is shown in Fig. 32.

5.9  Converging Cascade Circuits.   The convergence of two cascade circuits into one
circuit is accomplished by either of two methods.   One method is to have both of the
convergent circuits on a current-gain basis working into another current-gain device.
By this procedure the currents add to produce the current in the driven circuit.   An
example of this type of cascading is given in Fig. 33, where we obtain the current
addition

27
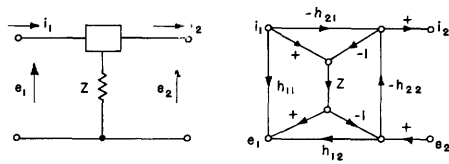
Fig. 28.   Base-loaded current-gain model.
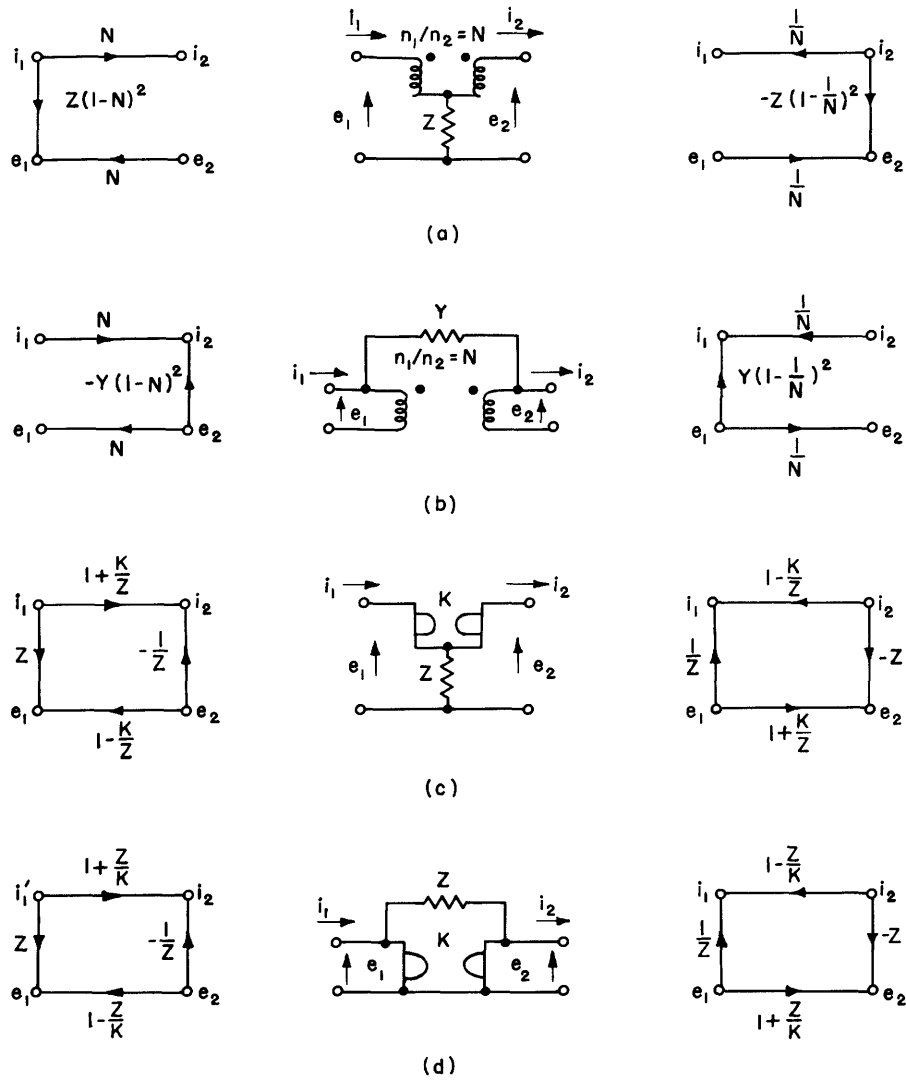


Fig. 29.   (a) Ideal transformer, base-loaded.  (b) Ideal transformer, shunt-loaded.  (c) Ideal gyrator, base-loaded.  (d) Ideal gyrator, shunt-loaded.

28
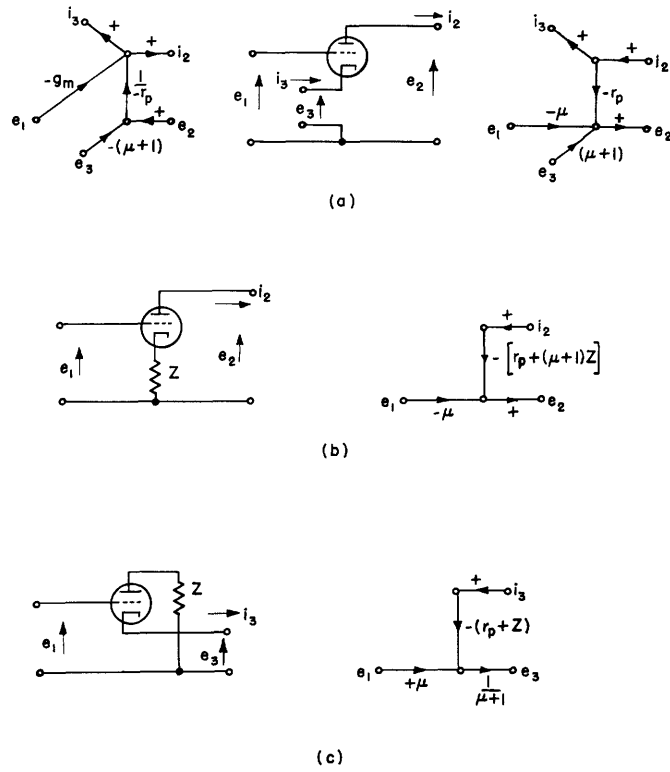
Fig. 30. (a) Voltage gain, cathode input. (b) Simple cathode feedback.
(c) Plate-loaded cathode follower.
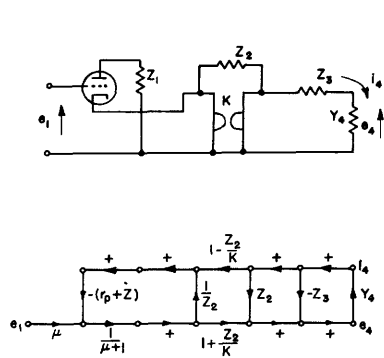


Fig. 31. Cascading flow-graph
elements.
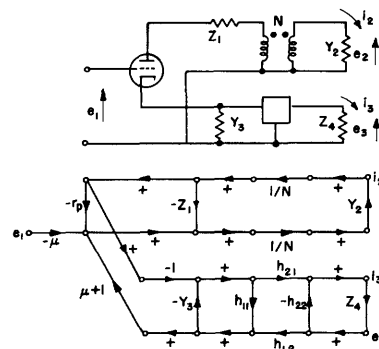


Fig. 32. Changing inputs to outputs.

29
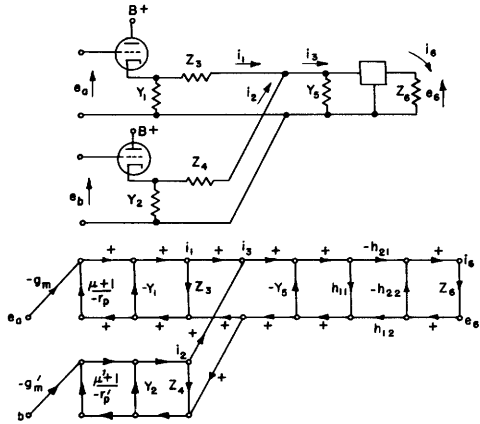
Fig. 33.  Converging-current circuits.
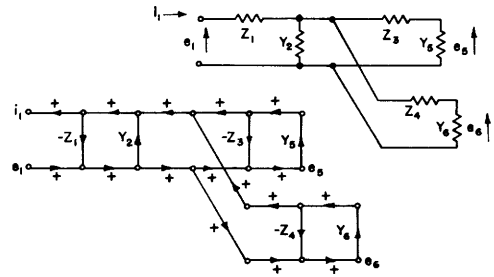


Fig. 34.  Diverging-voltage circuits.
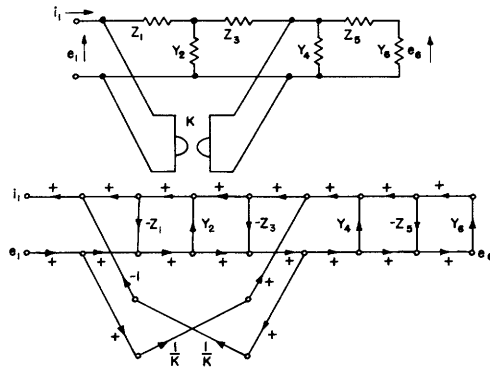


Fig. 35.  Split-T feedback.
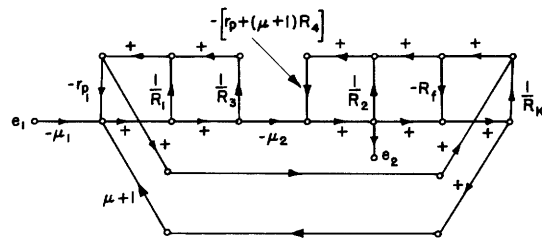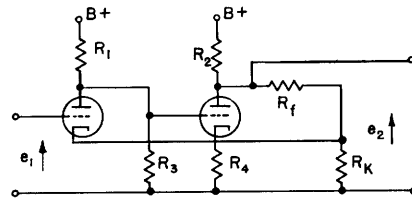


Fig. 36.  Twin-triode feedback amplifier.

$$i_3 = i_1 + i_2 \tag{13}$$

The current-gain output graphs are used for the cathode followers in order to get the subsequent graphs on a current-gain basis.

The second method of connecting the flow graphs of a convergent circuit involves the ideas of changing a current-gain output into a voltage-gain input. Converging circuits can be thought of as one voltage-gain circuit driving the other converging circuits and the common circuit as in the case of diverging circuits. This means that it is necessary to change the outputs of all but one of the converging circuits so that they appear to be a voltage-gain input. The change is made by changing the sign of the current in the output, which is essentially changing its direction.

5. 10 Diverging Cascade Circuits. The divergence of cascade circuits is another way of looking at converging cascade circuits only with the viewpoint reversed.

The simplest method of connecting flow-graph elements in a divergent system is to have the driving circuit on a voltage-gain basis working into the diverging circuits, which are also on a voltage-gain basis. For example, consider the network of Fig. 34 and its flow-graph formulation. Looking at the flow graph backwards we see that its structure is really that of two circuits converging to one, the current being added at the junction.

Another way of treating diverging circuits from a common circuit, is to think of them as a number of current-gain circuits converging on one of the diverging circuits. Of course, this requires changing the sign of the current so as to change the diverging voltage-gain circuits into converging current-gain outputs.

5. 11 Combined Diverging and Converging Circuits. By combining these two ideas of converging and diverging circuits we have the means of introducing feedback into a circuit. The feedback circuit essentially diverges from the main circuit at one point and then converges with it at another point. Consider the example in Fig. 35 of a loaded split-T network with a gyrator in its feedback arm. As a final example consider the standard twin-triode feedback circuit of Fig. 36. In this example the feedback is on the voltage-gain basis feeding back to the voltage-gain input of the cathode.

The object of this section has been to present a type of flow-graph element that is capable of representing the elements of cascade circuits. These flow-graph elements provide the means of setting up a mathematical formulation that has the same general appearance as the cascaded system. Having the physical elements in a one-to-one correspondence with the mathematical elements also provides a means of approximation in the reduction of the flow graph by removing or inserting elements. The facility of representing feedback in the cascade system, either within the element or from another part of the circuit, comes from the ideas involved in converging and diverging networks.

This type of formulation can be worked out rapidly with very little complication.

There are equally rapid methods for the flow-graph evaluation. The important thing is that there is an exact mathematical method of representing the problem which exhibits the structure of the mathematical setup and the structure of the system.

## 6. FORMULATION OF CONVENTIONAL NETWORK ANALYSIS

This section shows how standard network analysis is formulated in terms of flow graphs. The formulation results in a flow graph that not only represents the mathematics of the network but also retains some of the geometry of the network.

6.1 Use of Inversion. The work in the past has been directed towards procedures for setting up a consistent set of equations based on loop currents or node-pair voltages. By making use of these procedures and the principles of inversion of Section 4 it is possible to obtain a set of precise rules for rapid formulation of the flow graph for electrical networks. The rapid formulation is possible because of the ability to select a set of variables that will be sufficient to formulate the network. The loop-current formulation parallels the formulation for node-pair voltages. The only difference is in the method for finding the self- and mutual impedances or admittances.

The formulation of network analysis on the basis of loop currents is perhaps the most generally known and best understood. In general, it is first necessary to select an independent set of loop currents. With this set it is then possible to write down in matrix notation the equations of the system

$$Z I = E \tag{14}$$

In order to solve this set of equations with flow graphs it is necessary that E be strictly independent and I be dependent. This type of dependency can be obtained by using the ideas of the second inversion method, section 4.2. Thus, rewriting the equations so that I is dependent

$$E + (U-Z) I = I \tag{15}$$

permits the construction of the flow graph in the following manner (the quantity U represents the unit matrix):

1. Set up the collection of nodes representing the loop currents of I and the voltages E.

2. Place on each current node a self-loop equal to one minus the self-impedance.

3. Doubly connect each current node by branches equal to the mutual impedance between the loop currents, the sign being positive if the currents are in opposite directions, and negative if they are not.

4. Drive each current node by the effective source-voltage rise in the path corresponding to the loop current (current sources are considered as effective voltage sources or as independent loop currents).

6.2 Another Loop Current Formulation. By using the principle of the third inversion method of section 4.3, a slightly different type of flow-graph formulation is obtained. Instead of writing Eq. 14 in the standard form

$$\pm Z_{ij} i_1 + \ldots + Z_{jj} i_j + \ldots \pm Z_{nj} i_n = E_j \tag{16}$$

it could have been written in the form

$$\mp \frac{Z_{1j}}{Z_{jj}} i_1 + \ldots \mp \frac{Z_{nj}}{Z_{jj}} i_n + \frac{E_j}{Z_{jj}} = i_j \tag{17}$$

Thus, the second method of producing the flow graph is the following:

1. Set up the collection of nodes representing the loop currents and the voltage sources associated with the loops.

2. Doubly connect each current node by branches equal to the mutual impedance divided by the self-impedance associated with the loop current at the nose end of the branch. The sign is positive if the loop currents are in opposite directions in the mutual impedance; it is negative if they are in the same direction.

3. Drive each current node by the effective path-voltage rise divided by the self-impedance associated with the current node. A current source may be considered as an independent loop current.

These rules are the same for the node-pair-voltage formulation except that everything is in the dual (admittances for impedances and currents for voltages).


6.3 Selection of Variables. An effective means of setting up loop currents or node-pair voltages is based on the concepts of cut sets and tie sets of a network tree. These concepts permit the determination of the self- and mutual impedance or admittance by inspection of the electrical diagram.

A tree is defined as any set of connected branches that is just sufficient in number to connect all the nodes. The nodes are the connection points of the network elements. The branches are lines that interconnect the nodes.[*] Adding to the tree a branch corresponding to an element in the network forms a loop of branches and thus the path of the loop current.

The direction of the loop is assigned arbitrarily. In general practice the paths of the current loops are drawn directly on the network graph. The self-impedance of the loop is the impedance encountered along the path of the loop. The mutual impedance $Z_{ij}$ corresponds to the impedance encountered jointly by loops $i_i$ and $i_j$. Current sources in the network can be considered as equivalent voltage sources by associating with them a particular path between the two terminals of the source. The effective voltage in the branch of this path is then the impedance multiplied by the magnitude of the current source.

The self- and mutual admittance are found by cutting the tree. Cutting a branch of the tree divides the tree into two parts; cutting two branches of the tree

---

[*]It is unfortunate that two convenient definitions of nodes and branches exist, but the context will make clear which interpretation is intended: the flow graph or the network graph.

divides it into three parts.

A node-pair voltage is associated with each branch of the tree and its direction is assigned arbitrarily. Usually the tree of node-pair voltages is superimposed directly on the network graph with arrowheads indicating the direction of the node-pair voltages.

The self-admittance $Y_{jj}$ corresponds to the admittance between the two halves of the tree when the node-pair voltage $e_j$ is cut. The sign is always positive. The mutual admittance $Y_{ij}$ corresponds to the admittance directly between the outer parts of the tree when the node-pair voltages $e_i$ and $e_j$ are cut. The sign is positive if the two voltages are in the same direction in the tree.

The effective current source $I_j$ is the source driving the two halves of the tree determined by the node-pair voltage $e_j$. Voltage sources in series with a branch element may be considered as effective current sources across that branch element.

6.4 Tie-Set and Cut-Set Example. As an example consider the network of Fig. 37a. The solid lines of Fig. 37b represent a convenient tree; the dotted lines represent the branches that are singularly added to indicate the paths of the loop currents. Super-imposing the loops with assigned directions on Fig. 37a gives Fig. 38a. The use of the tree allows us to set up a set of independent variables. We can then pick out the self-impedances, the mutual impedances, and the effective source voltage rises directly from the network graph, and thus we are able to construct the flow graph directly. By the first method, the flow graph of the network of Fig. 38a is given in Fig. 38b. By using the second method and considering the numbers associated with the resistors of Fig. 39a as conductances, the flow graph of Fig. 39b is drawn by inspection of the cut set of Fig. 39a.

Both formulations have their advantages. The first has simple branch functions and symmetry but a more complex structure than the second. The second is simple in structure but has complex branch functions. It does have the advantage of being dimensionally consistent. This consistency makes it possible to apply physical reasoning to the formulation mechanics.

6.5 Conventional Loop Currents and Node-Pair Voltages. A very convenient degenerate case is that of conventional node-pair voltages and loop currents. The flow graph corresponding to these loop currents is constructed of nodes placed in the same relative position as the loop currents of the network graph. The flow graph then retains some of the geometry of the network graph while representing the loop equations.

For node-pair voltages the degenerate case is the case in which the tree is such that each branch extends from the same datum node to a network node. The flow graph corresponding to this tree then retains the geometry of the network graph by placing the nodes of the flow graph in the same relative position as the nodes of the network graph.

Both of the methods are formulated for these degenerate cases in such a way that the sign of the mutual impedances and admittances will always be positive when the
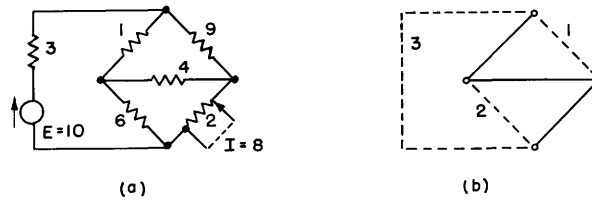
Fig. 37.   (a) Tie-set example.   (b) Network tree.
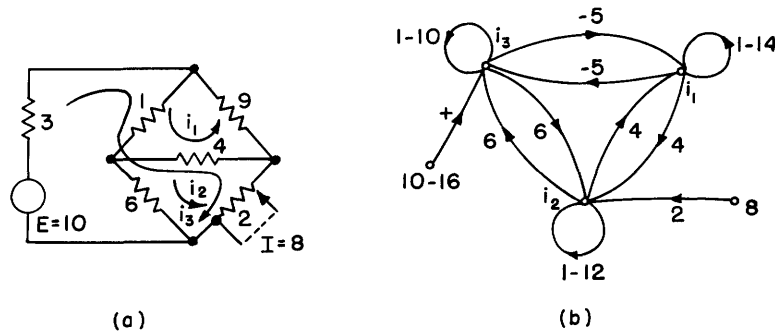


Fig. 38.   (a) Independent loop currents.   (b) Loop-current flow graph.
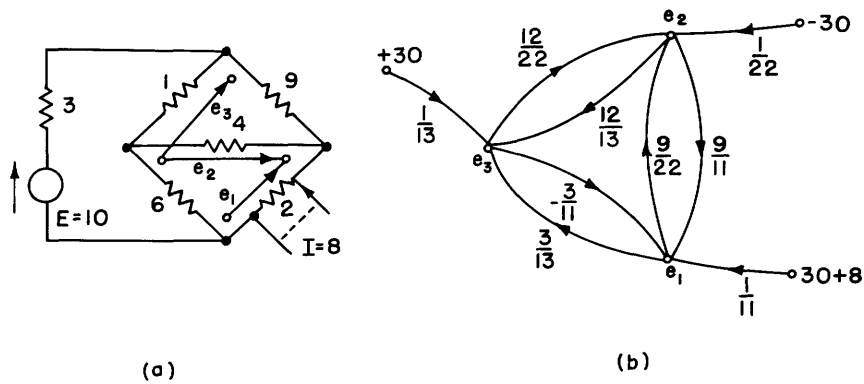


Fig. 39.   (a) Cut-set example.   (b) Node-pair voltage flow graph.

36
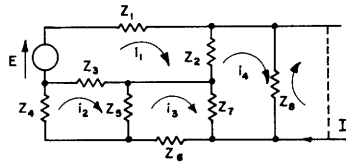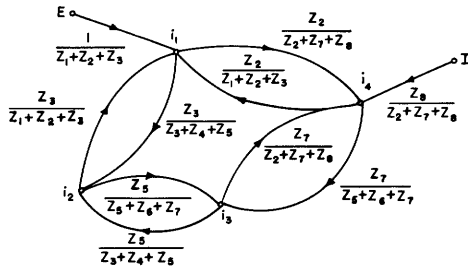
Fig. 40. (a) Conventional loop-current example. (b) Flow-graph formulation of Fig. 40a.
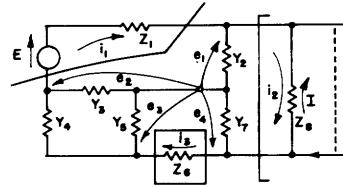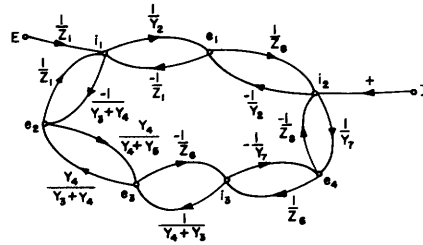


Fig. 41. (a) Combined use of loop currents and node-pair voltages. (b) Flow graph for taking out $Z_6$ or shorting out $Y_2$.

conventional loop currents and node-pair voltages are used. Placing the current nodes in the same relative position as the loop currents rapidly produces a graph having mutual impedance between the nodes equal to the mutual impedance between the corresponding loop currents. The same is true for node-pair voltages. An example is given in Fig. 40b, which represents the electrical network of Fig. 40a. Note that each loop current can be considered as an independent current source in relation to the other loop currents.

6.6 Joint Cut Sets and Tie Sets. It is not always desirable to solve a whole network on the basis of all node-pair voltages or all loop currents. By dividing the network into sections it becomes possible to formulate any one section by either node-pair voltages or loop currents. The other sections act on this section as though they are current or voltage sources, depending upon their formulation. This type of reasoning works particularly well when applied to the conventional node-to-datum voltage and loop-current formulation, since the node variables are placed in the same relative position as found in the electrical network graph.

6.7 Formulations Suitable for Approximations. Many times we are interested in formulating the flow graph so that we can "short out" or take out different electrical elements of the network. For this situation it is convenient to have the element appear alone in the flow graph and in such a form that the flow-graph branch becomes zero when

37

the electrical element is shorted out or taken out.

To formulate a flow graph with these requirements for shorting out an electrical element, formulate the element on the node-pair voltage basis and the rest of the surrounding network on the loop-current basis, placing the datum voltage node at one end of the element.

To take out an element, formulate the element on the loop-current basis and the surrounding network on the node-pair voltage basis.

Figure 41b is an example of a formulation of Fig. 41a suitable for shorting out $Y_2$ and taking out $Z_6$. When $Y_2$ is shorted out, two branches of the flow graph are made zero, so that $e_1$ becomes a strictly independent source and can be made zero.

This type of formulation is quite important, since it leads to a solution of the flow graph by which the effect of these elements can be separated from the rest of the solution. This separation is quite desirable in finding approximations and the effect of approximations without having to reformulate the mathematics to recognize the effect. That is, the flow graph of Fig. 41b could have been solved with $Y_2$ very large or essentially shorted out. The effect on the solution by this approximation that $Y_2$ is large can be investigated by replacing $Y_2$ and adding its correction terms to the main solution. This method of solution is discussed in the next section.

This section has presented a number of rapid methods of formulating the flow graph of networks containing two-terminal bilateral elements. It is important to have a number of methods available in order to obtain a solution in a desired form and with a minimum of complication. Since flow graphs exhibit the mathematical complexity, it is possible to see in advance which formulations are best suited for the desired results.

# 7. FLOW-GRAPH DETERMINANTS

Much of the work with linear sets of equations parallels the work with flow graphs. The determinant of a linear set of equations is paralleled by the determinant of a flow graph representing these equations.

This section shows that the determinant of a flow graph and the determinant of the representative linear set of equations are equal except for a sign factor, $(-1)^n$, where n is the number of dependent variables in the inversion. Subsequent development deals with the loop rule (corresponding to Cramer's rule), partitioning, and expansion.
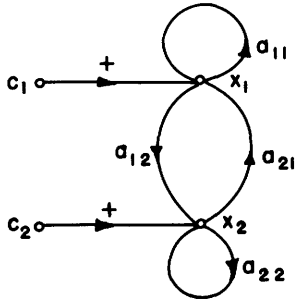


Fig. 42. Second-order set of equations.

It is both interesting and profitable to be able to calculate and understand these flow-graph determinants. It is profitable because the two determinants are closely related. It is interesting because we can easily see how to expand and factor a determinant. It is also important because it permits the use of Cramer's rule in reducing the flow graph.

7.1 Flow-Graph Equations. Flow graphs are the representation of algebraic sets of equations in which the nodes represent the known and unknown sets of variables and the branches represent the dependency among the variables. Drawing a flow graph is, in effect, writing a set of equations in the following form:

$$a_{11}x_1 + \ldots + a_{j1}x_j + \ldots + a_{n1}x_n + c_1 = x_1$$

$$- - - - - - - - -$$

$$a_{1j}x_1 + \ldots + a_{jj}x_j + \ldots + a_{nj}x_n + c_j = x_j \tag{18}$$

$$- - - - - - - - -$$

$$a_{1n}x_1 + \ldots + a_{jn}x_j + \ldots + a_{nn}x_n + c_n = x_n$$

A second-order set of equations in this notation would be that of Fig. 42. The order of the subscripts was purposely chosen so that it indicates the direction of dependency. The coefficient $a_{21}$ is associated with a branch from $x_2$ to $x_1$.

7.2 Loops and Ordered Sequences. The set of loops is an important geometric property of a flow graph. Here, a loop is defined as an ordered sequence of the nodes corresponding to either known or unknown variables, no variable appearing more than once in any one loop. The initial variable is immaterial as long as the order is preserved.

Two disjoint loops are sequences that have no common elements. The loop product is the product of the coefficients associated with the branches that go from one member of the sequence to the next; for example, the loop $x_3 x_5 x_9 x_7$ has the loop product $L = a_{35} a_{59} a_{97} a_{73}$. The loop product associated with the loop $x_i$ is $L = a_{ii}$. (When it is not ambiguous it is customary to shorten the term "loop product" to "loop.")

7.3 Coefficient Determinant. In the more conventional notation, Eq. 18 could have been written as

$$(a_{11} - 1)x_1 + \ldots + \quad a_{j1} x_j \quad + \ldots + \quad a_{n1} x_n \quad = -c_1$$

$$a_{1j} x_1 \quad + \ldots + (a_{jj} - 1)x_j + \ldots + \quad a_{nj} x_n \quad = -c_j \tag{19}$$

$$a_{1n} x_1 \quad + \ldots + \quad a_{jn} x_j \quad + \ldots + (a_{nn} - 1)x_n = -c_n$$

The coefficient determinant, D, is associated with the set of Eqs. 19.

$$D = \begin{vmatrix} a_{11} - 1 & \ldots & a_{j1} & \ldots & a_{n1} \\ a_{1j} & \ldots & a_{jj} - 1 & \ldots & a_{nj} \\ a_{1n} & \ldots & a_{jn} & \ldots & a_{nn} - 1 \end{vmatrix} \tag{20}$$

In respect to standard mathematical manipulations it is important to know the relation between this coefficient determinant and the flow graph from which it was produced. In relation to the representative flow graph the value of the coefficient determinant is given by the equation

$$D = (-1)^n \sum (-1)^q L_1 L_2 \ldots L_q (1 - a_{ii}) \ldots (1 - a_{jj}) \tag{21}$$

where each term of the summation is composed of disjoint loops and contains n and only n coefficients. The summation is over all possible combinations of disjoint loops. The factor L is a loop involving two or more variables, q being the number of such loops in the particular term.

7.4 General Validity of Determinant Equation. In order to show the general validity of the determinant equation, Eq. 21, it is necessary to note that numbering of the variables associated with the graph and determinant is immaterial as long as no two variables have the same number. That is, the determinant permuted in the same order of column and row has the value of the original determinant, and the loop products are dependent only upon the geometry of the flow graph. By renumbering, any term of Eq. 21 can be put in the form

$$(-1)^{n+q} \left( a_{12} a_{23} \cdots a_{s_1 1} \right) \left( a_{s_1+1 \, s_1+2} \cdots a_{s_1+s_2 \, s_1+1} \right) \cdots$$

$$\cdot \, (1 - a_{s+1 \, s+1}) \cdots (1 - a_{nn}) \tag{22}$$

where the first set of coefficients refers to loop $L_1$, the second refers to $L_2$, and so forth, ending with the coefficients of the self-loops. The loop product $L_j$ contains $s_j$ coefficients, so that a total of

$$s = s_1 + s_2 + \ldots + s_q \tag{23}$$

coefficients are involved in the part of the term not containing the self-loop coefficients.

The general term can be written in a slightly different form:

$$(-1)^{s-q} \left( a_{12} a_{23} \cdots a_{s_1 1} \right) \left( a_{s_1+1 \, s_1+2} \cdots a_{s_1+s_2 \, s_1+1} \right) \cdots$$

$$\cdot \, (a_{s+1 \, s+1} - 1) \cdots (a_{nn} - 1) \tag{24}$$

The value of $D$ can also be obtained from the coefficient determinant, Eq. 20, by standard procedures (15, 16):

$$D = \sum (-1)^d \, a_{1r_1} a_{2r_2} \cdots a_{nr_n} \tag{25}$$

where any diagonal term is $(a_{ii} - 1)$ and the second subscripts $r_1, r_2, \ldots, r_n$ run through all the $n!$ possible permutations of the numbers $1, 2, \ldots, n$. The exponent $d$ is the number of permutations in the second subscript.

Any term of this evaluation will have certain convenient properties associated with its subscripts. Starting with any off-diagonal coefficient $a_{ij}$ $(i \neq j)$ of the general term, it will be possible to find another coefficient $a_{jk}$ in the same general term. Since $j$ has already been used as a second subscript it is known that $j \neq k$. Thus, it is possible to find a sequence of coefficients that is such that the subscripts form a sequence of closed ordered pairs, as $a_{ij} a_{jk} \ldots a_{mi}$. You may be certain that it is closed, since $i \neq j$, and the number $i$ must be used as a second subscript for some one of the coefficients.

Thus any term of this evaluation can be put in the form of Eq. 24 by properly renumbering the variables. The sign is correct, since the number of permutations of the second subscript of each factor $L_i$ is $s_i - 1$ or a total of $s - q$ permutations for the whole term.

By virtue of the equality of Eqs. 22 and 24, each term of the flow-graph evaluation of the determinant appears in the standard evaluation and vice versa. Each term in each evaluation appears once and only once with a unity coefficient. Thus the general validity of the determinant equation, Eq. 21 is established.

7.5 Calculation of the Flow-graph Determinant. For convenience, a new quantity $\Delta$, which is called the flow-graph determinant, is defined:

$$\Delta = (-1)^n D \tag{26}$$

Usually both $D$ and $\Delta$ are called the determinant. When a distinction between the two meanings is necessary, the context will make the difference known.

The second subscript of the coefficients in each term of the determinant runs through all of the $n$ numbers of the unknown variables once and only once. On the representative flow graph the second subscript indicates the variable at the nose end of the branch. Thus all of the unknown variables are associated once with each term of the determinant; or, essentially, each term covers the set of unknown variables.

This idea of covering is useful in calculating the value of the determinant from the flow graph; in order to be a complete term, each term which is written down must cover the graph. This means that all of the nodes corresponding to unknown variables must be involved in each term of the determinant.

Thus, the following equation is valid for the calculation of the flow-graph determinant $\Delta$:

$$\Delta = 1 \cdot S_1 - \Sigma_r L_r^1 \cdot S_r + \Sigma_r L_r^2 \cdot S_r - \Sigma_r L_r^3 \cdot S_r + \ldots \tag{27}$$

where $L_r^m$ is the loop product of $m$ disjoint loops of two or more variables, and $S_r$ is the product of one minus the self-loops disjoint from the loops in $L_r^m$.

Equation 27 is a convenient one to use because it gives an orderly procedure of writing down the loop products:

First, the product of one minus the self-loops.

Second, the sum of the loops taken one at a time, each multiplied by any disjoint self-loops.

Third, the sum of the products of the disjoint loops taken two at a time, each multiplied by any disjoint self-loops; and so forth. (Note that the sign is changed from one step to the next.)

The largest combination of disjoint loops of two or more variables will be no more than $n/2$ and usually considerably less for practical problems.

For the special case of no loops in the graph, the determinant $\Delta$ is unity. A unity self-loop on the unknown variable $x_j$ introduces a zero factor into the terms in which it appears. Since each term covers the graph, only terms which have a loop passing through $x_j$ will be nonzero.

The determinant is zero when there is a node variable having a self-loop of unity and no loops passing through the node. The determinant is also zero when there are two nodes that have identical sets of either converging or diverging branch coefficients.

Since the determinant is directly dependent upon the loops of the system, the determinant is independent of any variables which are not associated with any of the loops.

By expanding Eq. 21 for $\Delta$ we can obtain the form

$$\Delta = 1 - \Sigma_r P_r^1 + \Sigma_r P_r^2 - \Sigma_r P_r^3 + \ldots \tag{28}$$

where $P_r^m$ is the $r^{th}$ possible combination of $m$ disjoint loops. Each summation is over all possible combinations.

If all of the self-loops are zero, Eqs. 27 and 28 are identical.

7.6 Examples of Determinant Calculation. By following the procedure outlined in the previous section, the determinant associated with Fig. 42 is, by the first method,

$$\Delta = (1 - a_{11})(1 - a_{22}) - a_{12}a_{21}$$

or, by the second method, it is

$$\Delta = 1 - a_{11} - a_{22} - a_{12}a_{21} + a_{11}a_{22}$$

A more practical example is offered by finding the determinant of a transistor oscillator in order to investigate the frequency and requirements of oscillation. Consider the transistor circuit driving a phasing network that drives the input of the transistor, as in Fig. 43a. This circuit has the flow graph of Fig. 43b, where the phasing network has been described in the same set of parameters as the transistor. By inspection of the flow graph the determinant can be written as follows:

$$\Delta = 1 + h_{21}a - h_{12}a + \frac{Z_1}{z_{22}} + \frac{h_{11}}{Z_2} - h_{21}h_{12}\frac{Z_1}{Z_2} - h_{21}h_{12}a^2 + \frac{h_{11}Z_1}{Z_1 z_{22}}$$

Another common example is the standard current and voltage feedback amplifier shown in Fig. 44a. Its representative flow graph is that of Fig. 44b. The determinant of the system is then

$$\Delta = 1 + (\mu_1 + 1)\left[\mu_2 + \left(\frac{1}{R_1} + \frac{1}{R_3}\right)(r_2 + R_f)\right] + r_1\left(\frac{1}{R_1} + \frac{1}{R_3}\right)$$

$$+ \frac{r_2 + R_f}{R_k} + \frac{r_2}{R_2} + r_1\left(\frac{1}{R_1} + \frac{1}{R_3}\right)\left(\frac{r_2 + R_f}{R_k} + \frac{r_2}{R_2}\right)$$

As a final example, consider the determinant of the following set of equations:

$$x_1 = \frac{1}{g_1}(y_1 - p_n x_n)$$

$$x_2 = \frac{1}{g_2}(y_2 - p_1 x_1)$$

$$x_n = \frac{1}{g_n}(y_n - p_{n-1} x_{n-1})$$

The representative flow graph is that of Fig. 45. From inspection of this flow graph
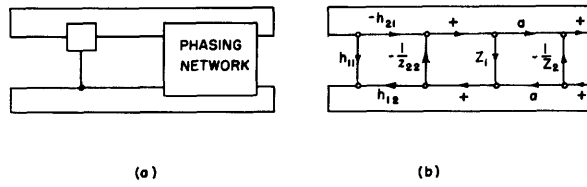
Fig. 43.   Transistor oscillator and flow graph.
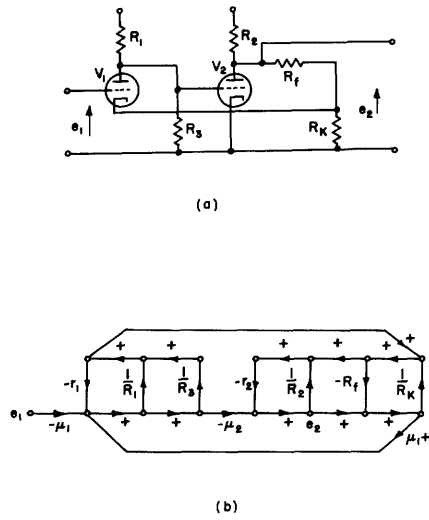


Fig. 44.   Twin-triode feedback amplifier.
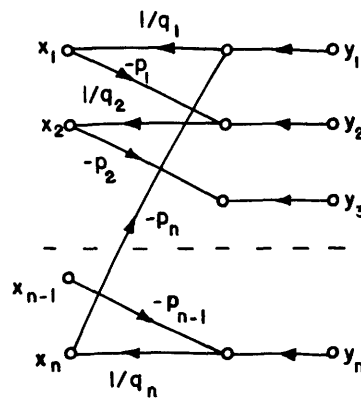


Fig. 45.   Flow graph of a set of equations.

we find that the determinant is

$$\Delta = 1 - (-1)^n \frac{\prod\limits_{i=1}^{n} p_i}{\prod\limits_{i=1}^{n} q_i}$$

7.7 Summary of Determinant Calculations. One of the chief advantages of the calculation of determinants with flow graphs is that it gives us the ability to see which determinants are going to be easy to calculate. This permits the generation of examples for which the final results are simple enough to understand. For instance, the last example refers to the input and output probabilities of a noisy information channel. The determinant indicates that the solution will always exist for n odd.

Many problems can be formulated directly in flow-graph notation without first writing out the representative algebraic equations. Thus, the determinant of the system can be produced by a one-step inspection process. In general, flow graphs represent another way of looking at the mathematical system and producing its properties by noting the characteristic interdependency of the variables.

The next sections deal with the application of flow-graph determinants. They present the loop rule parallel to Cramer's rule. Also they show how a graph may be partitioned in order to factor and expand the determinant into more desirable forms.

7.8 The Loop Rule. S. J. Mason first formulated the loop rule and proved it from purely flow-graph considerations (10). This section shows that the rule corresponds to Cramer's rule.

A general flow graph graphically represents the algebraic equations

$$a_{11}x_1 + \ldots + a_{j1}x_j + \ldots + a_{n1}x_n + c_1 = x_1$$

$$a_{1j}x_1 + \ldots + a_{jj}x_j + \ldots + a_{nj}x_n + c_j = x_j \qquad (29)$$

$$a_{1n}x_1 + \ldots + a_{jn}x_j + \ldots + a_{nn}x_n + c_n = x_n$$

Adding to this set the equation

$$x_j - x_j' = 0 \qquad (30)$$

the solution for $x_j$ is obtained by application of Cramer's rule:

45

$$x'_j = \cfrac{\begin{vmatrix} (a_{11} - 1) & \cdots & a_{j1} & \cdots & a_{n1} & -c_1 \\ \text{-------} \\ a_{1j} & \cdots & (a_{jj} - 1) & \cdots & a_{nj} & -c_j \\ \text{-------} \\ a_{1n} & \cdots & a_{jn} & \cdots & (a_{nn} - 1) & -c_n \\ 0 & \cdots & 1 & \cdots & 0 & (1-1) \end{vmatrix}}{\begin{vmatrix} (a_{11} - 1) & \cdots & a_{j1} & \cdots & a_{n1} & 0 \\ \text{-------} \\ a_{1j} & \cdots & (a_{jj} - 1) & \cdots & a_{nj} & 0 \\ \text{-------} \\ a_{1n} & \cdots & a_{jn} & \cdots & (a_{nn} - 1) & 0 \\ 0 & \cdots & 1 & \cdots & 0 & -1 \end{vmatrix}} \tag{31}$$

$$x'_j = \frac{\Delta_j}{\Delta} \tag{32}$$

where $\Delta$ is the determinant of the flow graph, and $\Delta_j$ is the determinant of a modified flow graph constructed by placing a self-loop of unity on $x'_j$ and branches having a coefficient of $-c_i$ from $x'_j$ to each $x_i$.

Due to the unity self-loop on $x'_j$, each nonzero term of the modified determinant involves a loop passing through $x'_j$. Each loop passing through $x'_j$ is associated with a set of disjoint loops or essentially a disjoint determinant or cofactor of the variables not included in the particular loop passing through $x'_j$.

Thus, the general loop rule for the reduction of a linear flow graph is

$$x'_j = \sum_{r=1}^{n} c_r T_{rj} \tag{33}$$

where

$$T_{rj} = \frac{\Sigma_k L_{k_{rj}} \Delta_{k_{rj}}}{\Delta} \tag{34}$$

and $L_{k_{rj}}$ is the $k^{th}$ path product of the path from node $c_r$ to the node $x'_j$. The quantity $\Delta$ is the determinant of the flow graph, and $\Delta_{k_{rj}}$ is the cofactor of the path associated
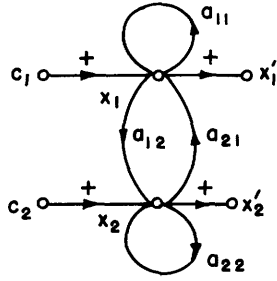
Fig. 46. Second-order set of equations.
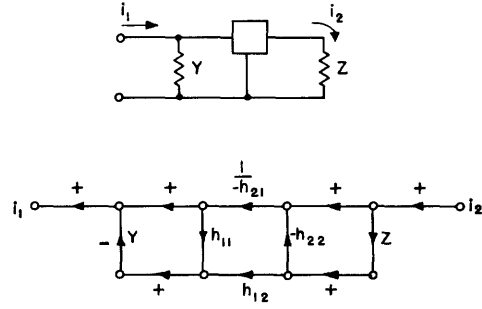


Fig. 47. Loaded transistor and inverted flow graph.

with the path product $L_{k_{rj}}$. The summation of $k$ is over all possible paths from node $c_r$ to $x'_j$. The quantity $T_{rj}$ is usually called the "dependency" of $x_j$ on $c_r$.

The solution of the flow graph of Fig. 46 corresponding to a second-order set of equations is

$$x_1 = \frac{c_1(1 - a_{22}) + c_2 a_{21}}{(1 - a_{11})(1 - a_{22}) - a_{12} a_{21}}$$

$$x_2 = \frac{c_1 a_{12} + c_2(1 - a_{11})}{(1 - a_{11})(1 - a_{22}) - a_{12} a_{21}}$$

It is interesting to note that the same dependency relation $T_{rj}$ is obtained between $x'_j$ and $c_r$ if all of the arrowheads are reversed and the dependency is calculated from $x'_j$ to $c_r$. Turning the arrowheads around only reverses the direction of the loops and paths but does not change the path products.

7.9 Special Cases of the Loop Rule. When there are no loops in the graph, the dependency $T_{rj}$ becomes the sum of all of the path products between $c_r$ and $x_j$. Very often it is possible to partially factor the dependency by breaking the dependency of a group of paths passing through some intermediate node $x_i$ into two parts, $T_{ri}$ and $T_{ij}$. Thus we have the result that

$$T_{rj} = T_{ri} \cdot T_{ij} + T'_{rj} \qquad (35)$$

where $T'_{rj}$ is the sum of the path products that do not pass through node $x_i$. Of course, this partial factoring can be extended to a number of nodes or individual branches in different combinations.

The loaded transistor, where the path between the input and output currents has been inverted, is an example of a graph containing no loops (Fig. 47). By using the individual branch $-1/h_{21}$ in the expansion, the dependency is obtained in partially factored form.

47

$$i_1 = i_2 \left\{ (1 + h_{22} Z) \left( \frac{1}{-h_{21}} \right) (1 + h_{11} Y) + h_{12} ZY \right\}$$

The advantage of partially factoring the dependency is that the dependency is calculated in parts. These parts are easy to locate in the flow-graph structure.

7.10 Determinant Expansions. Expansion of the flow-graph determinant offers a means of writing down its expression in a variety of ways. From this variety it is possible to select an expansion that will put the expression for the determinant in an easily understood form.

For example, we may be interested in a particular constant, $a_{ij}$, associated with one of the branches of the flow graph. Upon separating those terms of the determinant which do and do not contain $a_{ij}$ and factoring out $a_{ij}$, the following expansion is obtained

$$\Delta = \Delta_o - a_{ij} \Sigma_k L_{k_{ji}} \Delta_{k_{ji}} \tag{36}$$

where $\Delta_o$ is the determinant of the flow graph with the branch $a_{ij}$ removed, and $-a_{ij} \Sigma_k L_{k_{ji}} \Delta_{k_{ji}}$ is the sum of all terms involving a loop containing a sequence of the form $(\ldots x_i x_j \ldots)$. That is, we calculate $\Delta$ with $a_{ij}$ removed and add to it that part of $\Delta$ in which each term $a_{ij}$ forms part of a loop function.

This type of expansion can be extended to the removal of a number of branches. The determinant is first calculated with the desired set of branches removed. To this calculation is added that part of $\Delta$ involving the branches singularly (essentially singularly replaced in the graph). That part of $\Delta$ involving all possible combinations of any two branches of the set is next added, and so forth, with 3, 4, ... branches, until all combinations are treated.

A particular case of the loop rule occurs when the removed branches either all converge on a node or all diverge from a node. The cross terms are then nonexistent. For this special case

$$\Delta = \Delta_o - \Sigma a_{ij} \left\{ \Sigma_k L_{k_{ji}} \Delta_{k_{ji}} \right\} \tag{37}$$

where the first summation is over i or j depending upon whether the removed branches are convergent or divergent. This degenerate case corresponds to Laplace's expansion of a determinant by minors. Summing over i is an expansion of the row; summing over j is an expansion of the column. The self-term in Laplace's expansion is

$$(1 - a_{ii}) \Delta_o \tag{38}$$

Along this line of reasoning is the expansion about a loop product $L_i$ involving two or more variables. For this situation

$$\Delta = \Delta_o - L_i \Delta_i \tag{39}$$

where $\Delta_o$ is the determinant calculated with the loop excluded, and $\Delta_i$ is the cofactor of $L_i$.

In a similar manner this type of expansion can be extended to the removal of several loops. The degenerate case is where all of the removed loops are intersecting loops in all combinations.

An important property of these expansions is that the removal of a set of branches or loops may radically change the appearance of the graph and thus the way of looking at the graph.

7.11 Partitions. In practical applications, it will usually be found that the flow graph of a problem has a tendency to "string out" and to run in parallel. Sometimes it will be possible to partition the graph into smaller parts by drawing partition lines through the graph. A partition line is any line drawn around part of the graph in such a manner that all branches which cross the line are in the same direction, or a line drawn between partition lines which does not cross any of the branches of the flow graph.

The set of nonzero loop products associated with a partitioned graph has the property that each loop is entirely within a specific partition of the graph.

Subsequent calculation of the determinant of a graph divided into partitions shows that the determinant is equal to the product of the determinants of the partitions; that is,

$$\Delta = \Delta_A \cdot \Delta_B \cdot \ldots \cdot \Delta_N \tag{40}$$

As an example, consider determining the dependency of $e_2$ on $e_1$ of the two-stage vacuum-tube amplifier of Fig. 48. Since the graph for Fig. 48 can be divided into two partitions, the dependency can be written immediately by making use of the partitioning and the loop rule:

$$e_2 = e_1 \frac{\mu_1 \mu_2}{\left(1 + \dfrac{r_1}{Z_1}\right)\left(1 + \dfrac{r_2}{Z_2}\right)}$$

7.12 Partial Factoring. By using the ideas of expanding and factoring a determinant, there exists an excellent means of working a problem in isolated parts and then adding in the interaction of the parts.

By proper selection of the removed branches or loops in the expansion of a determinant, it may be possible to get the graph broken up into a number of partitions. This expansion can then take advantage of the partitions by expressing the $\Delta_o$ determinant as the product of smaller determinants which are easier to see and to write out. The expansion is especially profitable when there is a branch that is used in only one or two loop functions and its removal would split the graph up into a number of partitions.

A graph can be artificially partitioned by drawing a line around part of it and removing all the branches that cross the line in one direction. This artificial

partitioning is a method of seeing how to expand the determinant. It permits working on isolated parts of the graph and then adding in the interdependency of the parts.

A particular example might consist of finding the current gain of the loaded transistor of Fig. 49. Removal of the backward voltage-gain branch, $h_{12}$, partitions the graph so that the gain is

$$i_2 = i_1 \frac{-h_{21}}{\left(1 + \dfrac{h_{11}}{Z_1}\right)\left(1 + \dfrac{Z_2}{h_{22}}\right) - h_{12} h_{21} \dfrac{Z_2}{Z_1}}$$

**7.13 Parallel Partitions.** In certain cases the dependency, $T_{rj}$, between two nodes will be made up of a number of paths through parallel partitions extending from one node to the other. Parallel partitions between two nodes are partitions that extend from one node to the other with no interconnecting branches between the partitions.

For two parallel partitions $M$ and $N$ extending from node $c_r$ to node $x_j$, the equivalent dependency would be

$$T_{rj} = \frac{\Sigma_k L_{k_M} \Delta_{k_M} \Delta_N + \Sigma_k L_{k_N} \Delta_{k_N} \Delta_M}{\Delta_M \Delta_N}$$

$$= \frac{\Sigma_k L_{k_M} \Delta_{k_M}}{\Delta_M} + \frac{\Sigma_k L_{k_N} \Delta_{k_N}}{\Delta_N}$$

$$= T_{rj_M} + T_{rj_N} \tag{41}$$

The important thing to notice is that the dependencies of parallel partitions are added together to get the total dependency.

Consider finding the dependency of the input voltage upon the input current of the base-loaded transistor. The circuit and flow graph are given in Fig. 50. In this example the graph is divided into two parallel partitions. Taking advantage of the partitioning, we can write the dependency as the sum of two terms:

$$e_1 = i_1 \left\{ h_{11} + \frac{Z_1 \left[ (1 + h_{21})(1 - h_{12}) + \dfrac{Z_2}{z_{22}} \right] - h_{12} h_{21} Z_2}{1 + \dfrac{Z_1 + Z_2}{z_{22}}} \right\}$$

**7.14 Isolated Partitions.** Sometimes in the calculation of the dependency between two variables $c_r$ and $x_j$ by the loop rule, none of the paths between the two nodes pass through a given partition $Q$. This partition is then an isolated partition.

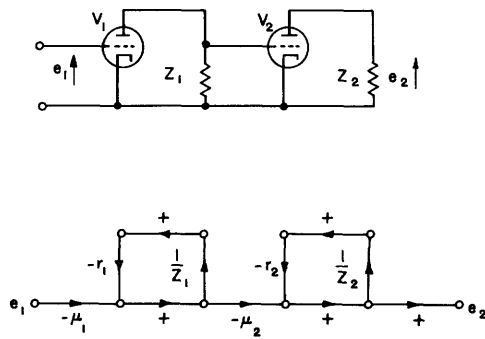Applying the ideas of factoring a determinant to this situation produces Eq. 42.

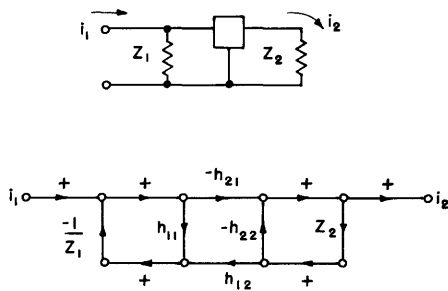Fig. 48. Vacuum-tube amplifier and flow graph.
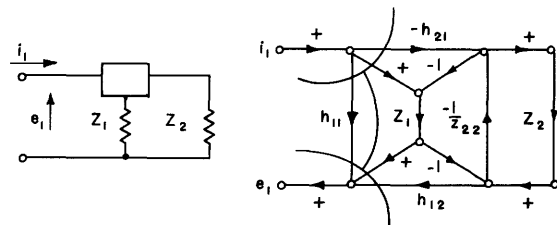


Fig. 49. Loaded transistor and flow graph.



Fig. 50. Base-loaded transistor and flow graph.
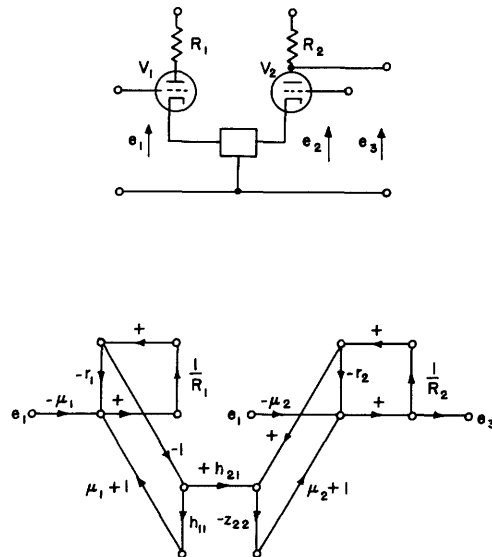


Fig. 51. Two vacuum tubes and a transistor.

51

$$T_{rj} = \frac{\Sigma_k \, L_{k_{rj}} \, \Delta'_{k_{rj}} \, \Delta_Q}{\Delta' \cdot \Delta_Q}$$

$$T_{rj} = \frac{\Sigma_k \, L_{k_{rj}} \, \Delta'_{k_{rj}}}{\Delta'} \qquad\qquad (42)$$

where $\Delta_Q$ is the determinant of the partition Q and $\Delta'$ and $\Delta'_{k_{rj}}$ are determinants of the remaining part of the graph.

The point is that isolated partitions may be excluded from the calculation of dependency between two nodes.

Consider the example in Fig. 51 of two vacuum tubes and a transistor, where the voltage gains from $e_1$ and $e_2$ to $e_3$ are of interest. In formulating the graph the backward voltage gain of the transistor is assumed to be negligible. By making use of the partitioning and the isolated partition for the gain from $e_2$, the following results are obtained directly from the flow graph.

$$e_3 = e_2 \frac{-\mu_2}{1 + \dfrac{r_2 + (\mu_2 + 1)\, z_{22}}{R_2}}$$

$$e_3 = e_1 \frac{-\mu_1 (\mu_2 + 1)\, h_{21} \dfrac{z_{22}}{R_1}}{\left[1 + \dfrac{r_1 + (\mu_1 + 1)\, h_{11}}{R_1}\right]\left[1 + \dfrac{r_2 + (\mu_2 + 1)\, z_{22}}{R_2}\right]}$$

7.15 Multiplication by a Constant. Quite often it is desirable to change either the numerical value or the sign of a number of branches in a flow graph. This section describes a method of making these alterations in linear flow graphs.

The dependency through any given region of a flow graph is solely determined by the paths and loops through the region and the loops within the region. Thus, a line drawn around a region intersects all loops an even number of times and all paths an even or odd number of times, dependent upon whether one or both of the end points of the paths are inside or outside the region. Multiplying all of the incoming branches to the region by a constant C and dividing all of the outgoing branches by C leaves the dependency relations through the region unchanged. Dependency relations into the region are multiplied by the constant C.

An important example is found in the application of flow graphs to reflection and image transfer loss and phase in network theory (1). The representative flow graph, Fig. 52b, of the network of Fig. 52a can be considerably simplified by eliminating all
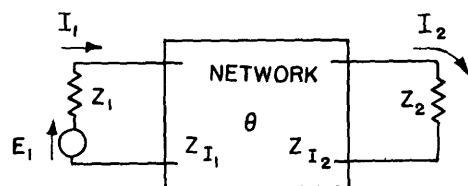
(a)
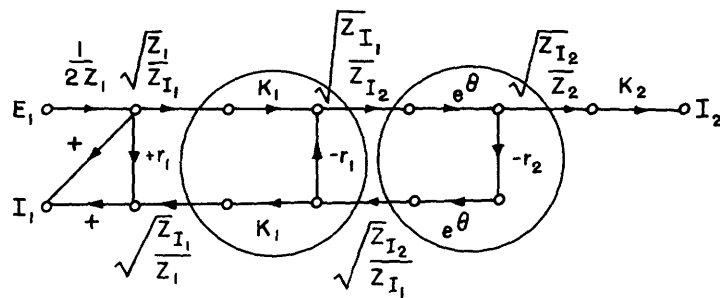


(b)
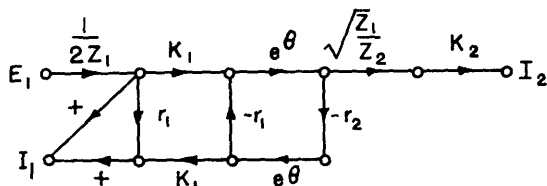


(c)

$$r_1 = \frac{Z_1 - Z_{I_1}}{Z_1 + Z_{I_1}} \qquad k_1 = \frac{2\left(Z_1 Z_{I_1}\right)^{1/2}}{Z_1 + Z_{I_1}}$$

$$r_2 = \frac{Z_2 - Z_{I_2}}{Z_2 - Z_{I_2}} \qquad k_2 = \frac{2\left(Z_2 Z_{I_2}\right)^{1/2}}{Z_2 + Z_{I_2}}$$

Fig. 52. Changing branches by a constant.

but one of the impedance level changes. The regions are shown on the graph of Fig. 52b. Modification then produces the simpler graph of Fig. 52c.

7.16 Maximum Number of Unity Branches. In certain situations it is helpful to manipulate the flow graph in order to introduce as many unity branches as possible. For a linear flow graph, where all of the branches have arbitrary constants, the largest number of branches that can be made unity is the number of nodes in the graph excluding those that are strictly independent or dependent, provided there is at least one. Otherwise, it is one less than the number of nodes. It has been assumed that dependencies are calculated from strictly independent nodes to strictly dependent nodes. C. E. Shannon (11) proved a very similar statement in a paper on analog computers.
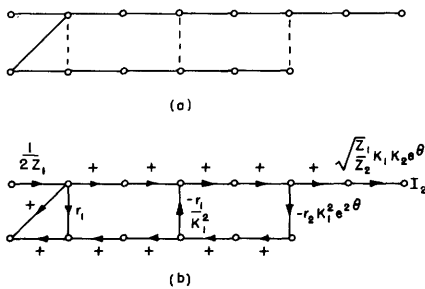


Fig. 53. Reduction of branches to unity.

By considering the graph of branches without arrowheads, it is seen that there exists a number of possible trees of branches. A tree of branches is a connected set of branches which is just sufficient to connect all of the nodes of the graph. For any tree all of the strictly independent and dependent nodes are connected to the tree by at least one branch.

Starting at any strictly independent or dependent node and passing along the tree it is possible to make each branch passed over equal to unity except those branches that lead to strictly independent or dependent nodes. The number of such operations is the number of nodes in the graph, excluding the strictly independent and dependent nodes.

The assumption that more than this number could be made unity implies that the operation must be applied twice at some node to produce two unity branches. This is a contradiction, since the branch constants are arbitrary. Thus the maximum is equal to the number of nodes.

As an example, the flow graph of Fig. 52a indicates that eleven branches can be made unity. A possible tree is that of Fig. 53a. Making the indicated alterations produces the graph of Fig. 53b.

7.17 Summary of Expansions and Partitions. Expansions and partitions represent a means of breaking the problem into its natural parts. You are able to see the expansions and partitions because you are able to see the structure of the mathematics. By observing the paths and loops of the mathematical structure, the problem is solved without the use of standard mathematical manipulation. Flow graphs solve the problem in parts with interaction — all steps being performed upon inspection of the mathematical structure.

# 8. SPECIAL TOPICS IN LINEAR EQUATIONS

Since linear flow graphs represent a method of solving linear sets of equations there exists a good deal of material pointing out how well-known methods of linear analysis are formulated in terms of flow graphs. This section deals first with the rank and positive definite properties of a determinant and then with the calculation of eigenvalues and eigenvectors.

8.1 Rank of a Determinant. The rank of a determinant is an indication of the largest number of independent variables which a consistent system of equations can specify in terms of the remaining independent and dependent variables.

The usual procedure in linear equation analysis is to start with a set of consistent equations. A subset of these equations is then solved for the independent variables in terms of the remaining independent and dependent variables. The largest number of independent variables for which the solution exists corresponds to the rank of the original set. Of course, the solution of a consistent set of linear equations exists when the determinant is nonzero.

In flow-graph notation the original set of equations corresponds to a graph in which all of the nodes are strictly dependent or independent. The solution is obtained by inversion of some of the paths and flow-graph reduction. Thus the rank is indicated by the largest number of inversions that can be performed without making the flow-graph determinant zero.

8.2 Positive Definite Determinants. For certain systems to be realizable it is necessary that the determinant of the system have the property generally known as "positive definite." Essentially this condition is that the quantity

$$\sum_{ij} x_i a_{ij} x_j \tag{43}$$

is greater than zero for any arbitrary set of x except when all x are zero. The determinant is negative definite when the quantity is less than zero for all sets except the zero set. It is well known that a real symmetric matrix is positive definite if each of the discriminants $\Delta'_m$ (m = 1, 2, ..., n) are positive (15).

This requirement on the discriminants corresponds to the following flow-graph requirement: For a flow graph in which each pair of variables is doubly connected with branches having equal real coefficients and any arbitrary numbering, the quantities $(-1)^m \Delta_m$ (m = 1, 2, ..., n) are all greater than zero. The quantity $\Delta_m$ is the flow-graph determinant of the system involving just the first m variables.

The sequence $(\Delta_1, \Delta_2, ..., \Delta_n)$ is usually quite easy to calculate according to the Laplace expansion in terms of flow graphs, Eq. 37.

8.3 Eigenvalues. In the process of setting up the normal coordinates of a linear system it is first necessary to obtain the natural frequencies or eigenvalues of the system. The characteristic equation

$$|A - \lambda I| = 0 \tag{44}$$

which leads to the eigenvalues can be formulated in terms of flow graphs by placing a self-loop of $-\lambda$ on each dependent variable. Calculation and solution of the flow-graph determinant

$$\Delta = 0 \tag{45}$$

for the roots of $\lambda$ produces the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.


8.4 Eigenvectors. The eigenvectors are obtained under the condition that $\Delta = 0$. This condition is obtained by setting $\lambda$ equal to one of the eigenvalues.

Two sets of eigenvectors are possible, a right-hand and a left-hand system. The right-hand system corresponds to the matrix equation in the form

$$A x = \lambda x \tag{46}$$

The left-hand system corresponds to the form

$$x A = \lambda x \tag{47}$$

For an eigenvalue not equal to any of the others, the right-hand vectors are obtained by ignoring all of the converging branches of one of the dependent nodes $x_i$ and calculating the dependency of the other nodes on this one node $x_i$. This set of dependencies is then the eigenvector in terms of the arbitrary scale factor $x_i$. The left-hand vector is obtained by ignoring all of the diverging branches of one of the dependent nodes $x_i$ and calculating the dependency of each of the other nodes to this one node $x_i$. The dependencies represent the eigenvector. Both of these calculations hinge on the cofactor of the node $x_i$ being nonzero. For an eigenvalue not equal to any of the others there will be at least one dependent variable which will have a nonzero cofactor.



$$A = \begin{bmatrix} -1 & 2 & -1 \\ 6 & -1 & 0 \\ 3 & 0 & -1 \end{bmatrix}$$

Fig. 54. Flow graph of $|A - \lambda I|$.

If there are two identical eigenvalues, two nodes $x_i$ and $x_j$ must be used to calculate the dependency. Of course, the determinant of the system formed by ignoring both of these nodes has to be nonzero. It will always be possible to find two such nodes for the case of equal eigenvalues. For higher numbers of equal eigenvalues, corresponding numbers of variables are selected and the dependencies calculated.
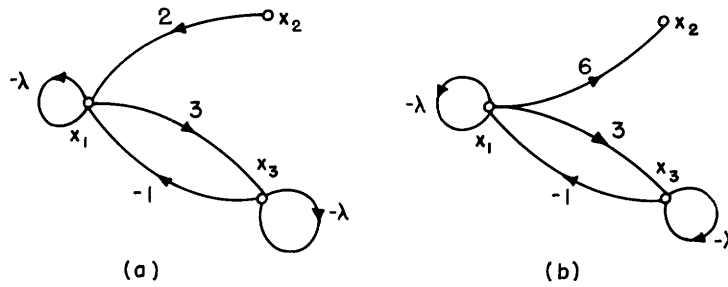
Right-hand eigenvector

|  | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|
| $\lambda_1$ | 0 | 1 | 2 |
| $\lambda_2$ | $\frac{1}{2}$ | 1 | $\frac{1}{2}$ |
| $\lambda_3$ | $-\frac{1}{2}$ | 1 | $\frac{1}{2}$ |

Left-hand eigenvector

|  | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|
| $\lambda_1$ | 0 | 1 | -2 |
| $\lambda_2$ | $\frac{3}{2}$ | 1 | $-\frac{1}{2}$ |
| $\lambda_3$ | $-\frac{3}{2}$ | 1 | $-\frac{1}{2}$ |

Fig. 55. Eigenvectors: (a) right-hand; (b) left-hand.

8.5 An Example of Eigenvalues and Eigenvectors. The following matrix, A, provides an example of the flow-graph manipulation involved in calculating eigenvalues and eigenvectors. The determinant $|A - \lambda I|$ is obtained from Fig. 54. Equating the determinant of the system

$$\Delta = (\lambda+1)(\lambda-2)(\lambda+4)$$

to zero produces the 3 eigenvalues

$$\lambda_1, \lambda_2, \lambda_3 : -1, +2, -4$$

The right- and left-hand vectors are calculated from Figs. 55a and 55b, respectively. Notice that it was not possible to use node $x_1$ in calculating the eigenvector for $\lambda = \lambda_1$ because of its zero cofactor.

# 9. SOLUTION OF MATRIX EQUATIONS

Flow graphs are ideal for working with a system of variables. This section deals with the flow-graph application of solving for the system of variables involved in matrix equations. The object is to set up a notation for matrix equations, demonstrate some of the simpler properties, and then show how to solve a system directly from the flow graph. The solution process will not be restricted to square matrices.
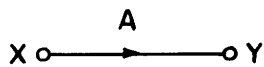
Fig. 56. Matrix equation.

P. C. Watson first recognized the application and performed some of the initial work. The actual mechanics was worked out by S. J. Mason and his graduate class. The material is included here for completeness.

9.1 The Left-Hand System. The left-hand notation for matrix equations is best suited for flow-graph representation. In this notation the equation

$$x A = y$$

has the flow-graph representation of Fig. 56.

On the basis of the representative equations, the equivalent flow graph of a cascade is the matrix product in the order of the cascade, Fig. 57a. Parallel branches are additive, as in Fig. 57b.

The calculation which justifies the equivalence relations of Figs. 58a and 58b shows that matrix equations are linear, so that superposition is applicable.

The solution of matrix equations rests on the solution of the equation

$$x A + y B = y$$

which has the flow graph of Fig. 59. This solution is

$$x \cdot A \cdot (1-B)^{-1} = y$$

or, in another notation,

$$x \cdot \frac{A}{1 - B} = y$$

where it is understood that quantities in the denominator are inverse-post-multiplied to the quantity in the numerator.

The introduction of the new variable $z$ in Fig. 59 provides a means of observing the loop dependence as a dependence of $z$ upon $y$ when the unit matrix branch is removed.

In order to calculate the equivalence relation of a path, the values of successive variables are calculated on the basis of the previous variable in the path, as in Fig. 59. In each succeeding calculation all of the paths through the previous variables are

58
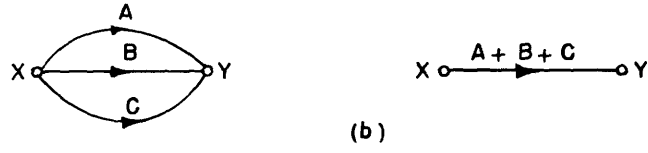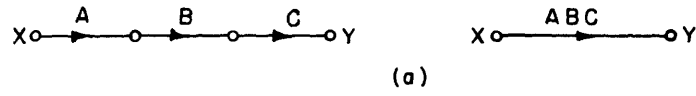
Fig. 57.  (a) Cascade and its equivalence.
(b) Parallel and its equivalence.
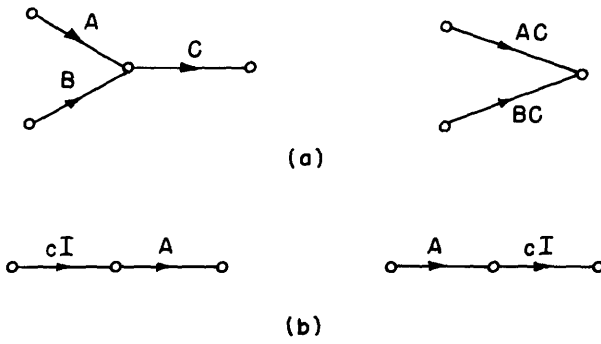


Fig. 58.  (a) Distributive law of matrix equations.
(b) Multiplication by a constant.
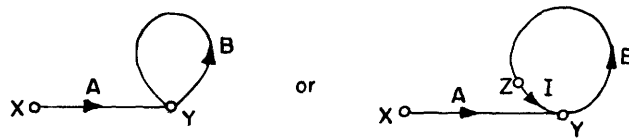


Fig. 59.  Basic matrix equation.
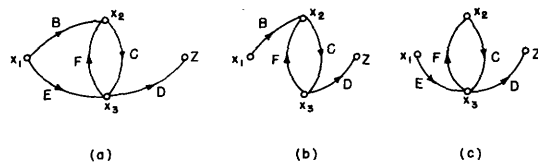


Fig. 60.  (a) Matrix example.  (b) and (c) Component matrix paths.

59

ignored, since these variables are no longer dependent upon the succeeding variables.

Because of superposition, the equivalence relation between a strictly independent variable and any dependent variable is the sum of the dependencies over all possible paths between the two variables.

9.2 Matrix Example.  Consider the example of Fig. 60a.  The two paths from $x_1$ to z are separated into Figs. 60b and 60c by superposition.  The dependency of the path $x_1 x_2 x_3 z$ is found by successively calculating $x_2$, $x_3$, and z on the basis of the immediately preceding variable in the path.

$$x_1 \cdot \frac{B}{1 - CF} = x_2 \qquad \begin{aligned} x_2 \, C &= x_3 \\ x_3 \, D &= z \end{aligned}$$

Notice how the path to $x_2$ is ignored when $x_3$ is calculated.  The dependency is then

$$x_1 \frac{B}{1 - CF} \cdot C \cdot D = z$$

The path $x_1 x_3 z$ has the dependency

$$x_1 \frac{E}{1 - FC} \cdot D = z$$

By superposition the total dependency of z on $x_1$ is

$$x_1 \left\{ \frac{B}{1 - CF} CD + \frac{E}{1 - FC} \cdot D \right\} = z$$

A more complicated example is that of Fig. 61, where the dependency of $x_1$ on $x_0$ is of interest.  The equivalent loop dependence is that of the last example, so that

$$x_0 \cdot \frac{A}{1 - \dfrac{B}{1 - CF} CD - \dfrac{E}{1 - FC} D} = x_1$$

9.3 Right-Hand System.  It is unfortunate that most matrix equations are expressed in a right-hand system.  In this system all of the relations are essentially backwards.  A convenient method of getting around this awkward situation is to trace out the paths backwards so that the expressions can be written in the customary left-to-right notation. Essentially, we look for the points of the arrowheads instead of the tails.  Here again
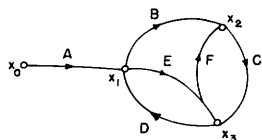


Fig. 61.  Matrix example.
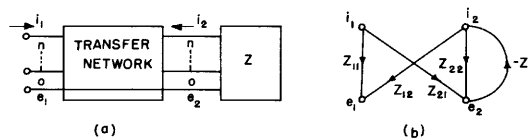


Fig. 62.  Transfer network and flow-
graph representation.

it is possible to use the notation that the inverse of the matrix in the denominator is to be post-multiplied to that of the numerator.

An actual example encountered in noise figures is the connection of a 2n-terminal transfer network to an n-terminal impedance network, Fig. 62a. The matrix equations are

$$e_1 = Z_{11}i_1 + Z_{12}i_2$$

$$e_2 = Z_{21}i_1 + Z_{22}i_2$$

$$i_2 = -Z^{-1}e_2$$

which are represented by the flow graph of Fig. 62b. The new input impedance matrix is then

$$e_1 = \left\{ Z_{11} - Z_{12} \frac{Z^{-1}}{1 + Z_{22}Z^{-1}} \cdot Z_{21} \right\} i_1$$

# 10. TRANSCENDENTAL FUNCTIONS

Included in this section are a number of elementary considerations dealing with the exponential and logarithmic functions, and the group operation of multiplication. With these elementary considerations it becomes possible to manipulate the flow graphs much as the representative mathematics would be manipulated.

10.1 Group Operation of Multiplication. In parallel with addition, multiplication of a number of variables is a group operation. A small square in flow-graph notation seems to be a good representation.

The equation $Z = w \cdot x \cdot y$ has the flow graph of Fig. 63. That is, convergence on the square indicates multiplication. Divergence from the square indicates the direction of dependency. Self-loops on the squares have no meaning.

A number of properties are of interest. A constant multiplier on any branch preceding the square can be slipped through the square to be multiplied into all of the diverging branches. Figure 64 shows two equivalent flow graphs. Of course, the operation works in reverse: dividing all of the diverging branches from a square multiplies any one of the converging branches by that constant.

Inversion of a path through a product square requires the same node-splitting ideas of the first inversion method for the additive operation (sec. 4.1). In contrast to the use of subtraction for the inversion of the additive operation, division is used for the inversion of the multiplication operation. Figure 65b represents the inversion of a path through the flow graph of Fig. 65a.

In the degenerate case of the inversion of multiplication there is only one branch entering or leaving the square.

10.2 Transcendental Functions. Transcendental functions, the exponential and logarithmic, have a property similar to one of the linear properties. The difference is an interchange of the group operation of multiplication and addition. This property is indicated in Fig. 66a for the exponential; in Fig. 66b, for the logarithm.

These two properties are of interest because they provide a means of manipulating a graph that involves transcendental functions and a means of deriving other flow-graph operations. For example, they provide a means of slipping forward an exponential function through a product square or slipping forward a logarithmic function through a summation node — operations that are indicated in Figs. 67a and 67b. They are obtained by introducing the unity operation of a logarithm cascaded with the exponential.

The slipping of an exponential function back through a summation node is indicated in Fig. 68a; the slipping of the logarithm back through a product square, in Fig. 68b. Two simple loop graphs that are easily reduced with these properties are those of Figs. 69a and 69b.
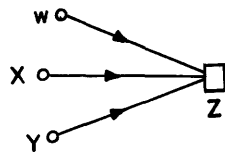
Fig. 63. Group multiplication.
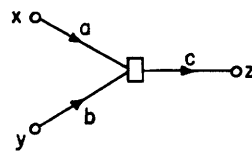


Fig. 64. Slipping a constant.
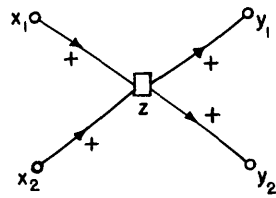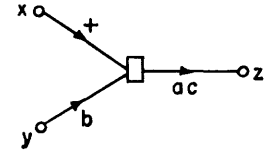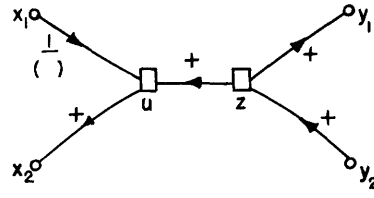


(a)                    (b)

Fig. 65. (a) General multiplication square. (b) Inversion of Fig. 65a.
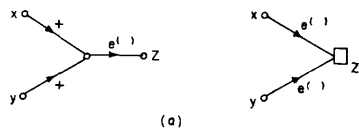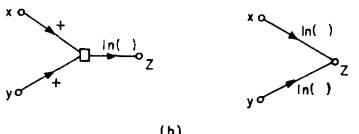


Fig. 66. (a) Equivalent exponential flow graphs. (b) Equivalent logarithmic flow graphs.
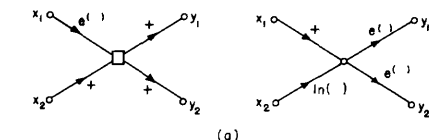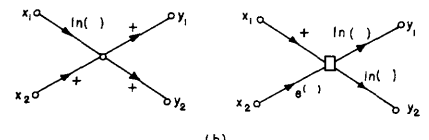


Fig. 67. (a) Slipping the exponential function forward through a product square. (b) Slipping the logarithmic function forward through a summation node.
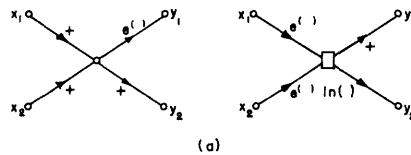


(a)
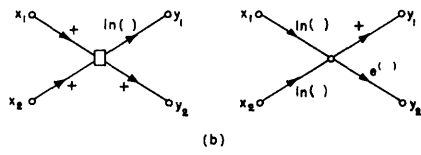
(b)

Fig. 68. (a) Slipping back the exponential. (b) Slipping back the logarithm.
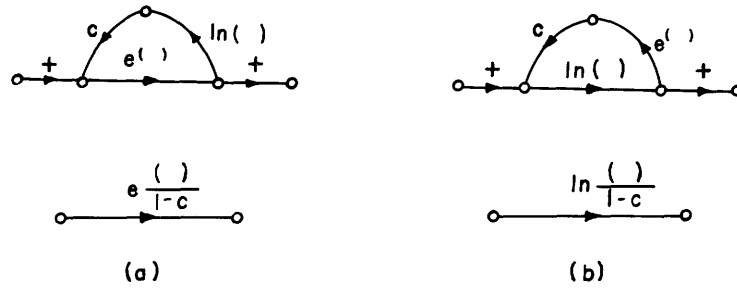
Fig. 69. Simple loop reduction.
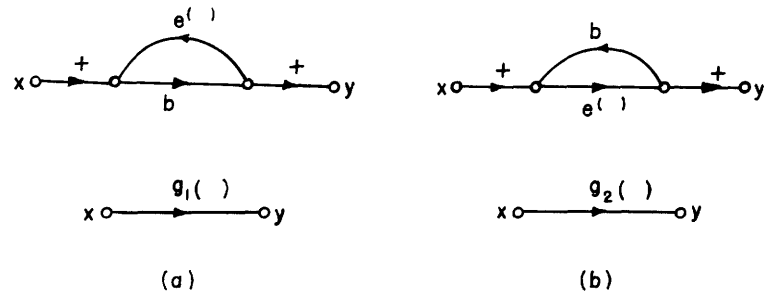


Fig. 70. Flow graphs of transcendental functions.
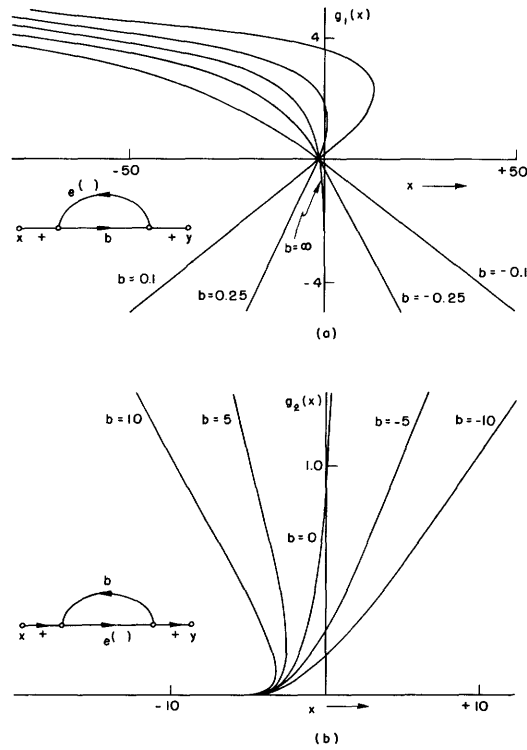


Fig. 71. Transcendental functions: (a) $y = g_1(x)$; (b) $y = g_2(x)$.

64
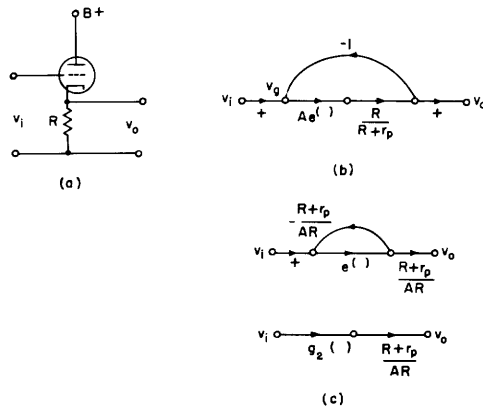
Fig. 72. Exponential amplifier.

10.3 Complex Transcendental Functions. Most transcendental sets of equations are not reduced as rapidly as those of Figs. 69a and 69b. However, the flow-graph notation provides an improved means of seeing the manipulations that are necessary for reducing the set to the desired properties.

Flow graphs also provide a means of specifying types of complex transcendental functions in terms of the inherent equations from which they are produced. Two simple functions, $g_1(x)$ and $g_2(x)$, are those produced by the flow graphs of Figs. 70a and 70b. Inversion of these graphs shows that $g_1(x)$ and $g_2(x)$ are the inverse functions of

$$x = -e^y + \frac{1}{b}y \quad \text{and} \quad x = \ln y - by$$

Typical plots of these functions are shown in Fig. 71.

It seems much easier to associate these functions with their flow graphs than with the inverse relations that actually specify the functions. The following example illustrates this point.

10.4 Exponential Amplifier. Consider the example of a cathode follower where the "$\mu v_g$" of the tube varies exponentially with the grid voltage: $\mu v_g = A \exp v_g$. A flow graph of the circuit of Fig. 72a is shown in Fig. 72b. Placing the flow graph in standard form produces the graph of Fig. 72c and the result

$$v_o = \frac{R + r_p}{A R} g_2(v_i) \qquad b = -\frac{R + r_p}{A R}$$

This final section on flow-graph theory has introduced some of the simpler properties of the group operation of multiplication and the transcendental functions. Since flow graphs are a means of depicting the functional dependency in sets of equations, they represent a means of manipulating the mathematics in a pictorial form from which the important properties of the dependency can be seen.

65

## II. STATISTICAL APPLICATIONS OF FLOW GRAPHS

### 1. INTRODUCTION

Pictorial graphs have long been used to demonstrate the structure of discrete statistical systems. The purpose of this part of the report is to show how this graphical structure is directly related to the mathematical structure of various statistical properties of the system.

A number of people have felt that it should be possible to obtain the properties of a Markov system directly from the graph which represents it. R. W. Sittler (17) demonstrated that the generating function of the transitional probabilities can be obtained by thinking of the system as a pulsed-data system.

Part II of this report treats the statistical graph as the mathematical structure of the system. This structure corresponds closely to the mathematical structure involved in deducing the probabilities of recurrent events, stationary distributions of recurrent systems, transient occurrences and durations, sensitivity and variation, and generating functions.

The mathematical structure of the statistical graph is the same mathematical structure found in flow-graph analysis. Much of the material of Part II is directed at showing this correspondence and showing that the statistical properties are closely linked together.

Part I dealt with the general theory of flow graphs and its applications to electrical networks. Part II will use the ideas of drawing out flow-graph equations, the inversion of dependencies, and the general reduction of linear graphs. The use of the flow graph will be principally to solve the systems of equations associated with properties and to display the interconnection between these properties.

1.1 Summary of Part II. Much of the material of Part II is directed towards Markov systems. The material is in the form of defining characteristic properties of the statistical system and showing how these properties are simply related to the Markov graph with the use of flow-graph techniques. This approach is important because it shows how various properties are tied together and tied in with the structure of the system.

The first section is an introduction to the general structure of Markov systems. Here the basic structure and terminology is introduced for those unfamiliar with this type of system. Many properties of a Markov system are the direct result of the first occurrence probabilities based on particular starting conditions. The second section shows that these probabilities, $P_{jk}$, can be calculated directly from the Markov graph by making the event represented by a state $s_k$ strictly dependent and driving the graph with a unit source at the initial starting state $s_j$.

With the first occurrence probabilities $P_{jk}$ and $P_{kk}$ it is then possible to calculate the following probabilities of occurrence of the state $s_k$ for the system

starting in state $s_j$:

$$P_{jk}(\text{at least } N) = P_{jk}P_{kk}^{N-1}$$

$$P_{jk}(\text{at most } N) = 1 - P_{jk}P_{kk}^{N}$$

$$P_{jk}(\text{exactly } N) = P_{jk}P_{kk}^{N-1}(1 - P_{kk})$$

We also obtain the result that the mean, $O_{jk}$, and variance, $V_{jk}$, of occurrence are given by the following formulas:

$$O_{jk} = \frac{P_{jk}}{1 - P_{kk}}$$

$$V_{jk} = \frac{P_{jk}(1 + P_{kk} - P_{jk})}{(1 - P_{kk})^2}$$

The first occurrence probability can also be used to calculate the probability of entering a particular closed set by considering the probabilities of the different transitions leading to the closed set.

The more interesting systems are those that run for a while and then stop. In the mean, each state in the system will occur a number of times. This number is the mean occurrence of the state and is identical with the definition of mean occurrence in the section on recurrent events.

Section 3 (transient systems) shows that the mean occurrence, $O_{jk}$, is given by the equation

$$O_{jk} = \sum_{n=0}^{\infty} p_{jk}^n$$

where $p_{jk}^n$ is the $n^{th}$ step transition probability from state $s_j$ to state $s_k$. Through the use of a recurrence relation it is then possible to obtain a set of equations that leads to the solution of $O_{jk}$. The flow graph of this set of equations corresponds exactly to the Markov graph, so that the property is a direct result of the graph. The mean duration of a transient system is the sum of the mean occurrences of the transient states

$$D_j = \sum_{k=1}^{m} O_{jk}$$

The report also shows how to obtain the mean duration to a particular closed set and the variance of the duration for the whole transient system and

to a particular closed set.

Irreducible recurrent Markov systems approach a stationary distribution irrespective of the initial state of the system. Section 4 presents four methods of using flow graphs to find this distribution directly from the Markov graph. The fourth method is particularly interesting in that it shows how this distribution is related to the recurrent event theory and to the mean duration of a transient system, $P(s_k) = 1/D_k$.

Engineers are usually quite interested in the effect of changes in the size of different components or criterions. For a Markov system these changes result in changes in the probabilities of the graph.

Because of the formulation of the Markov properties in flow-graph notation, properties are conveniently expressed in a bilinear form, which places the changing parameter in evidence. Since the denominator is the same for all terms of properties involving a summation, the properties are usually easy to express in this form:

$$P, O, D = \frac{\alpha p + \beta}{\gamma p + \delta}$$

The bilinear form conveniently shows the variation of the property as the parameter changes. The sensitivity is then given by the relation

$$S = \frac{(\alpha \delta - \beta \gamma) p}{(\alpha p + \beta)(\gamma p + \delta)}$$

The material on variation and sensitivity is in Section 5.

R. W. Sittler's thesis (17) dealt essentially with generating functions by treating the Markov system as a pulsed-data system with a corresponding delay function in each transitional branch.

Section 6 on generating functions has been included for two reasons: to make the material on Markov systems complete and to show the connection between generating functions and the previous work in the report. Starting with the definition of the generating function

$$O_{jk}(x) = \sum_{n=0}^{\infty} p_{jk}^n \cdot x^n$$

Section 6 develops the connecting relations to previously derived parameters and properties. Some of the important relations involve the first and second derivatives of the function. With the use of a recurrence relation it is shown that this generating function can be calculated by the solution of the set of equations

$$O_{jk}(x) = \delta_{jk} + \sum_{r=1}^{m} O_{jr}(x) \cdot p_{rk} \cdot x \qquad (k = 1, \ldots, m)$$

The flow graph of this set of equations corresponds to the Markov graph with each transitional probability multiplied by x. Thus, with a modification of the graph, these generating functions can be obtained directly from the Markov graph.

It is important to note that it is not as easy to calculate or to differentiate these generating functions as it may at first seem. It is usually easier to obtain certain desired properties directly from the graph rather than indirectly from the generating functions.

A simple type of Markov system is the discrete information channel. Muroga (20) posed and solved the problem of determining the optimum input distribution to the channel. With the flow graphs given in Section 7 it becomes possible to see how this solution comes about and to understand the treatment of exceptional cases in which sizes of the input and output alphabet do not match or when certain solutions do not exist. Being able to see the solution, we are in a good position to select examples that are workable and can be interpreted.

Section 8 deals with the association of a measure of information to the statistical process. This measure is essentially an inverse measure of the deterministic nature of the system. The treatment is carried out for transient and recurrent systems.

1.2 Discrete Markov Systems. Many statistical processes can be approximated by a number of states and transitional probabilities of moving from one state to another. This is a convenient representation because it permits us to follow the system as it moves from state to state and to formulate and solve for various properties of the system. Systems of this sort are commonly known as discrete Markov systems. This type of system model is often used in describing the operation of discrete waiting lines, reordering schemes, learning processes, component failure and test, discrete information channels, and language structure.

Discrete Markov systems are defined by a discrete set of states $(s_1, s_2, \ldots, s_m)$ with an associated matrix of conditional probabilities $[p_{jk}]$ of going from state $s_j$ to state $s_k$. Each of these conditional probabilities of going from one state to another is restricted to being a function of only the two states.

The graph representing this system is a collection of nodes representing the states $(s_1, s_2, \ldots, s_m)$ and branches representing the conditional transitions between the states. Branches representing zero probability are usually omitted. Along with the initial starting probabilities the system's statistical properties are completely determined in the sense that they exist. Each transition is called one move or one step. The $n^{th}$ step transitional probabilities, $p_{jk}^n$, are the probabilities of going from state $s_j$ to $s_k$ in exactly n steps by any path between $s_j$ and $s_k$. Because of the particular assumed independence of the system, the $n^{th}$ step transitional probabilities satisfy the recurrence relation of Eq. 1:

$$p_{jk}^{o} = \delta_{jk} \qquad p_{jk}^{n} = \sum_{r=1}^{m} p_{jr}^{n-1} \, p_{rk} \qquad\qquad (1)$$

where the summation is over the $m$ states of the system and $\delta_{jk}$ is the Kronecker delta.

Since $n$ moves will put the system in some state, the $n^{th}$-step probabilities also satisfy the relation of certainty, Eq. 2.

$$\sum_{k=1}^{m} p_{jk}^{n} = 1 \qquad\qquad (2)$$

1.3 Set Properties. Graphical representation of Markov processes usually leads to a better insight into the mechanisms of the statistical operation of the system. From the graph, certain properties of the states become evident upon inspection. Consider, for example, the Markov graph of Fig. 73.
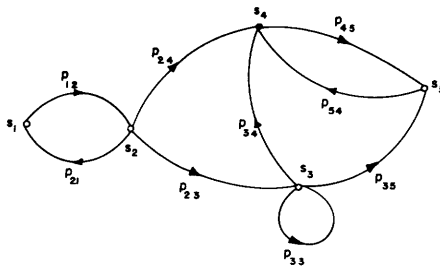


Fig. 73. Markov graph.

A "closed" set of states, $(s_3, s_4, s_5)$ or $(s_4, s_5)$, is a set of states out of which the process cannot move, once it has moved into one of the states. For the special case of a state $s_k$ being a closed state, the state is called an "absorbing" state. The whole system is "irreducible" if there is only one closed set of states corresponding to the set of all states. A "decomposable" system refers to a system which can be decomposed into a number of closed sets. "Transient" states $(s_1, s_2, s_3)$ are states to which the system may never return. "Periodic" states are states that can occur only at definite regular intervals. States $s_4$ and $s_5$ of Fig. 73 can occur only alternately. States $s_1$ and $s_2$ are also periodic. A "recurrent" state $(s_4$ or $s_5)$ is a state which, having occurred once, can always occur again. A special type of recurrent state is the "null" state which has the characteristic that its $n^{th}$-step transitional probability approaches zero.

$$p_{kk}^{n} \to 0 \qquad n \to \infty$$

All of these set properties except for the null state can usually be found by inspection of the Markov graph. They represent certain descriptive peculiarities about the system which are convenient in describing the system's operation as it moves from state to state.

## 2. RECURRENT EVENTS

In any experimental science, the experimenter is usually quite interested in observing events and the characteristic properties of their recurrence. This section deals with the probabilities of first and repeated occurrences of events and their use in calculating the mean and variance of occurrence.

2.1 Probability of First and Repeated Occurrence. For a system containing transient states it is of interest to know the probabilities associated with the occurrence of a particular transient event (transition or state) occurring "at least," "at most," or "exactly" N times before the system passes out of the transient states. These probabilities are calculated from what are generally known as first-occurrence probabilities of the particular event based on a specified initial starting distribution.

Let $P_{jk}^n$ represent the probability of the first occurrence of the state $s_k$ on the $n^{th}$ move based on the assumption that the system started in state $s_j$. Since the $n^{th}$ and the $n+1^{st}$ events are mutually exclusive events, the probability of first occurrence on any move, $P_{jk}$, is the infinite sum of the probability of first occurrence on the $n^{th}$ move.

$$P_{jk} = \sum_{n=0}^{\infty} P_{jk}^n \qquad (3)$$

If this probability should happen to be unity, the event is called "certain" in contrast to being called "uncertain" when the probability is less than unity.

The $n^{th}$-step first-occurrence probabilities differ from the $n^{th}$-step transitional probabilities in that the first-occurrence probabilities are restricted to those paths that do not pass through the event state $s_k$. In the particular case of the initial state $s_j$ being the event state $s_k$ the probabilities are known as the $n^{th}$-step recurrent probabilities.

A graphical model depicting the paths involved in the first-occurrence probabilities can be constructed by splitting the event state $s_k$ into two states: one with all of the branches diverging from $s_k$; the other, with all of the branches converging on $s_k$.

In this type of system the $n^{th}$-step first-occurrence probabilities can be calculated with the recurrence relation similar to Eq. 1 but different in the respect that paths do not go through the event state $s_k$.

$$P_{jk}^n = \sum_{\substack{r=1 \\ r \neq k}} P_{jr}^{n-1} P_{rk} \qquad j = 1, \ldots, k, \ldots, m \qquad (4)$$

$$P_{jr}^o = \delta_{jr} \qquad P_{jr}^n = \sum_{\substack{i=1 \\ i \neq k}}^{m} P_{ji}^{n-1} P_{ir} \qquad r = 1, \ldots, k, \ldots, m \qquad (5)$$

It may seem that it would be quite a task to perform the indicated calculations of Eqs. 3, 4, and 5. However, by defining a quantity called the mean occurrence $O_{jr}$ of a state $s_r$ for the system starting in state $s_j$, the calculations can be performed by inspection of the statistical graph.

$$O_{jr} = \sum_{n=0}^{\infty} p_{jr}^n \tag{6}$$

Substituting Eqs. 4 and 5 in Eqs. 3 and 6 and changing the order of summation produces the following set of equations.

$$P_{jk} = \sum_{\substack{r=1 \\ r \neq k}}^{m} O_{jr} P_{rk} \qquad j = 1, \ldots, k, \ldots, m \tag{7}$$

$$O_{jr} = \delta_{jr} + \sum_{\substack{i=1 \\ i \neq k}}^{m} O_{ji} P_{ir} \qquad r = 1, \ldots, k, \ldots, m \tag{8}$$

The solution of these equations for $P_{jk}$ gives the desired result. This solution can be found by solving a flow graph representing the equations. Moreover, the representative flow graph corresponds exactly to the modified statistical graph of the system where the initial state $s_j$ is driven with a unit source. That is, the variable $O_{jr}$ is associated with the state node $s_r$ and $P_{jk}$ is associated with the event state $s_k$. The statistical graph is then the flow graph of Eqs. 7 and 8. Solution of the flow graph for $P_{jk}$ gives the probability of occurrence of state $s_k$, the system having started in state $s_j$.

2.2 An Example. As a particular example consider the statistical system shown in Fig. 74, where we are interested in finding the probability of occurrence of state $s_1$ for the system having started in state $s_2$, $P_{21}$, and the probability of the recurrence of state $s_1$ for the system starting in state $s_1$, $P_{11}$.

The modified statistical graph is Fig. 75, which is used to find $P_{21}$. Figure 78 is used to find $P_{11}$. Note how the variables of Eqs. 7 and 8 are associated with the state
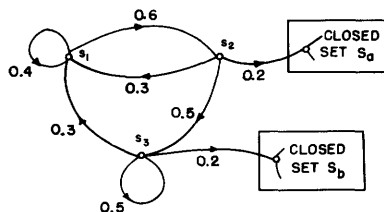


Fig. 74. A statistical system.
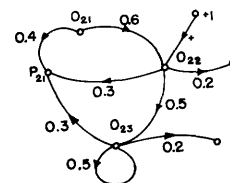


Fig. 75. Probability of occurrence of state $s_1$ from state $s_2$.
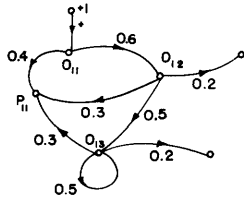
Fig. 76. Probability of occurrence
of state $s_1$ from state $s_1$.
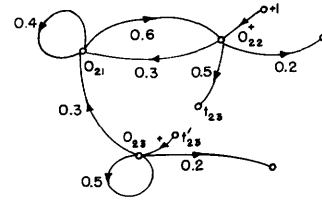


Fig. 77. Probability of transition
occurrence $t_{23}$.

nodes of the modified graphs of Figs. 75 and 76. From these graphs the following
solution for $P_{21}$ and $P_{11}$ is obtained:

$$P_{21} = 0.3 + \frac{(0.3)(0.5)}{1 - 0.5} = 0.6$$

$$P_{11} = 0.4 + 0.6 \left[ 0.3 + \frac{(0.3)(0.5)}{1 - 0.5} \right] = 0.76$$

2.3 Transitional Events. If the event to be observed is a particular transition, an
extra state can be introduced in the nose end of the branch that represents the particular
transition. For example, we may want to know the probability of the first and repeated
occurrences of the transition $t_{23}$ from $s_2$ to $s_3$ in the statistical system shown in
Fig. 74. By introducing the extra state $t_{23}$ and then splitting it, Fig. 77 is obtained,
and from it we obtain the following results:

$$P(t_{23} | s_2) = \frac{0.5(1 - 0.4)}{1 - 0.4 - (0.3)(0.6)} = \frac{5}{7}$$

$$P(t_{23} | t'_{23}) = \frac{(0.3)(0.6)(0.5)}{(1 - 0.5)(1 - 0.4 - (0.3)(0.6))} = \frac{3}{7}$$

The mechanics of finding these probabilities is then to isolate the desired event by
splitting its state node, to drive the initial state with a unit source, and then to calcu-
late the value of the isolated state node with flow-graph techniques.

In order to see how the probability of a particular event varies as the starting point
of the system varies, we merely "probe" the starting points with a unit source and
observe the probability of the event directly from the graph.

Where the system has an initial start distribution $\{P_j\}$, each node $s_j$ should be driven
with the source $P_j$ instead of one node being driven with a unit source.

2.4 Probability of Entering a Closed Set. The probability of entering a closed set
from a transient system is the sum of the probabilities of the transitions into the closed
set, since these events are mutually exclusive. The probabilities of entering the closed

73

set $S_a$ or $S_b$ from state $s_2$ of Fig. 74 are calculated on the basis of the probability of occurrence of the transitions into the sets.

$$P(S_a \,|\, s_2) = \frac{0.2(1 - 0.4)(1 - 0.5)}{(1 - 0.5)(1 - 0.4 - (0.3)(0.6)) - (0.5)(0.3)(0.6)} = \frac{1}{2}$$

$$P(S_b \,|\, s_2) = \frac{(0.5)(0.2)(1 - 0.4)}{(1 - 0.5)(1 - 0.4 - (0.3)(0.6)) - (0.5)(0.3)(0.6)} = \frac{1}{2}$$

Of course, the system will eventually go into one of the closed states $S_a$ or $S_b$, so that the sum of the mutually exclusive events of entering the closed sets is unity, as is indicated by the numerical calculations.

If the transient system has only one transition leading from the transient states, the probability of the occurrence of the transition is unity. In the same manner, if there is only one closed set, the probability of getting into this set is unity. These facts are a consequence of the definition that the states considered are transient.

If the original graph was an irreducible recurrent graph instead of a transient graph, the event to be observed will be the only closed set in the modified graph. Thus, events in an irreducible recurrent graph, represented by states, have a probability of first occurrence equal to unity.

2.5 Probability of At Least, At Most, or Exactly N Times. In observing an event as a recurrent event it is interesting to ask for the three probabilities of the event happening "at least N times," "at most N times," or "exactly N times." Knowing the probabilities of the first and repeated occurrences of the events $P_{jk}$ and $P_{kk}$, it is then possible to make the desired calculations as follows:

$$P_{jk}(\text{at least } N) = P_{jk} P_{kk}^{N-1} \tag{9}$$

$$P_{jk}(\text{at most } N) = 1 - P_{jk} P_{kk}^{N} \tag{10}$$

$$P_{jk}(\text{exactly } N) = P_{jk} P_{kk}^{N-1}(1 - P_{kk}) \tag{11}$$

The second equation comes from the fact that a particular event will occur at most N times or at least N + 1 times with a probability of unity.

$$P_{jk}(\text{at most } N) + P_{jk}(\text{at least } N+1) = 1 \tag{12}$$

The third equation represents the situation in which the event occurs at least N times and then ceases to occur. The equation may also be obtained by recognizing the fact that "at least N" is made up of the mutually exclusive events of "exactly N" and "at least N+1."

$$P_{jk}(\text{at least } N) = P_{jk}(\text{exactly } N) + P_{jk}(\text{at least } N+1) \tag{13}$$

By putting Eqs. 12 and 13 together, the following relation becomes valid:

$$P_{jk}(\text{at most } N-1) + P_{jk}(\text{exactly } N) + P_{jk}(\text{at least } N+1) = 1 \tag{14}$$

2.6 Mean and Variance of Occurrence. With the probability for exactly $N$ occurrences of a particular event it is possible to define and calculate the mean occurrence and the variance of the occurrence, as in Eqs. 15 and 16.

$$M_{jk} = \sum_{n=0}^{\infty} n\, P_{jk}(\text{exactly } n) = \frac{P_{jk}}{1 - P_{kk}} \tag{15}$$

$$V_{jk} = \sum_{n=0}^{\infty} n^2 P_{jk}(\text{exactly } n) - M_{jk}^2 = \frac{P_{jk}(1 + P_{kk} - P_{jk})}{(1 - P_{kk})^2} \tag{16}$$

To the person who is familiar with flow-graph calculations it will be quite evident that the two definitions of mean occurrence given in Eqs. 6 and 15 are the same. A general proof of the identity is given in the next section.

A physical example is an oscillator which runs erratically and which can be approximated by a Markov process. Consider pulsing the oscillator to get it started and then finding the probability that it will run through at least 10 cycles before it quits. The material in this section shows how to calculate this probability, the probability that it will stop running before it is pulsed again, and the mean and variance of the number of cycles it goes through each time it is pulsed. More complex conditional events can be treated by suitably modifying the structure of the graph.

The important idea is that these probabilities are obtainable directly from the Markov graph by making a slight modification in its structure in order to observe the desired event.

## 3. TRANSIENT SYSTEMS

Statistical transients form an interesting field of study. There is interest in setting up transient experiments and investigating the characteristic probabilities and running durations. This section deals with the mean occurrence of a particular transient state and with the mean and variance of the duration of the transient system. The objective is to show how these properties can be calculated directly from the descriptive Markov graph by the use of flow-graph techniques.

3.1 Mean Occurrence. This section answers the question of how often a particular transient event will occur before the system moves off into a closed set. Essentially, we are interested in the mean occurrence of a particular state. The previous section on recurrent events gave one method of treating this problem by calculating the occurrence probabilities and computing the mean occurrence directly from the standard definition

$$O_{jk} = \sum_{n=1}^{\infty} n \cdot P_r\{s_k \text{ exactly } n\} \tag{17}$$

We shall now present another convenient method for finding the mean occurrence.

Consider an ensemble of experiments of the transient system. The common statistical characteristic function $\xi_r^n$ is used to indicate the success $\left(\xi_r^n = 1\right)$ or failure $\left(\xi_r^n = 0\right)$ of the $n^{th}$ move in the $r^{th}$ experiment. In each experiment the number of occurrences of the particular state $s_k$ is given by the sum

$$O_r = \sum_{n=0}^{\infty} \xi_r^n \tag{18}$$

This summation is essentially "counting" the number of occurrences. The mean occurrence is then

$$O_{jk} = \lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} O_r = \lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} \sum_{n=0}^{\infty} \xi_r^n \tag{19}$$

However, reversing the order of summations gives the result

$$O_{jk} = \sum_{n=0}^{\infty} \lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} \xi_r^n = \sum_{n=0}^{\infty} p_{jk}^n \tag{20}$$

where $p_{jk}^n$ is the transitional probability from $s_j$ to $s_k$ in exactly $n$ moves.

Flow graphs provide a convenient method of computing this infinite sum directly from the Markov graph. By introducing the recurrence relation

$$p_{jk}^{o} = \delta_{jk} \qquad p_{jk}^{n} = \sum_{r=1}^{m} p_{jr}^{n-1} \, p_{rk} \qquad (21)$$

into the derived definition of mean occurrence and changing the order of summation we obtain the relation

$$O_{jk} = \delta_{jk} + \sum_{r=1}^{m} O_{jr} \, p_{rk} \qquad (22)$$

This equation represents a set of equations for $k = 1, 2, \ldots, m$ corresponding to the transient states. In flow-graph notation the set of equations corresponds exactly to the transient part of the Markov diagram where the variable $O_{jk}$ is associated with the state $s_k$ and the starting state $s_j$ is driven with a unit source. Solution of the graph for the variable $O_{jk}$ with flow-graph techniques then produces the mean occurrence of the state $s_k$ for the system having started in the state $s_j$.

3.2 An Example of Mean Occurrence. As an example consider the Markov system shown in Fig. 78a, where the transient starts in state $s_2$. The flow graph for computing the mean occurrence of the various states is then Fig. 78b, where the variable $O_{jk}$ has been associated with the state $s_k$. From this graph the following quantities are computed:

$$O_{21} = \frac{0.8}{1 - (0.4)(0.8) - (0.2)(0.8) - (0.5)(0.8)(0.8)} = 4$$

$$O_{22} = \frac{1}{1 - (0.4)(0.8) - (0.2)(0.8) - (0.5)(0.8)(0.8)} = 5$$

$$O_{23} = \frac{0.2 + (0.5)(0.8)}{1 - (0.4)(0.8) - (0.2)(0.8) - (0.5)(0.8)(0.8)} = 3$$

Usually the original Markov diagram is used as the flow graph by mentally driving the initial state with a unit source and considering only the transient part of the graph.
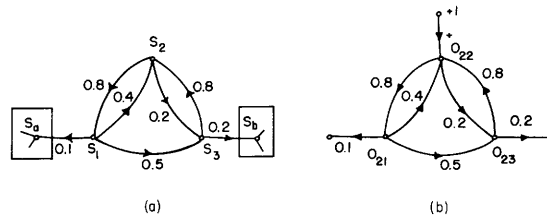


Fig. 78. (a) A Markov system. (b) Flow graph of Fig. 77a for computing mean occurrence.

3.3 Mean Occurrence to a Particular Closed Set. A conditional type of mean occurrence comes about by asking for the mean occurrence of a state $s_k$ when the transient ends in a particular closed set. Here we define one closed set $(s_a)$ as being a first entry state from the transient states. We are interested in the mean occurrence to the state $s_k$ on the assumption that when the system moves out of the transient states it moves directly to state $s_a$.

This situation physically corresponds to an experimental setup. The experimenter can observe that the system has been running only when it moves into a particular closed set where it stops. The experimenter is interested in finding the average number of occurrences of an event when he observes that the system stops.

By starting again with an ensemble of experiments, $r = 1, 2, \ldots$, and observing the conditional event on the $n^{th}$ move with a conditional characteristic function $\xi_{ra}^n$, the number of occurrences can be summed to give

$$O_{ra} = \sum_{n=0}^{\infty} \xi_{ra}^n \tag{23}$$

By taking the average over the ensemble and interchanging the order of summation the following result is obtained:

$$O_{jk_a} = \sum_{n=0}^{\infty} \lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} \xi_{ra}^n = \sum_{n=0}^{\infty} p_{jk}^n P_{ka} = O_{jk} \cdot P_{ka} \tag{24}$$

where $p_{jk}^n$ is the $n^{th}$-step transitional probability and $P_{ka}$ is the probability of the transient system terminating in the closed state $s_a$. Thus, the conditional mean occurrence has a simple relation to the actual mean occurrence and the probability of the closed state for the system starting at the observed event.

For the transient conditionally terminating in a set of states $\{s_a, \ldots, s_q\}$, the conditional mean occurrence is given by a summation over these states.

$$O_{jk_{a \ldots q}} = \sum_{\zeta=a}^{q} O_{jk} \cdot P_{k\zeta} \tag{25}$$

Summing over all first-entry closed states produces the obvious relation

$$O_{jk_{\text{all states}}} = O_{jk} \tag{26}$$

As an example of these conditional mean occurrences consider the system shown in Fig. 78a, where we are interested in the condition that the transient system stops in the closed set $S_a$.

By using the methods of a preceding section the closed-set probabilities are

$$P_{1s_a} = \frac{0.084}{0.2} = 0.42$$

$$P_{2s_a} = \frac{0.08}{0.2} = 0.40$$

$$P_{3s_a} = \frac{0.064}{0.2} = 0.32$$

Thus, the conditional mean occurrences of the states for the system starting in state $s_2$ and ending in the closed set $S_a$ are

$$O_{21_a} = O_{21} \cdot P_{1s_a} = 1.68$$

$$O_{22_a} = O_{22} \cdot P_{2s_a} = 2.00$$

$$O_{23_a} = O_{23} \cdot P_{3s_a} = 0.96$$

3.4 Unsolved Problem. The direct calculation of the variance of occurrence is an unsolved problem. In a previous section we showed how to calculate the variance of occurrence by considering the probabilities. Another approach through the characteristic function leads to the summation over the correlation between the transitional probabilities.

$$V_{jk} = \lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} O_r^2 - O_{jk}^2 \tag{27}$$

$$= \sum_{\substack{m=0 \\ n=0}}^{\infty} \lim_{R \to \infty} \sum_{r=1}^{R} \xi_r^m \cdot \xi_r^n - O_{jk}^2 \tag{28}$$

$$= \sum_{\substack{m=0 \\ n=0}}^{\infty} \phi(m,n) - O_{jk}^2 \tag{29}$$

This correlation function is simply the joint probability of a particular transition on the $m^{th}$ and $n^{th}$ step.

$$\phi(m,n) = \begin{cases} P_{jk}^m \, P_{kk}^{n-m} & n > m \\ \\ P_{jk}^m & n = m \end{cases} \tag{30}$$

As yet, no simple method exists for computing this double infinite sum directly from the graph. Possibly you can find a way of doing it. The simple expressions found in a preceding section would lead us to suspect that it can be done.

3.5 Transitional Events. The mean number of occurrences of a particular event is a direct consequence of the system and can be computed directly from the graph. In order to compute the mean occurrence of a particular transition, a state is introduced into the latter part of the transition branch. The mean occurrence of this new state is the desired result. A slightly easier procedure consists of calculating the mean occurrence of the state at the initial end of the branch and multiplying it by the probability of the transition.

In the preceding example, Fig. 78a, the mean occurrence of the transition between $s_1$ and $s_3$ would be

$$O_{2t_{13}} = O_{21} \cdot p_{13} = 4 \cdot 0.5 = 2$$

3.6 Mean Duration. It takes a number of transitions for a system in a transient state to move into any one of a number of closed sets of states. The average number of moves involved in a transient system is generally known as the mean duration of the transient system. Since it takes one move to get out of a transient state, the mean duration (average number of moves) is equal to the sum of the mean occurrences of the transient states.

$$D_j = \sum_{k=1}^{m} O_{jk} \tag{31}$$

where $D_j$ is the mean duration of a system starting in state $s_j$.

This result can also be obtained from the characteristic function, $\xi_r^n$, described in the preceding section or from the following consideration, which parallels Feller's derivation (18). For the system starting in state $s_j$, let $P_j^n$ be the probability that after n transitions the system is still in one of the transient states. Because of the mutually exclusive nature of the transient states, $P_j^n$ satisfies the relation

$$P_j^n = \sum_{k=1}^{m} p_{jk}^n \tag{32}$$

where $p_{jk}^n$ is the $n^{th}$-step transitional probability, and the summation is over all of the transient states.

Using the standard definition, the mean duration is defined as

$$D_j = \sum_{n=1}^{\infty} n \, Pr\{d_j = n\} \tag{33}$$

where the probability that the duration will be n moves is

$$Pr\{d_j = n\} = P_j^{n-1} - P_j^n \tag{34}$$

80

The mean duration is then

$$D_j = \sum_{n=1}^{\infty} n\left(P_j^{n-1} - P_j^n\right) = \sum_{n=0}^{\infty} P_j^n = \sum_{n=0}^{\infty} \sum_{k=1}^{m} p_{jk}^n = \sum_{k=1}^{m} O_{jk} \tag{35}$$

where

$$O_{jk} = \sum_{n=0}^{\infty} p_{jk}^n \tag{36}$$

In the example worked out in the previous section on mean occurrences we find that the mean duration of the system starting at state $s_2$ is 12 moves, which is the sum of the mean occurrences of the transient states.

The mean duration can also be thought of as the average amount of time spent in the transient states. Thus, we have a means of weighting the duration calculation for the systems in which the time interval spent in the various states between transitions is not uniform over the states. This produces a mean duration with the dimensions of time.

$$D_j = \sum_{k=1}^{m} W_k O_{jk} \tag{37}$$

where $W_k$ is the time spent by state $s_k$.

3.7 Mean Duration to a Particular Closed Set. In certain situations it is possible to observe that a transient system has been in operation only when it terminates in a particular closed set. Thus, it is of interest to determine the mean duration of a transient system on the condition that the transient terminates in a particular closed set.

By paralleling the reasoning for the mean duration that it takes one move to get out of a transient state, the following can immediately be set down.

$$D_{ja} = \sum_{k=1}^{m} O_{jk_a} = \sum_{k=1}^{m} O_{jk} P_{ka} \tag{38}$$

where the summation is over all the transient states. A more convincing (and longer) derivation is the following one.

The conditional mean duration is defined as

$$D_{ja} = \sum_{n=1}^{\infty} n\, P_r\{d_{ja} = n\} \tag{39}$$

where the probability of the conditional duration is given by the expression

$$Pr\{d_{ja} = n\} = \sum_{k=1}^{m} p_{jk}^{n-1} \, p_{ka} \qquad (40)$$

It has been assumed that the transient system starts in state $s_j$ and ends in the closed state $s_a$.

If instead of one state $s_a$, we allow the system to terminate conditionally in a number of states $s_a, \ldots, s_q$, the conditional duration of the set is the sum of the conditional durations of the individual states

$$D_{j(a\ldots q)} = \sum_{\zeta=a}^{q} D_{j\zeta} \qquad (41)$$

This is a result of the fact that the states $s_a, \ldots, s_q$ are mutually exclusive, so that the probability of duration is given by the expression

$$Pr\{d_{j(a\ldots q)} = n\} = \sum_{\zeta=a}^{q} \sum_{k=1}^{m} p_{jk}^{n-1} \, p_{k\zeta} \qquad (42)$$

By inserting Eq. 40 into Eq. 39 we obtain the flow-graph equation

$$D_{ja} = \sum_{k=1}^{m} D_{jk} \, p_{ka} \qquad (43)$$

where

$$D_{jk} = \sum_{n=1}^{\infty} n \cdot p_{jk}^{n-1} \qquad (44)$$

The quantity $D_{jk}$ is given the name of mean duration of the state $s_k$ for the system starting in state $s_j$. This quantity will be encountered in subsequent sections.

By using the transitional probability recurrence relation, Eq. 21, another set of flow-graph equations is obtained:

$$D_{jk} = O_{jk} + \sum_{r=1}^{m} D_{jr} \, p_{rk} \qquad (k = 1, \ldots, m) \qquad (45)$$

The equations indicated in Eqs. 43 and 45 represent a flow graph corresponding exactly to the transient part of the Markov graph except that each node is driven by a source equal to the mean occurrence of that node.

Another derivation shows that the quantity

$$\sum_{r=1}^{m} O_{jr} O_{rk}$$

satisfies Eq. 45. Since this quantity and the conditional duration are unique, we have the relation

$$D_{jk} = \sum_{r=1}^{m} O_{jr} O_{rk} \tag{46}$$

or

$$D_{ja} = \sum_{r=1}^{m} \sum_{k=1}^{m} O_{jr} O_{rk} P_{ka} \tag{47}$$

$$D_{ja} = \sum_{r=1}^{m} O_{jr} P_{ra} \tag{48}$$

where $P_{ra}$ is the probability of the closed set $s_a$ for the system starting at state $s_r$.

Equation 38 is interesting because it leads us into the physical interpretation that the conditional duration is equal to the sum over the transient states of the mean occurrences of the states times the probability of getting to the particular closed set $s_a$. Summing Eq. 38 over all "a" then leads back to the relation

$$D_j = \sum_{k=1}^{m} O_{jk} \tag{49}$$

3.8 An Example of Mean Duration to a Closed Set. Consider as an example the Markov diagram of Fig. 79. The flow graph corresponding to the conditional mean duration to the closed set $s_a$ is given in Fig. 80. From this graph we have

$$D_{1a} = \frac{15}{11}$$

Subsequent calculation shows that

$$D_{1b} = \frac{25}{11} \quad \text{and} \quad D_1 = \frac{40}{11}$$

Thus we have the consistent result that

$$D_{1a} + D_{1b} = D_1$$

3.9 Possible Graph Modifications. In trying to find ways of calculating the conditional duration we might suppose that it could be done by modifying the transitional probabilities. A plausible modification is to have all the transitions either lead back into the
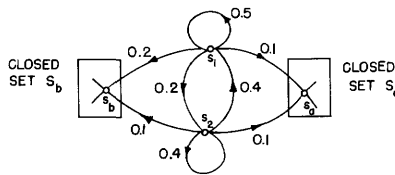
Fig. 79. Markov diagram.
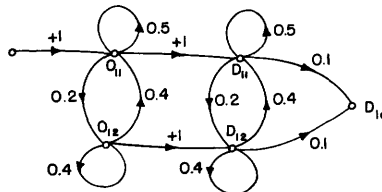


Fig. 80. Mean-duration solution to a particular closed set.

transient states or directly to the closed set $s_a$. This intuitive approach does not seem to be justifiable. Possibly you can find a means of making the modification.

The mean duration of a transient system is a direct result of the mean occurrence of all the transient states, while the conditional mean duration is the direct result of the mean occurrence of the transient states and the probability of the closed set from the various transient states.

3.10 Second Moment of the Duration. The second moment of the duration of a transient system represents a measure of the average spread in the duration averaged over a number of experiments. With the second moment of the duration, $D_j^2$, it is then possible to calculate the variance $\sigma^2$ and standard deviation of the duration.

$$\sigma^2 = D_j^2 - [D_j]^2 \tag{50}$$

For a system starting at the state $s_j$ the second moment of the duration is defined as

$$D_j^2 = \sum_{n=1}^{\infty} n^2 \, \Pr\{d_j = n\} \tag{51}$$

where the probability of the duration is defined, as before, in terms of the probability that after $n$ moves the system is still in a transient state.

$$\Pr\{d_j = n\} = P_j^{n-1} - P_j^n \tag{52}$$

$$\Pr\{d_j = n\} = \sum_{k=1}^{m} \left( p_{jk}^{n-1} - p_{jk}^n \right) \tag{53}$$

84

With this definition and a change of summation order we have the relation for the second moment in terms of the mean occurrence and duration:

$$D_j^2 = \sum_{k=1}^{m} \left(2D_{jk} - O_{jk}\right) \tag{54}$$

An example is that of Fig. 79. The desired quantities are obtained from Fig. 80.

$$O_{12} = \frac{30}{11} \qquad\qquad D_{12} = \frac{100}{11}$$

$$O_{22} = \frac{10}{11} \qquad\qquad D_{22} = \frac{50}{11}$$

Thus, the second moment and standard deviation of the duration of the whole system are

$$D_j^2 = \frac{260}{11} \qquad\qquad \sigma_j = \frac{(1260)^{1/2}}{11} \approx 3.2$$

as compared to a mean of 3.6 moves.

3.11 Conditional Second Moment of the Duration. It is also possible to calculate a conditional second moment of the duration to a particular closed set $s_a$ as was done for the conditional mean duration. The conditional second moment is defined as

$$D_{ja} = \sum_{n=1}^{\infty} n^2 \, \Pr\{d_{ja} = n\} \tag{55}$$

where (as before) the conditional duration probability is defined as

$$\Pr\{d_{ja} = n\} = \sum_{k=1}^{m} p_{jk}^{n-1} \, p_{ka} \tag{56}$$

As with conditional mean duration, the second moment of the system terminating in a number of states satisfies the additive relation

$$D_{j(a\ldots q)}^2 = \sum_{\zeta=a}^{q} D_{j\zeta}^2 \tag{57}$$

The Markov graph represents a flow graph from which this conditional moment can be obtained. Using the definitions of Eqs. 55 and 56, we obtain the expression

$$D_{ja}^2 = \sum_{k=1}^{m} D_{jk} \, p_{ka} \tag{58}$$

where

$$D_{jk}^2 = \sum_{n=1}^{\infty} n^2 \, p_{jk}^{n-1} \tag{59}$$

Breaking this equation into parts and changing the order of summation produces the flow-graph equation

$$D_{jk}^2 = (2D_{jk} - O_{jk}) + \sum_{r=1}^{m} D_{jr}^2 \, p_{rk} \tag{60}$$

Drawing this equation out in flow-graph notation produces a graph corresponding to the original Markov diagram, in which each state is driven with a source equal to the second moment of that state.

It can be shown that the conditional second moment is given by the relation

$$D_{ja}^2 = \sum_{r=1}^{m} (2D_{jr} - O_{jr}) \, P_{ra} \tag{61}$$

where $P_{ra}$ is the probability of the closed set $s_a$ from the state $s_r$.

Thus you see the parallelism between finding the conditional second moment and conditional mean duration.

In the example used in Fig 79, it was found that

$$2D_{11} - O_{11} = \frac{170}{11}$$

$$2D_{12} - O_{12} = \frac{90}{11}$$

The probabilities of the closed state $s_a$ from the transient states are

$$P_{1a} = \frac{5}{11}$$

$$P_{2a} = \frac{7}{22}$$

Thus the conditional second moment and standard deviation are

$$D_{ja}^2 = \frac{1165}{121} \qquad \sigma_{ja} = \frac{(940)^{1/2}}{11} \approx 2.8$$

Calculation of the second moment of the duration by these flow-graph methods is limited to fairly simple graphs because of the necessity of a triple solution. However, many problems must be simplified to fairly simple graphs in order to understand the basis of the operation.

Transient systems are usually found in actual practice and are usually interesting to work with. This section has shown how the mean occurrence of a particular state, the mean duration of the system, and the second moment of the system can be calculated

from the Markov graph by making use of flow-graph techniques. The important idea is that the properties are a direct consequence of the representative graph from the standpoint of existence and numerical calculation.

# 4. STATIONARY STATE DISTRIBUTION

A recurrent Markov system is a system of recurrent states in which the process continues without limit. The original starting transient effects are lost in the limiting process, so that all that is left is a certain probability of finding the process in a given state. The set of these probabilities is known as the stationary state distribution.

There are four methods of considering the process in order to be able to calculate the distribution. Two of these methods result in identical calculations. In each of the four methods the flow graph used to perform the calculations corresponds to the Markov graph except for minor changes in structure. These methods are based on three well-known equations:

$$\sum_{j=1}^{m} P(s_j) p_{jk} - P(s_k) = 0 \tag{62}$$

$$\sum_{k=1}^{m} P(s_k) = 1 \tag{63}$$

$$\sum_{k=1}^{m} p_{jk} = 1 \tag{64}$$

where $P(s_k)$ is the stationary state probability and $p_{jk}$ is the one-step transitional probability from state $s_j$ to state $s_k$. The system is assumed to be made up of aperiodic recurrent states.

4.1 Four Methods for Calculating Stationary Distributions. In the first method, $m - 1$ equations of Eq. 62 are solved with the use of a flow graph in terms of one state probability $P(s_j)$ to obtain the set of relations

$$P(s_k) = O_{jk} P(s_j) \tag{65}$$

With the help of Eq. 63, the desired solution is obtained:

$$P(s_\ell) = \frac{O_{j\ell}}{\sum\limits_{k=1}^{m} O_{jk}} \tag{66}$$

The second method uses Eq. 63 and $m - 1$ equations of Eq. 62 in flow-graph form. The state probabilities are obtained directly from the flow graph.

The third method essentially calculates the diagonal cofactors $\Delta_k$ of the transition matrix minus the unit matrix and uses Eq. 67 to calculate the desired probabilities.

$$P(s_k) = \frac{\Delta_k}{\sum\limits_{k=1}^{m} \Delta_k} \tag{67}$$

Calculation of the cofactor $\Delta_k$ is performed directly from the Markov graph with flow-graph techniques.

The fourth method is based on the duration $D_k$ of the recurrent event of the state $s_k$. Since the stationary probability is really the average frequency of occurrence, it is related to the duration of recurrence by the relation

$$P(s_k) = \frac{1}{D_k} = \frac{O_{jk}}{D_j} \tag{68}$$

The variance of the duration then gives an indication of the transient fluctuations around the average occurrence of the particular state.

4.2 An Example of the Four Methods. As an example, consider finding the stationary state probabilities of the Markov system of Fig. 81a.

In using the first method, all branches converging on a single-state node $s_j$ are removed so as to represent the $m-1$ equations of Eq. 62 as in Fig. 81b. The equivalence relations $O_{jk}$ from $s_j$ to $s_k$ are then calculated for $k = 1, \ldots, m$. Stationary state probabilities result from the application of Eq. 66. In this example,

$$O_{11}, \quad O_{12}, \quad O_{13} : 1, \quad \frac{7}{8}, \quad \frac{33}{32}$$

Thus the stationary state probabilities are

$$P(s_1), \quad P(s_2), \quad P(s_3) : \frac{32}{93}, \quad \frac{28}{93}, \quad \frac{33}{93}$$

The second method uses the $m-1$ equations of Eq. 62 and Eq. 63 in flow-graph form. All branches converging on a single-state node $s_j$ are first removed, so that the Markov diagram represents Eq. 62. Equation 63 is represented by attaching a branch having a coefficient of $-1$ from each state node to state node $s_j$ and driving-state node $s_j$ by a unit source as is done in Fig. 81c. The value of each state node as determined
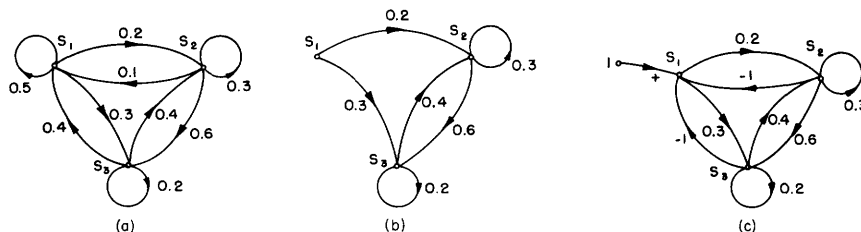


Fig. 81. (a) Markov diagram. (b) First method for calculating stationary distribution. (c) Second method for calculating stationary distribution.

by flow-graph manipulation is the state probability. From Fig. 80c the following calculations are made:

$$P(s_1), \quad P(s_2), \quad P(s_3) : \frac{32}{93}, \quad \frac{28}{93}, \quad \frac{33}{93}$$

The third method is based on properties of singular determinants having an eigenvalue of +1. Normally the solution of Eq. 62 is the cofactor of a row multiplied by an arbitrary constant of the matrix (P-I).

$$P(s_1), P(s_2), \ldots, P(s_m) : A_{1j}C, A_{2j}C, \ldots, A_{mj}C$$

However, in the calculation it is found that the row cofactors are equal to the diagonal cofactors. This is due to Eq. 64. The cofactor and minor are related by Eq. 69.

$$A_{ij} = (-1)^{i-j} M_{ij} \tag{69}$$

where $M_{ij}$ is the determinant (P-I) with the $i^{th}$ column and $j^{th}$ row removed. In the determinant (P-I), the $i^{th}$ row is reconstructed as the sum of all the rows in the minor $M_{ij}$. The new coefficient of the $i^{th}$ row and the $r^{th}$ column will then be

$$\sum_{\substack{k=1 \\ k \neq j}}^{m} (p_{rk} - \delta_{rk}) \qquad \begin{array}{l} r = 1, 2, \ldots, m \\ r \neq i \end{array}$$

which is equal to

$$-(p_{rj} - \delta_{rj})$$

Changing the sign of the $i^{th}$ row and permuting the $i^{th}$ row into the $j^{th}$ row produces the result

$$A_{ij} = (-1)^{i-j} M_{ij} = M_{ii} = A_{ii} \tag{70}$$

The cofactor $A_{ii}$ of the matrix (P-I) is then related to the disjoint determinant of the node $s_i$ by the relation

$$A_{ii} = (-1)^{m-1} \Delta_i \tag{71}$$

Subsequent use of Eq. 63 then gives the desired result of Eq. 67. From Fig. 81 the cofactors are determined.

$$\Delta_1, \quad \Delta_2, \quad \Delta_3 : 0.32, \quad 0.28, \quad 0.33$$

Applying Eq. 67 to these results then gives the state probabilities:

$$P(s_1), \quad P(s_2), \quad P(s_3) : \frac{32}{93}, \quad \frac{28}{93}, \quad \frac{33}{93}$$

In applying the fourth method we find it to be identical with the first method. The

duration of the recurrent event of the state $s_j$ is calculated by splitting the event state $s_j$ and calculating the mean occurrence of each state, $O_{jk}$, $k = 1, \ldots, m$. (See sec. 3.6.)

The state $s_j$ is considered as occurring once — not twice, as some may think. The sum of the mean occurrences represents the mean duration of the recurrent event, so that the probability of the state is given by the relation

$$P(s_j) = \frac{1}{\sum\limits_{k=1}^{m} O_{jk}} \tag{72}$$

which is the same relation that was used in the first method. The mean occurrence, $O_{jj}$, is unity. In the example

$$D_1 = 1 + \frac{7}{8} + \frac{33}{32} = \frac{93}{32}$$

so that

$$P(s_1) = \frac{1}{D_1} = \frac{32}{93}$$

A comparison of the four methods shows that each method has its advantages. The first and fourth methods simplify the diagram. The second requires only one calculation to obtain a single stationary probability. The third requires no modification of the diagram. Of course, each method has its disadvantages. The first, third, and fourth essentially require a solution for the whole set of probabilities before any one can be found. The second method is practical only if a great number of the states have a finite probability of moving to one state. It is hard to say that one method is better than the other without specifying the system under study. Each method has a particular field of application in which it works best.

4.3 Two-State Statistical Servosystem. As a final example of stationary distributions consider the two-state servosystem that has a unit delay in its reaction time. This is a system that statistically oscillates between its two states. By observing the state, the probabilities are biased so that the system tends to remain in permanent oscillation
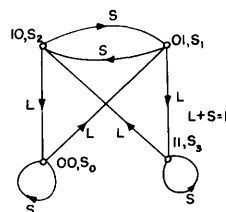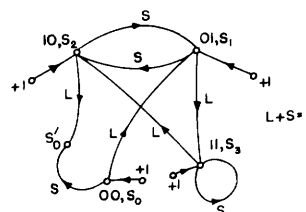


Fig. 82.  Two-state servosystem.



Fig. 83.  Second-moment calculation of the duration.

in contrast to remaining in a particular state. As for most systems, a finite reaction time is assumed. Here it is assumed to be the time of one move. The system is represented by Fig. 82, where the two states are zero and one. The number on the right indicates the present state of the system. The probabilities are biased either large, L, or small, S, depending upon the previous state.

If we use the third method, the cofactor determinants are

$$\Delta_o, \ \Delta_1, \ \Delta_2, \ \Delta_3 \ : \ L^2, \ L^2, \ L^2, \ L^2$$

so that the probability of each state is

$$P(s) = \frac{1}{4}$$

From this calculation it would seem that the system operates independently of how heavily the servosystem biases the probabilities. Use of the fourth method indicates that though the average is independent of the servo weighting, the second moment is a function of the weighting and indicates the best type of weighting to use.

Figure 83 is used to make the calculations of the mean occurrences.

$$O_{oo}, \ O_{o1}, \ O_{o2}, \ O_{o3} \ : \ 1, \ 1, \ 1, \ 1$$

The duration is then 4, which gives the state probability of 1/4. With these mean occurrences, the state durations are calculated from Fig. 83.

$$D_{oo} = 1$$

$$D_{o1} = 1 + \frac{1 + 2S}{L}$$

$$D_{o2} = 1 + \frac{3}{L}$$

$$D_{o3} = 1 + \frac{2 + 2S}{L}$$

Thus the second moment is

$$D_o^2 = 4\left(1 + \frac{3 + 2S}{L}\right) = 4\left(\frac{5}{L} - 1\right)$$

which has its minimum value at L = 1. Hence the most desirable situation is to have as much biasing as possible, even though it is a unit of time late in being applied.

# 5. SENSITIVITY AND VARIATION

The value of Markov systems stems from their correspondence to physical situations. In these physical situations there is always the question of the effect of changing certain standards and thus the corresponding probabilities of the system. For example, we may be interested in knowing how the probability of an unserved customer will change in relation to a change in the ordering and servicing system.

5.1 Bilinear Form. Having developed a method of solving for probabilities, mean occurrences, and mean durations with flow graphs, it is convenient to write the results in a standard form from which the variations can be observed. The standard form depends upon the manner in which the transitional probabilities change. In order to change one transitional probability from a state $s_k$, another must also be changed from the same state so that the divergent probabilities sum to unity.

For the special case of this second transitional probability going to a closed set, the calculations of the probabilities and mean occurrences simplify to a bilinear form.

$$P, O = \frac{\alpha p + \beta}{\gamma p + \delta} \tag{73}$$

The constants of Eq. 73 have simple interpretations in terms of loops and paths through the graph. Discussion of these interpretations will be found in Part I.

Since the duration is the sum of the mean occurrences and since the denominator of the mean occurrences will be the same, except in degenerate cases, the mean duration can be conveniently calculated in the form

$$D = \frac{\sum_{k=1}^{m} \alpha_k p + \beta_k}{\gamma p + \delta} \tag{74}$$

This bilinear form provides a convenient means for investigating the variations in the properties as one of the transitional probabilities changes.

A criterion of measure often applied to this type of problem is that of sensitivity, S, which is defined as the percentage incremental change in O by a percentage incremental change in p.

$$S = \frac{\frac{\Delta O}{O}}{\frac{\Delta p}{p}} \rightarrow \frac{p}{O} \frac{dO}{dp} \tag{75}$$

In terms of the bilinear constants

$$S = \frac{(\alpha \delta - \beta \gamma) p}{(\alpha p + \beta)(\gamma p + \delta)} \tag{76}$$
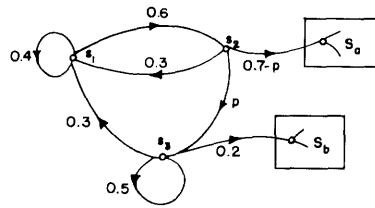
Fig. 84. Markov system.

5.2 Example. Consider the system in Fig. 84. This system could be a representation of a holding pattern for landing airplanes at two different runways, $S_a$ and $S_b$, or it could be a representation of a switching system hunting for either of two pieces of equipment, $S_a$ or $S_b$. Here we are interested in the variation and sensitivity of the transition probability $p$ in relation to the mean occurrence and duration. By assuming that the system starts in state $s_2$, the following calculations are made.

Mean Occurrences:

$$O_{21} = \frac{0.3p + 0.15}{-0.18p + 0.21}$$

$$O_{22} = \frac{0.3}{-0.18p + 0.21}$$

$$O_{23} = \frac{0.6p}{-0.18p + 0.21}$$

Mean Duration:

$$D_2 = \frac{0.9p + 0.45}{-0.18p + 0.21}$$

The sensitivity of the duration is then

$$S = \frac{0.27p}{(0.9p + 0.45)(-0.18p + 0.21)}$$

$$S\big|_{p=0.5} = \frac{5}{4}$$

5.3 Variation in General. In the more general situation the second probability returns to the transient system. Here, the expressions for the probabilities, the mean occurrences, and the mean duration are bilinear expressions of the change in the transitional probability, $\Delta p$.

$$P, O, D = \frac{\alpha \Delta p + \beta}{\gamma \Delta p + \delta} \tag{77}$$

This section has outlined an approach to calculating the variation and sensitivity of

Markov properties. These are interesting aspects of the theory because people are usually interested in making changes in a system in order to see what improvements can be obtained. Effect of a change can best be shown by calculating the properties in a bilinear form, an operation that is easy to perform with flow-graph techniques.

## 6. GENERATING FUNCTIONS

W. Feller (18) in his treatment of Markov systems and random walks introduces the generating function which enables him to solve a number of problems. This function is very much like the ones that are used to generate such functions as Legendre and Bessel functions. In this case, the function generates the $n^{th}$ step transitional probabilities.

R. W. Sittler (17) demonstrated that it was possible to obtain this function directly from the Markov graph by thinking of it as a pulsed-data system. A substantial part of his Sc.D. thesis was devoted to this subject. This section parallels Sittler's work in that it presents the basis for being able to calculate the generating functions and gives some of their uses.

**6.1 Definition of the Generating Function.** Consider forming a function $O_{jk}(x)$ as a power series of $x$, where the $n^{th}$ coefficient, $p_{jk}^n$, is the $n^{th}$-step transitional probability of moving from state $s_j$ to state $s_k$ in exactly $n$ transitions.

$$O_{jk}(x) = \sum_{n=0}^{\infty} p_{jk}^n \cdot x^n \tag{78}$$

Since the coefficients are bounded by unity, the series always converges for $x < 1$. Moreover, where $p_{jk}^n \to 0$, in the case of a transient state or a null state, the series converges for $x = 1$.

For transient states the following relations are valid:

Mean Occurrence:

$$O_{jk} = O_{jk}(x)\big|_{x=1} \tag{79}$$

Mean Duration:

$$D_{jk} = \frac{d}{dx} O_{jk}(x)\big|_{x=1} + O_{jk} \tag{80}$$

Second Moment of Duration:

$$D_{jk}^2 = \frac{d^2}{dx^2} O_{jk}(x)\big|_{x=1} + 3D_{jk} - 2O_{jk} \tag{81}$$

With these relations and the preceding material using these constants it is easy to see the versatility of the generating function.

**6.2 Flow-Graph Calculations.** For the general case, a convenient modification of the transitional probabilities of the Markov graph provides a means of calculating these generating functions directly from the graph with flow-graph techniques. By using the recurrence relation

$$p_{jk}^{o} = \delta_{jk} \qquad p_{jk}^{n} = \sum_{r=1}^{m} p_{jr}^{n-1} \, p_{rk} \qquad\qquad\qquad (82)$$

it becomes possible to derive the following set of equations for calculating the generating function.

$$O_{jk}(x) = \delta_{jk} + \sum_{r=1}^{m} O_{jr}(x) \cdot p_{rk} \cdot x \qquad\qquad k = 1, \ldots, m \qquad (83)$$

The flow graph that represents this set of equations corresponds to the Markov graph where each transitional probability has been multiplied by x, and the initial state is driven by a unit source. For example, consider the periodic Markov graph shown
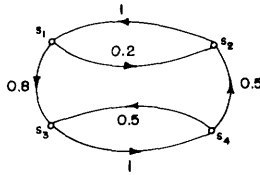


Fig. 85. Markov process.
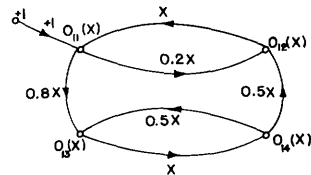


Fig. 86. Flow graph of the generating function $O_{1k}(x)$.

in Fig. 85. For the system starting in state $s_1$, we have the corresponding flow graph of Fig. 86.

Solution of the flow graph of Fig. 86 then produces the following generating functions:

$$O_{11}(x) = \frac{1 - 0.5x^2}{1 - 0.7x^2 - 0.3x^4} \qquad\qquad O_{12}(x) = \frac{0.2x(1 - 0.5x^2) + 0.4x^3}{1 - 0.7x^2 - 0.3x^4}$$

$$O_{13}(x) = \frac{0.8x}{1 - 0.7x^2 - 0.3x^4} \qquad\qquad O_{14}(x) = \frac{0.8x^2}{1 - 0.7x^2 - 0.3x^4}$$

6.3 Another Method of Modification. A variation in the above procedure for finding the generating function is provided by the equation of certainty.

$$\sum_{k=1}^{m} p_{jk}^{n} = 1 \qquad\qquad\qquad (84)$$

Multiplying by $x^n$ and summing over n produces the relation

$$\sum_{k=1}^{m} O_{jk}(x) = \frac{1}{1 - x} \qquad\qquad\qquad (85)$$

Fig. 87. (a) Sittler's example. (b) Sittler's formulation. (c) Another
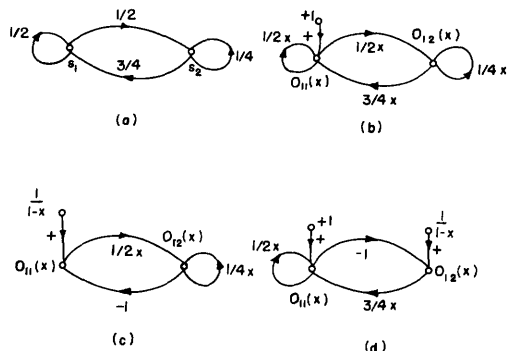formulation. (d) Another formulation.

In setting up the flow graph for calculating the generating functions it is then possible
to use the equation

$$O_{ji}(x) = \frac{1}{1-x} - \sum_{\substack{k=1 \\ k \neq i}}^{m} O_{jk}(x)$$ (86)

as one of the flow-graph equations.

Consider the example that Sittler (17) used in his thesis, our Fig. 87a. Three pos-
sible formulations are shown in Fig. 87b, c, and d. Figure 87b is the one used by
Sittler. These graphs produce the relations

$$O_{11}(x) = \frac{1 - 1/4x}{(1-x)(1 + 1/4x)}$$

$$O_{12}(x) = \frac{1/2x}{(1-x)(1 + 1/4x)}$$

Use of this variation somewhat simplifies the calculation when state $s_i$ is a central
return node and the process starts at state $s_j$. This is the result of branch values being
changed to $(-1)$ and the denominator of $O_{ir}(x)$ appearing partly factored.

6.4 Transitional Probabilities. The generating function represents a storehouse of
information. All of the transitional probabilities can be obtained directly from it.
Essentially there are three methods of obtaining these probabilities.

The first method is that of differentiation and evaluation at $x = 0$:

$$p_{jk}^{n} = \frac{1}{n!} \frac{d^n}{dx^n} O_{jk}(x)\Big|_{x=0}$$ (87)

In the second method, a long-hand division (in the right direction) is performed to

produce the power series of x. By means of division the following series for $O_{14}(x)$ of the example in Fig. 85 is obtained.

$$O_{14}(x) = 0.8x^2 + 0.56x^4 + 0.632x^6 + \ldots$$

Thus

$$p_{14}^n = 0 \qquad \text{n odd}$$

$$p_{14}^o = 0 \qquad\qquad p_{14}^4 = 0.56$$

$$p_{14}^2 = 0.8 \qquad\qquad p_{14}^6 = 0.632$$

Naturally, the even coefficients approach the stationary distribution probability $P(s_4) = 8/13$ as $n \to \infty$. The even nature of the function makes all of the odd coefficients zero, as might be expected from the periodic structure of the graph.

These first two methods are convenient for finding the low-order transitional probabilities, but they quickly become laborious for the higher-order transitions.

The third method produces a general functional form for the transitional probabilities $p_{jk}^n$. The procedure is to locate the roots of the denominator and then expand the generating function into a partial fraction expansion. Each term of the partial fraction expansion is then expanded into a power series, so that the transitional probability $p_{jk}^n$ is just the sum of the $n^{th}$ coefficients of the various terms:

$$O_{jk}(x) = k_o + \frac{k_1}{1 - a_1 x} + \ldots + \frac{k_q}{1 - a_q x} \tag{88}$$

$$\frac{k_i}{1 - a_i x} = k_i \left\{ 1 + a_i x + (a_i x)^2 + \ldots \right\} \tag{89}$$

$$p_{jk}^n = \sum_{i=0}^{q} k_i a_i^n \tag{90}$$

The expanded generating function $O_{14}(x)$ of Fig. 85 in terms of $x^2$ produces the relations

$$O_{14}(x) = \frac{8/13}{1 - x^2} + \frac{-8/13}{1 + 0.3x^2}$$

Thus

$$O_{14}(x) = \frac{8}{13}(1 + x^2 + x^4 + \ldots) - \frac{8}{13}(1 - 0.3x^2 + (0.3x^2)^2 + \ldots)$$

$$p_{14}^{2n} = \frac{8}{13}(1 - (-0.3)^n)$$

$$p_{14}^{o} = 0 \qquad p_{14}^{4} = 0.56$$

$$p_{14}^{2} = 0.8 \qquad p_{14}^{6} = 0.632 \qquad p_{14}^{2n} \rightarrow \frac{8}{13}$$

These relations correspond to those found by division.

6.5 Limit Calculations with Generating Functions. Quite often in more complicated problems it is not possible to determine the roots of the denominator of the generating function. This failure to determine all of the roots certainly prevents us from finding a general expression for $p_{jk}^{n}$. However, a desirable property of $p_{jk}^{n}$ is its behavior for large values of n. This behavior can be obtained by finding the roots that have the smallest magnitude $|x|$. Some of the roots may be complex. With them, the partial fraction terms can be found by finding the residues. Thus, a partial expansion is obtained. Since the other roots are larger, the coefficients that correspond to the larger roots in the series expansion die off faster and leave only the coefficients that correspond to the roots with the smallest magnitude.

For transitions to recurrent states, this smallest root will be $|x_1| = 1$. The next largest root then tells how fast the system approaches the stationary distribution. For transitions to transient states the root with the smallest magnitude will be greater than one, so that $p_{jk}^{n} \rightarrow 0$ as $n \rightarrow \infty$.

By the first two methods we are able to determine how the probabilities start off; by the third method we are able to determine the other extreme of the probabilities as n goes to infinity. If the roots can be obtained, a general expression for the nth-step probability can be obtained.

The order of the denominator of the generating function for a Markov process that consists of one set of recurrent states will be that of the number of states. In the general case the order is equal to or less than the number of states. This fact is the result of the correspondence between determinants of the flow graphs and the linear set of equations. As an example, consider the generating function associated with the systems in Figs. 85 and 87a, where the order is respectively 4 and 2.

6.6 Generating Functions of Recurrent Events. Generating functions provide a concise formulation of a wide set of properties. Three properties that are tied together by the generating function are the probability of first occurrence of the event state $s_k$, the mean duration to the event state $s_k$, and the second moment of the duration to the event state $s_k$.

The object of this section is to calculate the generating function of $P_{jk}^{n}$, the

probability of first occurrence of the state $s_k$ on the $n^{th}$ move for the system having started at state $s_j$.

$$P_{jk}(x) = \sum_{n=0}^{\infty} P_{jk}^n \cdot x^n \qquad (91)$$

With this generating function it is easy to verify that: The probability of first occurrence is

$$P_{jk} = P_{jk}(x)\Big|_{x=1} \qquad (92)$$

The mean duration to the first occurrence is

$$D_{jk} = \sum_{n=1}^{\infty} n \cdot P_{jk}^n = \frac{d}{dx} P_{jk}(x)\Big|_{x=1} \qquad (93)$$

And the second moment of the duration to the first occurrence is

$$D_{jk}^2 = \sum_{n=1}^{\infty} n^2 \cdot P_{jk}^n = \frac{d^2}{dx^2} P_{jk}(x)\Big|_{x=1} + D_{jk} \qquad (94)$$

By following the ideas of section 2.1 of calculating the transitional probabilities having no paths through the event state $s_k$, a set of equations involving the generating function $P_{jk}(x)$, similar to those of Eqs. 7 and 8, can be derived.

$$P_{jk}(x) = \sum_{\substack{r=1 \\ r \neq k}}^{m} O_{jr}(x) p_{rk} \cdot x \qquad (95)$$

$$O_{jr}(x) = \delta_{jr} + \sum_{\substack{i=1 \\ i \neq k}} O_{ji}(x) p_{ir} \cdot x \qquad (96)$$

Formulating these equations as flow graphs produces the original Markov graph with each transitional probability multiplied by $x$ and the event state $s_k$ split into two states of converging and diverging branches.

Thus, with flow-graph techniques and with a slight modification of the Markov graph, it is possible to calculate the generating function $P_{jk}(x)$ directly from the original graphical formulation.

6.7 An Example of Recurrent Events. As an example, consider calculating the generating function of the recurrent probabilities of the two-state servosystem described in section 4.3 and depicted in Figs. 82 and 83.

The recurrent generating function is given by the graph of Fig. 88, where the state
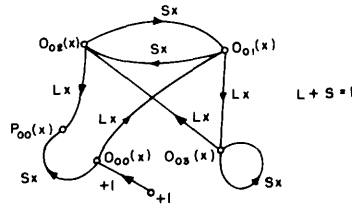
101

Fig. 88. Generating function of the two-state servosystem.

$s_0$ has been split and the transitional probabilities have been multiplied by x. From Fig. 88 we obtain the result

$$P_{00'}(x) = S_x + L^2 \frac{(L-S)x^4 + Sx^3}{S(S-L)x^3 - S^2x^2 - Sx + 1}$$

The probability of recurrence is

$$P_{00'} = P_{00'}(x)\big|_{x=1} = 1$$

which indicates that the recurrence is "certain." The mean duration is

$$D_{00'} = \frac{d}{dx} P_{00'}(x)\big|_{x=1} = 4$$

which agrees with previous calculations made in section 4.3. Also, the second moment of the duration is

$$D^2_{00'} = \frac{d^2}{dx^2} P_{00'}(x)\big|_{x=1} + D_{00'} = 4\left(\frac{5}{L} - 1\right)$$

6.8 Correlation. R. W. Sittler (17) assigns a measure to each state and then shows how to calculate a correlation function of the measure and how to obtain the power spectrum by using the generating function concepts. His ideas correspond to a system in which each state gives a particular magnitude output when the system is in that state, Of course, his ideas are also applicable to the situation in which each state is observed, as in the case of a number of binary outputs of different magnitudes.

More complex generating functions, based on the idea of assigning different varia- bles to the transitional probabilities, are possible.

6.9 A Direct Solution. The main reason for not always using a generating function in solving for the statistical properties of a Markov system is that we may not want the storehouse of information that the generating function contains. Sometimes all that we want is the answer to a certain problem. The quickest approach is usually a direct solution of the problem.

## 7. MUROGA'S PROBLEM

Flow graphs have an exceptional ability of making difficult systems problems quite easy to see and understand. Muroga's problem is an example.

Muroga's problem deals with finding the optimum input probabilities of a noisy discrete information channel. The problem has been formulated and solved, but in a form that is quite hard to understand and to teach. However, with flow graphs and a simple derivation, the mechanics of the solution become quite obvious. With the mechanics displayed in graphical form it is possible to see which examples will be easy to solve and which examples will not be easy to solve.

7.1 Information Channel. An information channel is a transformation from an input alphabet, $x$ ($x = 1, 2, \ldots, m$) to an output alphabet, $y$ ($y = 1, 2, \ldots, n$). In order to cut down the bulk of the notation, the associated probabilities are indicated by $x$ and $y$ instead of $P(x)$ and $P(y)$. Under stationary conditions there exists a matrix of transitional probabilities $p_{xy}$ from the input $x$ to the output $y$.

Figure 89 is an example of a channel. The input is on the left; the output is on the right. Branches indicate the transitional probabilities.

If we are given the input probabilities $\{x\}$, the output probabilities are given by the formula

$$y = \sum_{x=1}^{m} x \cdot p_{xy} \tag{97}$$

The graph of the channel is a flow-graph representation of Eq. 97. A solution of this equation is required in order to find the remaining probabilities when only a part of the input and output probabilities are given. In flow-graph notation, the solution is performed by first inverting the paths from unknown probabilities to known probabilities.

One method of inversion applicable for these channels consists of inverting the path of interest, changing its value to the reciprocal, and multiplying any incoming branches to the inverted path by $(-1)$. For example, the input probabilities in terms of the output probabilities of Fig. 89 are given by the flow graph of Fig. 90. From this figure

$$x_1 = \frac{1}{q_1} y_1$$

$$x_2 = -\frac{p_1}{q_1 q_2} y_1 + \frac{1}{q_2} y_2$$

$$x_3 = \frac{p_1 p_2}{q_1 q_2} y_1 - \frac{p_2}{q_2} y_2 + y_3$$

Of course such a simple solution will not, in general, guarantee that we started with
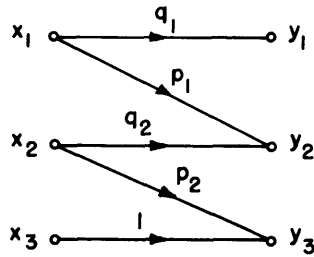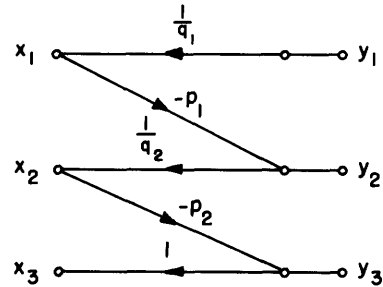
Fig. 89. Information channel.



Fig. 90. Inverted channel.

a valid set of probabilities. An arbitrary specification can lead to requiring that some of the probabilities be negative, as can be seen in the example above.

The only constraint that exists on the channel is the equation of certainty

$$\sum_{n=1}^{n} p_{xy} = 1 \tag{98}$$

This equation, along with Eq. 97, assures us that the equation of certainty is valid for x or y if it is valid for the other.

$$\sum_{y=1}^{n} y = \sum_{x=1}^{m} x = 1 \tag{99}$$

Thus if we constrain the summation of the output probabilities to be unity we know that the summation of the input probabilities is also constrained to equal unity.

7.2 Channel Capacity. Most people feel that there is a best way of doing almost anything. Thus it is interesting to investigate the best way to use the information channel in order to get the most information through it. This maximum is called the channel capacity, C. It may be obtained by a particular set of input probabilities.

The information between x and y is given by the equation

$$I(x;y) = \sum_{x=1}^{m} \sum_{y=1}^{n} x \cdot p_{xy} \log \frac{p_{xy}}{y} \tag{100}$$

It is our object to find a set of probabilities {x} such that I(x;y) is maximized. A somewhat more convenient form for Eq. 100 is

$$I(x;y) = \sum_{y=1}^{n} y \log \frac{1}{y} - \sum_{x=1}^{m} H_x \cdot x \tag{101}$$

where

104

$$H_x = \sum_{y=1}^{n} p_{xy} \log \frac{1}{p_{xy}} \qquad (102)$$

The set of quantities $\{H_x\}$ is the conditional entropy of the channel and is completely determined by the channel probabilities $p_{xy}$. The only constraint that exists is that the summation of the output probabilities equals unity.

$$\phi = \sum_{y=1}^{n} y - 1 = 0 \qquad (103)$$

7.3 Maximization. A convenient method of maximization is Legendre's method of undetermined multipliers. The function U is first constructed as

$$U = I(x;y) + \lambda_o \phi$$

$$= \sum_{y=1}^{n} y \left( \log \frac{1}{y} + \lambda_o \right) - \sum_{x=1}^{m} H_x \cdot x - \lambda_o \qquad (104)$$

where $\lambda_o$ is unspecified. The total derivative is then taken and set equal to zero.

$$dU = \sum_{y=1}^{n} dy \left\{ \log \frac{1}{y} + \lambda_o + \log \frac{1}{e} \right\} - \sum_{x=1}^{m} H_x \, dx$$

$$= \sum_{x=1}^{m} dx \left\{ \sum_{y=1}^{n} p_{xy} \log \frac{1}{\lambda y} - H_x \right\} \to 0 \qquad (105)$$

where

$$\lambda_o = \log \frac{e}{\lambda} \qquad (106)$$

Thus for an arbitrary change dx we must have the maximization relations

$$H_x = \sum_{y=1}^{n} p_{xy} \log \frac{1}{\lambda y} \qquad x = 1, \ldots, m \qquad (107)$$

in order to satisfy the relation

$$dU = 0 \qquad (108)$$

Use of the maximization relations, Eq. 107, in the original definition then produces the simple relation

$$C = I(x;y)_{max} = \log \lambda \qquad (109)$$

The fact that the second differential is negative shows that the solution actually is a maximum point.

$$d^2U = \sum_{y=1}^{n} -\frac{1}{y} \cdot \log e \, dy^2 \tag{110}$$

The solution for y and x are best obtained with flow-graph techniques. The method of starting with the set $\{y\}$ and obtaining the set $\{x\}$ was outlined in section 7.1.

The set $\{y\}$ can be obtained by solving the maximization relations, Eq. 107. These relations represent a flow graph; the inversion of the graph represents the solution.
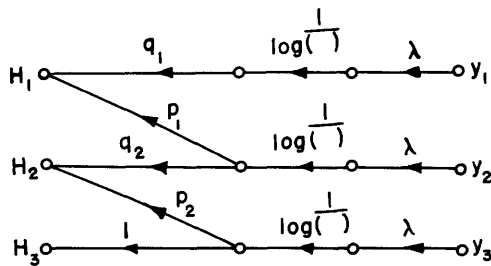


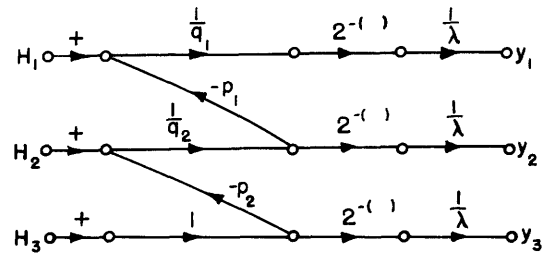Fig. 91. Maximization relations.



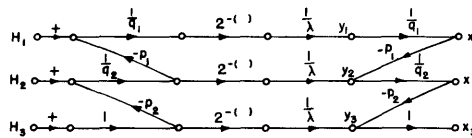Fig. 92. Solution of the maximization.



Fig. 93. Complete maximization solution.

The graph of the relations is similar to the channel flow graph except that all of the arrows are turned around. For example, the maximization relations for the channel of Fig. 89 are given by the flow graph of Fig. 91. The inversion of this graph is then the solution, Fig. 92.

Having the values of the y's we then solve another flow graph for the x's. By putting these two solutions together we obtain the flow graph of Fig. 93. As we can see from the flow graphs, the constant $\lambda$ is a scaling factor to get the sets $\{x\}$ and $\{y\}$ to sum to unity. For this special example we have the relations

$$H_1 = -\log p_1^{p_1} q_1^{q_1}$$

$$H_2 = -\log p_2^{p_2} q_2^{q_2}$$

$$H_3 = 0$$

106

By using these relations in the flow graph we obtain these results:

$$x_1 = \frac{1}{\lambda q_1}\, 2^{-\frac{H_1}{q_1} + \frac{p_1}{q_1}\frac{H_2}{q_2}}$$

$$= \frac{1}{\lambda}\left(\frac{p_1}{q_2\, p_2^{\,p_2/q_2}}\right)^{p_1/q_1}$$

$$x_2 = \frac{1}{\lambda q_2}\left\{2^{-\frac{H_2}{q_2}} - \frac{p_1}{q_1}\, 2^{-\frac{H_1}{q_1} + \frac{p_1}{q_1}\frac{H_2}{q_2}}\right\}$$

$$= \frac{1}{\lambda}\left(\frac{p_2/q_2}{p_2}\right)\left(1 - \left(\frac{p_1}{q_2\, p_2^{\,p_2/q_2}}\right)^{1/q_1}\right)$$

$$x_3 = \frac{1}{\lambda} - p_2 x_2$$

The use of the relation

$$\sum_{x=1}^{3} x = 1$$

gives a simple relation in $\lambda$ and thus the channel capacity.

7.4 Special Cases. The exceptions are usually difficult to understand and to treat. But they are often the most interesting problems. By having the mechanics of the solution displayed in flow-graph form it is possible to get a better understanding of the exceptions and an indication of how to treat them.

The next three sections treat the special cases of negative probabilities, unequal input and output alphabets, and singular solutions.

7.5 Negative Probabilities. Flow-graph reduction produces the exact mathematical solution of the formulated problem. It is unfortunate that there is lacking a method of requiring all the solutions to be positive — one of the two conditions sufficient for the numbers to be probabilities.

The only means of knowing that the solution represents a bona fide probability distribution is to calculate the distribution and see if each probability is positive. All of the output probabilities $\{y\}$ will be positive, as shown by the maximization relation, Eq. 107. The solution of the set of equations for $\log (1/\lambda y)$ may be positive or negative,

107

but the inverse of the logarithm is always positive.

In the previous example the input probability $x_2$ goes negative, where

$$p_1 > q_2 p_2^{p_2/q_2}$$

Plotting this inequality gives Fig. 94.

In order to correct the situation of negative input probabilities we must systematically eliminate members of the input alphabet, each time calculating the input probabilities and the channel capacity. The optimum is the distribution that gives the maximum channel capacity with a valid set of input probabilities. In the previous example it is easy to see that the input member to be eliminated is $x_2$.

7.6 Unequal Input and Output Alphabets. Another special case is the one in which the number of elements in the input and output are not equal. When the number of the input, m, is smaller than the number of the output, n, all of the necessary inversions are not possible. However, by performing the inversions that can be performed and representing the flow graph as in Fig. 95, the desired solution is obtained. The reduction of the flow graph will, of course, be more difficult because of the transcendental loop functions, but it will be exact and will indicate the solution procedure.

Eliminating $x_2$ (as is necessary when $x_2$ is negative) from the example given above and setting up the flow-graph solution produces Fig. 95. Reduction of this transcendental graph gives

$$x_1 = \frac{1}{\lambda} = \frac{1}{2}$$

$$x_3 = \frac{1}{\lambda} = \frac{1}{2}$$

This result could also have been obtained by the fact that the channel separates upon elimination of the element $x_2$.

The case in which there are more input symbols than output symbols indicates that it is not possible to satisfy all of the maximization relations, Eq. 107. However, if we solve the problem by ignoring the excess input symbols, the quantities

$$\sum_{y=1}^{n} p_{xy} \log \frac{1}{\lambda y} - H_x$$

for the excess input symbols may be negative, zero, or positive. If any of the quantities are positive, a new set of excess symbols must be chosen and the problem resolved. A valid solution exists when the excess quantities are all negative or zero. This valid set of excess symbols is then eliminated from the system.

In large complex systems, there is no end to the number of exceptions and counter-exceptions to these procedures. However, the flow-graph representation will show us

how to treat the simpler and more workable systems with an understanding far exceeding that obtained by straight mathematical analysis.

7.7 Singular Solutions. The third special case is the one in which, because of the singular nature of the determinant, the inversion does not exist. We must not ignore this situation. In the following example where the determinant is singular, the output is thoroughly confused as to what the input is doing. This would correspond to a case of zero channel capacity. Most engineers are looking for high capacity in their channels, but a few would like a total "blackout." For example, consider the reception of noise, and the problems of jamming and counter-measures.

An example of the singular determinant arises in the binary channel, Fig. 96a. The complete solution to the maximization problem is given in Fig. 96b. The nonsingular solution of this problem is

$$x_1 = \frac{q_2 - p_2 \, 2^{\eta_1 - \eta_2}}{(q_2 - p_1) \, 1 + 2^{\eta_1 - \eta_2}}$$

$$C = \log \left( 2^{-\eta_1} + 2^{-\eta_2} \right)$$

$$\eta_1 = \frac{q_2 H_1 - p_1 H_2}{q_2 - p_1}$$

$$\eta_1 - \eta_2 = \frac{H_1 - H_2}{q_2 - p_1}$$

These results correspond to those that R. A. Silverman (19) worked out.

On the other hand, the solution does not exist under the condition

$$1 - \frac{p_1 p_2}{q_1 q_2} = 0$$

which is equivalent to the relations

$$q_1 = p_2 \qquad q_2 = p_1$$

Using these relations in the original definition of information, we find

$$I(x;y) = 0$$

Thus if we are interested in shutting off a particular binary channel, we merely raise the noise until the determinant is singular.

Muroga's problem is an interesting example of what flow graphs can do to simplify a particular area so that further work can be performed. The advantage of the flow graph comes about from being able to see how the solution is constructed and thus being able to see where the exceptional cases are going to be and how to treat them.
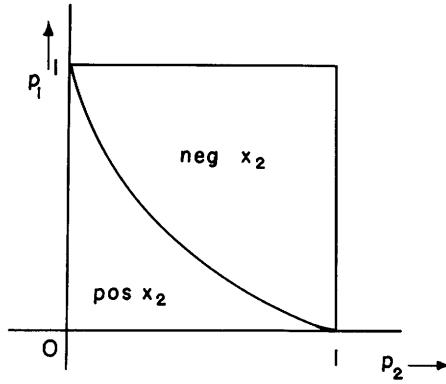
| $P_2$ | $q_2 P_2^{P_2/q_2}$ |
|-------|---------|
| 0 | 1.00 |
| 1/4 | 0.47 |
| 1/2 | 0.25 |
| 3/4 | 0.11 |
| 1 | 0.00 |

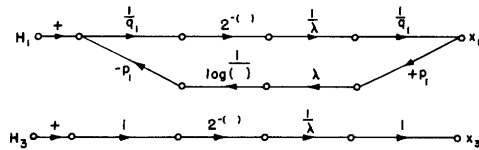Fig. 94. Negative-input probabilities.



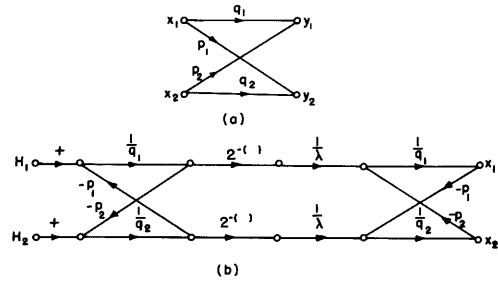Fig. 95. Elimination of the input $x_2$.



Fig. 96. (a) Binary channel. (b) Maximization of Fig. 95a.

The results obtained in the general solution of a particular channel are usually very complicated. It is almost imperative to work with simple structures in order to get results that can be interpreted numerically. By using flow graphs we can see in advance which structures are going to give the simpler results. Having the flow graph, we have the mathematical tools for solving the problem.

# 8. INFORMATION STRUCTURE

At each state of the Markov system the process must make a decision as to which state will be next. It is interesting to apply a measure of information to this type of decision-making in order to obtain an idea of how close the Markov system comes to describing a completely deterministic system.

For transient systems like the structure of languages, the total average information of the transient is calculated. In contrast, for recurrent systems (such as those that are found in continuous source generators), the average information per move is calculated.

8.1 Transient Information. For each transient state, $s_k$, of a Markov system we associate a measure of information $H_k$ that indicates the amount of decision that the system must make, once it moves into the state $s_k$.

$$H_k = \sum_{s=1}^{m} p_{ks} \log \frac{1}{p_{ks}} \tag{111}$$

For the $r^{th}$ experiment of this transient system it is convenient to use the characteristic function in order to keep track of how the system moves about.

$$\xi_{k_r}^{n} = \begin{cases} 1 & r^{th} \text{ experiment in state } s_k \\ & \text{on the } n^{th} \text{ move} \\ 0 \end{cases} \tag{112}$$

Then for any one experiment the total information is the sum over all the states and all the possible chances of being in these states.

$$I_r = \sum_{k=1}^{m} \sum_{n=0}^{\infty} \xi_{k_r}^{n} \cdot H_k \tag{113}$$

The average information over an ensemble of transient experiments would be

$$I = \lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} I_r \tag{114}$$

Since $H_k$ is positive for all k, we can invert the order of summation and obtain the simple result

$$I = \sum_{k=1}^{m} O_{jk} \cdot H_k \tag{115}$$

where $O_{jk}$ is the mean occurrence of the state $s_k$ for the system's having started in state $s_j$.

A preceding section showed how to calculate the mean occurrence directly from the Markov graph. It is also possible to calculate the information directly from the graph by inserting branches of value $H_k$ from the state node $s_k$ to a common node I. However, it is usually more efficient to use Eq. 115 rather than try to construct it in flow-graph form.

As an example, the system of Fig. 97 has these parameters:

$$H_1 = 0 \qquad\qquad O_{o1} = \frac{1}{p}$$

$$H_2 = p \log \frac{1}{p} + q \log \frac{1}{q} \qquad\qquad O_{o2} = \frac{1}{p}$$

$$I = \log \frac{1}{p} + \frac{q}{p} \log \frac{1}{q}$$

8.2 Normalized Transient Information. Usually it is possible to make the duration of the transient go to infinity faster than the information of the various states goes to zero. For some calculations it is convenient to normalize the information:

$$I' = \frac{\displaystyle\sum_{k=1}^{m} O_{jk} \cdot H_k}{\displaystyle\sum_{k=1}^{m} O_{jk}} \tag{116}$$

In the previous example the normalized information is

$$I' = \frac{1}{2}\left\{ p \log \frac{1}{p} + q \log \frac{1}{q} \right\}$$

8.3 Recurrent Information. Recurrent systems lose their identity with the initial starting state as the number of moves goes to infinity. Each state then obtains a certain probability of occurrence, so that the amount of information associated with the system is biased by the stationary probabilities.

In order to understand this reasoning consider the information associated with a recurrent system on the $n^{th}$ move in the $r^{th}$ experiment.

$$I_r^n = \sum_{k=1}^{m} \xi_{k_r}^n \cdot H_k \tag{117}$$

Averaging over the ensemble of experiments we obtain

$$I^n = \lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} I_r^n = \sum_{k=1}^{m} p_{jk}^n \cdot H_k \tag{118}$$

where $p_{jk}^n$ is the $n^{th}$-step transitional probability, $s_j$ to $s_k$. In the limit, the average

information per move is then

$$I = \lim_{n \to \infty} I^n = \sum_{k=1}^{m} P_k \cdot H_k \qquad (119)$$

where $P_k$ is the stationary probability of state $s_k$. This stationary probability is the same one that was calculated in the section on stationary distributions.

As a simple example, the following parameters of Fig. 98 lead to the stationary information:

$$P_1 = \frac{1}{3} \qquad\qquad H_1 = 1$$

$$P_2 = \frac{2}{3} \qquad\qquad H_2 = 2 - \frac{3}{4} \log 3$$

$$I = \frac{5}{3} - \frac{1}{2} \log 3$$

Note that this type of development would not have much meaning for a transient system, since $p_{jk}^n \underset{n \to \infty}{\to} 0$, so that in the transient case we have $I^n \underset{n \to \infty}{\to} 0$.
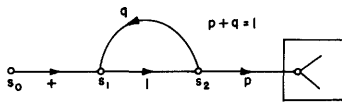


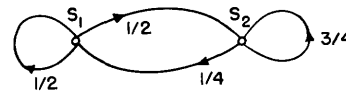Fig. 97. Simple information structure.



Fig. 98. Recurrent system.

The object of this section has been to show how a measure of information can be associated with a Markov system and then to show that it can be calculated by using an information entropy $H_k$ and the state properties. The state properties are those that can be calculated directly from the system diagram.

Acknowledgment

## References

1. C. S. Lorens, A new approach to reflection and image transfer loss and phase. L3 Coaxial System, Case 35552-124, Bell Telephone Laboratories, Inc., New York, Aug. 24, 1954.

2. C. S. Lorens, Applications of flow graphs to transistor circuits, Technical Information Series No. DF55ELP168, General Electric Company, Schenectady, N.Y., Dec. 20, 1955.

3. C. S. Lorens, A proof of the nonintersecting loop rule for the solution of linear equations by flow graphs, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Jan. 15, 1956, p. 97.

4. C. S. Lorens, Stationary state distribution of discrete Markov systems, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., April 15, 1956, p. 60.

5. C. S. Lorens, Basic theory of flow graphs — inversion, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., April 15, 1956, p. 64.

6. C. S. Lorens, Statistical application of flow graphs — mean occurrence of a Markov state, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., July 15, 1956, p. 32.

7. C. S. Lorens, Formulation of cascade circuits in terms of flow graphs, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., July 15, 1956, p. 35.

8. S. J. Mason, On the logic of feedback, Sc.D. Thesis, Department of Electrical Engineering, M.I.T., Jan. 6, 1951.

9. S. J. Mason, Feedback theory — some properties of signal flow graphs, Proc. IRE 41, 1144 (Sept. 1953); Technical Report 153, Research Laboratory of Electronics, M.I.T., Feb. 2, 1953.

10. S. J. Mason, Feedback theory — further properties of signal flow graphs, Proc. IRE 44, 920 (July 1956); Technical Report 303, Research Laboratory of Electronics, M.I.T., July 20, 1955.

11. C. E. Shannon, The theory and design of linear differential equation machines, OSRD Report 411, Jan. 1942.

12. A. Tustin, Direct Current Machines for Control Systems (Macmillan Company, New York, 1952).

13. J. G. Truxal, Automatic Feedback Control Systems Synthesis (McGraw-Hill Book Company, Inc., New York, 1955), Chap. 2, Signal-flow diagrams and feedback theory.

14. P. C. Watson, Sampled-data feedback systems and iteration procedures for simultaneous equations, S.M. Thesis, Department of Electrical Engineering, M.I.T., May 23, 1955.

15. F. B. Hildebrand, Methods of Applied Mathematics (Prentice-Hall, Inc., New York, 1952), p. 10, Determinants; p. 49, Discriminants and invariants.

16. H. Cramer, Mathematical Methods of Statistics (Princeton University Press, Princeton, N. J., 1946), p. 108, Determinants.

17. R. W. Sittler, Analysis and design of simple nonlinear noise filters, Sc.D. Thesis, Department of Electrical Engineering, M.I.T., 1955.

18. W. Feller, Probability Theory and Its Applications, Vol. I (John Wiley and Sons, Inc., New York, 1950).

19. R. A. Silverman, On binary channels and their cascades, Trans. IRE, vol. IT-1, no. 3 (Dec. 1955), p. 19; Technical Report 297, Research Laboratory of Electronics, M.I.T., May 25, 1955.

20. S. Muroga, On the capacity of a discrete channel I, J. Phys. Soc. Japan 8, 484-494 (July-August 1953).