THE SIGNIFICANCE OF SENSITIVITY IN FEEDBACK SYSTEM STUDIES

by

W. A. Lynch and J. G. Truxal

Research Report PIBMRI-866-60
Contract No. AF-18(603)-105

for
Office of Scientific Research
Washington 25, D. C.

June 1, 1961

MRI

POLYTECHNIC INSTITUTE OF BROOKLYN
MICROWAVE RESEARCH INSTITUTE

THE SIGNIFICANCE OF SENSITIVITY IN FEEDBACK SYSTEM STUDIES

by

W.A. Lynch and J.G. Truxal

Polytechnic Institute of Brooklyn
Microwave Research Institute
55 Johnson Street
Brooklyn 1, New York

Research Report PIBMRI-866-60
Contract No. AF-18(603)-105

June 1, 1961

Title Page
Acknowledgment
Abstract
Table of Contents
107 Pages of Text

W.A. Lynch
Professor

J.G. Truxal, Head
Electrical Engineering Dept.

Prepared for

Office of Scientific Research
Washington 25, D.C.

ACKNOWLEDGMENT

The work contained in this report was sponsored by the Office of Scientific Research under Contract No. AF-18(603)-105.

ABSTRACT

The sensitivity function defined originally by Bode and Mason is the fundamental measure of the effectiveness of feedback, in controlling the effects of both parameter variations and unwanted, noise signals. After a preliminary discussion of the significance of the sensitivity function, methods for the simple calculation of the sensitivity are presented (Section 5) -- methods valid whenever the overall transmission is a bilinear function of the varying parameter (thus, whenever the parameter is a circuit element or the transmittance of a controlled source).

The sensitivity function yields simple evaluations of the parameter margins (Section 6) -- the amount by which specific parameters must be varied to cause system instability. The parameter margins can be evaluated directly from the frequency-domain plots utilized in normal system design. This interpretation of the sensitivity function is of importance not only in the logical design for simultaneous control of several sensitivity functions, but also in the identification problem. Extensions to the study of nonlinear systems can be effected if a frequency-independent describing function can be defined.

TABLE OF CONTENTS

THE SIGNIFICANCE OF SENSITIVITY IN FEEDBACK SYSTEM STUDIES

In the analysis of physical systems, we persistently encounter examples which are conveniently described by considering them as feedback systems -- systems which, broadly speaking, are self-supervisory. In these systems, which are capable of performing their appointed tasks despite adverse influences, the output signal is compared with the desired value of the output in order to generate an "error" signal which can be used to change the output (usually in such a direction as to decrease the error toward zero). In this monograph, we attempt to summarize certain basic properties of feedback systems, and methods of analysis and design which are significant when engineers work with feedback. In short, in the following pages, we present the fundamental elements of feedback theory.

1. Definition of Feedback Systems

According to the definition given in the above paragraph, a feedback system includes at least the elements depicted in Figure 1-1, which also serves to introduce our notation. The output signal c is measured by the sensor to yield the feedback signal b. The comparator generates an error e which depends upon a comparison of the input or desired output r and the measured output b; in the simplest case, e is simply given by the difference

$$e = r - b \tag{1.1}$$

as depicted by the + and - signs in the figure more generally, the comparator might measure any function of r and b--e.g., $(r-b)^3$ --, but in this monograph we consider only comparators described by Equation (1.1) . Finally, the error signal e derives (or modifies) the output c through the process.

Illustrations of the basic system of Figure 1-1 are plentiful throughout our daily life. In a broad class of problems, the human being is an important element of the feedback system. For example, if a man reaches forward to pick up a pencil lying on a desk, a feedback system drives the grasping hand. In this case, the desired output (the desired position of the hand) is at the pencil; the actual output is the instantaneous position of the hand. The sensor is now the eye visualizing the present position of the hand; this position, b, is compared (in the brain) with the pencil position (also visualized by the eye), and an error signal actuates the muscles which move the hand (Figure 1-2).

The feedback aspect of the operation arises because of the continual

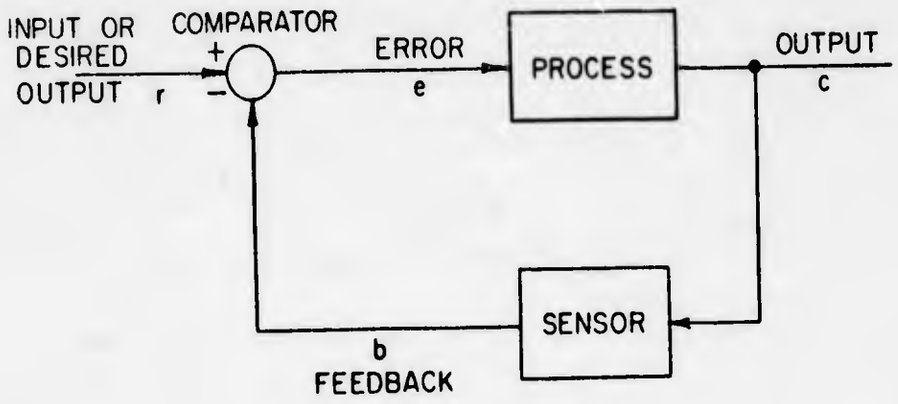


Figure 1-1 Simplest feedback system

measurement of the hand position and comparison with the measured pencil location; actually the eye functions to measure directly the error, which the brain then converts to the appropriate drive signal for the muscles. The improvement in performance which results from the feedback can be demonstrated if the same task is attempted with the eyes closed after only an initial measurement of the pencil position relative to the initial hand position.

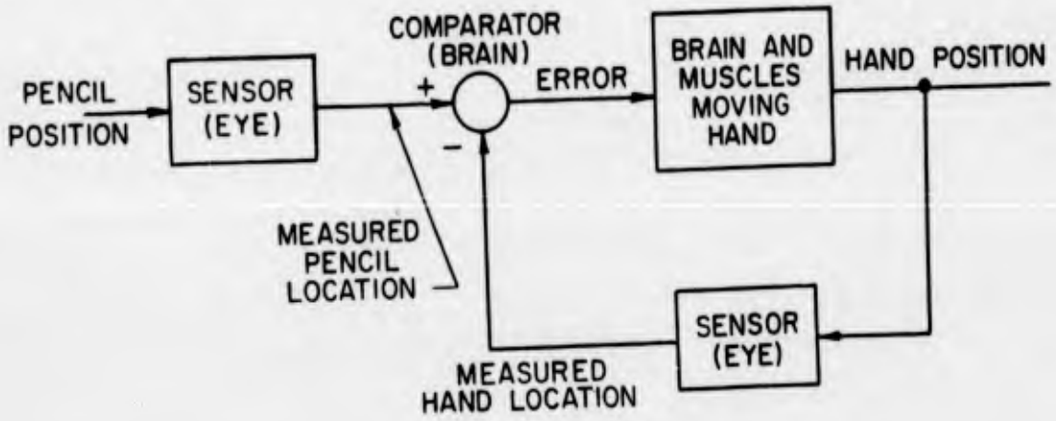


Figure 1-2 A human feedback system

Actually, the feedback concept we are discussing here has been important in engineering from the earliest times. The Romans during the days of the Empire used the same feedback control system we still employ today to maintain control over the water level in the reservoir of the common bathroom plumbing fixture. After the flushing is completed, the outlet valve is closed, and water rushes into the reservoir; as the water rises, a ball float senses the water level, closing the supply valve at the predetermined level. In this case, the block diagram is shown in Figure 1-3. The float position measures the actual water level; when the error exceeds a predetermined magnitude, the valve is open and water flows in to increase the actual water level. As soon as the error falls below the specified value, the valve is shut off, and the system remains in a steady condition. When the water level is abruptly lowered by a separate system, the cycle of operation starts once again.

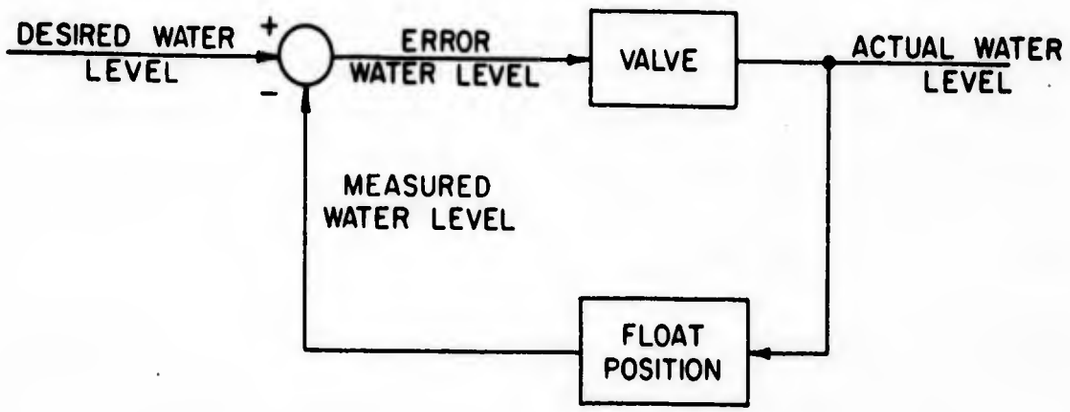


Figure 1-3 The plumbing feedback system

Both of the above systems demonstrate not only the phenomenon which we

have defined as feedback, but also several of the remarkable characteristics achievable with feedback. For example, each system possesses a reliability which is rarely surpassed in engineering design; in each system the performance is practically independent of system parameters. For example, in the case of the human pencil-grasper success is achieved by almost everyone, whether he be an Olympic weight-lifter, a ballet dancer, or a scholarly, bespectacled engineer. The plumbing fixture, similarly, operates almost regardless of the value of the water-supply pressure; the only common difficulty occurs when the auxiliary system emptying the water fails to shut off so that the water level is never able to rise. In view of the remarkable reliability of both systems, it is not surprising that they have been in use without modification for so many centuries.

In spite of the obvious merits of such feedback structures (and a primary objective of this monograph is to investigate analytically the quantitative reasons for such merits), it was not until very recently that engineers recognized the importance of feedback as a design tool. Feedback theory, as we know it today, was born in the 1920's with the work of the Bell Telephone Laboratories, and received its greatest impetus during the Second World War with the extensive work on automatic control systems for such military purposes as fire control (the positioning of gun turrets), radar antenna control (the positioning and measurement of the antenna), and aircraft control (for stabilization as well as navigation). During the last decade, major advances in feedback theory have been associated with missile guidance and control and industrial automation (for example, the automatic production line or the automatic adjustment of steam boiler operating conditions to maximize the economy of operation).

2. What Does Feedback Accomplish ?

Before we proceed further with the detailed analysis of various classes and types of feedback systems, it is useful to attempt to describe the properties of feedback systems in quantitative terms. In other words, we should pose the questions: quantitatively, what does feedback accomplish in systems similar to those described in Section 1? In what ways can we measure the effectiveness of the feedback?

In the vast majority of feedback systems which we encounter in engineering, the feedback is inserted in order to control the degree to which the overall transmission characteristics of the system depend upon the variation in the values of one or more of the system parameters. For example, in the plumbing system of

$$B = \beta C \qquad (2.2)$$

$$C = AE \qquad (2.3)$$

When the variables E and B are eliminated from these three equations (we shall see in Section 5 how to effect this elimination in a straightforward manner), the overall transmission is

$$\frac{C}{R} = \frac{A}{1 + \beta A} \qquad (2.4)$$

If β is selected (arbitrarily for the moment) as 99/1000, the realization of an overall gain of 10 requires that the quantity A be 1000; hence we have the feedback configuration shown in Figure 2.2.

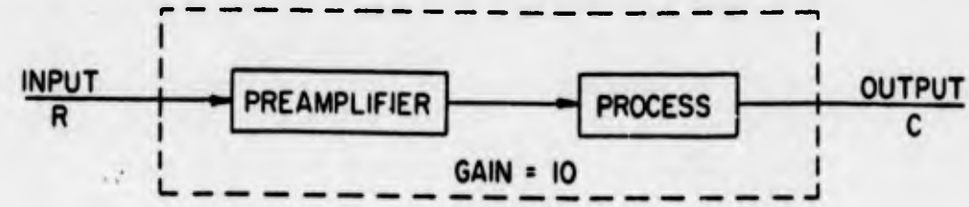
In order to realize our overall gain of 10, we now require an amplifier with a gain of 1000 (or three tandem amplifiers, each with a gain of 10). The introduction of feedback has obviously complicated our system design--not only is the number of required amplifiers raised from one to three, but we now need the sensor or β circuit and the comparator. What have we gained from this extra complexity?

The advantage of feedback is demonstrated if we consider now the effects of variation in the gain of the amplifiers. If the total amplifier gain of 1000 changes by 2 percent (e.g., if A drops to 980), the overall transmission of our complete system decreases to the value

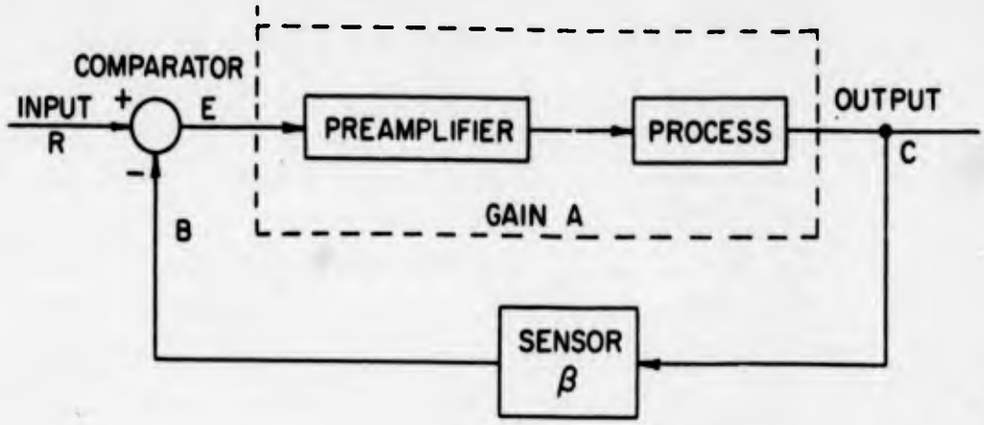
$$T = \frac{980}{1 + \frac{99}{1000} 980} = 9.998 \qquad (2.5)$$

As a result of the feedback, the overall gain drops only 0.02 percent, even though the gain of the amplifier falls 2 percent--we have achieved an improvement in performance of a factor of 100, in terms of making the system insensitive to the variations of system parameters.

It is natural to object that insertion of the feedback actually requires more amplifiers, hence the above improvement is deceptive. However, even if all three amplifiers change in the same direction by the full 2 percent in our example, the total amplifier gain changes by approximately 6 percent, and the gain of our overall system changes by only 0.06 percent--so that we still achieve a marked improvement in comparison to the system with no feedback.



(a) Open-Loop System



(b) Feedback system

Figure 2-1 Evaluation of feedback

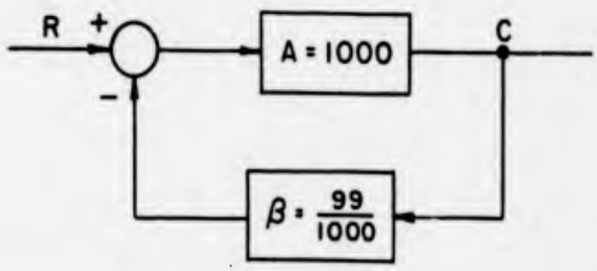


Figure 2-2 System for C/R = 10

3 Models

From the introductory ideas of Section 1, we have begun to see the potentialities of feedback configurations in the creation of systems that are relatively insensitive to the characteristics of the devices of which they are constituted. We are now ready to deal with feedback on a quantitative or analytical basis with a

view to establishing a theory for feedback systems that will be capable of accounting for all of the various aspects of system performance that are influenced by the presence of a feedback configuration. As an essential first step, we turn to the important question of how to describe a system analytically and to the corollary problem of how to develop the feedback point of view advantageously.

It is very easy to lose sight of the central fact of analysis--that we analyze an idealization or model of a system and not the system itself. Before we can embark upon any phase of analysis, we must somehow contrive to effect a transformation from the physical system to a corresponding (but not equivalent) description that is phrased mathematically and therefore is amenable to the methods of mathematical analysis. In addition to acepting the fact that our analytical representation of the system is a frank idealization involving various degrees of judicious approximation, we must accept the associated fact that the signals with which we drive or excite the analytical model must also be idealizations. Thus, the whole process of analysis is completely removed from the realm of reality and at the conclusion of our analyis we have the engineering task of using our results as a means for understanding the behavior of the physical system.

The very impressive achievements of present-day engineering. technology stem in no small measure from these three developments of analysis in which the model plays a dominant part:

(1) The recognition of a clear distinction between physical systems and signals and their corresponding models.
(2) The development of a wide variety of model concepts and associated symbolisms.
(3) The development of the several bodies of theory which guide our theoretical utilization of the various models.

In this section, in which we consider several different model concepts, we shall attempt to show that each has a sphere of applicability in which it may excel in a specific problem. We hope also to underline the suggestion that no single model can satisfy all of our analytical needs and therefore our best preparation for analysis and design is achieved when we have a working facility with all relevant models. The models that we shall consider can be placed in three categories: mathematical models, circuit models and flow diagrams.

(a) Mathematical models

Any set of mathematical relations that describe our analytical evaluation of a physical system constitutes a mathematical model. That such a model represents an interpretative concept is seen when, for example, we write the relation

$$e = ir \tag{3.1}$$

as a model of a physical resistor. In so doing, we have already decided that for our purposes we may neglect any residual effects which are capacitive or inductive. We are therefore idealizing the resistor in terms of a purely dissipative device. If we view r as invariant with the current, we are assuming a linear system and if we also regard r as invariant with time, we are assuming a linear, time-invariant system. Thus, in writing this very familiar expression, we make (although usually without conscious thought) many of the idealizations that are the ingredients of every model formulation. Because of the nearly ideal nature of electric circuit elements, we are actually on much firmer ground with the resistor model of Equation (3.1) than we are with the corresponding damper model of translation mechanical systems:

$$v = f(1/B) \tag{3.2}$$

where v and f are velocity and force and B is the damping constant.

(b) Circuit models

The familiar electric circuit model is a highly developed pictorial model which actually symbolizes a number of mathematical models. Its symbolism is such that each of the ideal elements of which it is constituted represents a physical law [such as that of Equation (3.1)] governing the terminal behavior in terms of voltage and current, and implying all of the previously mentioned aspects of idealization. It also symbolizes the energy-storage and dissipative properties of its elements. Therefore a configuration of ideal elements and ideal sources characterizing a model of an electric circuit represents sets of inter-relations that may be transcribed in terms of voltage or current equilibrium (i.e., loop or node equations), or alternatively in terms of energy or power equilibrium. Furthermore, we may transcribe a mathematical model which is descriptive of the transient or the steady-state behavior as desired, and all from the same circuit model.

Although the circuit model frequently serves as the basis for writing a set of equations describing an appropriate aspect of circuit behavior, it is sometimes more advantageous to transform the model itself into equivalent forms having a

simpler topology. Used in this way, the model retains its integrity and the reductions or simplification are accomplished without recourse to a mathematical version. Thus, it is useful to think in terms of an "algebra" of circuit models which consists in successive applications of network theorems (such as Thevenin's and Norton's theorems, the Substitution theorem, etc.,) and equivalent-circuit transformations, such as the ubiquitious tee-to-pi transformation. The ultimate exploitation of this kind of transformation brings the given circuit model to a state of simplification whereby the desired solution can be transcribed from the model by inspection.

(c) Flow diagrams

Quite apart from the familiar circuit models and their mathematical counterparts are the models based upon signal-flow concepts. These models which derive their chief analytical value from vivid pictorial properties, focus attention on a causal viewpoint expressing the analyst's visualization of various cause-and-effect relations with which the system can be described. The interjection of the cause-and-effect point of view forces the analyst to make certain decisions that he might not have to make in using a more conventional symbolism. For example, a cause-and-effect interpretation of Equation (3.1) could take the form of Figure 3-1. Here the current, i, is arbitrarily taken to be the cause, and the terminal voltage, e, the effect. The parameter, r takes the role of an operator and there is a clear progression of events from cause to effect, with the signal, i, being operated upon by the parameter, r, (here merely multiplication by r) with the emergent effect the terminal voltage, e.

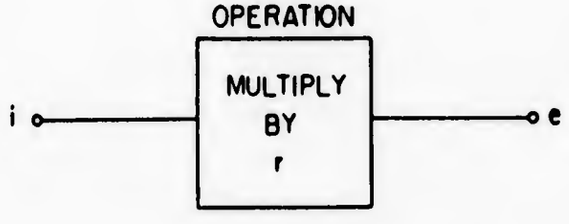


Figure 3-1 A cause-and-effect interpretation of Equation (3.1)

If we wish to portray the related expression

$$i = \frac{e}{r} \qquad (3.3)$$

a new flow diagram is required, because now the roles of cause and effect have become reversed. When formulating a circuit model of the resistor, we merely draw the familiar zig-zag symbol for the resistance element without taking heed of causal implications.

Any rebellion we may experience in having to cope with this seeming artificiality is more than offset when the benefits of causal representation are properly assessed. It is possible to visualize almost any situation in terms of cause-and-effect relations because this is a basic pattern of human experience. In many problems, such a view point produces no particular advantage over other representations; but in feedback systems, this way of portraying a system has tremendous advantages over mathematical and circuit model representations. The unifying motif of the systems with which this monograph is concerned is feedback. In a cause-and-effect portrayal, feedback is shown unmistakably as a closed causal chain which appears diagrammatically as a closed signal path or feedback loop.

The basic idea of showing cause-and-effect relations in terms of a graphical portrayal has provided the inspiration for several different, though related flow diagrams. The best known of these is the block diagram used in the illustrations of the first two sections. A more abstract interpretation of the flow diagram is the signal flow graph, which utilizes two very simple graphical elements instead of lines and blocks and summing elements to portray a set of simultaneous algebraic equations in a preconceived pattern of cause and effect. The third widely-used flow diagram is the analog simulation model, which in effect literally symbolizes a set of integro-differential equations in terms suitable for programing a physical analog computer.

(d) Block diagrams

We have already had some experience with block diagrams in Sections 1 and 2, and the illustrations of these sections are sufficient to show the symbolism. There are actually four graphical elements used: blocks, directed lines, summation devices (also referred to as comparators or differencing devices), and finally junctions or pick-off points. The block diagram has a great deal of appeal as a means for boxing off the principal component parts of a system, and indicating the way in which these main subdivisions are inter-related. As we have already seen,

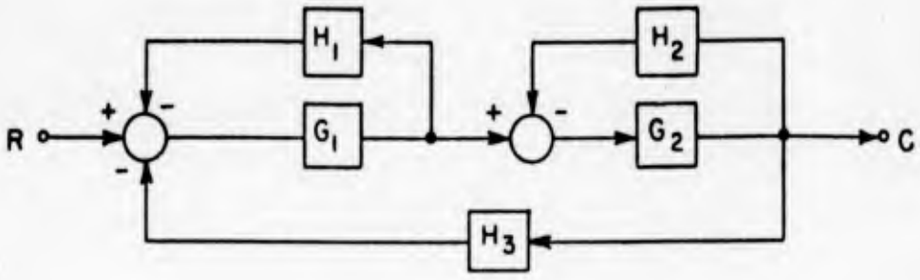
this kind of model lends itself to both qualitative and quantiative thinking. In Figure 1-2 illustrating a human feedback system, we were able to gain some insight into the human system and how it possibly functions without being able to describe the actions of the individual blocks quantitatively. Despite this qualitative description, various human endeavors, such as grasping an object, standing erect, driving an automobile, etc., are better understood in terms of a block-diagram representation.

When each of the component parts of a system is well enough understood to permit a model to be formulated, each block of a system representation can be labeled in terms of a pertinent mathematical model (a set of differential equations, or in a linear case, a transfer function). The block diagram is now a quantitative portrayal which still retains the features of segregating the non-interacting subdivisions of the system as well as retaining a strong visual appeal in terms of signal-flow paths and closed feedback loops.

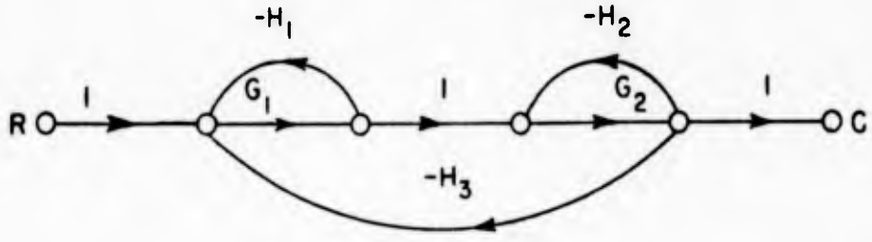
(e) Signal flow graphs

A signal flow graph might be thought of as a highly-stylized interpretation of a block diagram. Only two graphical elements are used: small circles, called nodes, and directed lines, called directed branches which interconnect the nodes. The branches replace the combination of directed lines and blocks of the block diagram, and the nodes replace both the summing devices and pick-off points. A block diagram is easily translated into an equivalent signal flow graph by transferring block labels to corresponding branches, replacing both the comparators and pick-off points by nodes, and finally introducing minus signs in appropriate branches when two signals are to be subtracted. In Figure 3-2, the details of two such translations are shown. It is noteworthy that the minor loops (involving $G_1 H_1$ and $G_2 H_2$) are non-interacting or non-touching in the models of parts (a) and (b) of the figure, whereas they are touching in the second example, [parts (c) and (d)] .

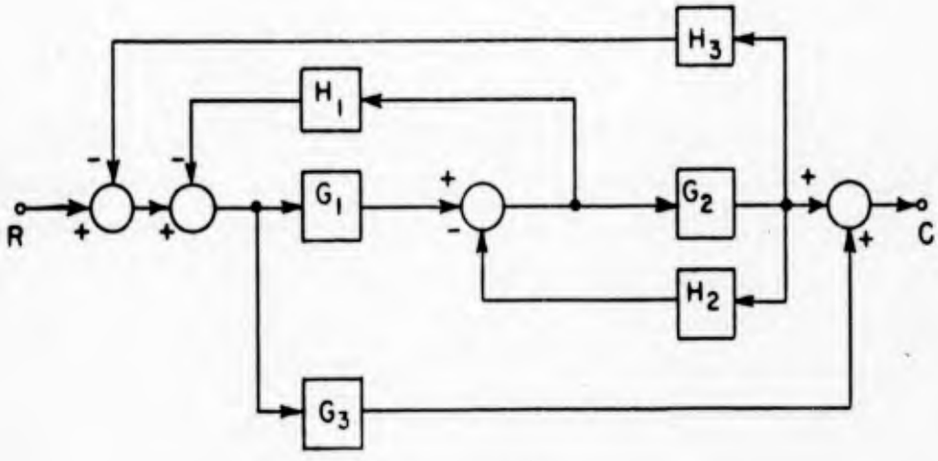
Although signal flow graphs serve a useful purpose as alternative ways of displaying the information of a block diagram, this is not their primary role. They were first devised as an aid to analysing electronic circuits employing feedback and were formulated directly from a mathematical model, or directly from a circuit model. In either case, they serve the very valuable purpose of placing the visualized feedback clearly in evidence. When used in this way, a signal flow graph may be regarded as a graphical portrayal of a set of simultaneous algebraic equations which are ordered in a specific cause-and-effect pattern.



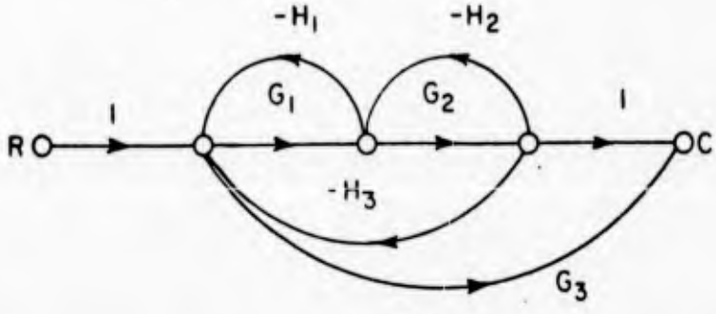
(a)



(b)



(c)



(d)

Figure 3-2 Two examples of conversions from block-diagram to signal-flow-graph symbolism

The nodes register the values of the variables of the equations, and the branches represent relations between pairs of variables. The arrows on the branches point away from cause toward effect and signals always flow in the arrow direction. As an example, the equation

$$ax_1 + bx_2 + cx_3 = x_4 \qquad (3.4)$$

is shown graphically in Figure 3-3. The ordering of the equation suggests x_4 as an effect arising from three independent causes. Each of the

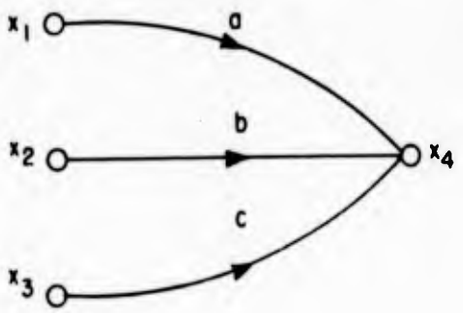


Figure 3-3 A graph of Equation (3.4)

terms of Equation (3.4) constitutes a branch signal, and the three branch signals incident upon node 4 are summed to produce the node signal x_4. This example illustrates the fact that each node symbol performs two separate tasks: it sums all incoming branch signals, and it serves as a source of its registered node signal for all outgoing branches. In general each node performs these tasks simultaneously--Figure 3-3 being a rather special case. The graph of Figure 3-4 shows node 4 as registering the node signal x_4, and in turn transmitting this signal along the outgoing branches d and g. At node 5, a second dependent node signal, x_5, represents a second equation.

$$dx_4 + ix_6 = x_5 \qquad (3.5)$$

Similarly at nodes 6 and 7, the node signals represent respectively the equation

$$gx_4 + ex_5 = x_6 \qquad (3.6)$$

$$fx_6 + hx_5 = x_7 \qquad (3.7)$$

Thus, the signal flow graph of Figure 3-4 is a graphical portrait of Equations (3.4), (3.5), (3.6) and (3.7) showing the particular indicated ordering. It is noteworthy that there is a feedback loop in the graph formed of branches e and i, which indicates

that x_5 is a function of x_4 and also of x_6, which in turn is a function of x_5. We may therefore designate any system that this model represents as being a feedback system, although such a designation may or may not prove to be a useful viewpoint.

Some concluding comments regarding the nodes may be useful. The nodes numbered 4, 5, 6 and 7 in Figure 3-4 are dependent nodes because each has at least one incoming branch symbolizing a dependency upon some other node. Nodes 1, 2 and 3 are independent or source nodes since clearly their node signals do not depend upon the signals at any of the other nodes. These source variables are therefore independently specified. Nodes 4, 5 and 6 serve simultaneously as summing points and as signal sources for outgoing branches. Node 7 which is called a sink node is unique in that it has only incoming branches.

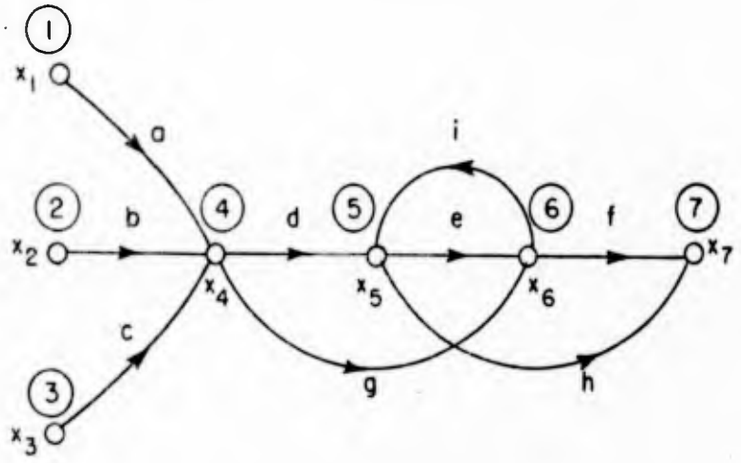


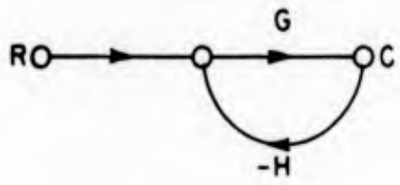
Figure 3-4 A flow graph representing a set of simultaneous algebraic equations

4. Four Definitions

Having dealt with the question of how to describe a system analytically, we turn next to the definition of four terms--four functions actually--which play a dominant role in feedback theory, and then in Sections 5 and 6 we develop various techniques which can be used in evaluating these functions.

(a) System Transfer Function or Transmittance

The transmittance is the overall transfer function relating the output to the input. For the single-loop system of Figure 4-1 we have seen in Section 2 that the transmittance T, defined as $C(s)/R(s)$ *, is simply



$$T(s) = \frac{G(s)}{1 + G(s)H(s)} \qquad (4.1)^{**}$$

Figure 4-1 Basic, single-loop system

In the next section we demonstrate the general formula for the transmittance of any complex feedback configuration.

(b) Loop Transmittance

The loop transmittance, L(s) is the transfer function all the way around the loop (with more than one loop, there are of course several loop transmittances). For example, in the system of Figure 4-1, if we break the loop at a-a', shown in Figure 4-2, L(s) is the transmittance from a' back to a, (R is assumed to be equal to zero). or

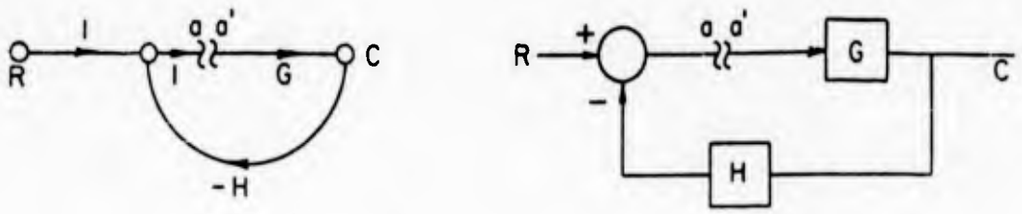
$$L(s) = - G(s) H(s) \qquad (4.2)$$

(The minus sign arises when we pass through the comparator in the block diagram. In the flow graph, the minus sign is explicitly associate with one of the branches of the loop.).

---

* Since in system analysis we are ordinarlily interested in system dynamics, we use the transfer function T(s), with the variable the complex frequency s. C(s) and R(s) are then Laplace transforms or the complex amplitudes of generalized exponential signals.

** In the last section, the expression derived was $A/(1 + \beta A)$, where A was the forward amplifier transfer function, $-\beta$ the feedback transfer function. In that example, both A and β were constants; more generally both are transfer functions.



(a) Flow graph (b) Equivalent block diagram

Figure 4-2 Loop broken at a-a'

(c) Return Difference

Return difference, F(s), can be defined conveniently with the aid of Figure 4-2. If a' is a unit signal, then a is a response signal having the value L(s) after having traversed the loop. The difference between a' and a is the return difference, and when there is a single loop, we may write in general

$$F(s) = 1 - L(s) \qquad (4.3)$$

and for the particular example of Figure 4-2,

$$F(s) = 1 - (-GH) = 1 + GH \qquad (4.4)$$

When there is more than one loop, there are several return differences (since there are then several different loop transmittances). It is necessary in such cases to identify in some way which of the several functions is being considered. This is usually accomplished by using a subscript on both L and F to identify the reference element to which the functions refer. (This point is discussed further in Appendix 5.1).

(d) Sensitivity

The sensitivity measures the effectiveness of the feedback in reducing the effects of variation of a specific parameter, which we shall denote x (x may be an amplifier gain, a time constant, a moment of inertia, a plate-supply voltage, or any other parameter affecting system performance). The sensitivity of the transmission T with respect to the parameter x, denoted S_x^T, is defined as

$$S_x^T = \frac{d \ln T}{d \ln x} \qquad (4.5)$$

Alternatively, from the known derivative of the logarithm of a function, we can write

$$S_x^T = \frac{dT/T}{dx/x} \qquad (4.6)$$

In other words, S_x^T is, for incrementally small changes, the percentage change in T divided by the percentage change in x.

Thus, a value of S_x^T of unity corresponds to an open-loop (non-feedback) system--if x changes by 2 percent, T does likewise. A sensitivity value of zero is the ideal system, since if x changes by a small amount, T remains constant. Practical feedback systems which we consider in the following sections frequently realize values of sensitivity as small as 0.001 for all signals of interest.

In our single-loop feedback system of Figure 4-1, we can calculate the sensitivity of T with respect to G by straightforward differentiation of the expression for T in Equation (4.1):

$$S_G^T = \frac{G}{T}\left[\frac{dT}{dG}\right] \qquad (4.7)$$

$$= \frac{G}{G/(1+GH)}\left[\frac{1}{1+GH} - \frac{GH}{(1+GH)^2}\right]$$

$$= \frac{1}{1+GH}$$

In other words, the sensitivity in this case is simply related to the overall transmission. As a numerical example, we can return to our feedback amplifier of Section 2, in which we had

$$G = 1000 \text{ and } H = 99/1000$$

or

$$S_G^T = \frac{1}{100}$$

Thus (as we saw previously), a change in G of 2 percent results in a change of T of only 0.02 percent.

The four terms which we have defined in this section describe fundamental concepts of the feedback engineer. The basic problem in feedback system design is the simultaneous satisfaction of the given specifications on overall transmittance T and sensitivity S: we are essentially trying to realize suitable dynamic transmission while controlling the effects of parameter variations. Thus, T and S are fundamental measures of significant system properties. In the calculation of both T and S, the loop gain L plays a central role; furthermore, we shall see in Section 7 that a rather general problem of design arises from the tendency of feedback systems to oscillate, the stability being determined by the loop transfer functions.

5. Evaluation of the Transfer Function of a Flow Diagram

Our objective in this section is to develop the procedure for evaluating by inspection a desired transfer function describing the ratio of any designated dependent node signal to any specified source. There are often several different transfer functions that are of possible interest in a given system and the inspection technique to be described therefore offers a significant advantage. For the inspection method to be applicable, it is necessary to represent the excitation in any graph transmittance evaluation as a pure source node.

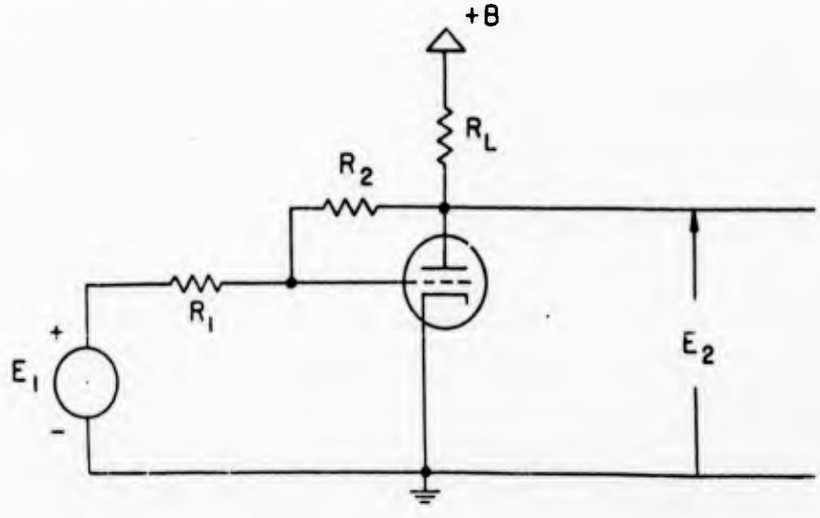


Figure 5-1 A triode amplifier with plate-to-grid feedback

As an example, let us consider the vacuum-tube amplifier of Figure 5-1. Our primary interest might lie in the voltage gain, E_2/E_1 and for this purpose a flow graph can be formulated with E_1 as the source and E_2 as the dependent node of interest. As an intermediate step, it is usually helpful to draw a linear incremental model as in Figure 5-2.

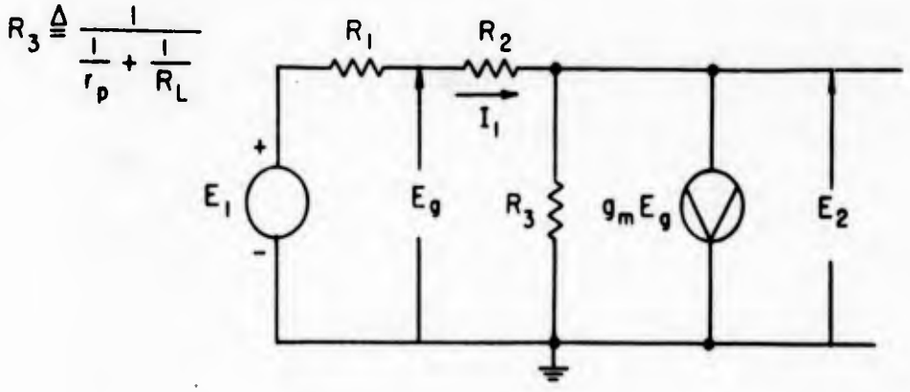


Figure 5-2 Linear incremental model of the triode amplifier

A flow graph may be based on any one of several possible sets of equations which can be transcribed from the graph. Writing node equations, for example,

$$(E_g - E_1)\frac{1}{R_1} + (E_g - E_2)\frac{1}{R_2} = 0 \qquad (5.1)$$

$$g_m E_g + E_2/R_3 - I_1 = 0 \qquad (5.2)$$

where

$$I_1 = (E_g - E_2)\frac{1}{R_2} = (E_1 - E_g)\frac{1}{R_1}$$

and

$$g_m E_g = g_m (E_g)$$

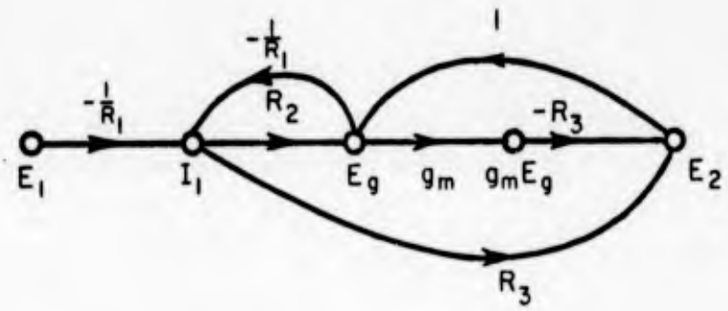


Figure 5-3 Flow-graph interpretation of the model of Figure 5-2

The resulting graph appears as in Figure 5-3, and from this we may evaluate the voltage gain, E_2/E_1, or the input-terminal admittance, I_1/E_1. Should we wish to determine the output-terminal impedance, we must modify the graph--though only in detail. We require a source node symbolizing the injection of the current I_2 at the output terminals. The presence of this new source simply adds one new term to Equation (5.2) which now reads

$$g_m E_g + E_2/R_3 - I_1 - I_2 = 0$$

from which (5.3)

$$E_2 = (I_1 + I_2 - g_m E_g) R_3$$

This shows E_2 to be the linear superposition of three branch signals. All but one of of the branches already exist, therefore only a branch R_3 from I_2 needs to be added. The modified graph is shown in Figure 5-4. With this graph available, we may evaluate any of the following transmittances*.

$$E_2/E_1,\ E_2/I_2,\ I_1/I_2,\ I_1/E_1 \qquad (5.4)$$

and we observe that the requirement that the quantity in each denominator must be represented as a source node in the graph has been met.

* It may strike the reader as unorthodox to regard quantities such as I_1/E_1 and E_2/I_2 as transfer quantities. However, in the cause-and-effect symbolism of a flow graph they become transmittances, and structurally are not distinguished from E_2/E_1, for example.

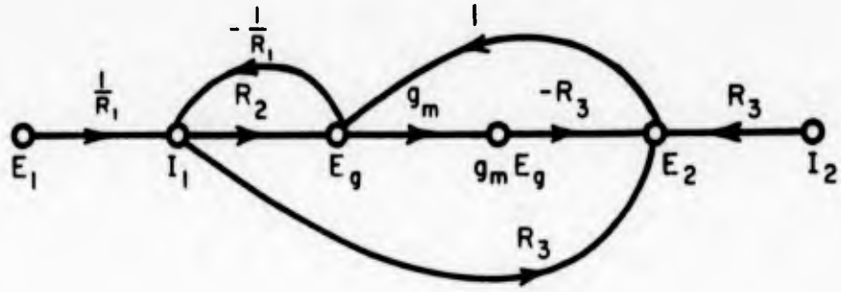


Figure 5-4 Inclusion of I_2 as a source

Similarly, it would be possible to evaluate the effect of a contaminating signal injected from the plate power supply. This could be represented as a second current source injected at the plate terminal and merely requires the addition of a new source node and one additional branch.

This discussion has revealed a rather important feature of signal flow graphs; additional information about the system may be obtained by making minor additions to an existing graph, without the necessity of repeating an analysis from the beginning.

(a) Step-by step reduction

Before presenting the reduction formula for the evaluation of a transmittance of a graph by inspection, it is useful to consider some reduction techniques which can be applied step-by-step. Algebraic reductions (such as the elimination of dependent variables for example) have their flow graph counterparts and several elementary simplifications suggest themselves as a result of the properties of the nodes and the assumption that the graphs are linear.

In general it is useful to think of graph reduction or simplification as involving the absorption of nodes or the elimination of loops or both. We shall first consider the reduction of single loops and then present a comprehensive step-by-step procedure for the absorption of any desired group of dependent nodes, with the end result, if desired,the reduction of the graph to a single branch. The transmittance of this final branch is, of course the graph transmittance or transfer

function being sought.

An elementary single-loop graph may be reduced to a single branch by dividing the forward path transmittance by one minus the loop transmittance. This simple rule which is illustrated in Figure 5-5 may easily be derived by reconstructing the set of equations which the graph represents and solving these for T; or it is possible to view a loop as generating a signal at node 2, as a result of a unit signal being applied at node 1, which takes the form of a power series. Thus, for $x_1 = 1$,

$$x_2 = P(1 + L + L^2 + L^3 + \ldots + L^n + \ldots) \qquad (5.5)$$

or

$$T_{12} = \frac{x_2}{1} = P\,\frac{1}{1 - L}$$

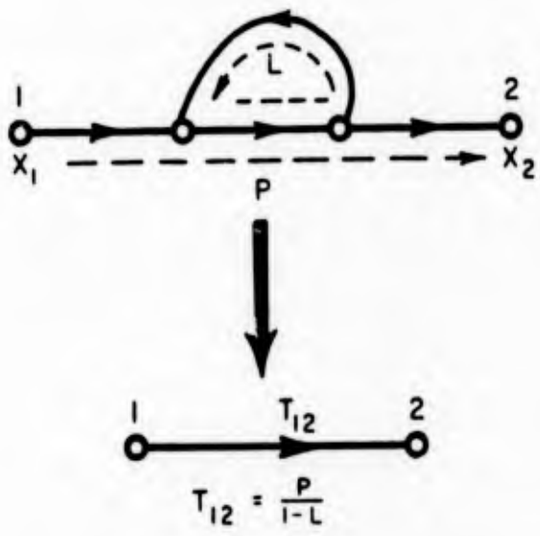


Figure 5-5 Reduction of a single loop

This simple rule for a single loop is applied to any paths which share nodes in common (i.e., touch) the loop. Thus, in Figure 5-6, there are two paths from node 1 to node 2 and a third path from node 1' to node 2. The two graph transmittances are therefore

$$T_{12} = \frac{abc}{1 - bf} + d = \frac{P_1}{1 - L} + P_2 \qquad (5.6)$$

and

$$T_{1'2} = \frac{ebc}{1 - bf} = \frac{P_3}{1 - L} \qquad (5.7)$$

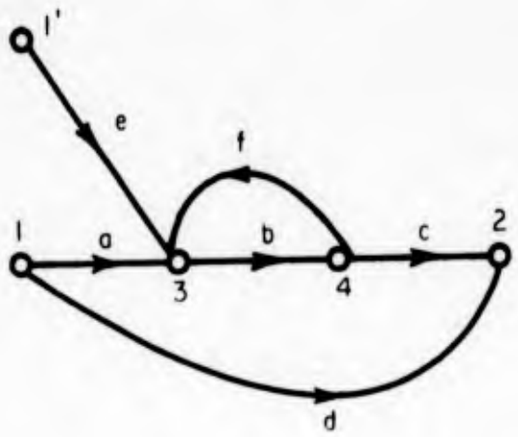


Figure 5-6 Illustrating touching and non-touching paths

The path *, d, does not touch the loop and is therefore not modified by it.

With the aid of the simple rule, any flow graph involving non-touching loops can be reduced to a cascade (non-feedback) graph and ultimately to a single branch. As an example, the transfer function

$$T = \frac{s + 1}{s^2 + 5s + 6} = \frac{s + 1}{(s + 2)(s + 3)} \qquad (5.8)$$

can be represented by either of the two graphs of Figure 5-7. In part (a) each loop is reduced using the simple loop rule with the result

$$T = \frac{T_2}{x_1} = \frac{P}{(1 - L_1)(1 - L_2)} = \frac{(s + 1)\frac{1}{2}}{(1 + \frac{2}{s})(1 + \frac{3}{s})} \qquad (5.9)$$

---

* We will usually not distinguish between the terms path and path transmittance or loop and loop transmittance. Thus, if we speak of the sum of two paths, we mean the sum of their transmittances.

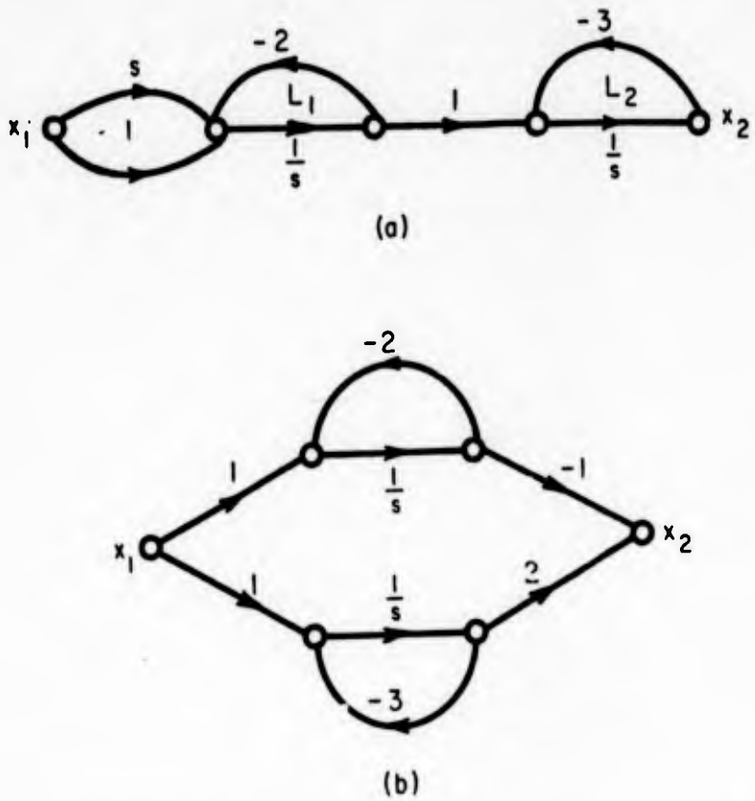


Figure 5-7 Two graphs of the transfer function (5.8)

In part (b), which is a representation of the partial-fraction expansion of (5.8), the graph transmittance is found to consist of the sum of two weighted path transmittances:

$$T = \frac{P_1}{1 - L_1} + \frac{P_2}{1 - L_2} = \frac{-\frac{1}{s}}{1 + \frac{2}{s}} + \frac{\frac{2}{s}}{1 + \frac{3}{s}} \qquad (5.10)$$

The graph of Figure 5-8 is used to illustrate in particular the interrelations between the paths and loops. There are three source-to-sink paths, and the graph transmittance is the sum of these paths with a proper weighting of each to account for touching loops. The path lal which we will designate P_1 touches only L_1, the second path, P_2, which is lbl touches only L_2 but the third path, P_3, or lacbl, touches both loops. The graph transmittance is written

$$T = \frac{x_2}{x_1} = \frac{P_1}{1 - L_1} + \frac{P_2}{1 - L_2} + \frac{P_3}{(1 - L_1)(1 - L_2)} \qquad (5.11)$$

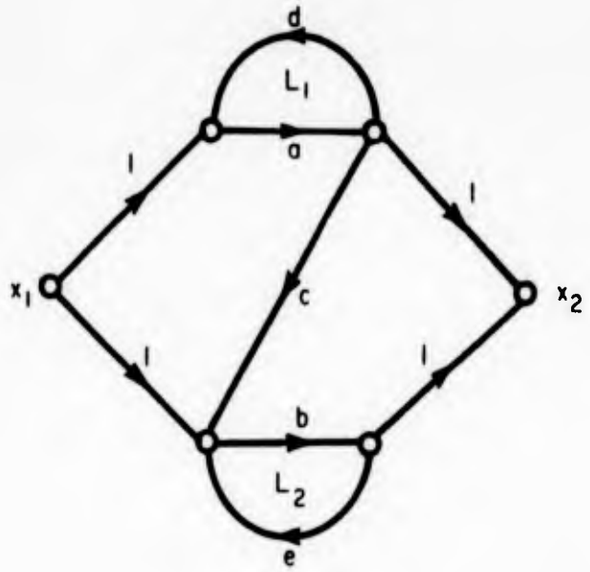


Figure 5-8 Another example of non-touching loops

When we attempt to extend these elementary ideas to include touching loops, the evaluations grow considerably more complicated and systematic procedures become necessary. We introduce therefore a systematic procedure based primarily on node absorption which is a completely general technique and one having considerable utility. It is based upon the fact that the two independent operational properties of a node can be segregated. This process, called node splitting, separates a node into its two constituent parts thereby creating source half and a sink half . The branches connecting to the node in question are rearranged so as to associate incoming and outgoing branches with the appropriate half-node. There is, of course, no transmission of signal through a split node.

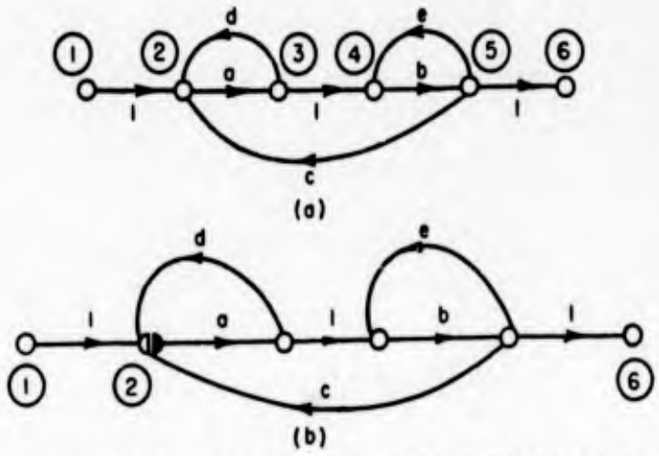


Figure 5-9 A graph prepared for node absorption

The node absorption scheme is very simple--the nodes of the graph which are to be retained are split, thereby forming a new set of source-to-sink paths as shown in the example of Figure 5-9. Here the source halves are shown in black and the sinks in white. In the example the arbitrary decision has been taken to absorb all nodes with the exception of the system source and sink (nodes 1 and 6) and the intermediate node, 2. Signal transmission between nodes 1 and 6 has been interrupted and the remaining signal paths are between nodes 1 and 2, nodes 2 and 6, and between the halves of node 2. A simplified graph or residue may now be constructed in which the branches are the source-to-sink paths of Figure 5-9(b). The branch evaluations of the residue are shown in Figure 5-10(a) and the reconstituted graph in part (b).

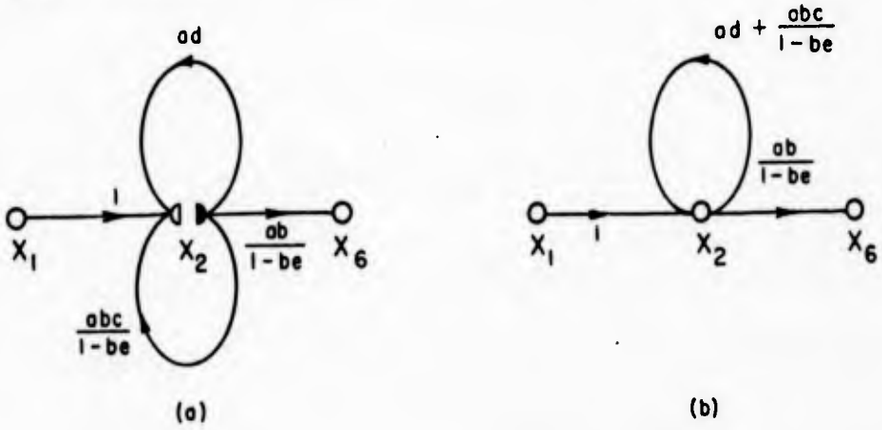


Figure 5-10 Showing the evaluation of the residue of Figure 5-9

The graph transmittance may now be evaluated by using the simple loop rule and

$$T_{16} = \frac{x_6}{x_1} = \frac{\dfrac{ab}{1 - be}}{1 - ad - \dfrac{abc}{1 - be}} = \frac{ab}{1 - be - ad - abc + adbe} \qquad (5.12)$$

It is noteworthy that the simple loop rule had to be employed in the evaluation of the residue branches, since two of the source-to-sink paths have a touching loop, eb , which has remained intact in the node splitting.

Sometimes it is desired to absorb as many nodes as possible, yet to preserve explicitly the integrity of the loop structure. To accomplish this we must open every loop in the graph by splitting an appropriate set of nodes. For the example of Figure 5-9 it is necessary to split at least two intermediate nodes in order to open the three loops. Several choices exist, and choosing nodes 3 and 5, the residue takes the form shown in Figure 5-11. This kind of graph in which the essential loop structure is preserved and described with a minimum number of nodes is called an index residue. The index, which is the number of nodes additional to the source and sink which is required, characterizes the essential complexity of the system and is a useful means for classifying systems.

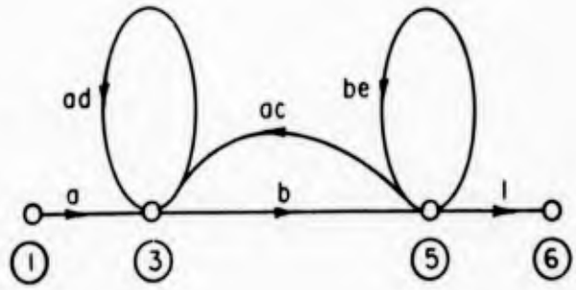


Figure 5-11 Index residue of Figure 5-9(a) which retains the essential loop structure

(b) Mason's reduction formula

With the aid of the foregoing examples it is possible to visualize a specified graph transmittance in terms of two basic topological elements--the open and closed paths. As we have already seen, the transmittance of a graph is the sum

of all different paths leading from a specified source to some chosen dependent node and each path transmittance becomes weighted by factors arising from the configuration of loops. When there are no loops, the weighting factor for each paths is unity.

Based upon this view of graph topology, Mason's reduction formula for graph transmittance * takes the form of summation:

$$T_{ij} = P_1 \frac{\Delta_1}{\Delta} + P_2 \frac{\Delta_2}{\Delta} + \ldots + P_n \frac{\Delta_n}{\Delta}$$

$$= \sum_k \frac{P_k \Delta_k}{\Delta} \qquad (5.13)$$

the summation being taken over all k different open paths from source i to dependent node j.

The common denominator, Δ, called the graph determinant , is a function of only the loops and the way in which they interact with one another. Formally Δ is described by the function

$$\Delta = \left[(1 - L_1)(1 - L_2) \ldots (1 - L_m) \right]^* \qquad (5.14)$$

where * indicates that, of the 2^m possible terms in the expansion of Δ, only those terms are to be retained which do not contain loops which touch in the graph. To follow the indicated procedure of Equation (5.14) would usually involve the generation of many terms which would eventually be discarded, therefore this approach is seldom used. It is easy to write the expansion in the following form however:

$$\begin{aligned} \Delta = 1 &- (\Sigma \text{ all different loops}) \\ &+ (\Sigma \text{ all different pairs of nontouching loops}) \\ &- (\Sigma \text{ all different triplets of nontouching loops}) \\ &+ - \ldots \end{aligned} \qquad (5.15)$$

---

* S. J. Mason, " Feedback theory - further properties of signal flow graphs ", PROC. IRE Volume 44, Number 7, pp 920-927; July, 1956.

The path factor, Δ_k, is a function of only the loops which do not touch the k^{th} path and is the determinant of the remaining graph after the k^{th} path (and its attached loops) has been deleted from the graph.

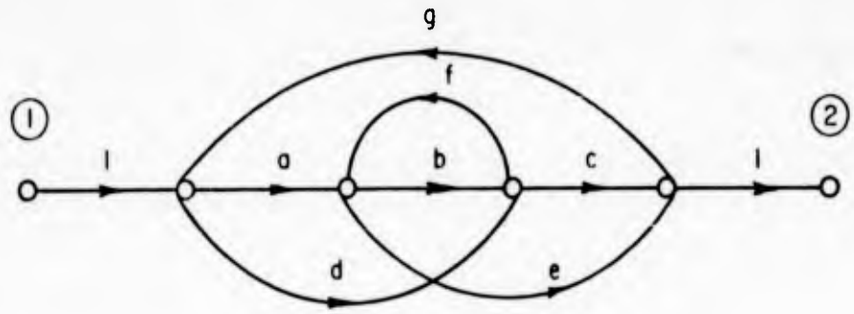


Figure 5-12 Example for evaluation

Before trying some examples we should observe that an open path is a succession of branches which never includes a loop therefore no node is ever encountered more than once. Similarly loops are simple closed paths in which no node is encountered more than once per cycle of traversal. With these agreements in mind, let us find the graph transmittance of the example shown in Figure 5-12.

There are four paths:

abc dc ae dfe

and five loops:

bf abcg dfeg dcg aeg

In this example, all of the loops touch each other and all paths touch all loops, therefore we write

$$T_{12} = \frac{abc + dc + ae + dfe}{1 - bf - abcg - dfeg - dcg - aeg} \qquad (5.16)$$

As a second example we consider the graph of Figure 5-13.

To evaluate the transmittance T_{12}, we observe that there are two paths:

ab gfab

and five loops:

ac gi abd ghj gfae

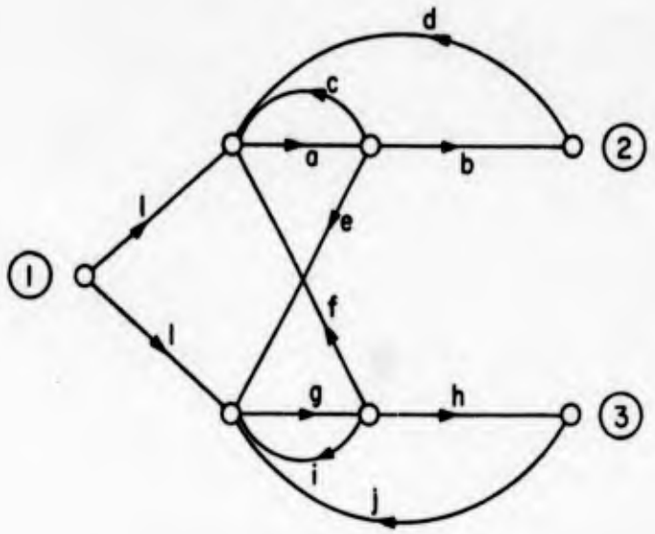


Figure 5-13 Second example for evaluation

The path gfab touches all loops, but ab fails to touch gi and ghj. Furthermore not all of the loops are touching and we can visualize four different pairs of non-touching loops, but no higher-order combinations. The graph transmittance may be written on the basis of these observations as follows:

$$T_{12} = \frac{ab(1 - gi - ghj) + gfab}{1 - ac - gi - abd - ghj - gfae + acgi + (abd)(ghj) + ac(ghj) + gi(abd)} \qquad (5.17)$$

Figure 5-13 represents a complicated system and yet the evaluation of the transmittance is quite straightforward. The evaluation of Δ in systems of this complexity becomes a bit tedious, but much of this difficulty can be circumvented by using the method described in Appendix 5.1.

Summary

In summary we have presented two basic techniques; one for the absorption of a selected group of nodes to effect a partial or controlled reduction of a graph, and the other to effect the complete reduction by inspection.

Appendix 5.1

The Evaluation of the Graph Determinant in Compact Form

The evaluation of Δ in expanded form (Eq. (5.15)) sometimes becomes tedious and involved. This is usually the case whenever a graph has several nontouching loops together with coupled loops so that the analyst is forced to concentrate in order not to overlook any of the nontouching loop sets. Recognizing this difficulty Mason proposed * a method of writing the determinant in factored form.

Although we shall not attempt a formal development and proof, we will offer an example from which the general procedure may be conjectured. Let us consider the familiar three-loop complex shown in Figure 5.1-1.

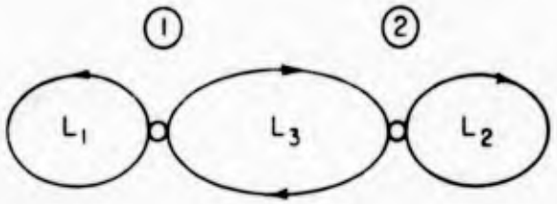


Figure 5.1-1 Loop configuration

Using the expansion technique of Equation (5.15) we write

$$\Delta = 1 - L_1 - L_2 - L_3 + L_1 L_2 \qquad (5.1.1)$$

which could be rearranged to read

$$\Delta = (1 - L_1)(1 - L_2) - L_3 \qquad (5.1.2)$$

This suggests that the graph might be partitioned by opening the central loop L_3. The determinant of the remaining loop structure is then simply the product of the separate determinants of the disparate loops L_1 and L_2. Thus, Equation (5.1.2) might be written

$$\Delta = \Delta_1 \Delta_2 - L_3 \qquad (5.1.3)$$

The idea suggested can be developed formally by inserting an interior node

---

* S. J. Mason, Op. Cit.

in L_3, splitting this node and then evaluating the resulting source-to-sink path, as in Figure 5.1-2.

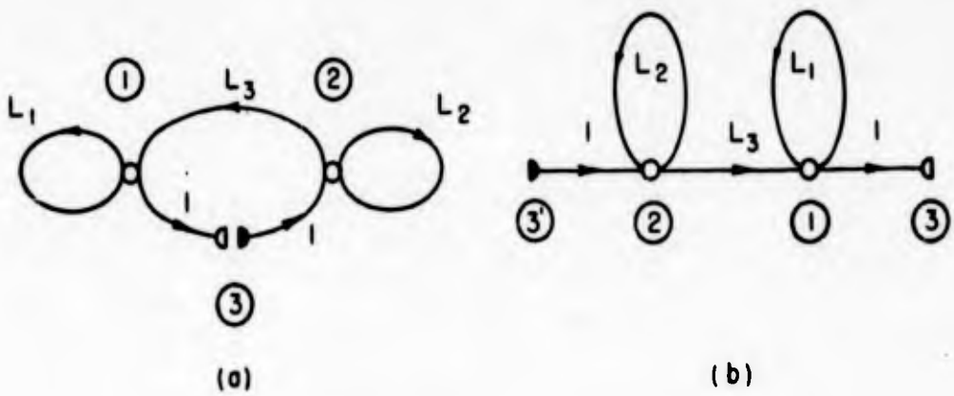


Figure 5.1-2 Loop transmittance of node 3

The transmittance $T_{3'3}$ which is the loop transmittance of node 3 is

$$T_{3'3} = \frac{L_3}{(1 - L_1)(1 - L_2)} \qquad (5.1.4)$$

The quantity $1 - T_{3'3}$ which is return difference of node 3 is found as

$$1 - T_{3'3} = \frac{(1 - L_1)(1 - L_2) - L_3}{(1 - L_1)(1 - L_2)} \qquad (5.1.5)$$

The numerator is Δ for the system with node 3 closed and the denominator is the determinant of the system after node 3 has been opened (or, in effect, removed). Designating the latter determinant by Δ^o, we note that

$$F_{3'3} = 1 - T_{3'3} = \frac{\Delta}{\Delta^o} = 1 - \frac{\sum_k L_k \Delta_k}{\Delta^o} \qquad (5.1.6)$$

where L_k is the k^{th} opened loop resulting from splitting the j^{th} node and Δ_k is the loop (path) factor. Solving for Δ we obtain

$$\Delta = \Delta^o - \sum_k L_k \Delta_k \qquad (5.1.7)$$

Before commenting on the significance of this relation, let us apply it to node 1 of Figure 5.1-1. Splitting node 1 and opening out the resulting source to sink paths (L_k) we obtain Figure 5.1-3.

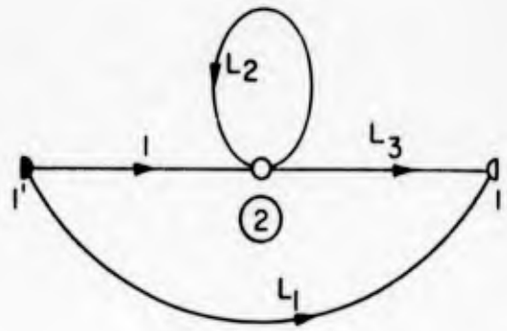


Figure 5.1-3 The result of splitting node 1

In this case, $\Delta^o = (1 - L_2)$ and the path summation, when subtracted from Δ^o yields

$$\Delta = (1 - L_2) - L_1 (1 - L_2) - L_3 \qquad (5.1.8)$$

which again is the known value of Δ.

It turns out that Equation (5.1.7) is general and may be applied with respect to any chosen node of a linear graph, whether that node be one in the given graph, or one inserted specifically for the evaluation of Δ.

Let us try the idea on a few examples which would ordinarily require rather complicated expansions in the determinant evaluation. In the first, shown Figure 5.1-4(a), there are seven loops, eleven separate nontouching pairs, six nontouching triplets and one nontouching quadruplet. Thus, the determinant is the algebraic sum of unity and 25 nontouching loop products. The real difficulty here, wholly aside from the tedium of writing this number of terms, is the imminence of danger from omitting or misinterpreting a term.

This example lends itself beautifully to the technique just described. The insertion of an interior node as shown in part (b) of Figure 5.1-4 permits the graph to be partitioned into two separate loop structures of the form shown in Figure 5.1-1.

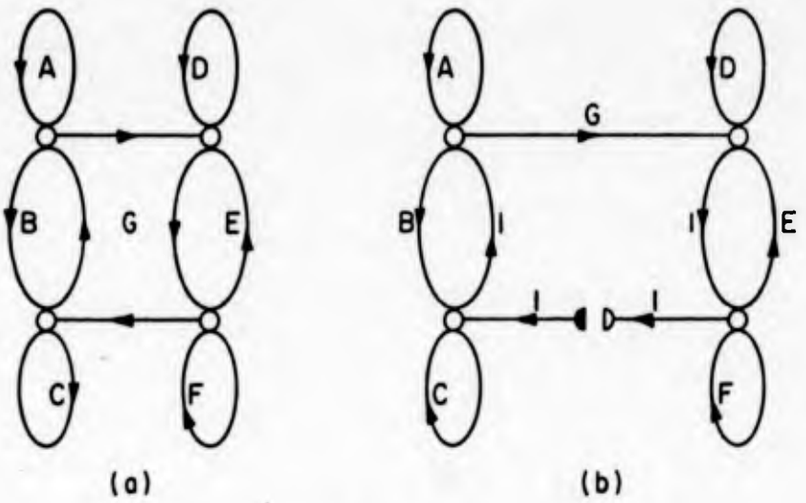


Figure 5.1-4 First example

Designating Δ_1 and Δ_2 as the determinants of the partitioned subgraphs, we employ Equation (5.1.7) to write by inspection

$$\Delta = \Delta_1 \Delta_2 - G$$

and substituting for Δ_1 and Δ_2,

$$\Delta = \Big[(1 - A)(1 - C) - B\Big] \Big[(1 - D)(1 - F) - E\Big] - G \qquad (5.1.9)$$

which represents a material saving in effort over the more conventional expansion.

For a second example, we consider the graph of Figure 5.1-5 which is clearly not partitionable. Here our best hope is to split a node which will open the greatest number of loops (thereby making the evaluations of Δ^o and Δ_k as easy as possible). A little study indicates that node (4) is a good choice since splitting it leaves intact only two loops, as shown in Figure 5.1-6.

Since the two remaining loops of Figure 5.1-6 are nontouching,

$$\Delta^o = (1 - H)(1 - CG) \qquad (5.1.10)$$

The forward paths, L_k, and their factors yield

$$\sum_k L_k \Delta_k = BGA + EA + (1 - H)(BF + BD + ECD + ECF) \qquad (5.1.11)$$

and the determinant of the system is

$$\Delta = (1 - H)(1 - CG - BF - BD - ECD - ECF) - BGA - EA \qquad (5.1.12)$$

In summary, the rules for the evaluation of the determinant of a system of coupled loops may be stated as follows:

1. Choose a single node, j, to be split (mentally) which will either partition the loop graph advantageously or open the majority of the loops. (The node may be one of the original nodes of the given graph, or a node inserted specifically for purposes of facilitating the evaluation).
2. Evaluate Δ^o as the determinant of all loops which do not contain node j.
3. Sum all different products of loops and loop factors which do contain node j.
4. Substitute values in

$$\Delta = \Delta^o - \sum_k L_k \Delta_k$$

It is unnecessary to split the chosen node in fact or to redraw the resulting graph, since the steps outlined above serve as a guide to the meanings of the various entries in the formulation of Δ.

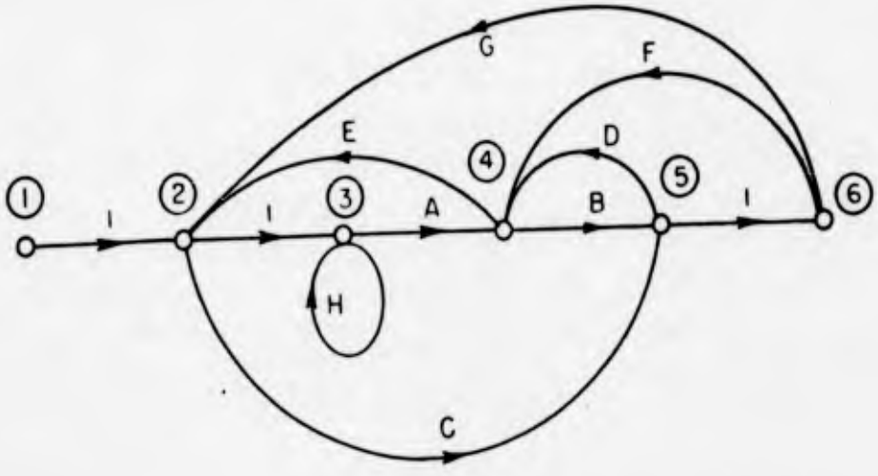


Figure 5.1-5 Second example for the evaluation of Δ

It should be noted finally that some systems contain loop subgraphs that are not coupled. When all branches that are not components of loops are deleted from the graph, there may remain several partitioned loop subgraphs. The foregoing procedure applies then to the evaluation of Δ for each of the several subgraphs.

The system determinant is then the product of the several subgraph determinants.

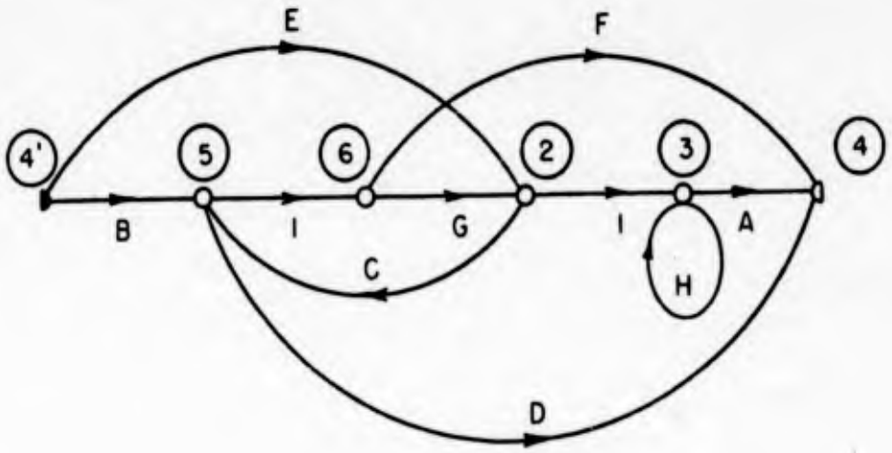


Figure 5.1-6 Graph with node 4 split

6. Sensitivity

As we have emphasized throughout the preceding sections, feedback is used primarily for the control of sensitivity--i.e., in order that the system may possess the self-calibration property of automatically compensating for parameter variations. Evaluation of a feedback system must then rely on a quantitative measure of this sensitivity; in Section 4 we have defined the sensitivity of the overall transmittance T to changes in a parameter x as

$$S_x^T = \frac{d \ln T}{d \ln x} = \frac{dT/T}{dx/x} \qquad (6.1)$$

In this section, we wish to consider first the sensitivity-control features of one typical feedback system (as an example); then we consider the techniques available for the calculation of sensitivity.

Example of sensitivity control with feedback

Examples of the importance of feedback in the control of sensitivity are available throughout the field of electronic instrumentation. As an illustration, we consider in this section an instrument for the measurement of temperature, involving the generation of an electric voltage proportional to the instantaneous temperature within an enclosure.

The basic, non-feedback system is depicted in Figure 6-1. The temperature-sensitive element is a thermistor, with the model, shown in part (b) of the figure, including a voltage source proportional to the change in temperature from a quiescent or normal operating value. In series with this voltage source there is a second voltage source dependent on the thermistor current, since an increase of the current causes an increase in temperature as a result of the power dissipated in this resistive element.

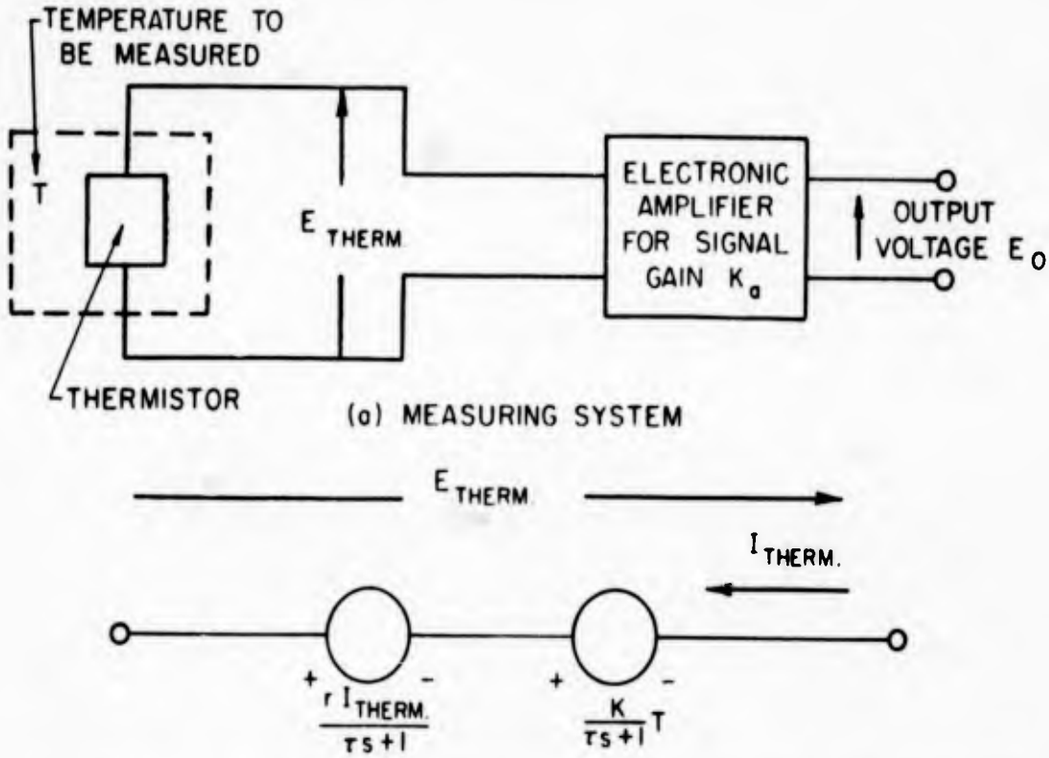


Figure 6-1 Temperature instrument

The voltage source is related to the temperature through the transfer function $-K/(\tau s + 1)$, where τ is the thermal time constant. In other words, when the ambient temperature changes abruptly in the oven or container within which the thermistor is located, the terminal voltage approaches exponentially the new value as depicted in Figure 6-2.

If the amplifier of Figure 6-1(a) draws no input current, the input voltage equals the voltage of the source in the thermistor model; consequently the output

E_0 is related to the temperature by the overall transfer function

$$\frac{E_0}{T} = -\frac{K_a K}{\tau s + 1} \qquad (6.1a)$$

Since the thermal time constant is typically of the order of magnitude of 2 seconds (or greater), the full reading of the output voltage is not approached until about 8 seconds (four time constants) after a change in temperature. In an alternative viewpoint, we can say that the

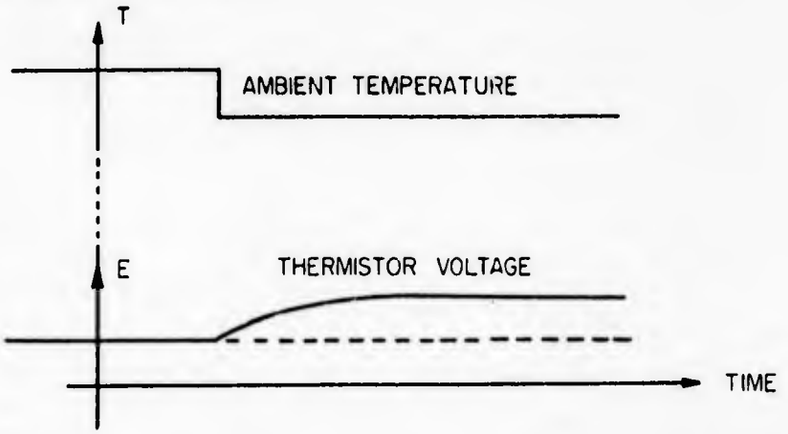


Figure 6-2 Change of thermistor voltage with temperature.

system does not respond to temperature variations at frequencies greater than 1/2 radian/ sec, (the reciprocal of the time constant), or about 1/ 12 cycle/ second.

In order to speed up the response of our instrument, we can pass the output voltage E_0 of Figure 6-1 through an electrical network, as depicted in Figure 6-3. If this compensation network possesses the

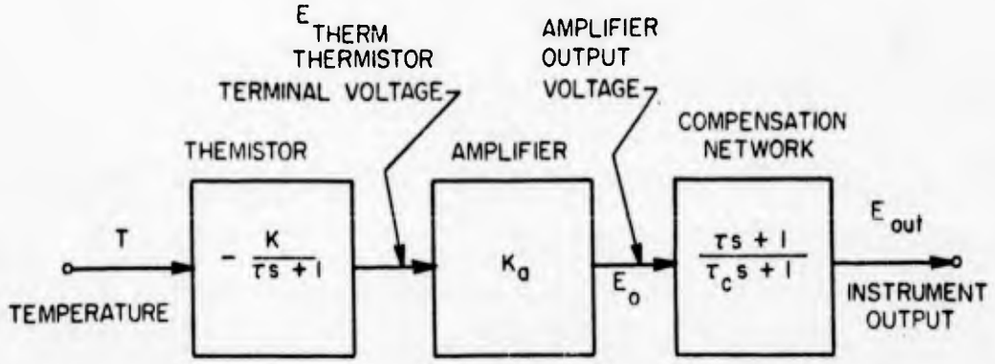


Figure 6-3 Improved instrument

transfer function $(\tau s + 1) / (\tau_c s + 1)$, the overall instrument transfer function is

$$\frac{E_{out}}{T} = - \frac{KK_a}{\tau s + 1} \frac{\tau s + 1}{\tau_c s + 1} = \frac{KK_a}{\tau_c s + 1} \qquad (6.2)$$

In other words, the thermal time constant τ is replaced by the electrical time constant τ_c. If τ_c is made 1/1000th of τ, we succeed in speeding up the instrument response by the factor of 100; in other words, the instrument responds to signals at frequencies 100 times greater.

We seldom achieve such significant results without an associated penalty; in the present case, the penalty is apparent if we consider the realization of the compensation network required in Figure 6-3. Search for a suitable network yields the system drawn in Figure 6-4; in order to affect a 100:1 decrease in the instrument time constant, we need to include an amplifier of gain 100. The greater the improvement demanded in dynamic response, the larger the gain required.

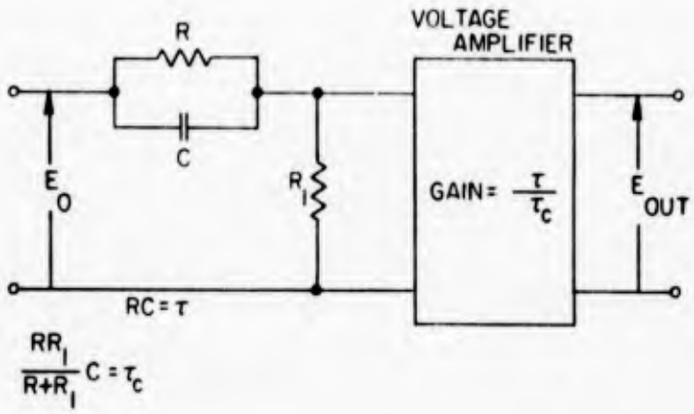


Figure 6-4 $\frac{E_{out}}{E_0} = \frac{\tau s + 1}{\tau_c s + 1}$

The instrument of Figure 6-3 is quite satisfactory for many purposes, but the modern systems engineer, armed with the tools of feedback theory, naturally asks whether, since an additional gain of τ / τ_c is required, better results might not be achieved through the use of feedback. A principal disadvantage of the above

instrument is that any changes in the gain of either amplifier result in corresponding (and equal) changes in the overall instrument transfer function. Hence, as amplifiers age or environmental conditions vary, the instrument requires continual recalibration. If feedback is used, not only can the speed of response be controlled - but also the re-calibration can be made automatic and self-contained.

An appropriate feedback configuration is shown in Figure 6-5, where a current proportional to the output voltage is passed through the thermistor. Under these conditions, the terminal voltage of the

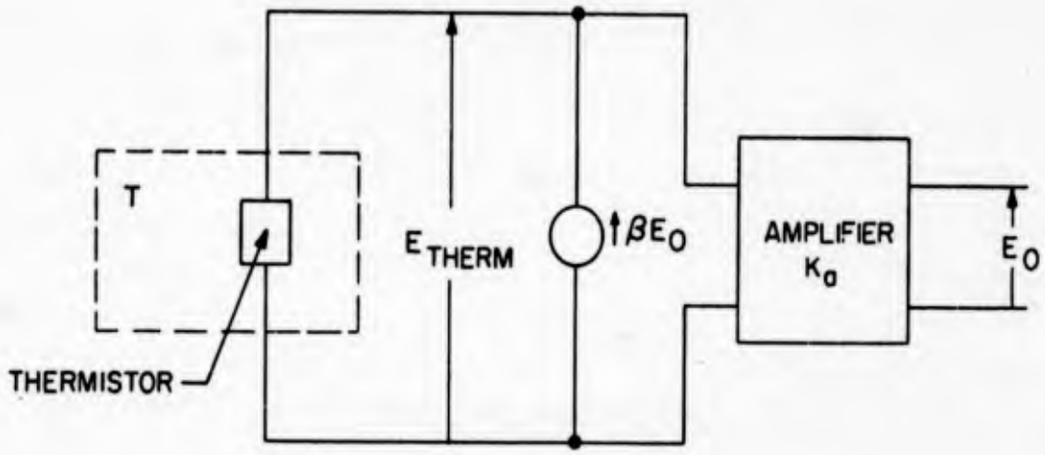


Figure 6-5 Instrument with feedback

thermistor is determined from Figure 6-1(b) as

$$E_{therm} = \frac{-K}{\tau s + 1} T - \frac{r}{\tau s + 1} \beta E_0$$

and the system is described by the block diagrams of Figure 6-6 [the second diagram is derived from the first if we recognize that addition after multiplication of each variable by $1/(\tau s + 1)$ is equivalent to addition followed by multiplication of the sum].

The dynamic characteristics of the feedback instrument are given by the overall transmittance, which can be written directly from the relations of Section 5:

$$\frac{E_0}{T} = (-K)\ \frac{\dfrac{K_a}{\tau s + 1}}{1 + \dfrac{r\beta K_a}{\tau s + 1}} \qquad (6.3)$$

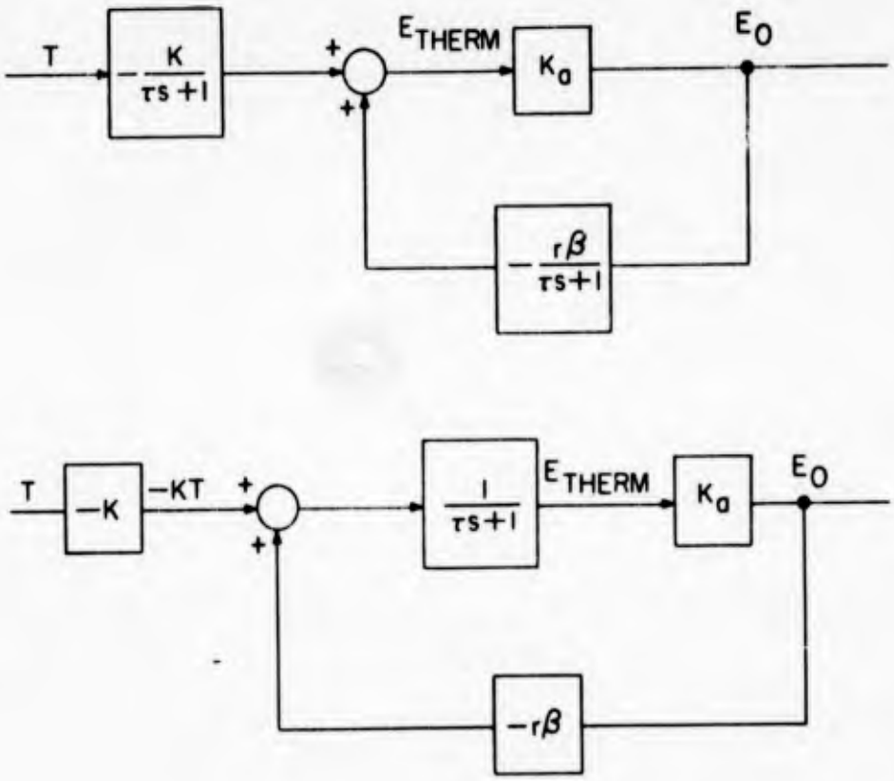


Figure 6-6 Equivalent block diagrams for the feedback instrument.

or, if we clear of fractions,

$$\frac{E_0}{T} = \frac{-KK_a}{\tau s + 1 + r\beta K_a} \qquad (6.4)$$

Equation (6.4) can be placed in somewhat more convenient form if we divide numerator and denominator by $(1 + r\beta K_a)$

$$\frac{E_0}{T} = \frac{\dfrac{-KK_a}{1 + r\beta K_a}}{\dfrac{\tau}{1 + \beta K_a} s + 1} \qquad (6.5)$$

The above equation reveals two important facts:

(1) The time constant of the feedback system is $\tau/(1 + r\beta K_a)$. In other words, if we wish to decrease the time constant by a factor of 100, we need only select β such that

$$1 + r\beta K_a = 100 \qquad (6.6)$$

Thus, feedback can be used to control the time constant, just as we did with the tandem compensation network.

(2) As a result of the feedback, the gain is reduced by division by $(1 + r\beta K_a)$; in other words, the gain is reduced in exactly the same ratio as the time constant. This result is rather startling, since it is identical with our earlier conclusions about the tandem compensation scheme. There, too, if we wished to decrease the time constant by a factor of 100, we had to insert an amplifier with a gain of 100 to restore the level of the output signal.

The surprising similarity between the two instruments ends at this point, however. In the tandem scheme, a 1% change in amplifier gain results in 1% change in output: i.e., $S_{K_a}^{E_0/T} = 1$. In the feedback configuration of Figure 6-5, however,

$$S_{K_a}^{E_0/T} = \frac{1}{1 + \dfrac{r\beta K_a}{\tau s + 1}} = \frac{\tau s + 1}{\tau s + 1 + r\beta K_a} \qquad (6.7)$$

at zero frequency (i.e., in steady-state conditions),

$$S_{K_a}^{E_0/T} = \frac{1}{1 + r\beta K_a} \quad \left(= \frac{1}{100} \text{ for our numerical example}\right) \qquad (6.8)$$

The sensitivity of the feedback system is reduced by the same factor as the time

constant and gain ! The feedback instrument contains automatic self-calibration ! *

Thus, the feedback system possesses all of the engineering-design flexibility of the tandem compensation scheme and, in addition, the possibility of design control over the sensitivity of the instrument transmission to changes in system parameters. Because of this added, significant advantage feedback systems are used extensively in instrumentation. The only penalty we must pay for this sensitivity control is the extra equipment (often trivial) to realize the feedback path--in Figure 6-5 the current generator with the value of current proportional to the output voltage.

Design Control of Sensitivity

The two control systems of Figure 6-7 illustrate two different design interpretations for achieving a specified overall transfer function with a particular process given. In case (a) the desired transfer function is achieved by utilizing a controller placed in tandem with the process; in (b) the controller function is supplied by a local feedback path encircling the plant and the tandem element is merely an amplifier. Our interest is centered in the comparison of the two systems from the point of view of how the overall transfer function is affected by variations in the plant parameters K and a over which we presumably have no direct design control. In other words, we want to investigate the sensitivity of $T(s)$ to both K and a for the two cases.

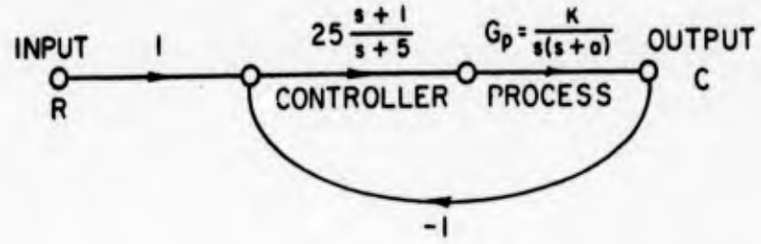
For case (a) the overall transmittance of the flow graph is

$$T(s) = \frac{C}{R} = \frac{25K(s+1)}{s(s+5)(s+a) + 25K(s+1)} \qquad (6.9)$$

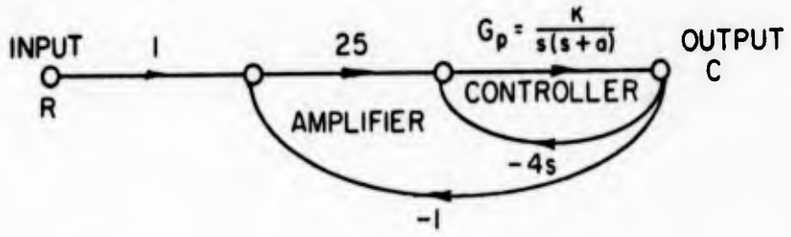
and if this is placed in the form

$$T(s) = \frac{1}{1 + \dfrac{s(s+5)(s+a)}{25K(s+1)}} \qquad (6.10)$$

the sensitivity S_K^T can be evaluated by direct differentiation



(a) TANDEM CONTROLLER



(b) MINOR-LOOP CONTROLLER

Figure 6-7 Control systems with normal process parameter values K = 1, á = 1.

to produce

$$S_K^T = \frac{s(s+5)(s+a)}{25K(s+1) + s(s+a)(s+5)} \tag{6.11}$$

and when a = 1 and K = 1, this reduces to

$$S_K^T = \frac{s(s+5)}{s^2 + 5s + 25} \tag{6.12}$$

Similarly, the sensitivity, S_K^T, for case (b) is carried out by operating with the transmittance

$$T(s) = \frac{\frac{25K}{s(s+a)}}{1 + \frac{4Ks}{s(s+a)} + \frac{25K}{s(s+a)}} \tag{6.13}$$

$$= \frac{25Ks(s+a)}{s(s+a) + K(25+4s)} \tag{6.14}$$

Dividing through by K, and then computing the sensitivity S_K^T we obtain

$$S_K^T = \frac{s(s+a)\,1/K}{\frac{1}{K}\,s(s+a)+25+4s} \qquad (6.15)$$

and again inserting the normal values of the process parameters, we find

$$S_K^T = \frac{s(s+1)}{s^2+5s+25} \qquad (6.16)$$

Thus, even though the two system transmittances are indentical, the two sensitivity functions differ. The significance of the difference is somewhat difficult to interpret, but one way of achieving a relative evaluation might be to plot $|S_K^T|$ versus the angular frequency ω (i.e., with $s = j\omega$ so that we are focusing attention on the system performance under sinusoidal excitation). The results shown in Figure 6-8 demonstrate the superiority case (b) (which utilizes local feedback around the process) when the angular frequencies are low. Thus, through the definition of the sensitivity function we have expressed in quantitative terms the degree to which the feedback accomplishes its appointed task and we have a firm basis for the comparision of alternate designs.

Evaluation of the competitive designs of Figure 6-7 might also require determination of the sensitivities which respect to the parameter a. By direct differentiation, we learn that for case (a)

$$S_a^T = \frac{-s(s+5)}{(s+1)(s^2+5s+25)} \qquad (6.17)$$

and for case (b)

$$S_a^T = \frac{-s}{s^2+5s+25} \qquad (6.18)$$

and plotting graphs of $|S_a^T|$ versus ω would again reveal the superiority of the minor-loop-compensation [case (b)] design at low frequencies.

In this section our objective is not, however, to consider a specific example in great detail, but rather to indicate the general way in which the feedback engineer attempts to make a quantitative assessment of the role of feedback in a system design--an assessment which is based on the sensitivity function.

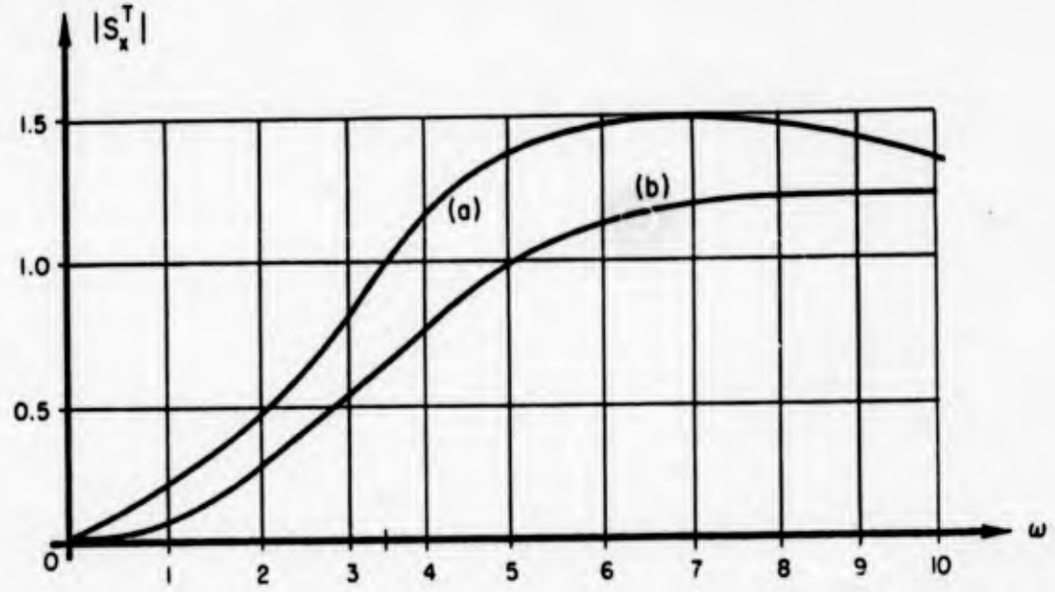


Figure 6-8 Sensitivities of the systems of Figure 6-7

Although the value of the sensitivity function is readily demonstrated, its computation has been shown to be tedious even in the simple examples that we have chosen. If we are to make any real use of this concept it is obvious that attention must be given to the problem of making sensitivity computations much simpler. Fortunately several solutions to the computation problem have been discovered and because there is a repertoire of these the remainder of this section is devoted largely to developing an illustrating the several different useful techniques and viewpoints.

Topological Interpretation of Sensitivity

The development of Mason's reduction theorem has been such a decided boon to the evaluation of graph transmittances that it was natural to seek a similar evalution of sensitivity. This is readily accomplished by carrying out the steps of the formal derivative definition on Mason's formula for transmittance. Thus

Thus

$$S_x^T = \frac{x}{T}\frac{\partial}{\partial x}\left\{\frac{\sum_k P_k \Delta_k}{\Delta}\right\} \qquad (6.19)$$

and for convenience, we symbolize the summation using the designation

$$\sum' \triangleq \sum_k P_k \Delta_k$$

Then

$$S_x^T = \frac{\Delta \frac{\partial \Sigma}{\partial x} - \Sigma \frac{\partial \Delta}{\partial x}}{\Delta^2} \cdot \frac{x\Delta}{\Sigma} \qquad (6.20)$$

$$= \frac{x\frac{\partial \Sigma}{\partial x}}{\Sigma} - \frac{x\frac{\partial \Delta}{\partial x}}{\Delta} \qquad (6.21)$$

We next take note of a significant property of the functions Δ and Σ; each is a linear function of any path or loop transmittance, and hence of any branch thereof.* In other words, in the expansions of Δ and Σ, no path nor loop appears more than once in any product term and this is a consequence of the restriction of such terms to products of non-touching loops and paths. This makes it possible to conclude that the functions

$$x\frac{\partial \Sigma}{\partial x} \quad \text{and} \quad x\frac{\partial \Delta}{\partial x}$$

contain only those product terms in which x was originally a factor. Therefore we might write

$$\boxed{x\frac{\partial \Delta}{\partial x} = \Delta - \Delta^0} \qquad (6.22)$$

---

* We are assuming here that the parameter x appears in the flow diagram only once, and then as a multiplier of a branch. If x is a circuit parameter (R, L, C, g_m, μ, α, or β, for example) we can always draw the flow diagram to satisfy this assumption, since the overall transmission T is a bilinear function of x. In the exceptional cases when the assumption can not be satisfied (e.g., when x is the ambient temperature, and hence affects several parameters simultaneously), we must return to the basic definition for the calculation of S.

and

$$x \frac{\partial \Sigma}{\partial x} = \Sigma - \Sigma^{o} \qquad (6.23)$$

where Δ^{o} and Σ^{o} signify the expansions of Δ and Σ evaluated with the branch x removed from the graph.

Returning to Equation (6.21) and substituting (6.22) and (6.23), we arrive at the following result:

$$S_x^T = \frac{\Sigma - \Sigma^{o}}{\Sigma} - \frac{\Delta - \Delta^{o}}{\Delta} \qquad (6.24)$$

which simplifies to

$$S_x^T = \frac{\Delta^{o}}{\Delta} - \frac{\Sigma^{o}}{\Sigma} \qquad (6.25)^{**}$$

Example

The middle-frequency behavior of a common-emitter transistor amplifier with collector-to-base feedback is described by the flow graph of Figure 6-9. Using Equation (6.25), the sensitivity

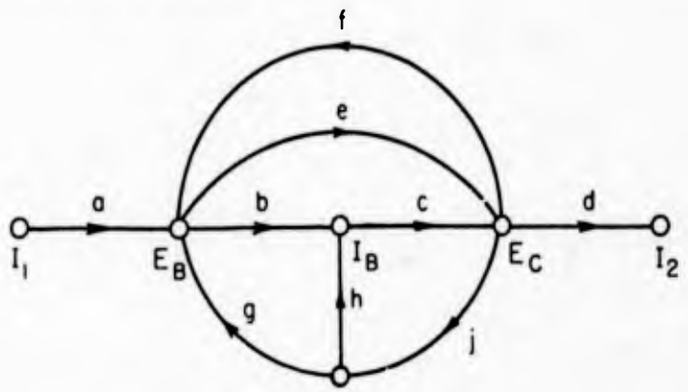


Figure 6-9 Example for sensitivity evaluation

of the current gain I_2 / I_1 can be determined readily with respect to any of the branches and to illustrate, we evaluate the sensitivity S_c^T . We observe first that there are two source-to-sink paths; abcd and aed each of which carries a unity path factor. Only the former is interrupted when the branch c is removed, therefore

*As noted in the previous footnote, Equation (6.25) is valid only if the signal flow diagram is drawn in such a way that x appears only as the multiplier of a single branch.

$$\frac{\Sigma^o}{\Sigma} = \frac{aed}{abcd + aed} \qquad (6.26)$$

We note further that there are five distinct loops (all touching one another): the the principal loop, bcf, involving the collector-to-base feedback and additionally the loops cjh, fe and ejg. Removing the branch c opens three of these and we have

$$\frac{\Delta^o}{\Delta} = \frac{1 - fe - ejg}{1 - bcf - bcjg - cjh - fe - ejg} \qquad (6.27)$$

The complete sensitivity function can now be written as

$$S_c^T = \frac{1 - e(f + jg)}{1 - c(bf + bjg + jh) - e(f + jg)} \cdot \frac{aed}{abcd + aed} \qquad (6.28)$$

A numerical evaluation would ordinarily eliminate a number of insignificant factors and further simplify the function.

The topological formulation of Equation (6.25) is amenable to further interpretation which sheds light on the significance of the sensitivity function. It will be recalled that the term Δ^o/Δ is the reciprocal of the return difference referred to a specified element. Thus,

$$F_x = \frac{\Delta}{\Delta^o} = 1 - L_x \qquad (6.30)$$

is the return difference referred to the element (branch) x or to be consistent with our previous reference to return difference (see Appendix 5.1), F_x is the return difference referred to an interior node introduced into branch x and L_x is the loop transmittance of the node. If we factor the term $1/F_x$, we may write the sensitivity in the form

$$S_x^T = \frac{\Delta^o}{\Delta} \left(1 - \frac{\Sigma^o}{\Delta^o} \cdot \frac{\Delta}{\Sigma} \right) \qquad (6.31)$$

which may be interpreted as

$$S_x^T = \frac{1}{F_x} \left(1 - \frac{T^o}{T} \right) \qquad (6.32)$$

where T is the graph transmittance and $T^o = T_{(x=0)}$. The form of Equation (6.32) is highly suggestive of the flow graph shown in Figure 6-10 which is referred to as the canonical sensitivity graph. In this portrayal, the reference element, x, is shown in an isolated branch and L_x and F_x are given by

$$L_x = x\,\delta$$

$$F_x = 1 - x\delta$$

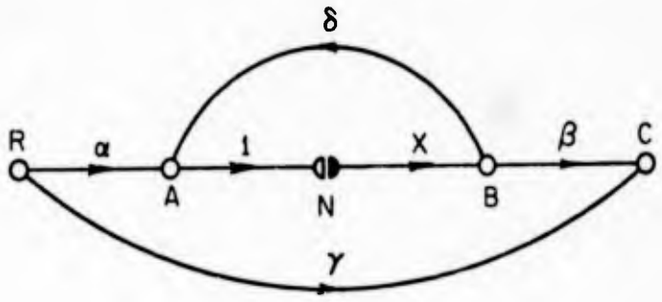


Figure 6-10 Canonical sensitivity graph

The transmittance of the graph (with node N closed) is

$$T = \frac{\gamma(1 - x\delta) + \alpha x \beta}{1 - x\delta} \qquad (6.33)$$

and the leakage transmittance (with node N open) is

$$T^o = \gamma$$

The canonical graph of Figure 6-10 is not important because we wish to convert every flow graph to this form; rather, the fact that this particular form in general permits us to visualize any system in this form with the focus of attention on the reference element. Thus,we gain some further insight into the meanings of sensitivity.

To demonstrate the generality of Figure 6-10, we write equations describing the three node signals A, B,and C:

$$\left.\begin{aligned} A &= \alpha R + \delta B \\ B &= xA \quad \text{(with N closed)} \\ C &= \gamma R + \beta B \end{aligned}\right\} \qquad (6.34)$$

From these relations we can express the branch transmittances (which in general are rational functions of the complex frequency) in the following forms:

$$\alpha = \left.\frac{A}{R}\right|_{B=0} \qquad (6.35a)$$

(The requirement that B = 0 can be met by opening node N so that no transmission can occur around the loop. Basically what is required in the evaluation of α is that no signal shall enter node A except that arriving via branch α. The way in which this requirement is satisfied is of secondary importance.)

$$\beta = \left.\frac{C}{B}\right|_{R=0} \qquad (6.35b)$$

(Here the implication is that unit signal is injected at node B and that no signal arrives at node C except that delivered by branch β.)

$$\delta = \left.\frac{A}{B}\right|_{R=0} \qquad (6.35c)$$

(Again the source R is made zero and unit signal is imagined to be injected at node B.)

$$\gamma = \left.\frac{C}{R}\right|_{B=0} \qquad (6.35d)$$

In each of the evaluations (6.35) it is helpful to think of a node signal (such as the signal at A, B or C) as representing a dependent variable resulting from the superposition of several independent variables acting through appropriate branches. The graph as a whole portrays a <u>set</u> of such equations--but in the case of each individual node equation, there is but one dependent variable--the node signal.

Furthermore, when we actually evaluate the four branch transmittances of Figure 6-10 from the original flow diagram, it is usually helpful to utilize the above equations with, in each case, x set equal to zero. Although x = 0 is not

necessary, visualization of δ as A/ B with R = 0 is ordinarily simpler if x = 0.

In order to illustrate the procedure of reduction to the canonical form we take the example of Figure 6-7(b) which is reproduced again as Figure 6-11. Let us assume that we wish to choose a as the

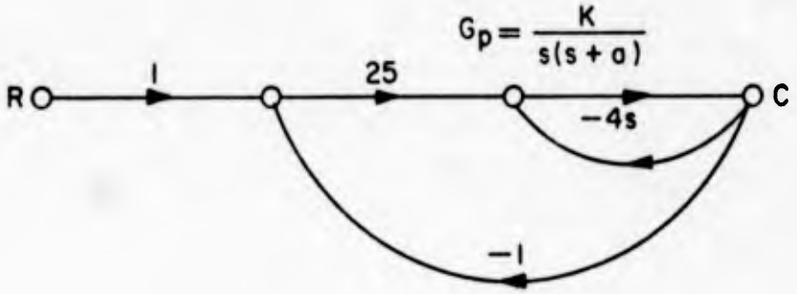


Figure 6-11 System of Figure 6-7(b) repeated

reference element. The first step is to modify the graph so that the element a appears alone in a branch. This is accomplished as shown in Figure 6-12.

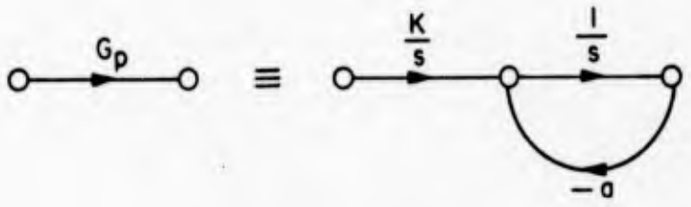


Figure 6-12 Isolation of branch a.

The modified graph is shown in Figure 6-13(a), but there is one more step required inasmuch as we do not wish any other branches or paths to appear in parallel with branch a.

The final isolation is accomplished in part (b) by arranging for branch a to occupy the central position in a tandem arrangement of three branches two of which are unity-transmission branches (actually only one isolator is required here, but if two are always used, all parallel-path difficulties are automatically avoided). With this form, the only branch leaving A is -a, the only branch entering B is -a.

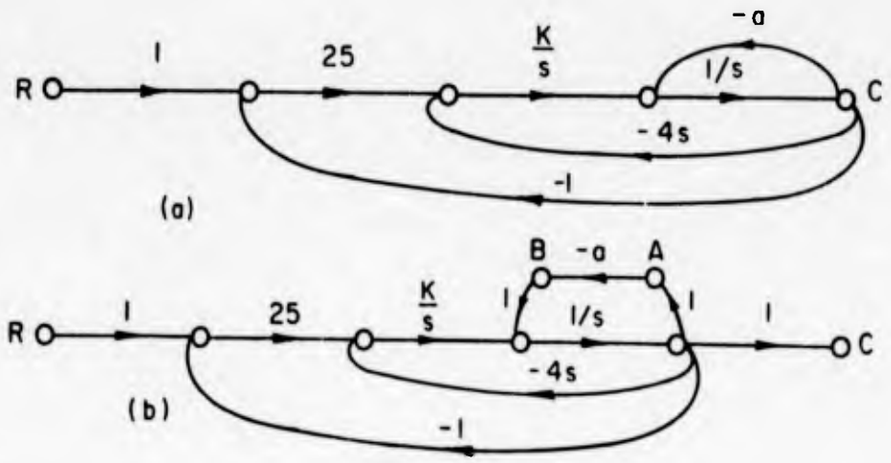


Figure 6-13(a) The modified graph; (b) further isolation of element

The evaluation of the canonical branches is now routine:

$$\left.\begin{aligned}
\alpha &= \frac{A}{R}\bigg|_{B=0} = \frac{\frac{25\,K}{s^2}}{\Delta} \\
\beta &= \frac{C}{B}\bigg|_{R=0} = \frac{\frac{1}{s}}{\Delta} \\
\delta &= \frac{A}{B}\bigg|_{R=0} = \frac{\frac{1}{s}}{\Delta} \\
\gamma &= \frac{C}{R}\bigg|_{B=0} = \frac{\frac{25K}{s^2}}{\Delta}
\end{aligned}\right\} \qquad (6.36)$$

where

$$\Delta = 1 + \frac{4K}{s} + \frac{25K}{s^2}$$

is the determinant of the flow graph of Figure 6-13 with a = o. The canonical graph appears finally as shown in Figure 6-14(a).

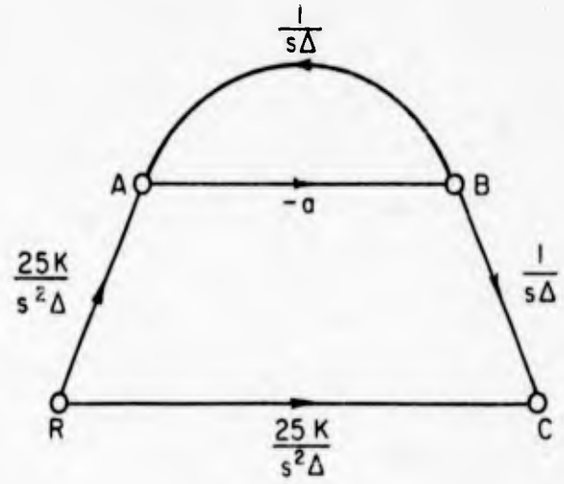


Figure 6-14(a) The canonical form

One final observation should be made while we are on the subject of topological interpretations of sensitivity. The form of Equation (6.25) may be expressed as

$$S_x^T = \frac{1}{F_x} - \frac{1}{F_x'} \qquad (6.37)$$

where $F_x' = \Sigma/\Sigma^o$ is called the null return difference referred to the element x. Null return difference is most conveniently evaluated in terms of the topological definition Σ/Σ^o, but can be computed independently. The normal return difference is defined with reference to a node (whether one of the given nodes of a graph or one inserted interiorly in a branch, as in the case of node N) as

$$F_N = 1 - L_N$$

where N is a node which has been split and where L_N is the source-to-sink transmittance from the source half back to the sink half and with all other excitation signals removed (R = 0 in this case). Here, since the node N has been inserted into branch -a, $F_N = F_{-a}$.

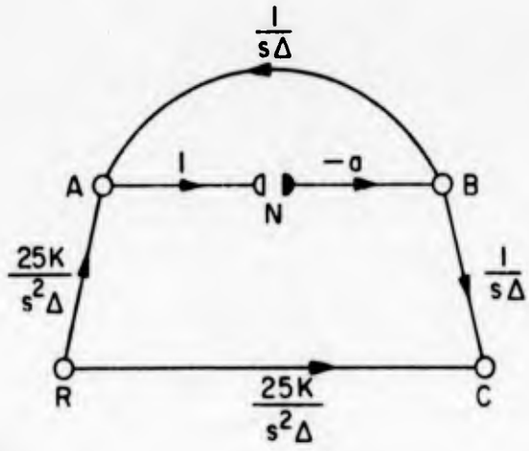


Figure 6-14b Graph for the definition of null return difference

The concept of null return difference is somewhat more complicated inasmuch as it involves not just the feedback loop, but the entire graph. It is again referred to a reference node and has the form

$$F_N' = 1 - L_N'$$

where L_N' is the signal at the sink half of the split node per unit signal at the source half and with the system source, R, adjusted to make the system sink, C, exactly equal to zero. Evidently null return difference depends on the existence of a leakage path (as also shown by the definition Σ/Σ^o) in order that a cancellation of signals through the leakage path and through the path from N to C may take place.

To illustrate the direct evaluation of null return difference referred to element -a of Figure 6-14(a), we use the graph with node N split (Figure 6-14(b)). To determine the null-signal value of R we note that

$$(-a)\frac{1}{s\Delta} + R\frac{25K}{s^2\Delta} = 0$$

and hence

$$R = \frac{as}{25K}$$

Then

$$F'_{-a} = 1 - \left[(-a)\frac{1}{s\Delta} + \left(\frac{as}{25K}\right)\cdot\left(\frac{25K}{s^2\Delta}\right) \right] = 1 \qquad (6.39)$$

which can be checked by evaluating Σ^o/Σ in the canonical graph or in Figure 6-13.

The importance of the concept of null return difference (and the recognition that Σ/Σ^o equals F') derives from the occasional possibility of calculating F' by inspection of the circuit diagram, without formulation of the flow diagram. For example, if the system is such that adjustment of the input R to yield zero output results in zero returned signal for the reference node or element under consideration, we can state at once that F' is unity.

Comprehensive Calculation Techniques

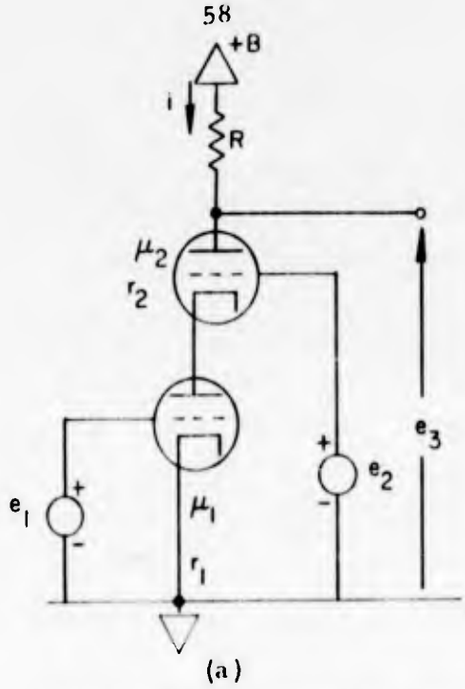
The topological formulation [Equation (6.25)] of the sensitivity function is convenient and efficient when the reference element is isolated in a single branch, and we have demonstrated the fact that this is a condition which can be brought about by transformation of the given graph. However, it is decidedly inconvenient to have to modify a given graph in order to be able to evaluate a sensitivity function. We are strongly interested therefore in exploring computational techniques which avoid the necessity for graph transformations.

Although the system of Figure 6-11 illustrates the need for a graph transformation in order to employ the topological formulation of sensitivity, we turn to a new example furnished by the casode amplifier of Figure 6-15(a), a flow graph for which is given in part (b) of the figure.

The calculation of the sensitivity of T_{13} to variation in r_2 presents a difficulty since r_2 is a term in the function G. We cannot employ Equation (6.25) directly unless we first modify the graph as we did in Figure 6-13(a), a step we now seek to avoid. We can, however, utilize a property of the partial derivative to write

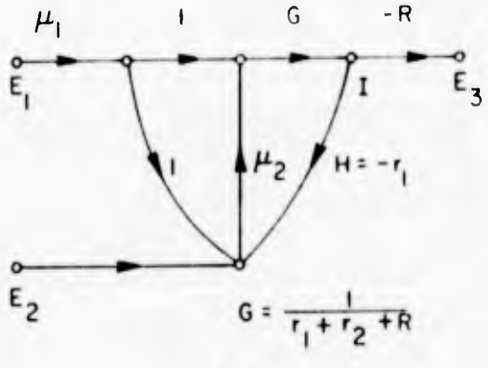
$$S^T_{r_2} = \frac{r_2}{T}\left(\frac{\partial T}{\partial r_2}\right) = \frac{r_2}{T}\left(\frac{\partial T}{\partial G}\cdot\frac{\partial G}{\partial r_2}\right) \qquad (6.40)$$

which, when rearranged becomes



(a)

Figure 6-15 (a) Cascode amplifier



(b)

Figure 6-15 (b) Signal flow graph

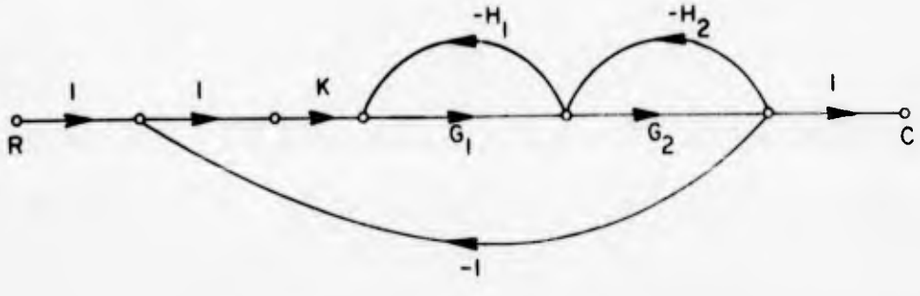


Figure 6-16 Flow graph of a position-control system

$$S_{r_2}^{T} = \left(\frac{\partial T}{\partial G} \cdot \frac{G}{T}\right)\left(\frac{\partial G}{\partial r_2} \cdot \frac{r_2}{G}\right) \tag{6.41}$$

and this is recognized to be

$$S_{r_2}^{T} = S_{G}^{T} \cdot S_{r_2}^{G} \tag{6.42}$$

The first factor on the right is very easily evaluated using the topological formulation of Equation (6.25) and since removal of the branch G opens both paths as well as the single loop,

$$S_{G}^{T} = \frac{1}{\Delta} = \frac{1}{1 + \mu_2 G r_1} \tag{6.43}$$

The second factor on the right can easily be found by straightforward differentiation to yield:

$$S_{r_2}^{G} = \frac{r_2}{G} \quad \frac{(-1)}{(r_1 + r_2 + R)^2} = \frac{-r_2}{(r_1 + r_2 + R)} \tag{6.44}$$

The desired sensitivity function, therefore is

$$S_{r_2}^{T} = \frac{-r_2}{r_2 + R + r_1(1 + \mu_2)} \tag{6.45}$$

The calculation of $S_{r_2}^{G}$ can be made somewhat easier by recalling that one of the original forms of the sensitivity function is

$$S_{x}^{T} = \frac{\partial \ln T}{\partial \ln x} \tag{6.46}$$

Therefore writing

$$\ln G = - \ln (r_1 + r_2 + R)$$

we have

$$S^{G}_{r_2} = \frac{\partial \ln G}{\partial \ln r_2} = \frac{-\partial\left[\ln(r_1 + r_2 + R\right]}{\partial \ln r_2} = - S^{1/G}_{r_2} \qquad (6.47)$$

Utilizing this convenient relation, we invert G and then since the differentiation is now trivially simple, we write by inspection

$$S^{G}_{r_2} = - S^{1/G}_{r_2} = \frac{r_2}{r_1 + r_2 + R}\ (-1) \qquad (6.48)$$

Incidentally, we should also recognize at this point another convenient relation:

$$S^{T}_{x} = - S^{T}_{1/x} \qquad (6.49)$$

The circuit of Figure 6-15 affords a second example of importance. When we attempt to evaluate the sensitivity of T_{13} with respect to r_1 we encounter a new difficulty in that r_1 appears in two different branches; G and H. Again we are anxious to avoid modification of the flow graph, therefore we consider the effect upon T_{13} of the variation of r_1 first in branch G and then separately in branch H and then we superpose the results. Phrased in terms of partial derivatives, what we propose is to write

$$S^{T}_{r_1} = \frac{r_1}{T}\left[\frac{\partial T}{\partial G} \cdot \frac{\partial G}{\partial r_1} + \frac{\partial T}{\partial H} \cdot \frac{\partial H}{\partial r_1}\right] \qquad (6.50)$$

which may be rearranged as

$$S^{T}_{r_1} = \left(\frac{\partial T}{\partial G} \cdot \frac{G}{T}\right)\left(\frac{\partial G}{\partial r_1} \cdot \frac{r_1}{G}\right) + \left(\frac{\partial T}{\partial H} \cdot \frac{H}{T}\right)\left(\frac{\partial H}{\partial r_1} \cdot \frac{r_1}{H}\right) \qquad (6.51)$$

But this is now recognized to be

$$S^{T}_{r_1} = S^{T}_{G} \cdot S^{G}_{r_1} + S^{T}_{H} \cdot S^{H}_{r_1} \qquad (6.52)$$

The first pair of factors on the right of Equation (6.52) yields

$$S_G^T \cdot S_{r_1}^G = \frac{1}{\Delta}\left(\frac{-r_1}{r_1 + r_2 + R}\right) \tag{6.53}$$

From Equation (6.43) we substitute the value of $1/\Delta$ to obtain

$$S_G^T \cdot S_{r_1}^G = \frac{-r_1}{r_2 + R + r_1(1 + \mu_2)} \tag{6.54}$$

The second pair of factors yields

$$S_H^T \cdot S_{r_1}^H = \left(\frac{1}{\Delta} - 1\right)(1) = \frac{-\mu_2 r_1}{r_2 + R + r_1(1 + \mu_2)} \tag{6.55}$$

and combining Equations (6.54) and (6.55) we obtain

$$S_{r_1}^T = \frac{-r_1(1 + \mu_2)}{r_2 + R + r_1(1 + \mu_2)} \tag{6.56}$$

From this example we observe that when an element appears in several branches of a flow graph, the sensitivity of a specified transmittance with respect to the element is a sum of sensitivities with each calculated as though the element varied in that branch and remained constant in all others. To express this situation somewhat generally, if the element x is imbedded functionally in several branches G_1, $G_2 \cdots$ H_1, $H_2 \cdots$

$$S_x^T = S_{G_1}^T \, S_x^{G_1} + S_{G_2}^T \, S_x^{G_2} + \cdots + S_{H_1}^T \, S_x^{H_1} + \cdots \tag{6.57}$$

A somewhat different situation is encountered in the flow graph of Figure 6-16 representing a position-control system.

Here it is specified that

$$G_2 = \frac{K_2}{s(s + c)} \cdot \frac{1 + \zeta\omega_n s}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{6.58}$$

and we wish to investigate the sensitivity of C/R to variations in the damping constant, ζ. We recognize the possibility of expressing the desired sensitivity in the form

$$S_{\zeta}^{T} = S_{G_2}^{T} \; S_{\zeta}^{G_2} \tag{6.59}$$

but the evaluation of $S_{\zeta}^{G_2}$ appears to be some what messy. In accordance with our earlier handling of a rational function, we may write

$$S_x^{F(s)} = \frac{\partial \ln F(s)}{\partial \ln x} = \frac{\partial \ln N(s)}{\partial \ln x} - \frac{\partial \ln D(s)}{\partial \ln x} \tag{6.60}$$

$$= S_x^{N(s)} - S_x^{D(s)} \tag{6.61}$$

where N(s) and D(s) are, respectively, the numerator and denominator of F(s). This makes the evaluation one that can be written by inspection:

$$S_{\zeta}^{G_2} = \left[\frac{\zeta}{1 + \zeta\omega_n s} (s\omega_n) - \frac{\zeta(2\omega_n s)}{s^2 + 2\zeta\omega_n s + \omega_n^2} \right] \tag{6.62}$$

The sensitivity $S_{G_2}^{T}$ is most easily obtained using the topological formulation of Equation (6.25):

$$S_{G_2}^{T} = \frac{\Delta^o}{\Delta} \tag{6.63}$$

$$= \frac{1 + G_1 H_1}{1 + G_1 H_1 + G_2 H_2 + K G_1 G_2} \tag{6.64}$$

Finally, substituting in Equation (6.59) we obtain

$$S_{\zeta}^{T} = \frac{(1 + G_1 H_1)}{(1 + G_1 H_1 + G_2 H_2 + K G_1 G_2)} \cdot \left(\frac{\zeta \omega_n s}{1 + \zeta\omega_n s} - \frac{2\zeta\omega_n s}{s^2 + 2\zeta\omega_n s + \omega_n^2} \right) \tag{6.65}$$

Summarizing Example

In order to show the effectiveness of some of the computational techniques we have developed in this section, we summarize by presenting an example of a fairly complicated system. Known as the cross-coupled inverter, the circuit of Figure 6-17 has been used extensively in instrumentation and audioamplifier practice. Before attempting a signal flow graph, it should be recognized that the circuit

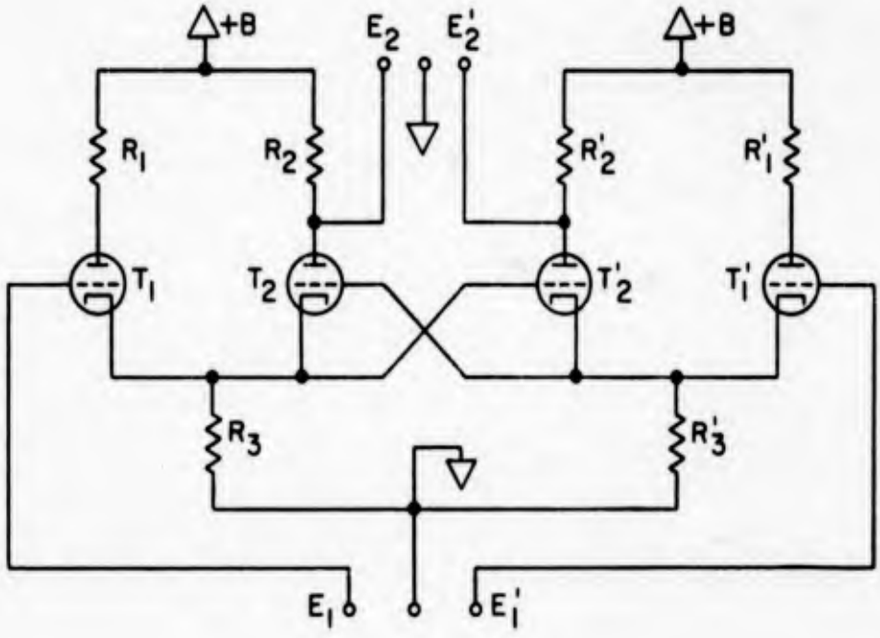


Figure 6-17 Cross-coupled inverter

can be partitioned into two cathode-coupled amplifiers by opening the cross couplings to grids 2 and 2' and temporarily grounding these grids. Flow graphs can now be developed for the separate cathode-coupled amplifiers and the cross links accounted for in a final flow graph.

In developing a flow graph for each amplifier, several possibilities exist, but perhaps the simplest graph results when the dependent controlled sources have been reduced in the circuit model. We show an original model and the reduced model in Figure 6-18.

It is now an easy matter to formulate a flow graph from Figure 6-18(b) such as that shown in Figure 6-19.

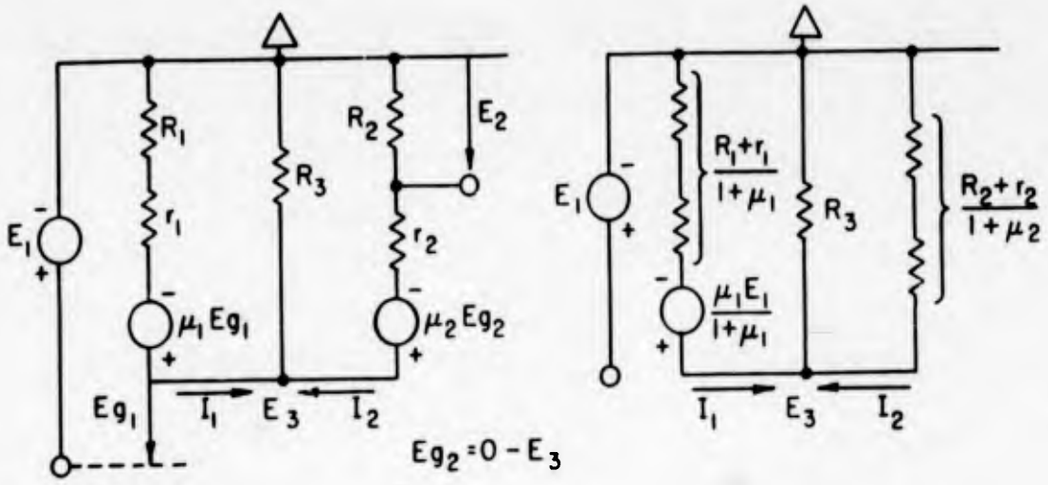


Figure 6-18 Amplifier circuit models

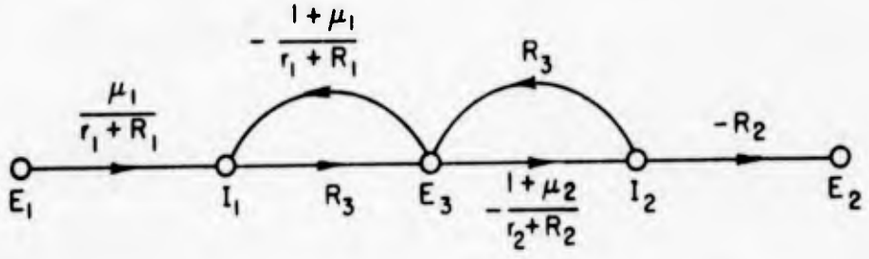


Figure 6-19 Flow graph of cathode-coupled amplifier

The cross coupling can next be taken into account by recognizing that I_2 is a function of E_3 and E'_3 ; in other words

$$I_2 = cE_3 + gE'_3 \tag{6.66}$$

from which we obtain

$$C = \frac{I_2}{E_3}\bigg|_{E'_3 = 0} = -\left(\frac{1+\mu_2}{r_2+R_2}\right) \tag{6.67}$$

which is the branch already in the graph, and

$$g = \left.\frac{I_2}{E_3'}\right|_{E_3 = 0} \tag{6.68}$$

From the circuit it is seen that

$$B = \frac{\mu_2}{r_2 + R_2} \tag{6.69}$$

and now the signal flow graph is completed in Figure 6-20.

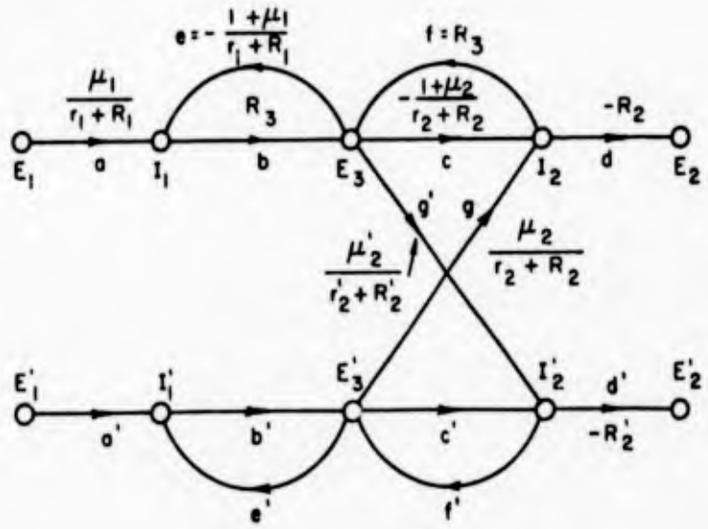


Figure 6-20 Completed flow graph

One of the conveniences of the cross-coupled inverter is its flexibility with respect to balanced and unbalanced input and output signals. It provides a choice of balanced or unbalanced outputs whether it is excited in a single-ended or push-pull mode. We therefore have an interest in examining the balance properties of the circuit and the dependence to these upon the circuit parameters-- and particularly on the tube parameters. One way of studying this aspect of circuit behavior is through the sensitivity functions. As an example, if we wish to learn how the balance of the output with single-ended input depends upon μ_1, we can evaluate the function

$$S_{\mu_1}^{T_{12}/T'_{12}} \qquad \text{which, for convenience we designate as} \qquad S_{\mu_1}^{T/T'} \tag{6.70}$$

In terms of the literal branch labels we note that

$$T \triangleq T_{12} = \frac{abcd\,(1 - b'e' - c'f') + abg'\,f'\,gd}{\Delta} \tag{6.71}$$

and

$$T' \triangleq T_{12}' = \frac{abg'd'\,(1 - b'e')}{\Delta} \tag{6.72}$$

Here is a case where there are numerous opportunities to become enmeshed in a long-winded development, particularly if we have a penchant for "straightforward" approaches. If we resolve Equation (6.70), writing

$$S_{\mu_1}^{T/\,T'} = S_{\mu_1}^{T} - S_{\mu_1}^{T'} \tag{6.73}$$

we may proceed to the expansion:

$$S_{\mu_1}^{T/\,T'} = S_a^T\, S_{\mu_1}^a + S_e^T\, S_{\mu_1}^e - S_a^{T'}\, S_{\mu_1}^a - S_e^{T'}\, S_{\mu_1}^e \tag{6.74}$$

$$= S_{\mu_1}^a \left(S_a^T - S_a^{T'} \right) + S_{\mu_1}^e \left(S_e^T - S_e^{T'} \right) \tag{6.75}$$

We should pause here to take stock and note that

$$S_a^T = S_a^{T'} \text{ and } S_e^T = S_e^{T'} \tag{6.76}$$

therefore

$$S_{\mu_1}^{T/\,T'} = 0 \tag{6.77}$$

which is an interesting (and comforting) result, since μ_1 can vary incrementally at will without affecting the balance of the outputs provided, of course, other quantities remain fixed.

We might have reached this conclusion more directly by exploiting the topological forms of the various functions. We note that Δ is the same for all graph transmittances, therefore in dealing with ratios of transmittances we need be con-

cerned only with Σ (which is not the same for T' as it is for T). We can always resolve the general sensitivity function into two component sensitivities:

$$S_x^T = S_x^\Sigma - S_x^\Delta \tag{6.78}$$

and in this example concentrate our attention on the sensitivity of Σ since the remaining sensitivity is invariant with the choice of source and sink and will therefore cancel out of any calculations of the sensitivity of a ratio of transmittances. We note that

$$S_a^T = S_a^{T'} = 1 \tag{6.79}$$

since removing $\underline{a}$ makes $\Sigma = 0$ but does not affect Δ. We therefore narrow our attention to the functions

$$S_e^\Sigma \quad \text{and} \quad S_e^{\Sigma'} \tag{6.80}$$

which may be written

$$\frac{\Sigma - \Sigma^o}{\Sigma} \quad \text{and} \quad \frac{\Sigma' - \Sigma^{o'}}{\Sigma'} \tag{6.81}$$

and inasmuch as Σ is invariant to the removal of branch e, the two sensitivities of Equation (6.80) are each identically zero. The evaluation then resolves to

$$S_{\mu_1}^{T/T'} = S_a^T - S_a^{T'} = 0 \tag{6.82}$$

The sensitivity of balance with respect to μ_2 is perhaps of greater interest inasmuch as μ_2 is involved in T_{12} and T'_{12} in a complicated way. Here again, since we are dealing with the ratio of transmittances, we need only evaluate the sensitivities of Σ with respect to μ_2. We therefore consider

$$S_{\mu_2}^{T/T'} = S_{\mu_2}^\Sigma - S_{\mu_2}^{\Sigma'} \tag{6.83}$$

but Σ' is not a function of μ_2 (though it is, of course, a function of μ'_2) therefore

$$S_{\mu_2}^{\Sigma'} = 0 \tag{6.84}$$

We have now reduced the evaluation to that of Σ alone, i.e.,

$$S_{\mu_2}^{T/T'} = S_{\mu_2}^{\Sigma} = S_c^{\Sigma} S_{\mu_2}^{c} + S_g^{\Sigma} S_{\mu_2}^{g} \tag{6.85}$$

We can next evaluate the constituents of Equation (6.85) and we note that

$$S_{\mu_2}^{c} = \frac{\mu_2}{1 + \mu_2} \tag{6.86}$$

and

$$S_{\mu_2}^{g} = 1 \tag{6.87}$$

Therefore we reduce the problem to a consideration of

$$S_{\mu_2}^{T/T'} = S_c^{\Sigma} \frac{\mu_2}{1 + \mu_2} + S_g^{\Sigma} \tag{6.88}$$

We might pause here again to assess the situation before embarking on an algebraic exercise. We have two different paths with yield the total function Σ. One path contains the branch c; the other the branch g. Therefore we can show that

$$S_c^{\Sigma} + S_g^{\Sigma} = 1 \tag{6.89}$$

since

$$S_c^{\Sigma} + S_g^{\Sigma} = \frac{\Sigma - \Sigma_c^o}{\Sigma} + \frac{\Sigma - \Sigma_g^o}{\Sigma} = 1 \tag{6.90}$$

because

$$\Sigma_c^o + \Sigma_g^o = \Sigma \tag{6.91}$$

However, Equation (6.88) indicates the S_c^{Σ} carries the factor $\mu_2/(1 + \mu_2)$, therefore the equation is not quite equal to unity, but the approximation is excellent when $\mu_2 >> 1$.

We might look for a moment at the conditions for a balance, i.e., for

$$T_{12} \doteq - T_{12}' \tag{6.92}$$

These can be shown (from Equations (6.71) and (6.72)) to be

$$cd \doteq -g'd' \tag{6.93}$$

and

$$cc' \doteq gg' \tag{6.94}$$

When interpreted in terms of the circuit parameters, these requirements imply that

$$R_2 = R'_2 \tag{6.95}$$

$$\mu_2 \doteq (1 + \mu_2) \tag{6.96}$$

$$\mu_2 = \mu'_2 \tag{6.97}$$

and within the approximation of Equation (6.96) we observe that

$$S_{\mu_2}^{T/T'} \doteq 1 \tag{6.98}$$

This is of course not a good situation since the balance depends directly upon the variations in μ_2 and if μ_2 changes by 1%, the balance also changes by the same percentage. This is a slightly artificial view because the quantity Σ' is sensitive to variations in the alternate amplification factor μ_2'. Perhaps a more realistic evaluation would result from a study of the variations in T_{12}/T_{12}' with respect to variations in μ_2/μ_2', with the assumption made that the normal value of the latter ratio is one, but this point is not pursued.

To conclude this example, we evaluate two important sensitivities,

$$S_{\mu_1}^{T_{12}} \quad \text{and} \quad S_{\mu_2}^{T_{12}}$$

numerically so as to give the reader some sense of scale. We assume the following typical parameter values:

$$\mu_1 = \mu_1' = 20\,, \quad r_1 = r_1' = 10\,\text{K}\,, \quad R_1 = R_1' = 20\,\text{K}$$

$$\mu_2 = \mu'_2 = 100\,,\ r_2 = r'_2 = 50\,K\,,\ R_2 = R'_2 = 200\,K,\ R_3 = R'_3 = 20\,K$$

The sensitivity $S_{\mu_1}^{T_{12}}$.

This may be expressed in expanded form as

$$S_{\mu_1}^{T} = S_a^T \cdot S_{\mu_1}^a + S_e^T \cdot S_{\mu_1}^e \qquad (6.99)$$

and we recognize that

$$S_a^T = 1,\ S_{\mu_1}^a = 1 \text{ and } S_{\mu_1}^e = \frac{\mu_1}{1+\mu_1} \qquad (6.100)$$

therefore

$$S_{\mu_1}^T = 1 + \frac{\mu_1}{1+\mu_1}\, S_e^T \qquad (6.101)$$

Applying the topological formulation (Equation (6.25)) we evaluate

$$S_e^T = \frac{\Delta^o}{\Delta} - 1$$

where from the flow graph of Figure 6-20 and using the method of Appendix 5.1

$$\Delta = (1 - be - cf)(1 - b'e' - c'f') - gfg'f' \qquad (6.102)$$

and

$$\Delta - \Delta^o = - be\,(1 - b'e' - c'f') \qquad (6.103)$$

The numerical values of the branches are

$$b = 20\,,\ c = -\frac{101}{250}\,,\ d = -200\,,\ e = -\frac{21}{30}\,,\ f = 20,$$

$$g = \frac{10}{25}$$

when cancellations of power of ten are made, therefore

$$\Delta = (1 + 14 + 8.08)(1 + 14 + 8.08) - 8^2$$

$$= (23.08)^2 - 64 = 468.7$$

and

$$\Delta - \Delta^{o} = 14(23.08) = 323.1$$

We now evaluate :

$$S_{e}^{T} = \frac{\Delta^{o} - \Delta}{\Delta} = -\frac{323.1}{468.7} = -0.69$$

and then finally

$$S_{\mu_1}^{T} = 1 - \frac{20}{21}(0.69) = 0.343$$

The sensitivity $S_{\mu_2}^{T}$.

Again we might begin with the formal expansion

$$S_{\mu_2}^{T} = S_{c}^{T} S_{\mu_2}^{c} + S_{g}^{T} S_{\mu_2}^{g} \qquad (6.104)$$

recognizing that

$$S_{\mu_2}^{c} = \frac{\mu_2}{1 + \mu_2} \quad \text{and} \quad S_{\mu_2}^{g} = 1 \qquad (6.105)$$

If again we assume $\mu_2/(1 + \mu_2)$ to be approximately one, we may split up the sensitivites as follows:

$$S_{\mu_2}^{T} \doteq (S_{c}^{\Sigma} - S_{c}^{\Delta}) + (S_{g}^{\Sigma} - S_{g}^{\Delta}) \qquad (6.106)$$

and rearranging

$$S_{\mu_2}^{T} \doteq (S_{c}^{\Sigma} + S_{g}^{\Sigma}) - (S_{c}^{\Delta} + S_{g}^{\Delta}) \qquad (6.107)$$

which, based upon earlier work, is

$$S_{\mu_2}^{T} \doteq 1 - (S_{c}^{\Delta} + S_{g}^{\Delta}) \qquad (6.108)$$

The sensitivities of Δ with respect to c and g may be written

$$S_c^{\Delta} = \frac{\Delta - \Delta_c^o}{\Delta} \quad \text{and} \quad S_g^{\Delta} = \frac{\Delta - \Delta_g^o}{\Delta} \tag{6.109}$$

hence

$$S_{\mu_2}^T \doteq 1 - 2 + \frac{\Delta_c^o + \Delta_g^o}{\Delta} \tag{6.110}$$

$$\doteq \frac{\Delta_c^o + \Delta_g^o}{\Delta} - 1 \tag{6.111}$$

Carrying out the operations indicated in Equation (6.111) gives us

$$S_{\mu_2}^T \doteq \frac{(1 - be)(1 - b'e' - c'f')}{\Delta} = \frac{(15)(23.08)}{468.7} \doteq 0.74$$

In these numerical evaluations we have tried to utilize a variety of approaches so as to emphasize the versatility of the various techniques we have presented.

Concluding remarks

In our discussion of sensitivity we have tried to underscore the importance of the concept of sensitivity as a basis for making a quantitative comparison of competetive feedback-system designs. We have then presented a repertoire of computational techniques so that the concept need never be lost in the complexities of its evaluation. Finally, although we have only shown this by implication, the importance of the various viewpoints of sensitivity to the feedback engineer is that from them he gains the required insight to enable him to have a sound, scientific basis for feedback-system design. By examining the separate constitutents of the sensitivity functions he has a basis for the placing of loops and paths that will appropriately control the sensitivities he wishes to reduce.

7. The Significance of Sensitivity

The sensitivity functions for a linear feedback system or active network

provide the fundamental evaluation of the effectiveness of the feedback. Specifically, when we measure the sensitivity of an overall transmission T to a parameter x, we determine the dependence of T on x in very general terms. In terms of the formal definition, S_x^T measures only the percentage change in T resulting from a given (small) percentage change in x; actually, however, we are measuring indirectly the relation between T and x, the allowable tolerances in design and manufacture, the effects of small nonlinearities in the element x, the effects of noise introduced at the point in the system where x appears.

In order to utilize the sensitivity concept in system evaluation and design, we need to develop correlations between the numerical value of the sensitivity and the more familiar criteria of system performance. For example, if the S_x^T is evaluated as -0.4, what is the significance in terms of system characteristics ? From the definition of sensitivity, we can state at once that a 1% increase in x results in a 0.4% decrease in T. As we indicate below, however, there are several other important deductions possible from this value of -0.4.

The problem of interpreting S_x^T is emphasized if we recognize that in almost all cases of significance the sensitivity is a function of the complex frequency s. For example, the simple system of Figure 7-1 is characterized by

$$S_K^T = \frac{s(s+1)}{s^2 + s + K} \qquad (7.1)$$

At zero frequency, the sensitivity is zero (the steady-state value of the step-function response is independent of K as long as K is greater than zero); when $s = -1$, the sensitivity is likewise zero. At any value of $s = j\omega$ other than $\omega = 0$, S involves both a magnitude and a phase angle, or both real and imaginary parts. Our basic question is: what is the significance of the complex-frequency dependence of s ?

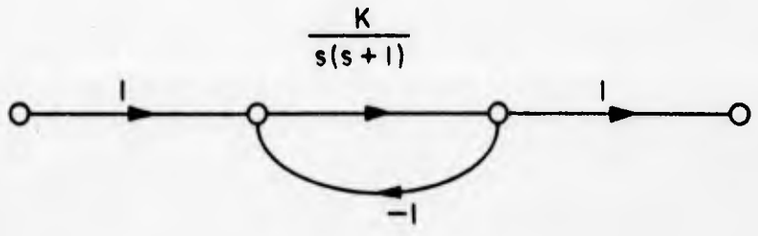


Figure 7-1 Simple feedback system

"Parameter margins"

In system design, x is allowed to assume its normal value x_n. One or more parameters are then adjusted to yield a feedback system with satisfactory overall characteristics. After the design is completed, we determine $S_x^T(s)$. In this section, we demonstrate how S_x^T can be used to determine simply the change in x which results in system instability.

The above concept is illustrated by the simple system of Figure 7-2. If the normal values of a and b are each 2, the system design involves the choice of K to yield satisfactory overall system performance.

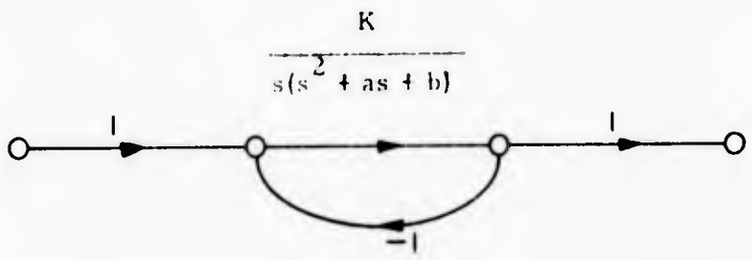


Figure 7-2 Third-order feedback system

By use of root-loci techniques or Bode diagrams, we might select the value of K as 2, with result

$$T = \frac{2}{s^3 + 2s^2 + 2s + 2} \qquad (7.2)$$

(a, b, K)

With the system design completed, evaluation of the effectiveness of the feedback must include consideration of the effects of variation of a and b (perhaps due to manufacturing tolerances, aging, environmental changes, etc.).

In this trivial example, such an investigation can be initiated very simply·

The Routh test, for example, indicates that the system becomes unstable if a is reduced from the normal value of two to a value of unity (and the corresponding frequency of oscillation is $\sqrt{2}$ rad/sec). In order to describe this change quantitatively, we can state that the "parameter margin" for a is -6 db --i.e., a 6-db decrease in a causes instability. The use of the term "parameter margin" here is exactly analogous to the familiar term gain margin: in our example, an increase of K from the value 2 to 4 (i.e., a 6-db increase) causes instability; hence, we say the system possesses a gain margin of 6 db.

Thus, the parameter margin for x is the required change in x to cause system instability. Conventionally, we measure the parameter margin in decibels.

In the trivial example of Figure 2, the parameter margins for both a and b are easily calculated. In more significant feedback systems, however, the Routh test is of little help, and evaluation demands extensive plotting of Bode diagrams or root loci as well as the factoring of high-degree polynomials (or tedious calculation). The sensitivity-function concept provides a simple, convenient basis for the evaluation of the parameter margins. In order to develop this approach, we consider first the special case with the null return difference equal to unity.

Relation between S_x^T and parameter margin for x if $F'_x = 1$

If the null return difference is unity, the overall transmission and sensitivity are given by expressions of the form

$$T = \frac{\Sigma_0}{\Delta_0 + x\Delta_1} \qquad S_x^T = \frac{\Delta_0}{\Delta_0 + x\Delta_1} - 1 \qquad (7.3)$$

where the determinant Δ of the flow graph is written as a linear function of x:

$$\Delta = \Delta_0 + x\Delta_1 \qquad (7.4)^*$$

The numerator of T is simply Σ_0 (rather than the general $\Sigma_0 + x\Sigma_1$) since

*Throughout this analysis we make the assumption that T can be written as a bilinear function of x. Such an assumption is valid if x is any circuit parameter (R, L, C, M), the multiplier of a controlled source (β, α, μ, or g_m), or a single coefficient in any one of the transfer functions describing elements of the system.

$$F' = \frac{\Sigma_0 + x\Sigma_1}{\Sigma_0} \qquad (7.5)$$

and we are considering only the situation in which $F' = 1$ (or $\Sigma_1 = 0$). Thus, the case $F'_x = 1$ corresponds to the special situation in which the numerator of T is independent of the parameter.

As x is varied, the transmission T of Equation (7.3) changes, with the poles described by the relation

$$\Delta_0 + x\Delta_1 = 0 \text{ or } x = -\frac{\Delta_0}{\Delta_1} \qquad (7.6)$$

Since the poles of T are continuous functions of the parameter x, the transition from stability to instability occurs when Equation (7.6) yields poles on the $j\omega$ axis at ω_1, or when

$$x_1 = -\frac{\Delta_0(j\omega_1)}{\Delta_1(j\omega_1)} \qquad (7.7)$$

where x_1 is the value of x resulting in oscillation at the frequency ω_1. Thus, instability occurs when

$$\Delta_1 = -\frac{\Delta_0}{x_1} \qquad (7.8)$$

or, if we substitute in the S_x^T expression of Equation (7.3), when

$$S_x^T = \frac{\frac{x}{x_1}}{1 - \frac{x}{x_1}} = -\frac{x}{x_1 - x} = \frac{1}{\left(\frac{x_1 - x}{x}\right)} \qquad (7.9)$$

The denominator of the last term of Equation (7.9) is simply the fractional change in x required to cause instability.

From the above discussion, we can state the following theorem:

If $F'_x = 1$, the necessary and sufficient condition for a change in x to cause instability at a frequency ω_1 rad/ sec is that $S_x^T(j\omega_1)$ should be real. The required fractional change in x is

$$\frac{x_1 - x_n}{x_n} = \frac{1}{S_x^T(j\omega_1)} \tag{7.10}$$

where x_n is the normal design value of x and x_1 is the value which results in instability.

The significance of the theorem can be illustrated by the simple example depicted in Figure 7-3, where $\underline{a}_n$, the normal value of $\underline{a}$, is two. In this case, the sensitivity of T to $\underline{a}$ is calculated directly by the techniques of the last section:

$$S_a^T = S_G^T S_a^G$$

$$= \frac{-1}{1 + G(s)} S_a^{s^2 + as + 2} = -\frac{as}{s^2 + as + 2} \frac{1}{1 + G} \tag{7.11}$$

$$S_a^T = -\frac{as^2}{2} \frac{G}{1 + G}$$

Thus, the condition for oscillation in this case is that $G/(1 + G)$ should be real (since s^2 is real when $s = j\omega$).

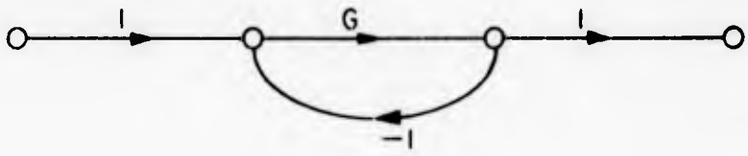
In the specific system of Figure 7-3,

$$\frac{G}{1 + G} = \frac{2}{s^3 + as^2 + 2s + 2} \tag{7.12}$$

If this function is to be real, the denominator polynomial must be real:

$$s^3 + 2s = 0 \qquad \begin{matrix} \omega_1 = 0 \\ \omega_2 = \sqrt{2} \end{matrix} \tag{7.13}$$

$$G = \frac{K}{s(s^2 + as + 2)} \qquad K = 2$$



$K = 2$

Figure 7-3 Example for theorem

At these two frequencies at which S_a^T is real, the sensitivity assumes the values

$$\omega_1 = 0 \quad S_a^T(j\omega_1) = 0 \qquad \omega_2 = \sqrt{2} \quad S_a^T(j\omega_2) = \frac{a}{1 - a} = -2 \tag{7.14}$$

The fractional changes in a required to cause instability are the reciprocals of the sensitivity values, or:

$$\omega_1 = 0 \quad \frac{a_1 - 2}{2} = \infty \qquad \omega_2 = \sqrt{2} \quad \frac{a_2 - 2}{2} = -\frac{1}{2} \ (-6 \text{ db}) \tag{7.15}$$

$$a_1 = \infty \qquad a_2 = 1$$

Thus, the only solution of interest (that for ω_2 and a_2) indicates instability at a frequency $\sqrt{2}$ rad/sec results if a is changed - 6 db from the normal value of 2.

The above result is, of course, no different from that obtained previously from the Routh test; indeed, in this trivial example, the Routh test is considerably simpler that the analysis via the sensitivity function. The important feature of the method, however, lies in the form of Equation (7.11), which states

$$S_a^T = -\frac{as^2}{2}\,\frac{G \rightarrow}{1+G} \qquad (7.16)$$

The parameter margin for a is determined entirely by the value of S_a^T when this function is real.

The parameter margin for a in the much more complicated (and interesting) system of Figure 7-4, in which the $G_1(s)$ block has been added to the earlier system of Figure 7-3, is determined from an equation of almost the same form. In this case

$$S_a^T = -\frac{as^2}{2}\,\frac{G_2 \rightarrow}{1+G}$$

For this fifth-order system, the Routh test immediately embroils the engineer in depressingly complicated calculations culminating the determination of the real zeros of a quartic polynomial.

$$G_1 = K_1\,\frac{(s^2+\alpha s+\beta)}{s^2+\gamma s+\delta} \qquad G_2 = \frac{2}{s(s^2+as+2)}$$

$$G = G_1 G_2$$

Figure 7-4

Once the gain and phase plots of $G_2/(1+G)$ are available [actually we only need the plots in the vicinity of the frequencies at which $G_2/(1+G)$ is real], determination

of the parameter margin for a is trivial. Furthermore, the plots of $1/(1 + G)$ and G_2 are ordinarily drawn in the frequency-domain design of the system, so that the plots for $G_2/(1 + G)$ can be drawn without a significant amount of extra work.

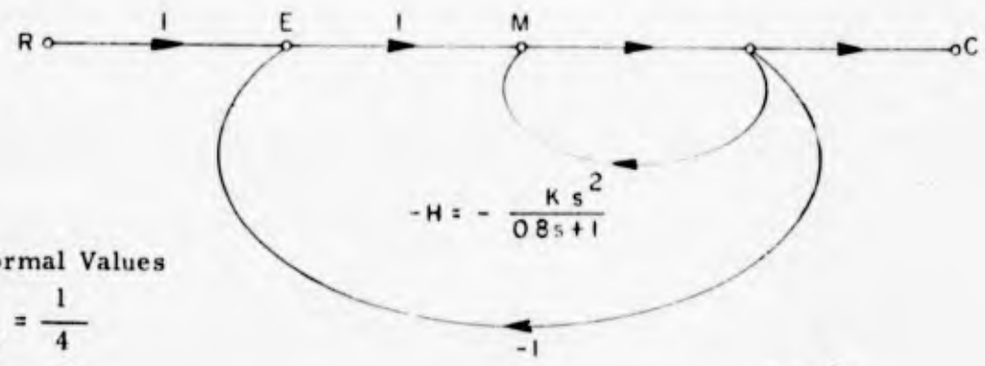
Thus, if the system is initially designed on the basis of frequency analysis, the determination of the parameter margins is simple and straightforward if we exploit the sensitivity concept.

A Non-Trivial Example

As indicated in the preceding paragraphs, the method described above is primarily useful in the study of reasonably complex systems. Once we plot the pertinent transfer functions for determination of the sensitivity (plots which ordinarily are developed in the course of normal system design), the evaluation of the various parameter margins are straightforward.

In order to illustrate the potency of the method, we consider the system in Figure 7-5.* Conventional, frequency-domain design in terms of the Bode plots and the closed-loop characteristics determined from the Nichols chart results in the selection of the parameter values shown in Figure 7-5

$$G = \frac{100}{s(\tau_1 s + 1)(\tau_2 s + 1)}$$



Normal Values

$\tau_1 = \frac{1}{4}$

$\tau_2 = \frac{1}{16}$

$K = -\frac{1}{6}$

$T = \frac{C}{R}$

Figure 7-5 Control system

* The system and closed-loop characteristics are taken from the book by H. Chestnut and R. Mayer "Servomechanism and Regulating System Desgin," John Wiley and Sons, New York, New York, 2nd edition, 1958, p. 373.

The resulting closed-loop gain and phase characteristics for $\frac{C}{R}(s)$ are shown in Figure 7-6.

In order to evaluate the final system which has been designed, we next need to determine the effects of nonlinearity, of environmental changes, and of aging and manufacturing tolerances. From the expected variations and our understanding of the Physical principles underlying the operation of the various elements of the system, we might decide that primary interest lies in the effects resulting from variation of the three parameters τ_1, τ_2 and K. A simple measure of the sensitiveness of system performance to each parameter is the parameter margin or the fractional parameter change required to cause instability.

The first step in the evaluation of the three parameter margins is the determination of the three sensitivity functions. On the basis of the methods described above, we can write

$$S^T_{\tau_1} = S^G_{\tau_1} \; S^T_G$$

$$= - S_{\tau_1}^{\tau_1 s + 1} \; S^T_G$$

$$= - \frac{\tau_1 s}{\tau_1 s + 1} \cdot \frac{1}{1 + G + GH} \qquad (7.17)$$

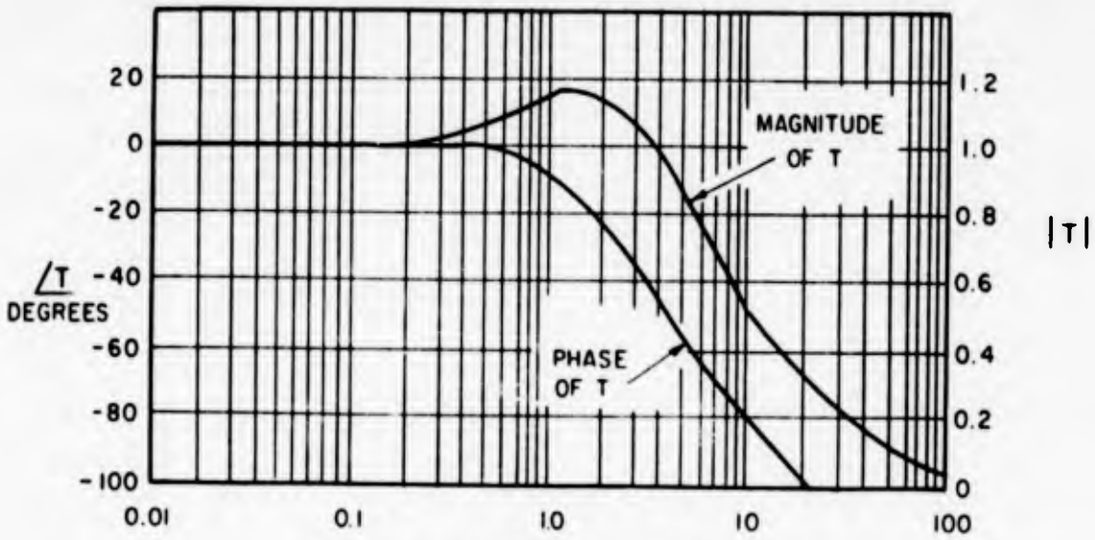
$$= - \frac{\tau_1 s^2 (\tau_2 s + 1)}{100} \; \frac{G}{1 + G + GH} = \frac{\tau_1 s^2 (\tau_2 s + 1)}{100} \; T$$

At the normal operating values $(\tau_1 = \frac{1}{4}, \; \tau_2 = \frac{1}{16})$,

$$S^T_{\tau_1} = - \frac{s^2 (s + 16)}{.6400} \; T \qquad (7.18)$$

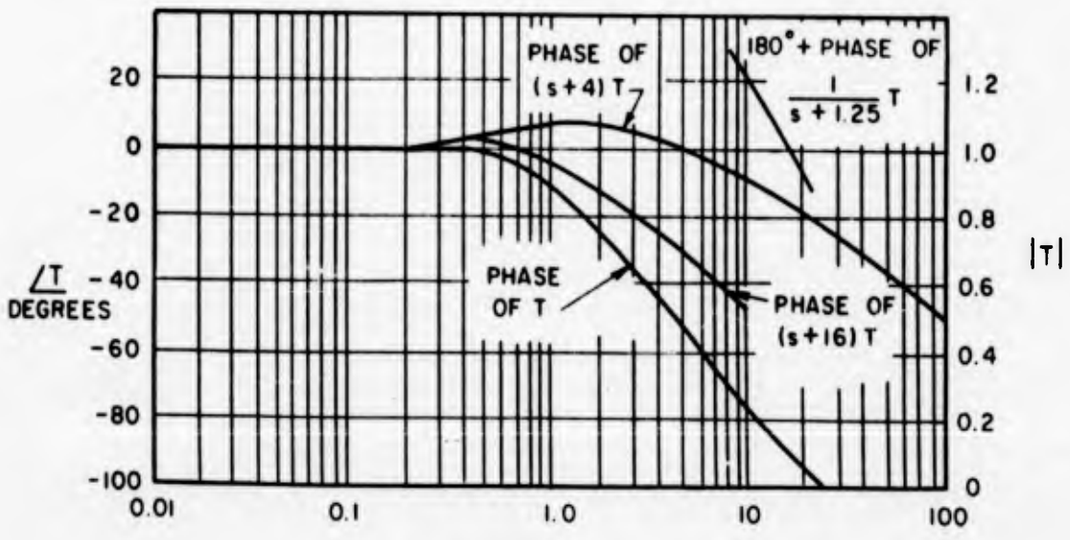
Correspondingly,

$$S_{\tau_2}^{T} = -\frac{s^2(s+4)}{6400}\,T \qquad (7.19)$$



ω in radians per second

Figure 7-6



ω in radians per second

Figure 7-7

The sensitivity of T to K is

$$S_K^T = S_K^H \; S_H^T$$

$$= S_H^T \qquad (7.20)$$

$$= \frac{\Delta(H = 0)}{\Delta} - 1$$

since the null return difference with reference to H is unity be inspection of the flow diagram of Figure 7-5. Thus, with Δ simply 1 + G + GH, we have

$$S_K^T = \frac{- GH}{1 + G + GH} = -HT \qquad (7.21)$$

or, at the normal value of K of 1/6,

$$S_K^T = - \frac{1}{4.8} \cdot \frac{s^2}{s + 1.25} \; T \qquad (7.22)$$

In order to determine the three parameter margins, we first note that each null return difference is unity. According to the theorem, then, we need only find the value of the sensitivity when S is real. Thus, in the three cases, we seek real values of

$$S_{\tau_1}^T = - \frac{s^2}{6400}(s+16)T \qquad S_{\tau_2}^T = - \frac{s^2}{6400}(s+4)T \qquad S_K^T = - \frac{s^2}{4.8} \; \frac{1}{s+1.25} T \qquad (7.23)$$

when $s = j\omega$. Rather than considering the entire functions, we can seek real values of :

For τ_1: $(s+16)T$ For τ_2: $(s+4)T$ For K: $\dfrac{1}{s + 1.25} T$

The simplest procedure is to sketch, directly on the plot of Figure 7-6, the phase angles of $(s + 16)T$, $(s + 4)T$, and $1/(s + 1.25)T$ in the vicinity where these phase angles are 0° or 180°. (We note that a high degree of accuracy is not at all

necessary since the magnitude of T is not changing rapidly).

The three sketches, shown in Figure 7-7, indicate that for each parameter there is only one frequency at which the relevant function is real. The frequencies and the corresponding values of $|T|$ are:

For τ_1	For τ_2	For K
$\omega_1 = 0.7$ rad/sec	$\omega_2 = 4.3$ rad/sec	$\omega_3 = 18$ rad/sec
$\lvert T(j\omega_1)\rvert = 1.07$	$\lvert T(j\omega_2)\rvert = 0.93$	$\lvert T(j\omega_3)\rvert = 0.33$

(7.24)

We are now ready to calculate the three values of sensitivity from the expressions of Equation (7.23). Substitution of the above ω (or s) and $|T|$ values gives:

For τ_1	For τ_2	For K
$S^T_{\tau_1}(j\omega_1) = \frac{0.5}{6400}\,16(1.07)$	$S^T_{\tau_2}(j\omega_2) = \frac{(4.3)^2\lvert 4+j4.3\rvert}{6400}\,0.93$	$S^T_K(j\omega_3) = \frac{-18}{4.8}\,0.33$
$= \frac{1}{748}$	$= \frac{1}{63}$	$= -\frac{1}{0.8}$

(7.25)

Therefore, we can move the system to the borderline of instability with any one of the following fractional changes in a parameter:

$\frac{\tau_{1\,osc} - \tau_1}{\tau_1} = \mp 748$	$\frac{\tau_{2\,osc} - \tau_2}{\tau_2} = 63$	$\frac{K_{osc} - K}{K} = -0.8$
Increase of 57.5db	Increase of 36 db	Decrease of 14 db

(7.26)

If we prefer, we can solve each of these solutions for the three parameter values--i.e., for the values, any one of which causes instability if the other two parameters assume normal values:

$$\tau_{1_{osc}} = 187 \qquad \tau_{2_{osc}} = 4 \qquad K_{osc} = \frac{1}{30} \tag{7.27}$$

Several aspects of the above example are noteworthy in conclusion:

(1) The results, Equation (7.27) or (7.26) are obtained with surprisingly little work once the initial, conventional design is completed for the control system.

(2) The effort involved is, to a large extent, independent of the complexity of the system. For example, if the system is complicated by the inclusion of G_1 block as depicted in Figure 7-8, the three sensitivity functions,

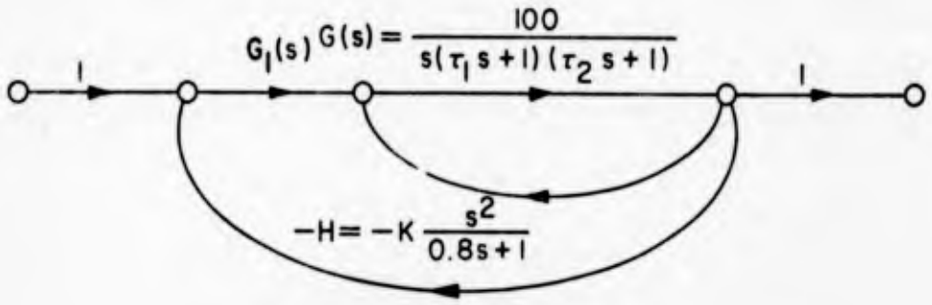


Figure 7-8 More complex system

given previously in Equation (7.23), become

$$S^T_{\tau_1} = -\frac{s^2}{6400}(s+16)\frac{T}{G_1}$$

$$S^T_{\tau_1} = -\frac{s^2}{6400}(s+4)\frac{T}{G_1}. \tag{7.28}$$

$$S^T_K = -\frac{s^2}{4.8}\,\frac{1}{s+1.25}\,\frac{T}{G_1}$$

In other words, instead of the T we used previously, we now require the gain and phase characteristics for T/G_1. But in the course of the normal system design, the gain and phase of G_1 would have been plotted; hence T/G_1 can be obtained simply by combining the T and the G_1 characteristics. Once T/G_1 is evaluated, the entire analysis for the determination of the parameter margins proceeds exactly as in the above example. The complexity of G_1 is irrelevant !

(3) The accuracy required in the evaluation of the frequency at which the sensitivity is real depends on the value of S and on the behavior of T in the neighborhood of the frequency in question. For example, in the above example with the three sensitivities

$$S^T_{\tau_1} = \frac{-s^2(s+16)}{6400}\,T \qquad S^T_{\tau_2} = \frac{-s^2(s+4)}{6400}\,T \qquad S^T_K = \frac{-s^2}{4.8(s+1.25)}\,T \tag{7.29}$$

the first two sensitivities are very small (i.e., very large fractional changes in τ_1 and τ_2 are required to cause instability). Under these circumstances, we would ordinarily be relatively unconcerned about accuracy. Furthermore, S is real in the frequency range in which T is changing rather slowly, so that the principal effect of inaccurate frequency determination is evidenced in error in the s factors, $s^2(s+4)$. In the case of the K variation, however, the sensitivity is close to unity; in this case, reasonable accuracy is required. For example, if S^T_K is measured as $-1/0.9$ rather than $-1/0.8$, the K for instability is $1/60$, or half of the value actually obtained. This demand for accuracy is largely offset, however, by the fact that the large sensitivity magnitude is associated with a frequency range well outside the bandwidth of the system; hence the phase of T is changing rapidly in the frequency range of interest , and an accurate determination of the frequency at which S^T_K is real is possible. Thus, when the feedback is effective in reducing sensitivity, poor accuracy is permissible; when the feedback is ineffective and accuracy is required, we can evaluate with reasonable precision.

(4) Finally, we should emphasize once more the condition originally imposed; the null return difference must be unity if the theorem is to be applicable (we next consider the case when F'_x is not unity). Unfortunately, there does not seem to be any simple technique for determination from the signal-flow graph of the conditions under which the null return difference is unity. From the basic definition, we know that F'_x is unity if the input of the x branch is proportional to the output when the

flow diagram is drawn in such a way that x appears only once and as the multiplier of a single branch.*

More generally, from the formula of Section 6, we know that

$$F'_x = \frac{\Sigma}{\Sigma_0} = \frac{\Sigma P_i \Delta_i}{\Sigma P_i \Delta_i \,|x=0} \tag{7.30}$$

Hence, F'_x is unity when x does not appear in any direct path (any P_i above) or in any loop which does not touch all of the direct paths.

Relation between S_x^T and parameter margin if $F'_x = \infty$

The entire discussion above is restricted to the special case when F'_x equals unity -- in other words, when

$$\Sigma = \Sigma_0 \tag{7.31}$$

where Σ represents $\sum_i P_i \Delta_i$ in Mason's formula and Σ_0 is the value of Σ when $x = 0$. The general form of the overall transmission T for a feedback system is

$$T = \frac{\Sigma_0 + x\Sigma_1}{\Delta_0 + x\Delta_1} = \frac{\Sigma}{\Delta} \tag{7.32}$$

where Σ_0, Σ_1, Δ_0, and Δ_1 are independent of x.** The special case we have been considering ($F'_x = 1$) corresponds to $\Sigma_1 = 0$, or

$$T = \frac{\Sigma_0}{\Delta_0 + x\Delta_1} \tag{7.33}$$

* The null return difference F'_x is unity minus the returned signal when the x branch is broken at the input, a unit signal is transmitted through x, and the system input is adjusted to yield zero output. The "returned signal" is the signal at the receiving side of the break.

** Equation (7.32) is valid only when T is a bilinear function of x, but this constraint does not seriously limit the applicability of the entire discussion here.

An equally important special case occurs when $\Sigma_0 = 0$, or

$$T = \frac{x\Sigma_1}{\Delta_0 + x\Delta_1} \tag{7.34}$$

Under these circumstances, there is no leakage transmission: when x is set equal to zero, the overall transmission becomes zero. The null return difference, F'_x, is infinite and the sensitivity is simply the reciprocal of the return difference with reference to x:

$$S_x^T = \frac{1}{F_x} \tag{7.35}$$

When the leakage transmission is zero, the sensitivity is again directly related to the corresponding parameter margin. The theorem is derived in exactly the same manner as in the previous case. Instability again depends upon the change in x moving the zeros of the denominator of T

$$T = \frac{x\Sigma_1}{\Delta_0 + x\Delta_1} \tag{7.36}$$

onto the imaginary axis. For this value of x and the particular frequency ω_1

$$\Delta_0(j\omega_1) + x\Delta_1(j\omega_1) = 0 \tag{7.37}$$

The corresponding value of sensitivity is, since the leakage is zero,

$$S_x^T = \frac{\Delta_0(j\omega_1)}{\Delta_0(j\omega_1) + x_n\Delta_1(j\omega_1)} \tag{7.38}$$

where x_n is the normal value of x. Substitution of Equation (7.37) in (7.38) gives:

$$S_x^T = \frac{-x\Delta_1(j\omega_1)}{-x\Delta_1(j\omega_1) + x_n\Delta_1(j\omega_1)} \tag{7.39}$$

If $\Delta_1(j\omega_1) \neq 0$, the sensitivity is real and has the form:

$$S_x^T = \frac{x}{x - x_n} \tag{7.40}$$

where x is the value which causes instability. From Equations (7.37) and (7.40), we can then state the following theorem:

> If $F'_x = \infty$, the necessary and sufficient condition for a change in x to cause instability at a frequency ω_1 rad/ sec is that $S_x^T(j\omega_1)$ should be real, and the required fractional change in x is
>
> $$\frac{x_1 - x_n}{x_n} = \frac{1}{S_x^T(j\omega_1) - 1} \tag{7.41}$$
>
> where x_n is the normal design value of x and x_1 is the value which results in instability.

Thus, the only difference between this case ($F'_x = \infty$) and the previous special case ($F'_x = 1$) resides in the form of the equation for the fractional change in x. In the $F'_x = 1$ case, the fractional change is $1/ S_x^T(j\omega_1)$, in contrast to Equation (7.41). Thus, from a design viewpoint, when we wish a large required fractional change in x, we seek near-zero values of S when $F'_x = 1$, but near-unity values of S when $F'_x = \infty$.

The analysis when $F'_x = \infty$ follows exactly the steps described in the examples illustrating the theorem when $F'_x = 1$.

Relation between S_x^T and parameter margin in general

We have now developed two general theorems covering the special cases of unity and infinite null return differences. In the general situation, F'_x assumes neither value. In general, however, T can be written as

$$T = \frac{\Sigma_0 + x\Sigma_1}{\Delta_0 + x\Delta_1} = \frac{\Sigma_0}{\Delta_0 + x\Delta_1} + \frac{x\Sigma_1}{\Delta_0 + x\Delta_1} \tag{7.42}$$

In other words, T can always be considered as the sum of two transmissions: T_1 in which $F'_x = 1$, and T_2 in which $F'_x = \infty$. Instability which occurs because, as x changes, $\Delta_0 + x\Delta_1$ possesses a zero on the $j\omega$- axis results in both T_1 or T_2 exhibiting this oscillatory behavior. Consequently, we can investigate either T_1 or T_2 in order to determine the parameter margin.

Alternatively, if we recognize that it is only the character of $\Delta_0 + x\Delta_1 = \Delta$ which we seek to study, we can study any system which possesses the same determinant as the original system. The technique is illustrated by the example depicted in Figure 7-9, in which the parameter of interest is a .

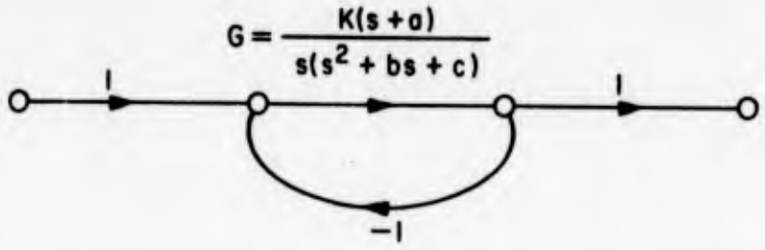


Figure 7-9 Simple system with $F'_x \neq 1$, $F'_x \neq \infty$

In order to simplify the determination of an appropriate system for evaluation of the parameter margin, we re-draw the flow diagram as shown in Figure 7-10.

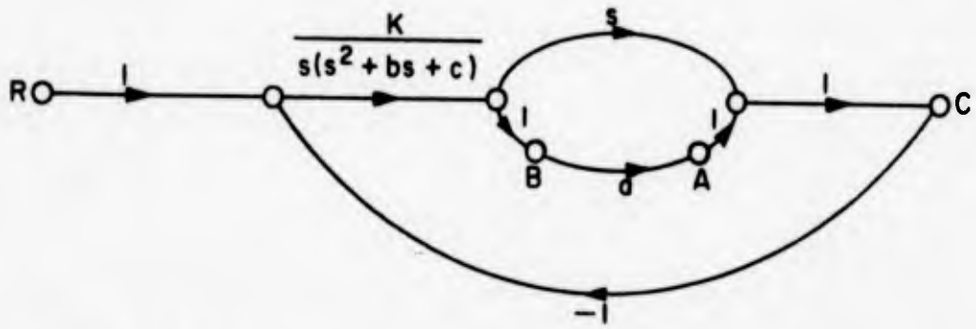


Figure 7-10 Modified form of Figure 7-9 with a appearing alone as multiplier of one branch.

If we now consider point A as the output of the system, we are dealing with the special case of no leakage ($F'_a = \infty$), so that the second theorem above is directly applicable. Alternatively, we can consider B as the output, with the result that $F'_a = 1$ and the Δ is again unchanged.

To complete this example, we choose the latter alternative (B as the output). The resulting sensitivity of the transmission, B/ R, to changes in a is determined by the same determinant as that Δ characterizing the original system:

$$S_a^{B/R} = \frac{\Delta(0)}{\Delta} - 1$$

$$= \frac{1 + \dfrac{K(s+a)}{s(s^2+bs+c)}}{1 + \dfrac{Ks}{s(s^2+bs+c)}} - 1 \qquad (7.43)$$

$$S_a^{B/R} = \frac{-aK}{s^3 + bs^2 + (c+K)s + aK} \qquad (7.44)$$

$$S_a^{B/R} = \frac{-a}{s+a} \quad \frac{G}{1+G} \qquad (7.45)$$

Equation (7.45) is directly useful in the general case in which $G/(1+G)$ is already plotted; in this simple example, we can proceed analytically from Equation (7.44).

The sensitivity is real at the frequency

$$-\omega^3 + (c+K)\omega = 0 \quad \text{or} \quad \omega_1 = \sqrt{c+K} \qquad (7.46)$$

The corresponding value of sensitivity is

$$S_a^{B/R} = \frac{-aK}{aK - b(c+K)} \qquad (7.47)$$

Hence, the value of a causing instability (i.e., a_1) is given by

$$\frac{a_1 - a}{a} = \frac{aK - b(c+K)}{-aK}$$

or

$$a_1 = b \frac{c + K}{K} \tag{7.48}$$

Concluding comments

In these paragraphs we have outlined a straightforward analytic and graphical technique for the determination of the parameter margins. The simplicity of the method derives from:

(1) The ease with which the various sensitivity functions can be evaluated.

(2) The fact that the various sensititivites are closely related to the functions which are plotted in normal closed-loop system design.

(3) The simplicity of the two theorems--in particular the fact that we need only determine the value of the sensitivity when this S is real. Even in the most general cases we are required to evaluate S only in the small portion of the frequency band.

Thus, the parameter margins can be determined in the course of the normal system analysis and design.

The parameter margins are important for a variety of reasons:

(1) In system evaluation, the margins measure the sensitiveness of system stability to specific parameters. Hence, we are provided with quantitative measures of the effectiveness of feedback in achieving self-calibration and the insensitivity of system dynamics to particular parameter variations.

(2) The simplicity of the expressions for the parameter margins allows the guiding of feedback-system design toward the realization of specified minimum values for the margin of those parameters which vary significantly, because of either manufacturing tolerances or environmental changes.

(3) The simplicity of the analysis also permits evalution of the relative values of various parameter margins. In most cases,

the different sensitivity functions possess common factors as well as the distinctive factors, so that the designer can see at once the possibility of controlling either simultaneously or independently the significant parameter margins. As a simple example, the difference in the two theorems pinpoint the difficulty of achieving simultaneously large parameter margins for two elements, one of which has $F'_x = \infty$, the other $F'_x = 1$. In this case, we must choose a configuration resulting in significantly different forms for the two sensitivity functions.

(4) The theory presented here provides a possible approach to the identification problem. Experimental measurement of the sensitivity and the transmission, for example, permits evaluation of specific coefficients or parameters in the internal transfer function, even though we may not be free to vary a parameter to the point at which the system becomes oscillatory.

(5) Finally, the theory is directly applicable to the identification and evaluation of multiple nonlinearities within the system. When the nonlinearity can be represented by a parameter variation with signal level (e.g., by a describing function), the theory presented above yields quantitative measures of the tendency of each nonlinearity to cause instability, the relative significance of the different nonlinearities, or the magnitude of the nonlinearity from the measured frequency and amplitude of oscillation.

8. Multi-loop System

Throughout the preceding sections our principal interest has focussed on feedback systems which are single-loop: there is only one feedback loop or closed loop around which signals may circulate. In Section 5 we considered multi-loop systems briefly as we presented the general techniques for the calculation of transmission; in this section we wish to look further at the properties and characteristics of such multi-loop systems.

As we noted in Section 2, the primary motivation for the use of feedback arises, if we exclude the special case of oscillators, from the possibility of controlling the sensitivity of the overall transmission to changes in various system parameters. Alternatively, we can consider this advantage of feedback as the

possibility of controlling two different transmissions independently. The equivalence of the two viewpoints is illustrated by Figure 8-1. Here we have a process (or known system); at the input of this process, there enters a corruption or noise signal which affects the output; we desire an overall system in which a specified relation between C and R is realized, and at the same time the effect of U on the output C is controlled.

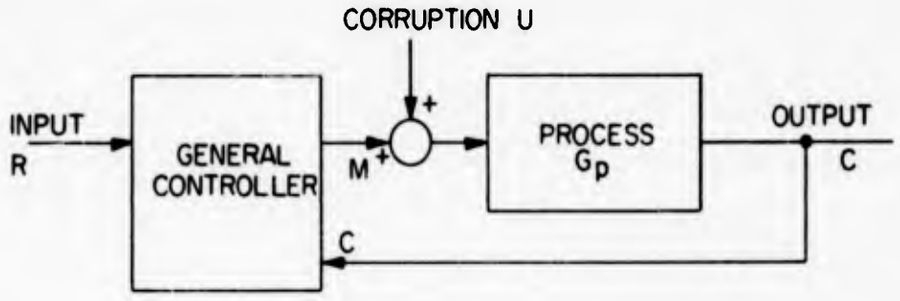


Figure 8-1 System with two inputs

Independent control of the two transmissions, C/ R and C/ U, is possible if we insert a general, linear controller as shown (the inputs to the controller are R and C, the output M). For this system, we can write the relation

$$C = G_p (M + U) \tag{8.1}$$

If the system is linear, M is some linear function of R and C; in other words

$$M = AR - BC \tag{8.2}$$

where A and B are two transfer functions which depend upon the particular form of the controller. Substitution of Equation (8.2) into Equation (8.1) and solution for C yields

$$C = \frac{AG_p}{1 + BG_p} R + \frac{G_p}{1 + BG_p} U \tag{8.3}$$

Equation (8.3) demonstrates the equivalence of control over sensitivity and control over the effects of corrupting signals, since the feedback results in reduc-

tion of the effect of U by division of C/U by the factor $(1 + BG_p)$, and the sensitivity of C/R with respect to G_p is likewise

$$S_{G_p}^{C/R} = \frac{1}{1 + BG_p} \tag{8.4}$$

In both cases, the return difference $(1 + BG_p)$ (i.e., unity minus the loop gain) measures the effectiveness of feedback.

Equation (8.3) indicates at once the possibility of independent control over the two transmissions, C/R and C/U. If we select both functions, the controller function A is given simply by

$$A = \frac{C/R}{C/U} \tag{8.5}$$

and B can be determined from either C/R or C/U. For example,

$$B = \frac{1}{C/U} - \frac{1}{G_p} \tag{8.6}$$

The concept of such independent control is intriguing, but the design Equations (8.5) and (8.6) are somewhat misleading in their simplicity. For example, if we wish to have our system reject completely the effects of U (or have zero sensitivity to changes in G_p), we must make C/U identically zero--a value which, with G_p given, can be achieved only if B is infinite. But, with B infinite, C/R is non-zero only if A likewise is infinite so that C/R reduces to

$$\frac{C}{R} \rightarrow \frac{A}{B} \tag{8.7}$$

Whether such an ideal situation can be realized in actual practice demands a more careful investigation of the available techniques for realizing the controller of Figure 8-1.

Postive and Negative Feedback

The realization of a suitable controller to achieve zero sensitivity concur-

rently with a specified transmission can be accomplished theoretically by a combination of at least two feedback paths with different signs on the loop gains,* as depicted in one simple case in Figure 8-2. Here the controller consists of the three separate transfer functions, G_1, G_2 and G_3; the overall system contains two loops, with the loop gains, $-G_2 G_p G_3$ and $+ G_2 G_1 G_3$ (one with a positive sign, the other with a negative sign, or the former negative feedback and the latter positive feedback in the vernacular).

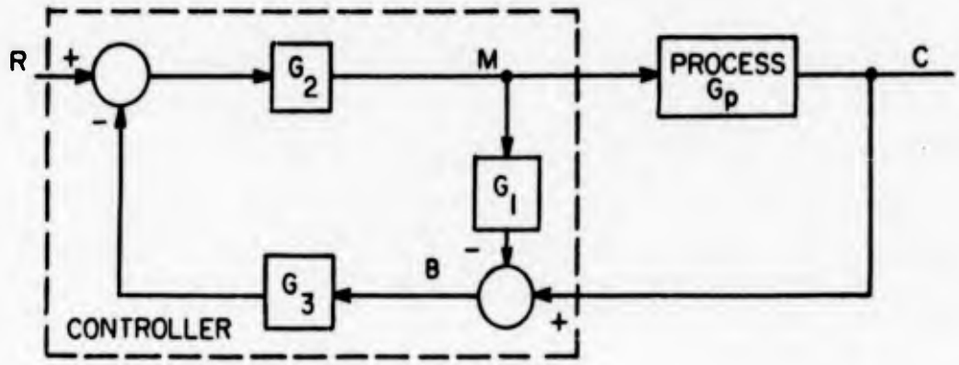


Figure 8-2 A two-loop system

The characteristics of the system are clear if we write the expressions for overall transmission and sensitivity with respect of G_p:

$$\frac{C}{R} = \frac{G_2 G_p}{1 + G_2 G_3 (G_p - G_1)} \qquad (8.8)$$

$$S_{G_p}^{C/R} = \frac{1}{1 + \dfrac{G_2 G_3 G_p}{1 - G_1 G_2 G_3}} = \frac{1 - G_1 G_2 G_3}{1 + G_2 G_3 (G_p - G_1)} \qquad (8.9)$$

* The electronics circuit designer often refers to such systems as having combined positive and negative feedback, with positive feedback referring to a loop gain which has a positive sign multiplying various transfer functions. Since the phase shift associated with the usual transfer function varies with frequency (and may pass through 180°), it is difficult to make a general definition of positive and negative feedback. In most cases, however, the implication of the terms is clear.

Thus, if we make

$$G_1 G_2 G_3 = 1 \tag{8.10}$$

we can realize a sensitivity of zero with an overall transmission

$$\frac{C}{R} = \frac{1}{G_3} \tag{8.11}$$

In other words, this configuration permits the realization of zero sensitivity simultaneously with the achievement of any desired overall transmission, subject only to the requirement that G_3, G_2, and G_1 be realizable transfer functions.

The situation is even more exciting than we have suggested so far. If we consider the denominator of C/ R (or S), which determines the stability problem, we find that the choice of G_1 as

$$G_1 = G_p \tag{8.12}$$

results in a remarkable simplification: the denominator is now just unity, so there is no possibility of poles lying in the right half plane. In other words, stability is assured. Hence, if G_3, G_1, and G_2 can be chosen to satisfy the relations

$$G_3 = \frac{1}{C/R} \quad G_1 = G_p \quad G_2 = \frac{1}{G_1 G_3} \tag{8.13}$$

We have designed a system which simultaneously has:

An arbitrary overall transmission
Zero sensitivity to changes in the process
No stability problem.

Before we consider the pitfalls associated with this magical configuration of Figure 8-2, which simultaneously satifies all conditions one might want in control-system design, let us consider what we have done in terms of the qualitative operation of the system. First, by selecting $G_1 = G_p$, we have made the signal B zero under normal operation (since B is $G_p M - G_1 M$); thus, there is no feedback at all unless G_p varies from the value for which G_1 was designed. In other words, with the normal value of G_p, the system is open-loop and not a feedback structure.

With regard to variations of G_p, however, the feedback does exist; indeed we have adjusted $G_1 G_2 G_3$ equal to unity, so that the feedback around G_p is infinite (the loop of $G_1 G_2 G_3$ is on the verge of instability); by this technique we have achieved the zero sensitivity. Insofar as overall transmission and sensitivity are concerned, the system we have designed is equivalent to that of Figure 8-3; we have simply discovered a "simple" way to realize the infinite gain through the medium of positive feedback.

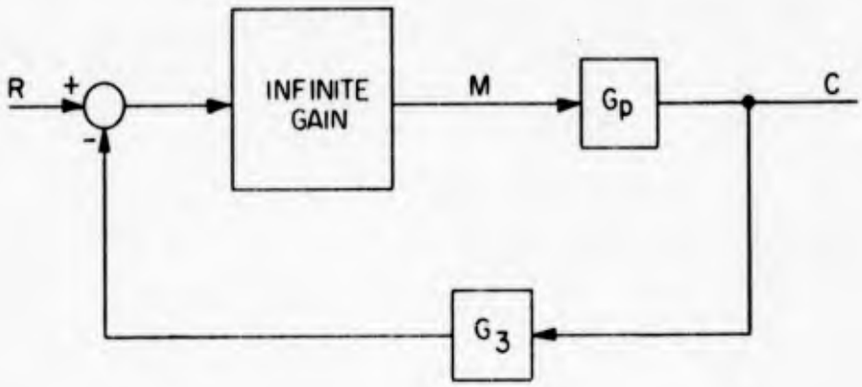


Figure 8-3 System equivalent to our two-loop configuration

When we turn to the actual problem of constructing the system of Figure 8-2, however, we find that this "simple" solution contains a few difficulties. First of all, in order to understand the practical limitations, we should consider typical transfer functions for the various blocks. For example, let us assume the process is second order, with

$$G_p = \frac{2}{s(s+5)} \qquad (8.14)$$

If we simply closed a single loop around such a process and adjusted the gain for a closed-loop system with a relative damping ratio of 0.5, we could obtain an undamped natural resonant frequency of 5; hence let us select, as the desired overall transfer function, the same

$$\frac{C}{R} = \frac{25}{s^2 + 5s + 25} \qquad (8.15)$$

What are the required three transfer functions: G_1, G_2 and G_3 ?

G_1 must be a model of the process, or

$$G_1 = \frac{2}{s(s+5)} \tag{8.16}$$

the feedback transfer function, G_3, is to be the reciprocal of C/R:

$$G_3 = \frac{1}{25}\, s^2 + \frac{1}{5}\, s + 1 \tag{8.17}$$

and G_2 is determined from the requirement that $G_1\, G_2\, G_3$ be unity, or:

$$G_2 = \frac{25}{2}\, \frac{s(s+5)}{s^2 + 5s + 25} \tag{8.18}$$

Two difficulties arise in the practical implementation of our system:

(1) The requirement that G_3 include a second-derivative term causes trouble if the signal C contains any high-frequency noise (since differentiation tends to accentuate high-frequency components). A practical compromise ordinarily involves realizing, not the G_3 of Equation (8.17), but a function which is approximately equal to this G_3 over the frequency band of interest: e.g.

$$G_{31} = \frac{\frac{1}{25}\, s^2 + \frac{1}{5}\, s + 1}{(\frac{1}{25}\, s + 1)(\frac{1}{40}\, s + 1)} \tag{8.19}$$

Even this presents problems since at high frequencies the transfer function is 40, meaning that an amplifier gain of 40 is required; if the amplifier gain falls off as s increases, G_3 may have to include an additional denominator factor. In other words, we encounter very major, practical design difficulties in constructing a minor loop which possesses a loop gain sufficiently close to unity over the band of frequencies of interest, particularly when we take into account the inevitable variations in the transfer functions with time, temperature, supply voltages, and environmental effects.

(2) When G_p changes, even slightly, we no longer realize the cancellation of feedback paths, and stability problems arise. For example, if

$$G_p = \frac{K}{s(s+5)} \tag{8.20}$$

where K varies about the normal value of 2, and if G_3 is not precisely equal to the desired value at high frequencies, the system may become unstable or at least undesirably underdamped. This difficulty is particularly poignant when we realize that in practice we never know G_p with great precision; indeed a primary purpose of of using feedback is to permit satisfactory system design with G_p only crudely known Particularly troublesome are the nonlinearities which may be present in the process, especially if we are not aware of the specific nature of the nonlinearities at the outset of the design.

In view of these difficulties, the usefulness of our two-loop configuration of Figure 8-2 depends markedly upon the specific transfer functions which are used; in the design of complex systems, we frequently require extensive analog-computer studies before we can evaluate the complete performance characteristics of the system. Our purpose here, however, is not to present a design theory, but rather to illustrate the flexibility achievable with multi-loop feedback systems.

Feed-Forward Systems

A somewhat different type of feedback configuration is illustrated in Figure 8-4; although this is single-loop, the system possesses many of the properties associated with our two-loop system of Figure 8-2 and hence is conveniently discussed briefly at this point. The system differs from a conventional single-loop system as a result of the feed-forward path from R through G_3; the signal R not only affects B through the $G_1 G_p G_2$ path, but also via $-G_3$.*

The characteristics of the system are apparent from the expressions for the overall transmission and the sensitivity of C/ R with respect to G_p:

$$\frac{C}{R} = \frac{(1 + G_3) G_1 G_p}{1 + G_1 G_p G_2} \qquad (8.21)$$

$$S_{G_p}^{C/R} = \frac{1}{1 + G_1 G_p G_2} \qquad (8.22)$$

* The system is termed "feed-forward" because R enters the main loop not only at the comparator, but also forward in the system after $G_1 G_p G_2$. We might equivalently consider the system in which R was passed through a block G_4 and then added to the output of G_1.

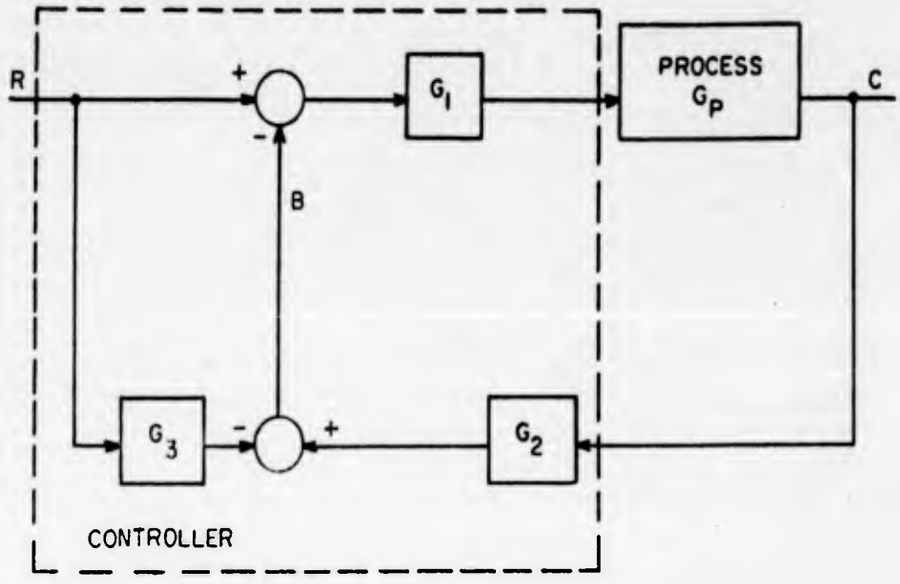


Figure 8-4 Ham-Lang system

If we select G_3 to satisfy the relation

$$G_3 = G_1 G_p G_2$$

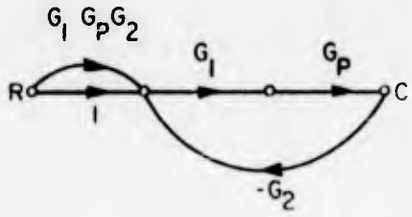
The overall transmission becomes

$$\frac{C}{R} = G_1 G_p \qquad (8.23)$$

and again we essentially possess an open-loop system. The two signals entering the lower comparator of Figure 8-4 are equal and cancel, B = 0, and there is no feedback, no chance for instability. The sensitivity, however, is exactly that of a conventional single-loop system (with the G_3 path omitted). Thus, we obtain the advantages of feedback without the associated difficulties of stabilization (again, of course, only as long as we can satisfy the cancellation condition implied by the above formula for G_3).

Figure 8-4 is merely one of several equivalent configurations in which the feed-forward technique is utilized to simplify the problem of designing a feedback configuration with, simultaneously, a high degree of relative stability and a low sensitivity. Figure 8-5 shows two equivalent configurations; additional schemes can be derived from the basic idea that, having chosen a configuration to realize the sensitivity specifications, we utilize an insertion of a function of the input signal

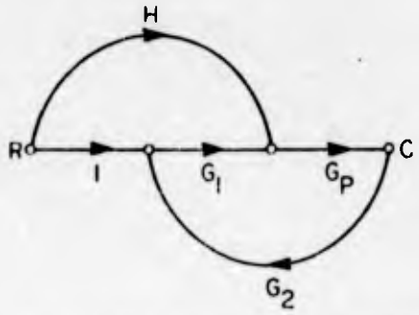
into the loop in such a way as to simplify the stability problem.



(a) Ham-Lang system of Figure 8-4.

$$S_{G_p}^{C/R} = \frac{1}{1 + G_1 G_p G_2}$$

$$\frac{C}{R} = G_1 G_p$$



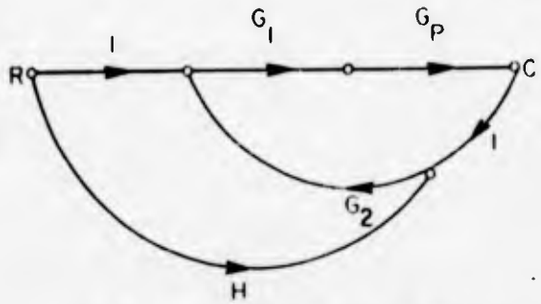
(b) Feed-forward to process input.

$$S_{G_p}^{C/R} = \frac{1}{1 + G_1 G_p G_2}$$

$$\frac{C}{R} = \frac{(H + G_1) G_p}{1 + G_1 G_2 G_p}$$

If $H = G_1^2 G_2 G_p$, same as (a)

If $H = \frac{1}{G_2 G_p}$, $\frac{C}{R} = \frac{1}{G_2}$



(c) Feed - forward into feedback path.

$$S_{G_p}^{C/R} = \frac{1}{1 + G_1 G_p G_2}$$

$$\frac{C}{R} = \frac{G_1 G_p (1 + H G_2)}{1 + G_1 G_p G_2}$$

If $H = G_1 G_p$, same as (a)

Figure 8-5 Simple feed-forward schemes

value of $1/H_3$.*

Example 4 Stabilization of Nonlinear Oscillation

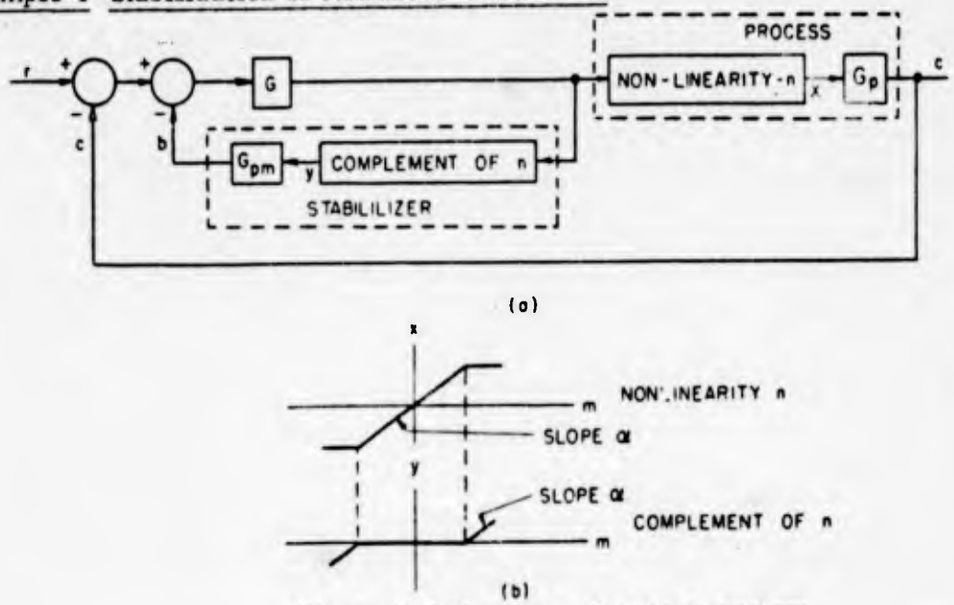


Figure 8-7 Nonlinear feedback system

The feedback control system of Figure 8-7 (omitting the stabilization section) includes a nonlinear process consisting of saturation followed by a linear section. When the controller G is designed to give appropriate small-signal performance and the loop is closed, the system is found to be unstable for large signals (when the nonlinearity affects the operation). Thus, the instability can be directly associated with the presence of the nonlinearity.

Under such circumstances, the system is stabilized by the insertion of the nonlinear stabilizer network consisting of two parts:

(1) A nonlinear element which complements the process nonlinearity such that the parallel combination of n and its complement represent a linear component;

(2) A linear block with a transfer function G_{pm} which equals, to the best of our knowledge, the linear G_p. If G_{pm} accurately models G_p and the two nonlinearities are complementary, the total feedback signal (b + c) is exactly that which would be present in the linear feedback system depicted in Figure 8-8. Thus, even though the nonlinearity n affects the output c, n has no effect on the stability of the closed-loop system. The stabilizer loop removes the insta-

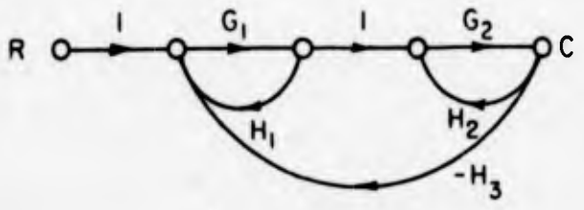
*This configuration, when utilized for a transistor amplifier for example, usually seems to result in a conditionally stable system: as one transistor gain falls, the overall system eventually becomes unstable for slight decreases in the gain of the other transistor. Such a characteristic results in the necessity for considerable care in the choice of actual transfer functions in order to realize satisfactory operation over wide operating ranges.

The above examples illustrate the general technique of using parallel transmission paths or multi-loop structures in order to achieve in design at least some degree of independence in the control of the two functions: the overall transmission and the sensitivity (or a subsidiary transmission). In more complicated situations (in which there may be many loops inherent in the structure before the feedback engineer initiates his phase of the design, or in which there may be many inputs and outputs), the determination of the characteristics of the feedback system or the design selection of various loops and transfer functions often taxes the ingenuity and theoretical background of the engineer; such problems constitute the core of linear feedback theory.

In this monograph, however, our interest is merely indicating the scope of the capabilities of feedback systems, and we leave to advanced texts the details of the general theory. Thus, we conclude this section with an extremely brief description of three multi-loop systems which are of considerable practical importance and which are related to the two systems we have discussed in more detail above.

Example 3 Double Control of Sensitivity

The two-loop, zero-sensitivity system of Figure 8-2 can be extended when two active elements, G_1 and G_2, are present (Figure 8-6). In this configuration, if H_1 and H_2 are selected to satisfy the criteria



$$H_1 = \frac{1}{G_1} \qquad H_2 = \frac{1}{G_2} \tag{8.24}$$

Figure 8-6

over the frequency band of interest, the overall transmission and sensitivities are

$$\frac{C}{R} = \frac{1}{H_3} \qquad S_{G_1}^{C/R} = 0 \qquad S_{G_2}^{C/R} = 0 \tag{8.25}$$

If either G_1 or G_2 changes, the overall transmission is not affected; if both change from the design value, the overall transmission than does vary from the design

value of $1/H_3$.*

Example 4 Stabilization of Nonlinear Oscillation

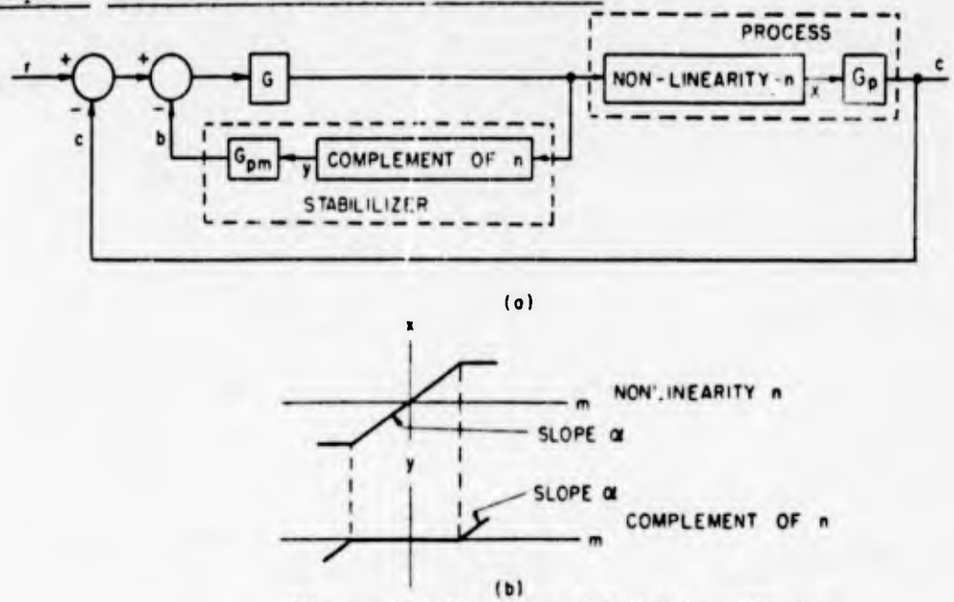


Figure 8-7 Nonlinear feedback system

The feedback control system of Figure 8-7 (omitting the stabilization section) includes a nonlinear process consisting of saturation followed by a linear section. When the controller G is designed to give appropriate small-signal performance and the loop is closed, the system is found to be unstable for large signals (when the nonlinearity affects the operation). Thus, the instability can be directly associated with the presence of the nonlinearity.

Under such circumstances, the system is stabilized by the insertion of the nonlinear stabilizer network consisting of two parts:

(1) A nonlinear element which complements the process nonlinearity such that the parallel combination of n and its complement represent a linear component;

(2) A linear block with a transfer function G_{pm} which equals, to the best of our knowledge, the linear G_p. If G_{pm} accurately models G_p and the two nonlinearities are complementary, the total feedback signal (b + c) is exactly that which would be present in the linear feedback system depicted in Figure 8-8. Thus, even though the nonlinearity n affects the output c, n has no effect on the stability of the closed-loop system. The stabilizer loop removes the insta-

*This configuration, when utilized for a transistor amplifier for example, usually seems to result in a conditionally stable system: as one transistor gain falls, the overall system eventually becomes unstable for slight decreases in the gain of the other transistor. Such a characteristic results in the necessity for considerable care in the choice of actual transfer functions in order to realize satisfactory operation over wide operating ranges.

bility resulting from the nonlinearity.

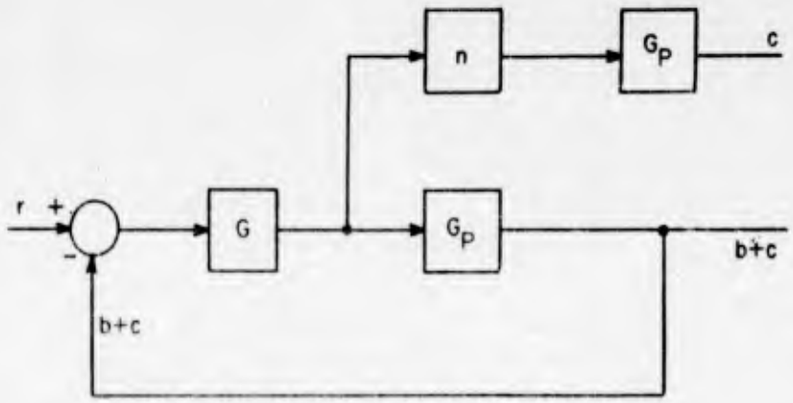


Figure 8-8 Equivalent of Figure 8-7 with perfect model

The basic nonlinear-compensation technique illustrated in Figure 8-7 can be extended to the stabilization of oscillating systems involving a wide variety of nonlinearities (backlash, hysteresis, static friction, etc.). Actually, it can be demonstrated that the model G_{pm} need be only a very crude approximation of G_p and the nonlinear compensator may deviate significantly from the true complement of n. Stabilization of this type has been used extensively in operating control systems (e.g., in removing the low-amplitude oscillations often associated with nonlinearities in the value of an hydraulic power amplifier).

Example 5 Data-Recovery System

In radar tracking, the radar receiver provides an output signal measuring the error between target and antenna positions (in azimuth, for example). For the computer used to calculate future target position, we require data on the present target location. Thus, the tracking system works under two requirements:

(1) The antenna must follow the target with sufficient accuracy to insure the target never leaves the antenna beam.

(2) The system must provide a separate data output with as accurate target position data as possible.

The system of Figure 8-9 is widely used in this application.

The basic control system consists of all the elements except H_2: G_p is the

power amplifier and antenna system, G_1 a pre-amplifier controller, and H_1 a minor feedback path to linearize and improve the performance of the power elements. If the data-recovery block H_2 is inserted and we choose H_2 so that

$$H_1 H_2 = K_1 \tag{8.26}$$

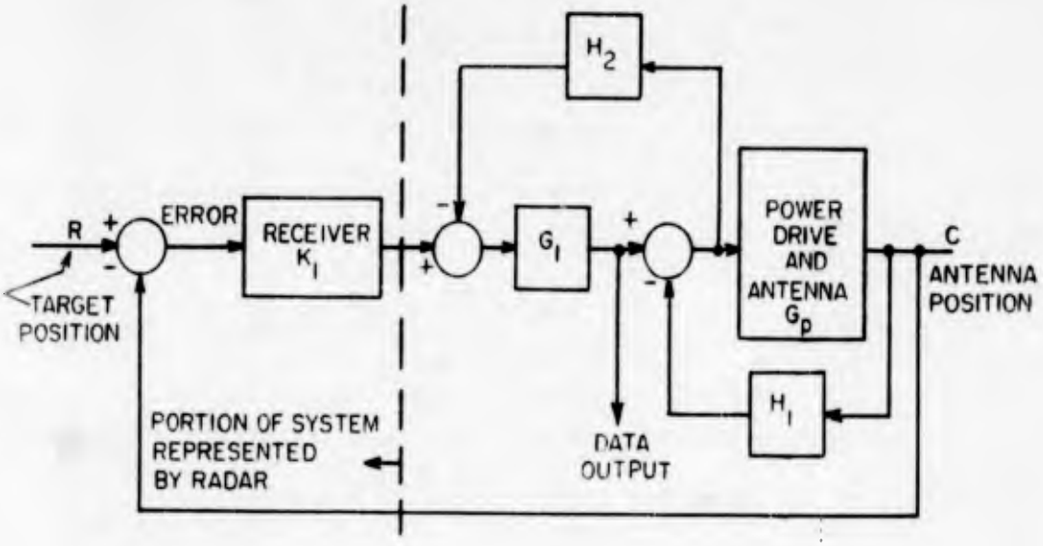


Figure 8-9 Data-recovery system

the feedbacks of C to the input of G_1 along the two possible paths ($-K_1$ through the main feedback path, the input comparator, and K_1; and $+H_1H_2$ through H_1, the comparator, H_2, and the second comparator) exactly cancel. Consequently, the only two closed loops in the system are those with the loop gains

$$-G_1 H_2 \qquad -G_p H_1$$

and the system is equivalent to the configuration of Figure 8-10. In other words, we have inserted H_2 in order that the "negative" feedback of C to the input of G_1 (feedback arising because of the nature of the radar operation) may be cancelled by the "positive" feedback through H_1 and H_2.

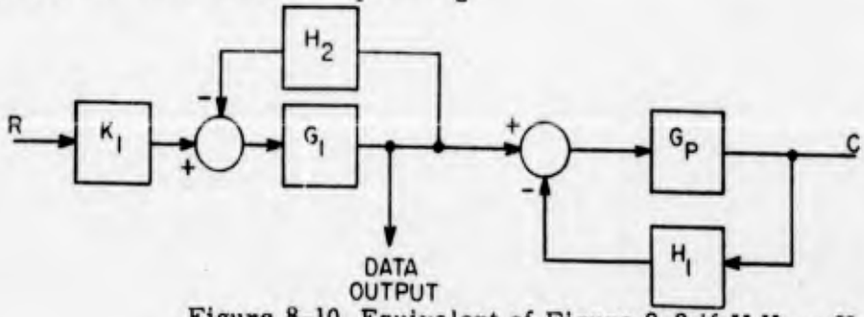


Figure 8-10 Equivalent of Figure 9-9 if $H_1H_2 = K_1$

The data on target position R can be removed from the system at the output of G_1 (in either Figure 8-10 or 8-9); since the preamplifier G_1 needs to provide almost no power, the $G_1 H_2$ feedback system can be a wide-band and high-accuracy system, so that the output of G_1 is an accurate reproduction of R.

There are two practical difficulties with such a system, both of which have to be investigated in terms of the specific transfer functions when we are designing an actual system. First, Equation (8.26) requires that H_2 track K_1: as K_1 varies (because of signal fading, for example), Equation (8.26) is violated, and we must evaluate the effects of this deviation on system performance. Second, G_p is often highly nonlinear (as we might expect because of the power requirements); in an actual system design, we must investigate the effects of such nonlinearities on the accuracy of the data at the output of G_1.

The importance of the data-recovery configuration of Figure 8-9 and the several, closely related configurations is evidenced by the widespread adoption of such feedback schemes in tracking systems.

21 July 1961

TECHNICAL AND FINAL REPORT DISTRIBUTION LIST

PHYSICAL SCIENCES DIRECTORATE

GOVERNMENTAL

No. of Copies	Agency
3	Commander AF Office of Scientific Research ATTN: SRY Washington 25, D. C.
2	Commander AF Office of Scientific Research ATTN: SRGL Washington 25, D. C.
4	Commander Aeronautical Systems Division (ASD) Wright-Patterson Air Force Base Ohio
1	Commander AF Cambridge Research Laboratories ATTN: CRRELA L. G. Hanscom Field Bedford, Massachusetts
2 (Unclassified Only)	Commander EOAR The Shell Building 47 Rue Cantersteen Brussels, Belgium
1	P. O. Box AA Wright-Patterson Air Force Base Ohio
1	Aeronautical Research Laboratories ATTN: Technical Library Building 450 Wright-Patterson Air Force Base Ohio
10	Armed Services Technical Information Agency ATTN: TIPCR Arlington Hall Station Arlington 12, Virginia

1 Director of Research and Development
Headquarters, USAF
ATTN: AFDRD
Washington 25, D. C.

1 Office of Naval Research
Department of the Navy
ATTN: Code 420
Washington 25, D. C.

1 Director, Naval Research Laboratory
ATTN: Technical Information Officer
Washington 25, D. C.

1 Chief of Research and Development
ATTN: Scientific Information Branch
Department of the Army
Washington 25, D. C.

1 Chief, Physics Branch
Division of Research
U. S. Atomic Energy Commission
Washington 25, D. C.

1 U. S. Atomic Energy Commission
Technical Information Extension
P. O. Box 62
Oak Ridge, Tennessee

1 National Bureau of Standards Library
Room 203, Northwest Building
Washington 25, D. C.

1 Physics Program
National Science Foundation
Washington 25, D. C.

1 Director, Army Research Office, Durham
Box CM, Duke Station
Durham, North Carolina

1 AEDC (ADOIM)
ATTN: AEDC Library
Arnold Air Force Station
Tullahoma, Tennessee

1 Commander
AF Flight Test Center
ATTN: FTOTL
Edwards Air Force Base
California

1		Commander AF Missile Development Center ATTN: HDOI Holloman Air Force Base New Mexico
1		Commander Army Rocket & Guided Missile Agency ATTN: ORDXR-OTL Redstone Arsenal Alabama
1		Commandant Air Force Institute of Technology (AU) Library MCLI-LIB, Bldg. 125, Area B Wright-Patterson Air Force Base Ohio
2		Commander Air Force Systems Command ATTN: SCRS Andrews Air Force Base Washington 25, D. C.
1		Commanding General U. S. Army Signal Corps Research and Development Laboratory ATTN: SIGFM/EL-RPO Ft. Monmouth, New Jersey
6		National Aeronautics & Space Administration Washington 25, D. C.
1		Advanced Research Projects Agency Washington 25, D. C.
1		Rand Corporation 1700 Main Street Santa Monica, California
1	(Unclassified Only)	Chairman Canadian Joint Staff For DRB/DSIS 2450 Massachusetts Ave., N. W. Washington 25, D. C.
1		Ames Research Center (NASA) ATTN: Technical Library Moffett Field, California
1		High Speed Flight Station (NASA) ATTN: Technical Library Edwards AFB, California
1		Langley Research Center (NASA) ATTN: Technical Library Langley AFB, Virginia

1	Lewis Research Center (NASA) ATTN: Technical Library 21000 Brookpark Road Cleveland 35, Ohio
1 (Unclassified Only)	Institute of the Aeronautical Sciences 2 East 64th Street New York 21, New York
1 (Unclassified Only)	Applied Mechanics Reviews Southwest Research Institute 8500 Culebra Road San Antonio 6, Texas
1 (Unclassified Only)	Linda Hall Library ATTN: Document Division 5109 Cherry Street Kansas City 10, Missouri
1	AFOSR (SRLTL) Holloman Air Force Base New Mexico

SUPPLEMENTAL DISTRIBUTION LIST

Control Theory, Information Theory, and Network Synthesis

Professor Gregory Young
Engineering Center
University of Southern California
Los Angeles 7, California

Professor C T Leondes
Department of Engineering
University of California
Los Angeles, California

Electronics Research Laboratory
College of Engineering
University of California
Berkeley 4, California

Dr. E. I. Jury
Department of Electrical Engineering
University of California
Berkeley 4, California

Dr. Gene F. Franklin
Electronics Research Laboratory
Stanford University
Stanford, California

Dean J. R. Ragazzini
College of Engineering
New York University
University Heights, New York

Columbia University
Electrical Engineering Department
ATTN: G. Kranc
New York 27, New York

University of Illinois
Department of Electrical Engineering
ATTN: M. E. Van Valkenburg
Urbana, Illinois

Polytechnic Institute of Brooklyn
Microwave Research Institute
ATTN: J. G. Truxal
55 Johnson Street
Brooklyn 1, New York

RIAS
1712 Bellona Avenue
Baltimore 12, Maryland

Johns Hopkins University
Department of Mechanics
ATTN: Library (Route to Prof F. H. Clauser)
Baltimore 18, Maryland

9/15/61

UNCLASSIFIED