Automatic Control Engineering

Francis H. Raven

ASSISTANT PROFESSOR OF MECHANICAL ENGINEERING
UNIVERSITY OF NOTRE DAME

McGRAW-HILL BOOK COMPANY, INC.

New York *Toronto* *London*

1961

AUTOMATIC CONTROL ENGINEERING

 Library of Congress Catalog Card Number 61-10470

IV

51228

THE MAPLE PRESS COMPANY, YORK, PA.

To Therese

PREFACE

In recent years, automatic control systems have been rapidly advancing in importance in all fields of engineering. The applications of control systems cover a very wide scope, ranging from the design of precision control devices such as delicate instruments used for inertial guidance to the design of massive equipment such as that used for controlling the manufacture of steel or other industrial processes. New applications for automatic controls are being continually discovered.

This text is the outgrowth of the notes developed by the author to teach control engineering at the University of Notre Dame. The author has endeavored to give the principles a thorough presentation and yet make them clear and easy to understand. It is presupposed that the reader has the general maturity and background of a third- or fourth-year engineering student, but no previous training in control engineering.

Although the principles of feedback control systems are presented in a manner which is appropriate to the interests of mechanical engineers, this text has also been successfully used to teach students in other fields of engineering. In addition, the author has taught night courses for practicing engineers. In the light of their enthusiastic comments, it is felt that this book will be of much value to the engineer in industry who did not have the opportunity to take such a course while in college.

The basic principles and fundamental concepts of feedback control systems are presented in the first portion (Chapters 1 through 10). The latter portion correlates basic theory with the more practical aspects involved in the design of control systems. Because the usual pattern encountered is that new engineers are strong in theory but weak in practice, it is felt that this latter portion will help to bridge the gap between theory and practice.

In particular, the study of control engineering is begun by showing how typical control systems may be represented by block diagrams. This is accomplished by first demonstrating how to represent each component or part of a system as a simple block diagram. Next, it is explained how these individual diagrams may be connected to form the over-all block diagram, just as the actual components are connected to form the complete control system. Because actual control systems frequently contain

nonlinear components, considerable emphasis is given to such components. The preceding material is presented in the first three chapters. In the fourth chapter, it is shown that much important information concerning the basic or inherent operating characteristics of a system may be obtained from a knowledge of the steady-state behavior.

This introduction to control theory differs from the usual "black-box" approach, in which the block diagram for a system is given outright. This black-box approach permits introducing Laplace transforms and other methods for system analysis at an earlier stage. However, it has been the author's experience that, if the student is first familiarized with the physical significance of feedback controls, then he is better able to appreciate the value of the more specialized techniques used in system analysis, and thus arrives at a far deeper understanding.

In Chapters 5 through 10, the various methods and techniques used for determining the performance of control systems are thoroughly described. In particular, in Chapter 5 it is shown how linear differential equations which describe the operation of control systems may be solved algebraically by the use of Laplace transforms. Chapter 6 explains how the roots of the characteristic equation govern the transient response, and in Chapter 7 it is shown how these roots may be ascertained by use of the root-locus method. Application of the analog computer for simulating control systems is presented in Chapter 8. The use of frequency-response techniques for evaluating dynamic performance is explained in Chapters 9 and 10.

More specialized considerations which arise in the design of hydraulic systems, pneumatic systems, and electrical systems and in inertial navigation are treated in Chapters 11 through 14. Chapter 15 presents techniques which are applicable for the analysis of certain types of nonlinearities and provides a general introduction to the field of nonlinear control systems.

The author wishes to acknowledge the excellent suggestions of Professor Ferdinand Freudenstein, Columbia University; Professor John E. Gibson, Purdue University; Professor Thomas J. Higgins, University of Wisconsin; Dr. Kenneth W. Kohlmeyer, Sperry-Gyroscope Company; Professor Richard M. Phelan, Cornell University; Professor John R. Ragazzini, New York University; Professor John G. Truxal, Polytechnic Institute of Brooklyn; and Mr. William E. Vannah, former editor of *Control Engineering*.

Appreciation is expressed for the many fine comments of the author's former students who used the original notes from which this text has been developed. Although it is not possible to acknowledge the contribution of each student, particular recognition is due Paul W. Beiter, Frank D'Souza, John Grace, William Howard, and Thomas McCarey.

Sincere gratitude is extended for the encouragement and suggestions of the author's colleagues at the University of Notre Dame, especially to Dr. M. J. Goglia, Dr. M. K. Newman, and Brother C. Albert Welsh, F.S.C., D.Sc. Appreciation is also expressed to Mrs. Ella Levee for her typing of the many revisions of the notes from which this text has been developed.

The author's wife, Therese, has faithfully worked with him throughout the development of this text. She has made innumerable suggestions and has been a constant source of encouragement.

Francis H. Raven

CONTENTS

CHAPTER 1

INTRODUCTION TO AUTOMATIC CONTROLS

1.1. Historical Development. Early man had to rely upon his own brute strength or that of beasts of burden to supply energy for doing work. By use of simple mechanical devices such as wheels and levers, he accomplished such feats as the building of high pyramids and Roman highways and aqueducts. He first supplemented his energy and that of beasts by utilizing power from natural sources such as the wind for powering sailing vessels and windmills, and waterfalls for turning water wheels. The invention of the steam engine was a milestone in man's progress because it provided him with useful power that he could harness at will. Since then, man has devised many different means for obtaining abundant and convenient sources of energy. Engineering effort is primarily concerned with the practical applications of using power to serve the purposes of man. That is, the engineer designs and develops machines and equipment by which man can utilize power.

Early machines and equipment had controls which were predominantly of a manual nature, and the adjustments had to be reset frequently in order that the desired output or performance could be maintained. The design of newer equipment with greater usefulness and capabilities is bringing about an ever-increasing growth in the development of control equipment. The reason is twofold. First, automatic controls relieve man of many monotonous activities so that he can devote his abilities to other endeavors. Second, modern complex controls can perform functions which are beyond the physical abilities of man to duplicate. For example, an elaborate automatic control system operates the engine of a modern jet airplane with only a minimum amount of the pilot's attention so that he is free to maneuver and fly his airplane.

It is interesting to note that, as the applications and uses for controls have increased, so also have the demands upon the performance of these systems increased. There is no doubt that a major concern of the engineer today, and even more so in the future, is, and will be, the design and development of automatic control systems.

1.2. Feedback Control Systems. The controlling of temperature is a typical example of a feedback control system. The position of the

temperature dial sets the desired temperature (i.e., the reference input). The actual temperature of the system is the controlled variable (i.e., the quantity which is being controlled). The thermostat, or comparator, compares the actual temperature with the desired temperature in order to measure the error. This error signal is the actuating signal, which is then sent to the heating units in order to correct the temperature. For example, if the actual temperature is less than the desired temperature, the actuating signal causes the control elements to supply more heat. If there is no error, the control elements do not change the amount of heat which is being supplied. When the actual temperature is greater than the desired value, then the actuating signal calls for a decrease in the amount of heat.

For a system to be classified as a feedback control system, it is necessary that the controlled variable be fed back and *compared* with the reference input. In addition, the resulting error signal must *actuate* the control elements to change the output so as to minimize the error. A feedback control system is also called a closed-loop system. Any system which incorporates a thermostat to control temperature is a feedback, or closed-loop, system. Well-known examples are electric frying pans, irons, refrigerators, and household furnaces with thermostatic control.

For speed control systems, the device which subtracts the feedback signal from the reference input (i.e., the comparator) is usually a centrifugal governor. The governor serves the same purpose that the thermostat does for temperature controls. That is, the governor compares the actual speed which is to be controlled with the desired value and measures the error. This error signal then actuates the control elements. The same basic concepts apply to all types of feedback control systems, whether the controlled variable be temperature, speed, pressure, flow, position, force, torque, or any other physical quantity.

In an open-loop system there is no comparison of the controlled variable with the desired input. Each setting of the input determines a fixed operating position for the control elements. For example, for a given input temperature setting, the heating units are positioned to supply heat at a fixed rate. (Note that there is no comparator, or thermostat, which measures the error and resets the heating units.) The disadvantage of such a system is illustrated by the fact that, for a fixed rate of heat supplied to a house, the inside temperature varies appreciably with changes in the outside temperature. Thus, for a given set input to an open-loop system, there may be a big variation of the controlled variable depending on the ambient temperature.

In this example, the ambient temperature is an external disturbance. By an external disturbance is meant something external to the system which acts to change or disturb the controlled variable. A major

advantage of employing feedback control is that, because of the comparator, the actuating signal continually changes so that the controlled variable tends to become equal to the reference input regardless of the external disturbance. Another consideration is that with feedback one can generally use relatively inexpensive components and yet obtain better control than is possible by using very expensive components in an open-loop system. The primary effort of this text will be devoted to feedback control systems.

1.3. System Representation. The mathematical relationships of control systems are usually represented by block diagrams. These diagrams have the advantage of indicating more realistically the actual processes which are taking place, as opposed to a purely abstract mathematical representation. In addition, it is easy to form the over-all block diagram for an entire system by merely combining the block diagrams for each component or part of the system.

A comparator subtracts the feedback signal from the reference input r. For the case in which the controlled variable c is fed back directly

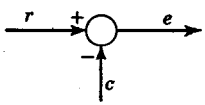


FIG. 1.1. Block diagram of a comparator.

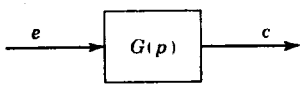


FIG. 1.2. Block diagram of the control elements.

(i.e., for unity-feedback systems), the signal coming from the comparator is $r - c$, which is equal to the actuating signal e. The mathematical relationship for this operation is

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

A circle is the symbol which is used to indicate a summing operation, as is illustrated in Fig. 1.1. The arrowheads pointing toward the circle indicate input quantities, while the arrowhead leading away signifies the output. The sign at each input arrowhead indicates whether the quantity is to be added or subtracted.

The relationship between the actuating signal e, which enters the control elements, and the controlled variable c, which is the output of the control, is expressed by the equation

$$c = G(p)e \tag{1.2}$$

where $G(p)$ represents the operation of the control elements. In Chaps. 2 and 3, it is shown how the actual values of $G(p)$ for specific control systems are obtained. The block-diagram representation for the preceding equation is shown in Fig. 1.2. A box is the symbol for multiplica-

tion. In this case, the input quantity e is multiplied by the function in the box $G(p)$ to obtain the output c. With circles indicating summing points and boxes, or blocks, indicating multiplication, any linear mathematical expression may be represented by block-diagram notation.

The complete block diagram for an elementary unity-feedback control system is obtained by combining Figs. 1.1 and 1.2 to yield Fig. 1.3. This

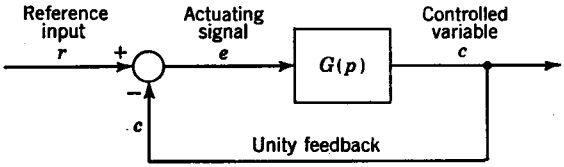


FIG. 1.3. Block diagram of an elementary unity-feedback control system.

diagram shows the controlled variable c being fed back to the summing point, where it is compared with the reference input r. This diagram pictorially shows why a feedback control system is also called a closed-loop system.

When the controlled variable is fed back to the comparator, it is usually necessary to convert the form of the controlled variable to a form that is suitable for the comparator. For example, in a temperature

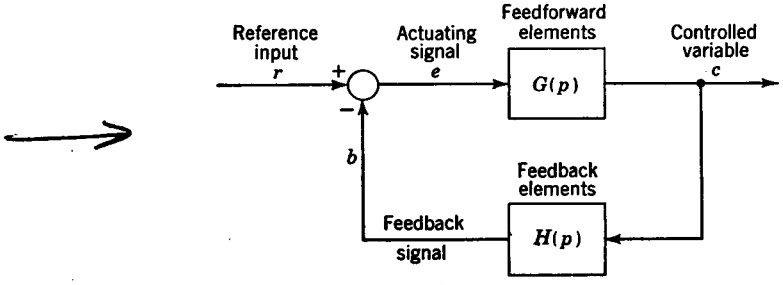


FIG. 1.4. Block diagram of an elementary feedback control system.

control system the controlled temperature is generally converted to a proportional force or position for use in the comparator. This conversion is accomplished by feedback elements $H(p)$. The block-diagram representation for this more general case of a feedback control system is shown in Fig. 1.4. The signal which is fed back is

$$b = H(p)c \tag{1.3}$$

The elements represented by $H(p)$ are called the feedback elements because they are located in the feedback portion of the control. The control elements represented by $G(p)$ are the feedforward elements

because of their location in the feedforward portion of the loop. The actuating signal e is now $r - b$. This actuating signal e is a measure or indication of the error.

The term "feedback control system" is a general term which applies to any system in which the controlled variable is measured and fed back to be compared with the reference input. The terms "servomechanism" and "regulator" are distinguished as follows: A servomechanism is a particular type of feedback control system in which the controlled variable is a mechanical position (e.g., the angular position of a shaft). A regulator is distinguished as a feedback control system in which the reference input, although adjustable, is held fixed, or constant, for long periods of time (e.g., most temperature controllers).

$$C = \frac{\text{OPEN-LOOP FUNCTIONS (PRODUCTS)}}{1 + \text{CLOSED-LOOP FUNCTIONS (PRODUCTS)}}$$

CHAPTER 2

REPRESENTATION OF CONTROL COMPONENTS

2.1. General. To investigate the performance of control systems, it is necessary to obtain the mathematical relationship $G(p)$ relating the controlled variable c and the actuating signal e of the feedforward elements. This is accomplished by first obtaining the mathematical representation for each component between the actuating signal and the controlled variable and then expressing each of these equations as a block diagram. The combination of the block diagrams for each component yields the desired representation for $G(p)$. The value of $H(p)$ is obtained by applying this same technique to the components in the feedback portion of the control.

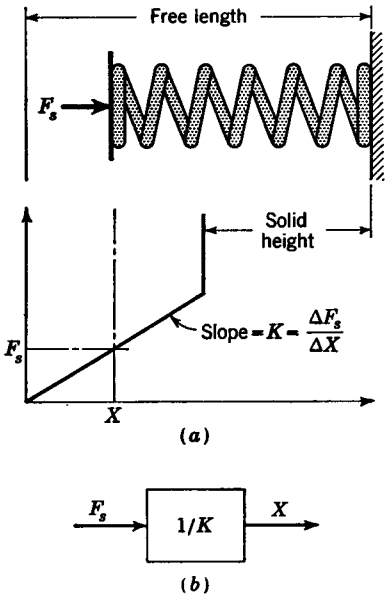


FIG. 2.1. Spring characteristics.

The quantity $G(p)$ could be obtained by writing the mathematical equation describing the operation of each component between e and c and then combining these individual equations algebraically to obtain the over-all relationship between e and c. However, for all but the simplest systems, this procedure proves cumbersome because of the interaction between the various components in a typical control system. In addition, the block-diagram method gives one a better understanding of the system because of its visual representation.

The obtaining of block diagrams for typical elements used in control devices is illustrated in this chapter. In the next chapter, it is shown how these individual diagrams are combined to form entire control systems.

2.2. Translational Mechanical Components. The load-deflection characteristics for a mechanical spring are shown in Fig. 2.1*a*. The spring

force F_s required to compress a spring X in. from its free length is given by the equation

$$F_s = KX \tag{2.1}$$

where K, the spring rate, is a constant which is equal to the slope of the curve of the load F_s versus deflection X. The input to a spring is usually the force F_s, and the output is the deflection X, so that the block-diagram representation for Eq. (2.1) is as shown in Fig. 2.1*b*.

For a viscous damper as illustrated in Fig. 2.2*a*, the force F_d required to move one end of the dashpot at a velocity V relative to the other

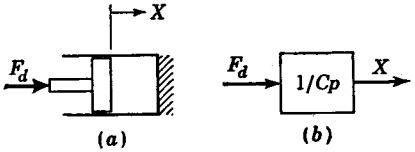


FIG. 2.2. Linear viscous damper.

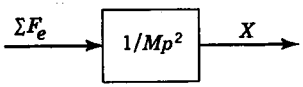


FIG. 2.3. Acceleration of a mass.

end is equal to the product of the damping coefficient C and the velocity. That is,

$$F_d = CV = C\frac{dX}{dt}$$

The substitution of the operator symbol $p = d/dt$ into the preceding expression yields

$$F_d = CpX \tag{2.2}$$

With the force F_d as the input and the displacement X as the output, the block-diagram representation for Eq. (2.2) is shown in Fig. 2.2*b*.

By Newton's first law of motion, it follows that the summation of the external forces ΣF_e acting on a mass is equal to the product of the mass and acceleration.

$$\sum F_e = MA = M\frac{d^2X}{dt^2} = Mp^2X$$

The displacement X is given by the equation

$$X = \frac{1}{Mp^2}\sum F_e \tag{2.3}$$

This is represented diagrammatically in Fig. 2.3.

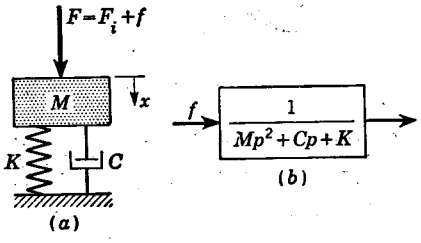


FIG. 2.4. Series mass-spring-damper combination.

For the mass-spring-damper combination shown in Fig. 2.4*a*, the spring force and damper force are opposed to, or resist, the motion caused by the applied load F. The summation of the forces acting on

the mass is

$$\Sigma F_e = F + Mg - F_s - F_d = Mp^2X$$

or

$$F = (Mp^2 + Cp + K)X - Mg \tag{2.4}$$

For control work, it is often more convenient to make measurements with respect to some initial or reference operating point. A lower-case letter is used to designate the variation or change in displacement x from the reference position X_i so that $x = X - X_i$. Because X_i is a constant, then $pX = p(X_i + x) = px$ and $p^2X = p^2x$. As would be expected, velocity and acceleration are independent of the reference position from which displacement is measured. Equation (2.4) may be written in the form

$$F = (Mp^2 + Cp + K)x + KX_i - Mg$$

When the system is at rest or equilibrium at the reference operating point (i.e., when $p^2x = px = x = 0$), then the value of the applied force is $KX_i - Mg$. Substitution of $F_i = KX_i - Mg$, which is a constant in the preceding equation, gives

$$F - F_i = f = (Mp^2 + Cp + K)x \tag{2.5}$$

where f is the change in applied force from the value F_i required for equilibrium at the reference operating point. Although x and f are measured from the reference operating point, Eq. (2.5) is a general equation describing the dynamic behavior of the system. It is not necessary that the system be initially at this reference operating point or that the system be initially at rest. As is later explained, it is usually much easier to obtain the equation of operation with respect to some convenient reference point rather than using absolute values. When absolute values are desired, it is an easy matter to add the reference value to the variation. The block diagram for Eq. (2.5) is shown in Fig. 2.4*b*.

2.3. Operational Notation. In the preceding section, it was found convenient to use the Heaviside operational notation

$$p^r \equiv \frac{d^r}{dt^r} \qquad r = 1, 2, 3, \ldots, n \tag{2.6}$$

The operator p^r is a symbol which indicates that certain operations of differentiation are to be performed. For example, if x and y are functions of time, then

$$p(x + y) = \frac{d}{dt}(x + y) = \frac{dx}{dt} + \frac{dy}{dt} = px + py$$

Thus, the operator p obeys the distributive law, i.e.,

$$p(x + y) = px + py \tag{2.7}$$

and may be factored as though it were an algebraic quantity.

It may also be shown that, if a and b are constants, then

$$\begin{aligned}(p+a)(p+b)x &= (p+a)\left(\frac{d}{dt}+b\right)x \\ &= \left(\frac{d}{dt}+a\right)\left(\frac{dx}{dt}+bx\right) \\ &= \frac{d}{dt}\left(\frac{dx}{dt}+bx\right)+a\left(\frac{dx}{dt}+bx\right) \\ &= \frac{d^2x}{dt^2}+(a+b)\frac{dx}{dt}+abx \\ &= [p^2+(a+b)p+ab]x\end{aligned}$$

Hence, it follows that the commutative law also holds, i.e.,

$$\begin{aligned}(p+a)(p+b)x &= (p+b)(p+a)x \\ &= [p^2+(a+b)p+ab]x\end{aligned} \tag{2.8}$$

Similarly, it may be shown that the law of exponents holds,

$$p^n p^m x = p^{n+m}x \tag{2.9}$$

where n and m are any two *positive* integers.

The meaning of the reciprocal of p is obtained as follows: From calculus

$$x(t) = \int [pf(t)]\,dt = f(t) + C \tag{2.10}$$

where $pf(t) = (d/dt)f(t)$ and C is the constant of integration. Differentiation of the preceding expression with respect to time yields

$$px(t) = pf(t)$$

Solving the preceding expression for $x(t)$ gives

$$x(t) = \frac{1}{p}[pf(t)] \tag{2.11}$$

Comparison of Eqs. (2.10) and (2.11) shows that

$$\frac{1}{p}[pf(t)] = \int [pf(t)]\,dt = f(t) + C \tag{2.12}$$

Therefore, the symbol $1/p$ indicates integration.

The operator p obeys all algebraic laws except for the following case, which arises in the cancellation of operators: This can be shown by writing Eq. (2.10) in the form

$$x(t) = \frac{1}{p}[pf(t)] = f(t) + C = f(t) - f(t_0) + x(t_0) = f(t)\Big|_{t_0}^{t} + x(t_0) \tag{2.13}$$

where C is obtained by evaluating Eq. (2.10) at the initial condition, i.e.,

$$C = [x(t) - f(t)]_{t=t_0} = x(t_0) - f(t_0)$$

The cancellation of operators gives the erroneous result

$$x(t) = \frac{1}{p} pf(t) = f(t) \qquad \text{invalid result} \tag{2.14}$$

In general, it may also be shown that

$$\frac{1}{p^m} p^n f(t) \neq p^{n-m} f(t) \tag{2.15}$$

The algebraic cancellation of the operator in Eq. (2.14) did not regard the constant of integration that arises from the integration indicated by Eq. (2.12). Thus, it is not possible to cancel operators when integration is the last operation to be performed unless there is no constant of integration, as is the case when the initial conditions are all zero.

Because no initial-condition terms arise from the differentiation process, it does follow that

$$p^n \frac{1}{p^m} f(t) = p^{n-m} f(t) \tag{2.16}$$

Thus, one may cancel operators when differentiating an integral, but not when integrating a differential unless the initial conditions are zero. In deriving the equations for control systems, one seldom has occasion for canceling operators.

In this text, the Heaviside operator notation is employed as an aid in obtaining the differential equation for control components and systems. In Chap. 5, it is shown how these differential equations may be solved algebraically by the Laplace transformation technique. These differential equations could be solved by use of Heaviside operational methods,[1-3] but this is considerably more tedious and complicated than the Laplace transform method.

2.4. Rotational Mechanical Components. A torsional spring is characterized by the equation

$$T_s = K_s \theta \tag{2.17}$$

where T_s = torque tending to twist spring
K_s = torsional spring rate
θ = angular displacement of spring

[1] H. S. Carslaw and J. C. Jaeger, "Operational Methods in Applied Mathematics," Oxford University Press, New York, 1941.

[2] N. W. McLachlan, "Complex Variable and Operational Calculus," Cambridge University Press, New York, 1953.

[3] G. W. Carter, "The Simple Calculation of Electrical Transients," The Macmillan Company, New York, 1945.

A well-known example of a torsional spring is a shaft as shown in Fig. 2.5. The right end of the shaft is displaced an angle θ with respect to the left end because of the twisting torque T_s. For a straight shaft, the torsional spring rate is

$$K_s = \frac{\pi D^4 G}{32L} \tag{2.18}$$

where G = modulus of elasticity in shear
D = diameter of shaft
L = length of shaft

The torque T_d required to overcome viscous friction of a rotating member is

$$T_d = C_v\omega = C_v\frac{d\theta}{dt} = C_v p\theta \tag{2.19}$$

where C_v = coefficient of viscous friction
ω = angular velocity

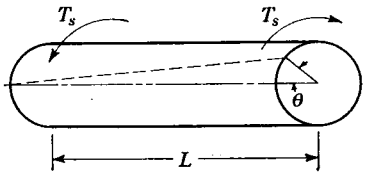


FIG. 2.5. Shaft acting as a torsional spring.

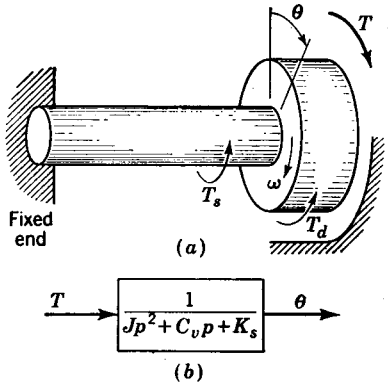


FIG. 2.6. Torsional inertia-spring-damper combination.

A disk rotating in a viscous medium and supported by a shaft is shown in Fig. 2.6*a*. The applied torque tending to rotate the disk is T. The shaft torque and viscous friction oppose the motion so that

$$\Sigma T_e = T - T_s - T_d = J\alpha = Jp^2\theta \tag{2.20}$$

where ΣT_e is the summation of external torques acting on the disk. The substitution of T_s from Eq. (2.17) and T_d from Eq. (2.19) into Eq. (2.20) yields

$$T = (Jp^2 + C_v p + K_s)\theta \tag{2.21}$$

The block-diagram representation for this system is shown in Fig. 2.6*b*.

2.5. Electrical Components. The resistor, inductor, and capacitor are the three basic components of electrical circuits. The equation for the voltage drop E_R across a resistor is

$$E_R = RI \tag{2.22}$$

where R is the resistance in ohms and I is the current flowing through the resistor in amperes.

For an inductor, the voltage drop E_L is given by the equation

$$E_L = L\frac{dI}{dt} = LpI \tag{2.23}$$

where L is the inductance in henrys.

Similarly, the voltage drop E_C across a capacitor is

$$E_C = \frac{1}{Cp}I \tag{2.24}$$

where C is the capacitance in farads.

The diagrammatic representations of Eqs. (2.22) to (2.24) are shown in Fig. 2.7.

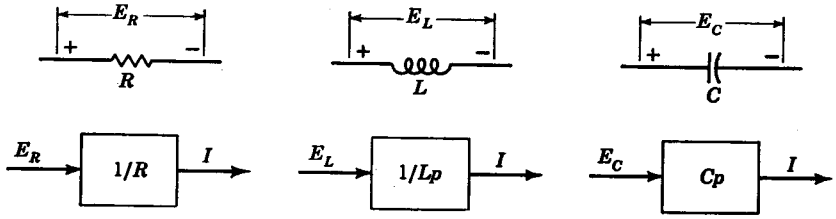


FIG. 2.7. Representation for resistor, inductor, and capacitor.

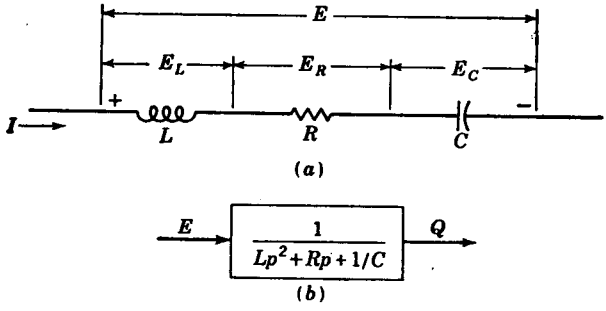


FIG. 2.8. *RLC* series circuit.

For the series *RLC* circuit shown in Fig. 2.8*a*, the total voltage drop E is the sum of the voltage drop across the inductor E_L, plus that across the resistor E_R and that across the capacitor E_C.

$$E = E_L + E_R + E_C = \left(Lp + R + \frac{1}{Cp}\right)I \tag{2.25}$$

The charge Q is the time integral of the current, that is, $Q = (1/p)I$. By noting that $LpI = Lp^2(I/p) = Lp^2Q$, $RI = Rp(I/p) = RpQ$, $1/C\ (I/p) = (1/C)Q$, Eq. (2.25) becomes

$$E = \left(Lp^2 + Rp + \frac{1}{C}\right) Q \tag{2.26}$$

The over-all block-diagram representation for this *RLC* circuit is shown in Fig. 2.8*b*.

2.6. Series and Parallel Combination of Elements. Often, many elements are connected in either a series or a parallel arrangement. Much simplification in arriving at the equation for such systems is afforded by the use of the theorem for series and the theorem for parallel combinations.

Series Electrical Circuits. A general series circuit is shown in Fig. 2.9*a*. In a series circuit, *the total voltage drop E is the sum of the individual*

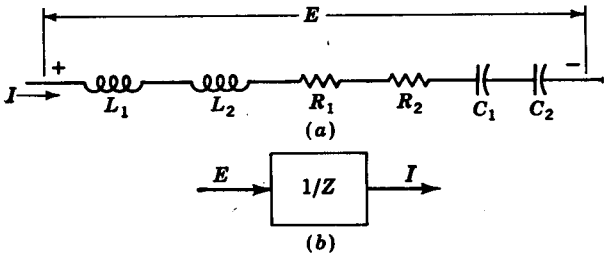


FIG. 2.9. General series circuit.

voltage drops across each element and *the same current I flows through each element.* The equation for the summation of the voltage drops is

$$E = \left(L_1p + L_2p + R_1 + R_2 + \frac{1}{C_1p} + \frac{1}{C_2p}\right) I = ZI \tag{2.27}$$

Thus, the equivalent impedence Z for elements in series is

$$Z = L_1p + L_2p + R_1 + R_2 + \frac{1}{C_1p} + \frac{1}{C_2p} \tag{2.28}$$

so that the equivalent representation is shown in Fig. 2.9*b*.

Parallel Electrical Circuit. A general combination of electrical elements in parallel is shown in Fig. 2.10*a*. The distinguishing features of a parallel arrangement are that *the voltage drop E across each element is the same,* and *the total current I flowing into the system is the sum of the currents flowing through each element.* Thus

$$I = \frac{E}{L_1p} + \frac{E}{L_2p} + \frac{E}{R_1} + \frac{E}{R_2} + \frac{E}{1/C_1p} + \frac{E}{1/C_2p} \tag{2.29}$$

or

$$E = \frac{1}{1/L_1p + 1/L_2p + 1/R_1 + 1/R_2 + C_1p + C_2p} I = ZI$$

The equivalent impedence Z for elements in parallel is

$$Z = \frac{1}{1/L_1p + 1/L_2p + 1/R_1 + 1/R_2 + C_1p + C_2p} \tag{2.30}$$

so that the equivalent representation is shown in Fig. 2.10*b*.

Illustrative Example 1. For the circuit shown in Fig. 2.11, let it be desired to determine the equation relating the output voltage E_2 to the input voltage E_1.

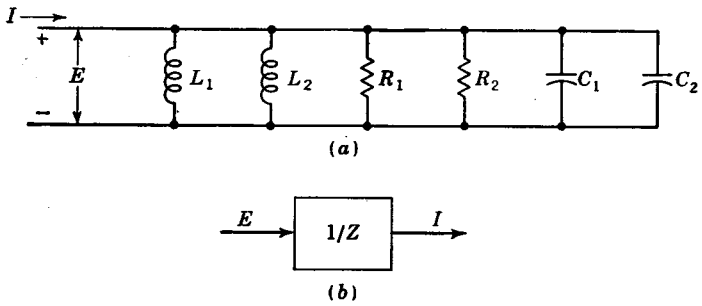


FIG. 2.10. General parallel circuit.

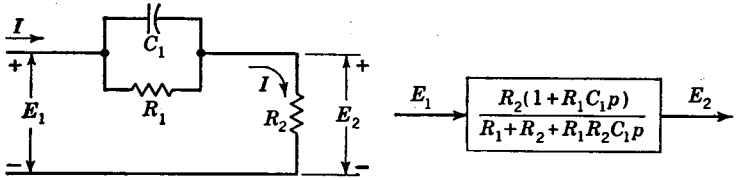


FIG. 2.11. Electrical circuit.

SOLUTION. The parallel combination of R_1 and C_1 is in series with R_2, so that the total impedance Z is

$$Z = Z_1 + R_2 = \frac{1}{1/R_1 + C_1p} + R_2 = \frac{R_1}{1 + R_1C_1p} + R_2 \tag{2.31}$$

The voltage E_1 is given by the equation

$$E_1 = ZI = \frac{R_1 + R_2 + R_1R_2C_1p}{1 + R_1C_1p} I \tag{2.32}$$

and similarly E_2 is

$$E_2 = R_2I \tag{2.33}$$

The substitution of I from Eq. (2.32) into Eq. (2.33) yields the desired answer

$$E_2 = \frac{R_2(1 + R_1C_1p)}{R_1 + R_2 + R_1R_2C_1p} E_1 \tag{2.34}$$

Series Mechanical Elements. A series arrangement of linear mechanical elements is shown in Fig. 2.12*a*. In general, it is better to use the equivalent "grounded-chair" representation for a mass, as shown in Fig. 2.12*b*, rather than the more common representation of Fig. 2.12*a*. The fact that the mass is in series with the other elements is more readily

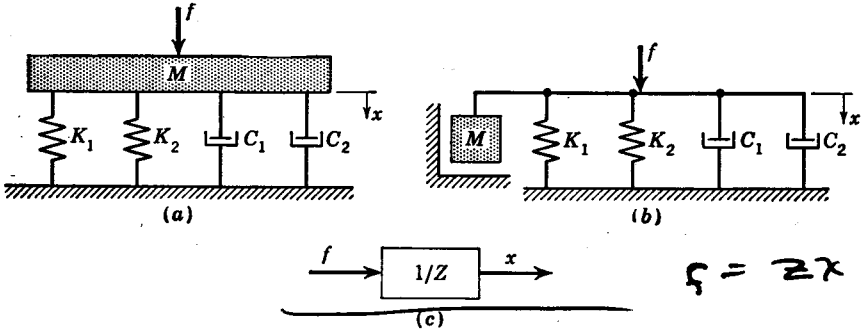


FIG. 2.12. Mechanical elements in series.

seen from Fig. 2.12*b* than from Fig. 2.12*a*. In determining inertia force, the acceleration of a mass is always taken with respect to the earth. Thus, providing the grounded chair to indicate motion relative to ground is a more justifiable representation than Fig. 2.12*a*, which shows better the actual physical arrangement of the elements in the system. For series mechanical elements, *the force f is equal to the summation of the forces acting on each individual component,* and *each element undergoes the same displacement.* Thus

$$f = (K_1 + K_2 + C_1p + C_2p + Mp^2)x = Zx \quad (2.35)$$

where x and f are measured from a convenient reference operating point. The equivalent impedence for mechanical elements in series is

$$Z_s = K_1 + K_2 + C_1p + C_2p + Mp^2 \quad (2.36)$$

The equivalent representation for this system is shown in Fig. 2.12*c*.

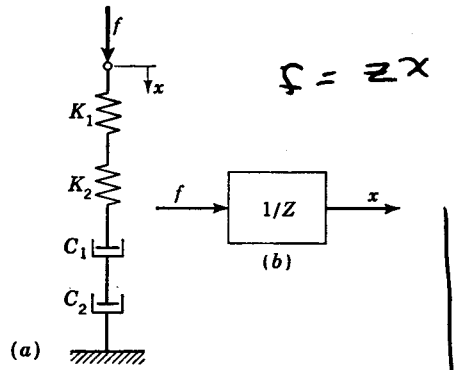


FIG. 2.13. Mechanical elements in parallel.

Parallel Mechanical Elements. A parallel combination of mechanical elements is shown in Fig. 2.13*a*. For parallel elements, *the force f is transmitted through each element.* In addition, *the deflection x is seen to be the sum of the individual deflections of each element.* Thus

$$x = \frac{f}{K_1} + \frac{f}{K_2} + \frac{f}{C_1 p} + \frac{f}{C_2 p} \tag{2.37}$$

or

$$f = \frac{1}{1/K_1 + 1/K_2 + 1/C_1 p + 1/C_2 p} x = Zx$$

The equivalent impedance for mechanical elements in parallel is

$$Z = \frac{1}{1/K_1 + 1/K_2 + 1/C_1 p + 1/C_2 p} \tag{2.38}$$

The equivalent representation is shown in Fig. 2.13*b*.

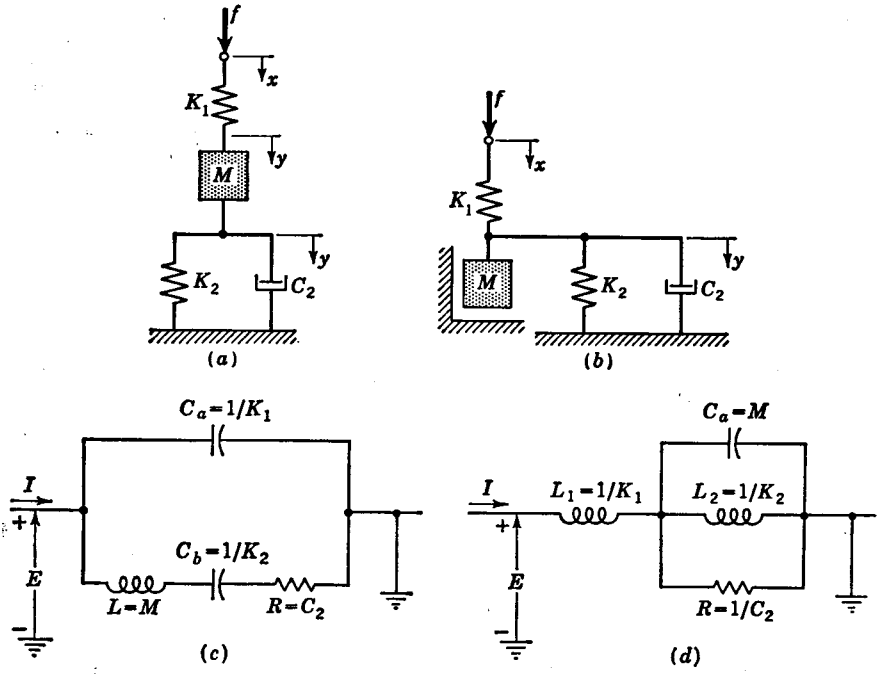


FIG. 2.14. Mechanical system.

A necessary condition for parallel elements is that the force be transmitted through each element. Springs and dampers satisfy this condition because the force is the same on both sides. However, this is not the case for a mass such as that shown in Fig. 2.14*a*, because the difference in forces acting on both sides of a mass is utilized in acceleration. Thus, a mass located between other elements cannot be in parallel with them. A mass can be in parallel only if it is the last element, as shown in Fig. 2.15. For this system, the displacement x is

$$x = (x - y) + (y - z) + z = \left(\frac{1}{K} + \frac{1}{Cp} + \frac{1}{Mp^2}\right) f \tag{2.39}$$

Parallel and series rules for rotational mechanical components may also be developed by extending the preceding techniques.[1,2]

Illustrative Example 2. For the mass-spring-damper combination shown in Fig. 2.14*a*, determine the equation relating f and x, the equation relating f and y, and the equation relating x and y.

SOLUTION. The first step is to draw the equivalent grounded-chair system, in which the motion of the mass with respect to ground is clearly indicated as shown in Fig. 2.14*b*. The spring K_1 is in parallel with the series combination of M, K_2, and C_2. Thus

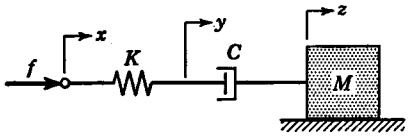


FIG. 2.15. Parallel mass-spring-damper combination.

$$f = Zx = \frac{1}{1/K_1 + 1/Z_2}\,x = \frac{K_1 Z_2}{K_1 + Z_2}\,x \tag{2.40}$$

or
$$\frac{f}{Z_2} = \frac{f}{Mp^2 + C_2 p + K_2} = \frac{K_1 x}{Mp^2 + C_2 p + K_1 + K_2} \tag{2.41}$$

where $Z_2 = Mp^2 + C_2 p + K_2$

The force f is transmitted through the spring K_1 and acts upon the series combination of M, K_2, and C_2. Thus, the equation of motion for this part of the system which relates f and y is

$$\frac{f}{Mp^2 + C_2 p + K_2} = y \tag{2.42}$$

Equating (2.41) and (2.42) yields the desired relationship between x and y. That is,

$$y = \frac{K_1}{Mp^2 + C_2 p + K_1 + K_2}\,x \tag{2.43}$$

Analogies. The equation of operation for the series mechanical system of Fig. 2.4*a* is given by Eq. (2.5), and the equation for the series electrical circuit of Fig. 2.8*a* is given by Eq. (2.26). Comparison of corresponding terms in Eqs. (2.5) and (2.26) shows that the differential equation of operation for each system has the same form. The terms which occupy corresponding positions are called analogous quantities. This particular analogy is referred to as the force-voltage analogy. The analogous quantities for a force-voltage analogy are shown in Table 2.1.

The total force acting on a group of mechanical elements in series is equal to the sum of the forces exerted on each element. Similarly, the

[1] H. F. Olson, "Dynamical Analogies," 2d ed., D. Van Nostrand Company, Inc., Princeton, N.J., 1958.

[2] R. L. Sutherland, "Engineering Systems Analysis," Addison-Wesley Publishing Company, Reading, Mass., 1958.

TABLE 2.1. ANALOGOUS QUANTITIES IN A FORCE-VOLTAGE ANALOGY

Translational mechanical system	Force	Mass	Viscous damping coefficient	Spring constant	Displacement	Velocity
	f	M	C	K	x	$\dot{x} = px$
Electrical system	Voltage	Inductance	Resistance	Reciprocal of capacitance	Charge	Current
	E	L	R	$\frac{1}{C}$	Q	$I = pQ$

total voltage drop across a group of electrical elements in series is equal to the sum of the voltage drops across each element. Thus, in constructing a force-voltage analogy, series mechanical elements are replaced by analogous series electrical elements.

For parallel mechanical elements, the force acting on each element is the same, and for parallel electrical elements the voltage drop across each element is the same. Thus, in a force-voltage analogy, parallel mechanical elements should be replaced by equivalent electrical elements in parallel.

Another type of analogy which is commonly employed is the force-current analogy. To construct a force-current analog, it should first be noted that the *total current* flowing through a group of electrical elements in parallel is the *sum* of the currents in each element. This is analogous to the fact that the *total force* acting on a group of mechanical elements in series is the *sum* of the forces acting on each element. Thus, to construct a force-current analog, series mechanical elements must be replaced by parallel electrical elements. Similarly, in a force-current analogy, it may be shown that parallel mechanical elements should be replaced by series electrical elements.

Analogous quantities for a force-current analogy may be determined by comparing the equation of operation for the parallel mechanical system of Fig. 2.15 with that for the series electrical system of Fig. 2.8*a*. The equation of operation for the parallel mechanical system of Fig. 2.15 is given by Eq. (2.39). Multiplication of both sides of Eq. (2.39) by p gives

$$\dot{x} = \left(\frac{p}{K} + \frac{1}{C} + \frac{1}{Mp}\right) f \tag{2.44}$$

The operation of the series electrical circuit of Fig. 2.8*a* is described by Eq. (2.25), which has the same form as Eq. (2.44). Comparison of corresponding terms in Eqs. (2.25) and (2.44) yields the analogous quantities for a force-current analog that are shown in Table 2.2.

TABLE 2.2. ANALOGOUS QUANTITIES IN A FORCE-CURRENT ANALOGY

Translational mechanical system	Force f	Velocity $\dot{x}$	Spring constant K	Damping coefficient C	Mass M
Electrical system	Current I	Voltage E	Reciprocal of inductance $\frac{1}{L}$	Reciprocal of resistance $\frac{1}{R}$	Capacitance C

Illustrative Example 3. Let it be desired to determine the electrical analogy for the mechanical system of Fig. 2.14*b* by using (*a*) analogous force-voltage terms, (*b*) analogous force-current terms.

SOLUTION. (*a*) The electrical force-voltage analogy for the mechanical system of Fig. 2.14*b* is shown in Fig. 2.14*c*. Note that the capacitor C_a is in parallel with the series combination of L, C_b, and R, just as the spring K_1 of Fig. 2.14*b* is in parallel with the series combination of M, K_2, and C_2. The equation of operation for the electrical circuit of Fig. 2.14*c* is

$$E = ZI = \frac{1}{\dfrac{1}{1/C_a p} + \dfrac{1}{Lp + 1/C_b p + R}} I$$

or

$$E = \frac{C_b(Lp^2 + Rp + 1/C_b)}{C_a C_b[Lp^2 + Rp + (1/C_a + 1/C_b)]} \frac{I}{p}$$

or

$$\frac{E}{Lp^2 + Rp + 1/C_b} = \frac{1/C_a}{Lp^2 + Rp + (1/C_a + 1/C_b)} Q \qquad (2.41a)$$

Comparison of corresponding terms in Eqs. (2.41) and (2.41*a*) verifies the force-voltage analogies given in Table 2.1.

(*b*) The resulting force-current analogy for Fig. 2.14*b* is shown in Fig. 2.14*d*. In Fig. 2.14*d*, it is to be noted that the inductor L_1 is in series with the parallel combination of C_a, L_2, and R, whereas in Fig. 2.14*b* the spring K_1 is in parallel with the series combination of M, K_2, and C_2. The equation of operation for the electrical circuit of Fig. 2.14*d* is

$$E = ZI = \left(L_1 p + \frac{1}{C_a p + 1/L_2 p + 1/R}\right) I$$

or

$$I = \frac{RL_2[C_a p^2 + (1/R)p + 1/L_2]}{RL_1L_2[C_a p^2 + (1/R)p + (1/L_1 + 1/L_2)]} \frac{E}{p}$$

or

$$\frac{I}{C_a p^2 + (1/R)p + 1/L_2} = \frac{1/L_1}{C_a p^2 + (1/R)p + (1/L_1 + 1/L_2)} \frac{E}{p} \qquad (2.41b)$$

Comparison of corresponding terms in Eqs. (2.41) and (2.41*b*) verifies the force-current analogies given in Table 2.2. In the force-current analogy, velocity is analogous to voltage, and thus displacement is analogous to the integral of voltage.

Comparison of corresponding terms in Eqs. (2.5) and (2.21) shows that a torque-force analogy may be developed in which series translational mechanical elements are replaced by series rotational elements. Similarly, parallel translational elements should be replaced by parallel rotational elements. Analogies may be developed for other phenomena such as fluid flow, thermal processes, etc.

A major use for analogs is that sometimes it is easier to study experimentally one type of system rather than another. For example, it may be easier to change a resistance rather than the coefficient of viscous friction. Whenever possible, it is best to work with the system directly rather than to consider the operation of an analogous system. This eliminates the chance for error in construction of the analogy. Also, when carried far enough, analogies usually break down because things which are physically possible for one component may be impossible for the analogous component.

Degrees of Freedom. By degrees of freedom is meant the number of coordinates required to specify the position of all the elements in a mechanical system. Thus, the system shown in Fig. 2.14*b* has two degrees of freedom, while that shown in Fig. 2.15 has three degrees of freedom.

Rather than using the series and parallel laws, an alternate method of determining the equation of operation of mechanical systems is to write the force balance at each coordinate. For example, the force equation at the x coordinate of Fig. 2.14*b* is

$$K_1(x - y) = f \tag{2.45}$$

The compression of the spring K_1 is $x - y$, and as x increases, so does the spring force, but as y increases, the spring force decreases. The force balance at the y coordinate is

$$K_1(x - y) = (K_2 + C_2p + Mp^2)y \tag{2.46}$$

The two preceding force equations may be written in the form

$$\begin{aligned} K_1x \qquad\qquad\quad &- K_1y = f \\ K_1x - (K_1 + K_2 + C_2p + Mp^2)y &= 0 \end{aligned} \tag{2.47}$$

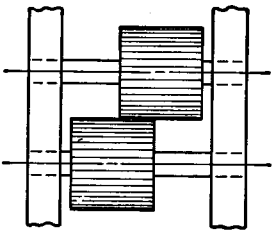
These two equations may be solved simultaneously to yield any of the desired relationships between x, y, or f. For example, solving for y gives

$$y = \frac{\begin{vmatrix} K_1 & f \\ K_1 & 0 \end{vmatrix}}{\begin{vmatrix} K_1 & -K_1 \\ K_1 & -(K_1 + K_2 + C_2p + Mp^2) \end{vmatrix}}$$

$$= \frac{K_1 f}{K_1[(K_1 + K_2 + C_2p + Mp^2) - K_1]} = \frac{f}{K_2 + C_2p + Mp^2} \qquad (2.48)$$

Suppose that it is desired to consider y positive for upward motion rather than for downward motion. (Downward motion is shown in Fig. 2.14*b*.) This reversal would change the sign of each y term in Eqs. (2.45) and (2.46), which would change the sign of y in the resultant expression given by Eq. (2.48). In effect, reversing the positive sense of a coordinate merely changes its sign.

In applying the series or parallel laws to mechanical elements, care must be exercised to take the positive sense of motion of each coordinate (this is indicated by the arrow at each coordinate) in the same direction as that of the applied force. It is a simple matter to later change the positive sense of a coordinate by merely changing the sign of the corresponding term in the derived equation.



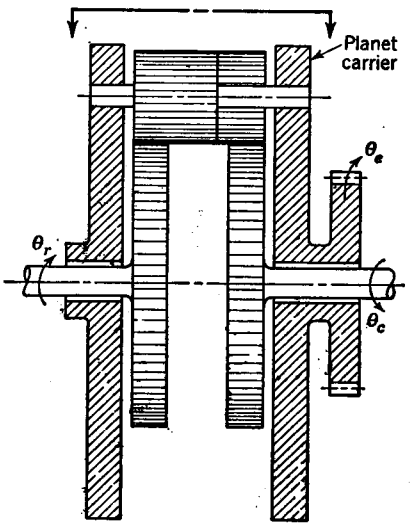


FIG. 2.16. Differential gear train.

2.7. Comparators for Rotational or Linear Motions. When the reference input and the feedback signal are each represented by the angular position of a shaft, then a differential gear train may be used to measure the difference. A schematic representation of such a gear train is shown in Fig. 2.16, and its equation of operation is

$$\theta_e = \frac{\theta_r - \theta_c}{2} \qquad (2.49)$$

where θ_e is the angular position of the planet carrier or cage which is a measure of the error, the shaft position θ_r is the reference input, and θ_c is the feedback angular position. Such a device might be used as the comparator of a system used for the remote control of the angular position

θ_c of a large mass such as a radar tracking antenna. The use of two differentials, one for azimuth and one for elevation, is required for orientation of an object in space.

For a typical control system, the planet carrier is connected to a power-amplifying device such that when θ_e is positive a torque is transmitted to increase θ_c and when θ_e is negative a torque is applied to decrease θ_c. When θ_e is zero, then the value of θ_c remains constant. For this case, it is seen from Eq. (2.49) that $\theta_c = \theta_r$.

To understand the operation of this device better, suppose initially that $\theta_e = 0$ and also $\theta_c = \theta_r = 0$. If the reference input position is instantaneously increased by 10°, then from Eq. (2.49) it follows that θ_e changes by 5°. This in turn causes a torque to be transmitted to increase θ_c. As θ_c increases, the planet carrier gradually returns to its initial position. Thus, when $\theta_c = \theta_r = 10°$, then θ_e is again zero.

For subtracting linear motions, one could use a rack and pinion to convert the linear motions to rotations and then use a differential gear train. However, the device shown in Fig. 2.17 subtracts linear motions directly. With the position of the lower rack x_c held fixed, then it is seen that the motion x_e is one-half the motion of the upper rack x_r. Similarly, with the upper rack held fixed, the motion x_e is one-half that of x_c. However, it should be noticed that x_e decreases as x_c increases. Because it makes no difference whether the movements of x_r and x_c occur at different times or simultaneously, the total movement of x_e is that due to a change in x_r plus that due to a change in x_c. Thus the equation for this mechanism is

$$x_e = \frac{x_r - x_c}{2} \tag{2.50}$$

It should be noted that changing the positive sense of motion for x_c in Fig. 2.17 changes the sign in front of x_c in Eq. (2.50) so that a summing device results.

2.8. Integrating Devices. In control systems, the error signal coming from the comparator is often fed into an integrating device. The reason for this is that because of friction, backlash, etc., the system might not detect a very small error, but the integral of a small error continually increases with time so that the system eventually detects it.

A device for integrating mechanically is shown in Fig. 2.18. This is called a ball-and-disk type of integrator. A differential rotation $d\theta$ of the input position θ produces a linear motion $r\,d\theta$ which is transmitted through the two balls to the output shaft of radius R and angular position ϕ. The term r is the distance from the centerline of the input shaft to the balls. The reason for the two balls is to ensure that pure rolling exists between the elements even when the position r is being varied.

There could not be pure rolling if there were only one ball. The differential equation of operation for this device is

$$R\,d\phi = r\,d\theta \tag{2.51}$$

Because this is a continuously acting device, it sums up or integrates all incremental motions. Thus, integration of the preceding expression gives

$$\phi = \frac{1}{R}\int r\,d\theta \tag{2.52}$$

The value of r is varied in proportion to the function that is to be integrated. The application of this device to integrate an error signal is demonstrated as follows: Let the distance r be varied in proportion to

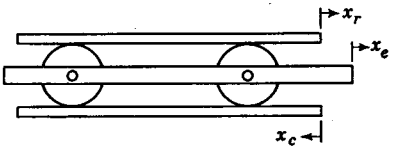


FIG. 2.17. Translational differential mechanism.

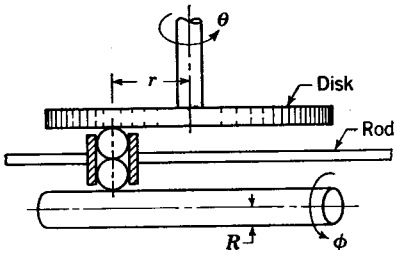


FIG. 2.18. Ball-and-disk integrator.

the error signal ($r = C_1 e$), and let the input shaft be driven at a constant angular velocity $\omega = d\theta/dt$ or $d\theta = \omega\,dt$, so that Eq. (2.52) becomes

$$\phi = \frac{1}{R}\int C_1 e\omega\,dt = \frac{C_1\omega}{R}\int e\,dt \tag{2.53}$$

Thus the position ϕ of the output shaft is seen to be proportional to the time integral of the error. A more thorough treatment of computing mechanisms may be obtained by referring to other publications.[1–3]

A hydraulic valve-and-piston combination which in effect integrates hydraulically is shown in Fig. 2.19*a*. The position of the valve is designated by x, and the position of the large piston which moves the load is y. This type of valve is called a balanced valve because the pressure forces acting on it are all balanced so that it requires little force

[1] W. W. Soroka, "Analog Methods in Computation and Simulation," McGraw-Hill Book Company, Inc., 1954.

[2] G. W. Michalec, Survey and Evaluation: Analog Computing Mechanisms, *Machine Design*, vol. 31, no. 6, pp. 157–179, Mar. 19, 1959.

[3] H. H. Mabie and F. W. Ocvirk, "Mechanisms and Dynamics of Machinery," pp. 175–194, John Wiley & Sons, Inc., New York, 1957.

to change its position. When the valve is moved upward, the supply pressure admits oil to the upper side of the piston and the fluid in the lower side of the piston is returned to the drain, where it is recirculated in the system through the pump. For the reverse process, the valve is moved downward so that the supply pressure is connected to the bottom side of the big piston. The upper side of this piston is connected to the upper drain to permit return flow to the pump.

For a constant pressure drop across the valve, the rate of flow to the piston is proportional to the area uncovered by the valve, which is seen to be proportional to the position x. Thus

$$\frac{\text{in}^3}{\text{min}} = C \text{ in} \qquad q = C_1 x \quad C = \frac{\text{in}^2}{\text{min in}} \tag{2.54}$$

where q is the rate of flow through the valve into the piston chamber.

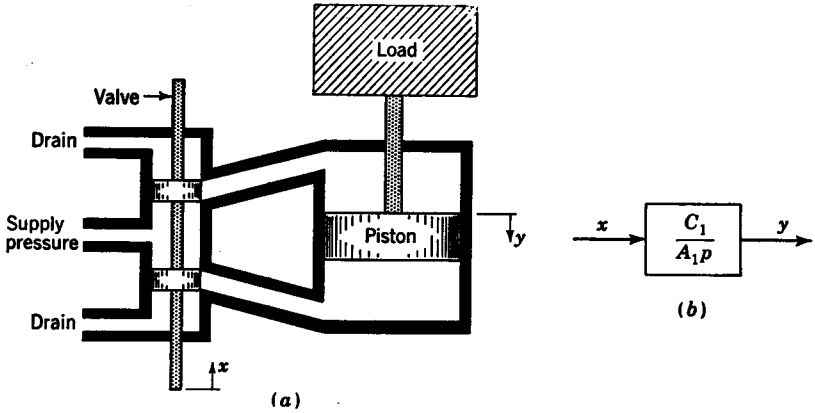


FIG. 2.19. Hydraulic valve and piston.

This rate of flow q into the piston chamber is equal to the rate of change in volume of the chamber, which is equal to the piston velocity py times the area A_1 of the piston.

$$q = A_1 py \tag{2.55}$$

Equating the preceding expressions for q and solving for y gives

$$y = \frac{C_1}{A_1 p} x \tag{2.56}$$

The block-diagram representation for this hydraulic integrator is shown in Fig. 2.19*b*.

2.9. Nonlinear Devices. In Fig. 2.20*a* is shown a mechanical linkage which is used for obtaining the square of a number. The point B is the corner of the right-angle linkage ($\alpha + \beta = 90°$) and is constrained to move in the vertical track. One leg of this right-angle linkage passes

through the slider, which pivots at point A. The other leg of the linkage must always pass through the slider at point C, which in turn is constrained to move in the horizontal track. Because of the geometry of this device, triangles BOC and AOB will always be similar. Therefore

$$\frac{Y}{X} = \frac{X}{K}$$

or

$$Y = \frac{X^2}{K} \tag{2.57}$$

where K is a constant.

Considering X as the input quantity, the position of point C and thus Y, the output, will vary as the square of X. If the operation were reversed so that Y were the input and X the output, this mechanism would be a square-root device in which $X = \sqrt{KY}$. Depending on

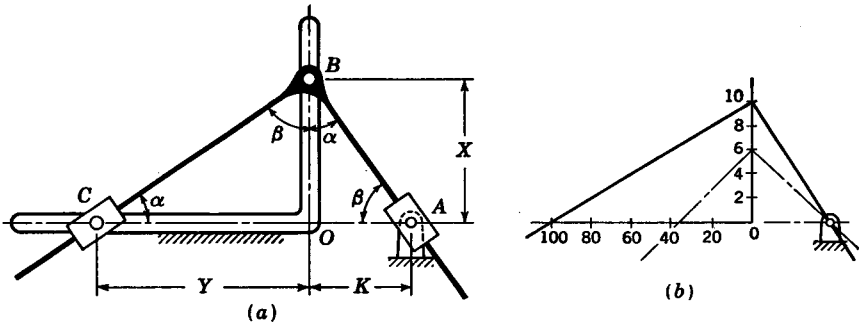


FIG. 2.20. (*a*) Mechanical squaring device; (*b*) skeletal representation of squaring device.

which scale is used for the input, this device may be used for obtaining squares or square roots, as is illustrated by the skelton diagram of Fig. 2.20*b*. The scale of this diagram is for the case in which $K = 1$.

For an equation to be linear, each variable term $X_1, X_2, X_3, \ldots, X_n$ in the equation for Y must be of the first power, and the contribution of each term must be added independently as in the general linear equation shown below.

$$Y = C_1X_1 + C_2X_2 + C_3X_3 + \cdots + C_nX_n \tag{2.58}$$

In the following section, it is shown that, for small changes of the variable terms, nonlinear relationships may be approximated by linear equations. Linear methods of analysis may then be applied for the analysis of such linearized systems. Additional methods for treating nonlinear phenomena are discussed in Chap. 15.

The most powerful methods of system analysis have been developed

for linear control systems. By a linear control system is meant one in which all the relationships between the variables can be expressed by linear differential equations, usually with constant coefficients. The reason why differential equations are obtained is that time is always a variable in feedback control systems. For example, in controlling temperature, the actuating signal causes a change in heat flow, and time is required for this added heat to bring the temperature to its desired value. In speed control systems, the actuating signal causes a change in power of the prime mover, but it takes time to accelerate or decelerate the engine to its desired speed. Similarly, in pressure control systems, it takes time to bring the pressure in a chamber to some desired value.

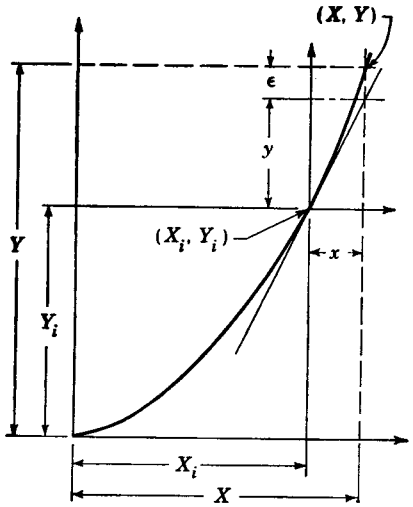


FIG. 2.21. Graph of function $Y = X^2/K$.

2.10. Linear Approximation of a Nonlinear Function. A plot of the nonlinear relationship given by Eq. (2.57) is shown in Fig. 2.21. It is to be noticed that, in the vicinity of the point of interest (X_i,Y_i), the nonlinear function $Y = X^2/K$ is closely approximated by the tangent to the function. For example, consider a new operating point (X,Y) on the curve of the nonlinear function. The abscissa X is seen to be displaced a distance x from X_i. This abscissa X intersects the nonlinear function a vertical distance $y + \epsilon$ from Y_i, and it intersects the tangent a distance y from Y_i. The equation for Y is

$$Y = Y_i + y + \epsilon \approx Y_i + y \tag{2.59}$$

Lower-case letters indicate the variation of the capital-letter parameters from the point of interest or the reference point. From the

geometry of Fig. 2.21, it is seen that the slope of the tangent line is

$$\frac{y}{x} = \left.\frac{dY}{dX}\right|_i = \text{slope at point } (X_i, Y_i)$$

The symbol $|_i$ means that the derivative is to be evaluated at the reference condition. Thus

$$y = \left.\frac{dY}{dX}\right|_i x = \left.\frac{d}{dX}\left(\frac{X^2}{K}\right)\right|_i x = \frac{2X_i}{K} x \tag{2.60}$$

Substitution of y from Eq. (2.60) into Eq. (2.59) yields the following linear approximation for Y:

$$Y \approx Y_i + \frac{2X_i}{K} x \tag{2.61}$$

Illustrative Example 1. Effect a linear approximation for the equation $Y = X^2$ for values of X in the neighborhood of 10, and find the error when using this approximation for $X = 11$. The reference values are $X_i = 10$ and $Y_i = X_i^2 = 100$. Substitution of $x = X - X_i = 1$ and $K = 1$ into Eq. (2.61) gives

$$Y \approx 100 + \frac{2(10)(1)}{1} = 120 \tag{2.62}$$

The exact value is $Y = X^2 = 121$; thus the error is 1 part in 121, less than 1 per cent.

A more general procedure for obtaining a linear approximation is to use the expression derived in calculus[1,2] for approximating the incremental variation ΔY for a function $Y = F(X_1 X_2, \ldots, X_n)$ of n independent variables. That is,

$$\Delta Y = \left.\frac{\partial Y}{\partial X_1}\right|_i \Delta X_1 + \left.\frac{\partial Y}{\partial X_2}\right|_i \Delta X_2 + \cdots + \left.\frac{\partial Y}{\partial X_n}\right|_i \Delta X_n \tag{2.63}$$

By using the lower-case letters to represent variations from the reference conditions, it follows that

$$\begin{aligned} Y - Y_i &= \Delta Y = y \\ X_1 - X_{1_i} &= \Delta X_1 = x_1 \\ X_2 - X_{2_i} &= \Delta X_2 = x_2 \\ X_n - X_{ni} &= \Delta X_n = x_n \end{aligned} \tag{2.64}$$

Thus, the general expression for obtaining a linear approximation for a nonlinear function is

[1] Louis A. Pipes, "Applied Mathematics for Engineers and Physicists," 2d ed., McGraw-Hill Book Company, Inc., New York, 1958.

[2] I. S. Sokolnikoff and E. S. Sokolnikoff, "Higher Mathematics for Engineers and Physicists," 2d ed., McGraw-Hill Book Company, Inc., New York, 1941.

$$y = C_1x_1 + C_2x_2 + \cdots + C_nx_n \tag{2.65}$$

where $C_1 = \left.\frac{\partial Y}{\partial X_1}\right|_i \quad C_2 = \left.\frac{\partial Y}{\partial X_2}\right|_i$ etc.

Evaluation of these partial derivatives at the reference conditions yields constants.

The application of the general expression Eq. (2.65) to the nonlinear equation $Y = (1/K)X^2$ is effected as follows: The independent variable is X, which corresponds to X_1 in the general equation. Thus

$$C_1 = \left.\frac{\partial Y}{\partial X}\right|_i = \left.\frac{\partial}{\partial X}\left(\frac{X^2}{K}\right)\right|_i = \left.\frac{2X}{K}\right|_i = \frac{2X_i}{K}$$

From Eq. (2.65)

$$y = C_1x = \frac{2X_i}{K}x$$

The absolute value of Y is

$$Y \approx Y_i + y = Y_i + \frac{2X_i}{K}x \tag{2.66}$$

This is the result obtained by the preceding geometric interpretation and given by Eq. (2.61).

The need for linearizing nonlinear relationships is frequently encountered in control engineering. For example, most mechanical speed control systems incorporate a flyball governor for sensing the speed error. This is a centrifugal device, so that a force is obtained which is proportional to the square of the speed. In the design of hydraulic equipment in which the working medium is an incompressible fluid, one encounters the nonlinear equations which govern such fluid flow. Similarly, the working medium for pneumatic equipment is air, whose flow is described by nonlinear relationships.

Illustrative Example 2. Effect the linear approximation for P in the equation of state $PV = WRT$. The reference conditions are $P_i = 100$ lb_f/ft^2, $V_i = 100$ ft^3, $W_i = 10/53.3$ lb_m, and $T_i = 1000°R$. Determine the per cent error in using this approximation for P when $V = 110$ ft^3, $T = 1200°R$, and W remains the same. The constant R is 53.3 $ft\text{-}lb_f/lb_m/°R$.

SOLUTION. From the equation of state and the fact that W remains constant, it is seen that P is a function of the independent variables T and V, or $P = F(T,V)$. Application of Eq. (2.65) to obtain the variation p of the pressure from its reference value yields

$$p = \left.\frac{\partial P}{\partial T}\right|_i t + \left.\frac{\partial P}{\partial V}\right|_i v \tag{2.67}$$

The partial derivatives are evaluated from the equation of state as

follows:

$$\left.\frac{\partial P}{\partial T}\right|_i = \left.\frac{\partial}{\partial T}\left(\frac{WRT}{V}\right)\right|_i = \left.\frac{WR}{V}\right|_i = \frac{(10)(53.3)}{(53.3)(100)} = 0.10$$

$$\left.\frac{\partial P}{\partial V}\right|_i = \left.\frac{\partial}{\partial V}\left(\frac{WRT}{V}\right)\right|_i = \left.\frac{-WRT}{V^2}\right|_i = -1.0$$

The linearized approximation for P is

$$P \approx P_i + p = P_i + 0.1t - v \tag{2.68}$$

From the given information, it follows that

$$v = V - V_i = 110 - 100 = 10$$
$$t = T - T_i = 200$$

Thus
$$P \approx 100 + (0.1)(200) - (10) = 110 \text{ lb}_f/\text{ft}^2 \tag{2.69}$$

The exact value of P is

$$P = \frac{WRT}{V} = \frac{(10)(53.3)(1{,}200)}{(53.3)(110)} = 109.1$$

Therefore, the per cent error is

$$\frac{(109.1 - 110)100}{109.1} = 0.82\%$$

2.11. Geometric Evaluation of Error Introduced by a Linear Approximation. From the linear approximation for the area of a rectangle, it

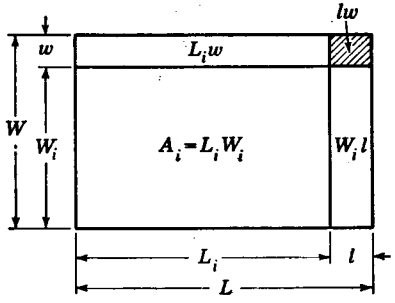


FIG. 2.22. Geometric representation of error incurred from the linear approximation for a rectangle.

is possible to represent geometrically the error which is introduced. In Fig. 2.22 is shown a rectangle in which the reference length is L_i and the width W_i. The area of a rectangle is a function of the length L and width W, so that the variation in the area from its reference size is obtained as follows:

$$A = F(L,W)$$
$$a = \left.\frac{\partial A}{\partial L}\right|_i l + \left.\frac{\partial A}{\partial W}\right|_i w \tag{2.70}$$

The preceding partial derivatives are evaluated from the equation $A = LW$,

$$\left.\frac{\partial A}{\partial L}\right|_i = W_i \quad \text{and} \quad \left.\frac{\partial A}{\partial W}\right|_i = L_i \tag{2.71}$$

Thus the linearized expression for the area A is

$$A \approx A_i + a = A_i + W_i l + L_i w \tag{2.72}$$

Each term in the preceding expression is represented by an area in Fig. 2.22. The difference between this approximation and the actual area LW is the small shaded portion lw.

Application of this linearization technique to Eq. (2.4) to obtain the equation for operation about some reference point gives $f = \left.\partial F/\partial X\right|_i x$, where from Eq. (2.4) it follows that

$$\left.\frac{\partial F}{\partial X}\right|_i = Mp^2 + Cp + K$$

Thus, the result previously obtained in Eq. (2.5) follows directly. As illustrated by this example, for equations which are already linear the substitution of lower-case letters for the capital-letter variable terms and dropping out the constant terms (for example, $-Mg$) yields directly the linear equation of operation about some reference point. Constant terms drop out because the partial derivative of a constant is zero.

2.12. Linearization of Operating Curves. In the preceding section, it was shown how equations which are nonlinear could be linearized. However, for many components encountered in control systems, the operating characteristics are given in the form of general operating curves rather than equations. For example, Fig. 2.23 shows a typical family of operating curves for an engine. Usually such curves are determined experimentally, and it would be quite tedious and difficult to express these curves as equations. The linearized equation for the operation of the engine about some reference operating point is obtained as follows: From Fig. 2.23, it is seen that the speed N is a function of the rate of fuel flow Q and the engine torque T, thus

$$N = F(Q,T) \tag{2.73}$$

Linearization gives

$$n = \left.\frac{\partial N}{\partial Q}\right|_i q + \left.\frac{\partial N}{\partial T}\right| t \tag{2.74}$$

The term $\partial N/\partial Q\Big|_i$ is the change in speed per change in fuel flow with all other parameters held constant (in this case with T constant). Thus, this partial derivative is equal to the reciprocal of the slope of the line of constant torque evaluated at the reference point. That is,

$$\frac{\partial N}{\partial Q}\bigg|_i = \frac{2{,}400 - 1{,}600}{32 - 20} = 66.7 \tag{2.75}$$

The partial derivative $\partial N/\partial T\Big|_i$ is the change in speed per change in torque with Q held constant. This is evaluated from a horizontal interpolation of the characteristic operating curves as follows:

$$\frac{\partial N}{\partial T}\bigg|_i = \frac{2{,}730 - 1{,}530}{80 - 160} = -15 \tag{2.76}$$

The minus sign indicates that for a constant Q the speed decreases as the

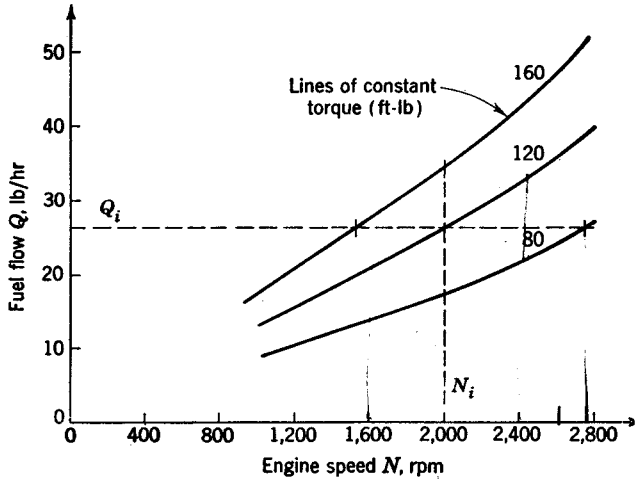


FIG. 2.23. Characteristic curves for an engine.

torque increases. Thus, for operation in the vicinity of the point $N_i = 2{,}000$, $Q_i = 26$, and $T_i = 120$, the linearized approximation for N is

$$N \approx N_i + n = N_i + 66.7q - 15t \tag{2.77}$$

The main difference in working with characteristic operating curves for a component rather than equations is that the partial derivatives are evaluated from a physical interpretation of the curves rather than mathematically from the equation.

CHAPTER 3

REPRESENTATION OF CONTROL SYSTEMS

3.1. Introduction. In this chapter it is shown how to obtain the over-all block-diagram representation for some typical control systems. In brief, the method employed is to obtain the block diagram for each component or process and then "hook up," or connect, the corresponding inputs and outputs for each diagram to obtain the one over-all representation for the system. The techniques which are presented in later chapters

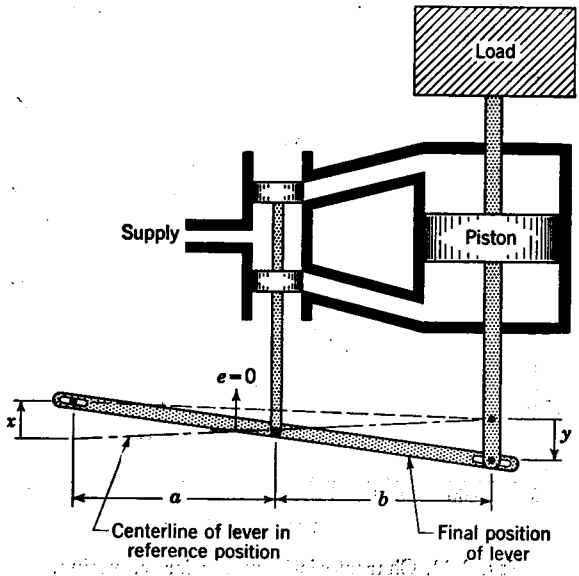


FIG. 3.1. Hydraulic servomotor.

for determining the operating characteristics of control systems are based on a knowledge of the over-all block-diagram representation for the system.

3.2. Hydraulic Servomotor. A hydraulic servomotor is shown in Fig. 3.1. A linkage called a "walking beam" connects the input position x, the valve position e, and the piston position y. The centerline of the lever when the servomotor is in its reference position is indicated in Fig. 3.1. The variations in x, e, and y from their reference positions are

also indicated. When e is zero, the valve is "line on line" and no flow can go to or from the big piston.

The operation of this servomotor may be visualized as follows: When the input x is changed from the reference position, the walking beam first pivots about the connection at y because the large forces acting on the piston hold it in place temporarily. This position of the walking beam is shown by the dashed line in Fig. 3.1. Because of the corresponding movement of e, the valve now admits fluid to the big piston to move it in the direction which makes e zero. The final position of the walking beam, in which e is again zero and the piston has moved a distance y, is indicated in Fig. 3.1. For steady-state operation ($e = 0$), the relationship between the input x and the output y is

$$y = \frac{b}{a} x \tag{3.1}$$

The over-all block diagram which describes the dynamic as well as the steady-state operation of this servomotor is obtained as follows: The

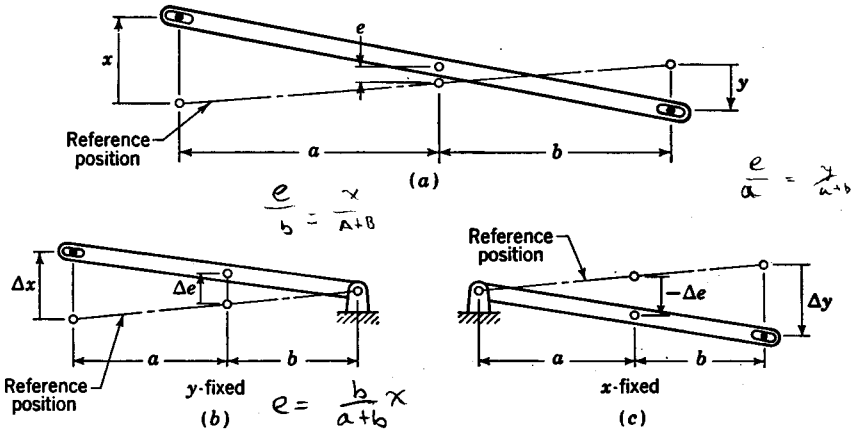


FIG. 3.2. Walking-beam linkage.

walking-beam linkage, as shown in Fig. 3.2*a*, is actually a summing point. The value of e in Fig. 3.2*a* is seen to be a function of the independent variables x and y. That is,

$$e = f(x,y)$$

The application of Eq. (2.65) to evaluate e yields

$$e = \left.\frac{\partial e}{\partial x}\right|_i x + \left.\frac{\partial e}{\partial y}\right|_i y \tag{3.2}$$

The value of $\left.\partial e/\partial x\right|_i$ is obtained by finding the ratio of the change in e

for a change in x with all other parameters fixed (in this case y). Figure 3.2*b* illustrates the linkage with y fixed. From similar triangles

$$\left.\frac{\partial e}{\partial x}\right|_i = \lim_{\substack{\Delta e \to 0 \\ \Delta x \to 0}} \left.\frac{\Delta e}{\Delta x}\right|_i = \frac{b}{a+b} \tag{3.3}$$

Similarly, from Fig. 3.2*c*, in which x is fixed,

$$\left.\frac{\partial e}{\partial y}\right|_i = \lim_{\substack{\Delta e \to 0 \\ \Delta y \to 0}} \left.\frac{\Delta e}{\Delta y}\right|_i = \frac{-a}{a+b} \tag{3.4}$$

The minus sign arises because e decreases as y increases. The substitution of the preceding results into Eq. (3.2) yields the following expression for the walking-beam linkage:

$$e = \frac{b}{a+b}x - \frac{a}{a+b}y \tag{3.5}$$

The preceding result could have been obtained directly by a closer examination of Fig. 3.2*a*. It is apparent that the motion of e is the sum of the contribution due to changing x with y fixed, that is, $[b/(a+b)]x$, and that due to changing y with x fixed, that is, $[-a/(a+b)]y$.

For the case in which $a = b$

$$e = \frac{x-y}{2} \tag{3.6}$$

The block-diagram representation for Eq. (3.6) is shown in Fig. 3.3*a*.

The equation for the valve-and-piston combination is given by Eq. (2.56), in which x is replaced by e. Thus

$$y = \frac{C_1}{A_1 p}e \tag{3.7}$$

The block-diagram representation for the preceding expression is shown in Fig. 3.3*b*. Combining Fig. 3.3*a* and *b* yields the over-all block diagram for the servomotor as shown in Fig. 3.3*c*.

The over-all relationship between the input x and the output y is obtained as follows from the block diagram of Fig. 3.3*c*:

$$(x-y)\frac{C_1}{2A_1 p} = y \tag{3.8}$$

or

$$\left(1 + \frac{2A_1}{C_1}p\right)y = x$$

or

$$(1+\tau p)y = x \tag{3.9}$$

where

$$\tau = \frac{2A_1}{C_1}$$

Equation (3.9) is the differential equation relating x and y. For steady-state operation, both x and y are constant. The quantity $py = dy/dt$ is zero when y is constant, and thus for steady-state operation Eq. (3.9) becomes

$$y = x \tag{3.10}$$

To determine the transient response of y for a given change in x, it is necessary to solve Eq. (3.9), which is a first-order linear differential equation with constant coefficients. If, at some arbitrary time $t = 0$, the input x changes instantaneously from its reference position in which $x = 0$

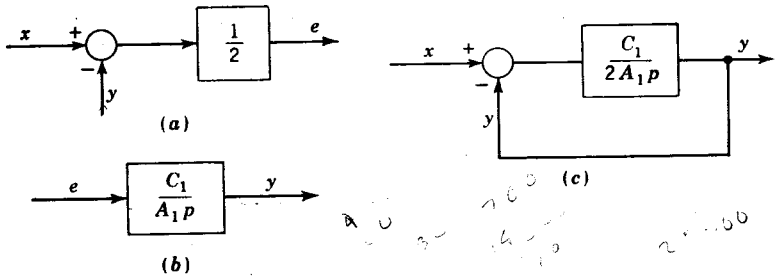


FIG. 3.3. (*a*) Block diagram for walking-beam linkage; (*b*) block diagram for valve and piston; (*c*) block diagram for servomotor.

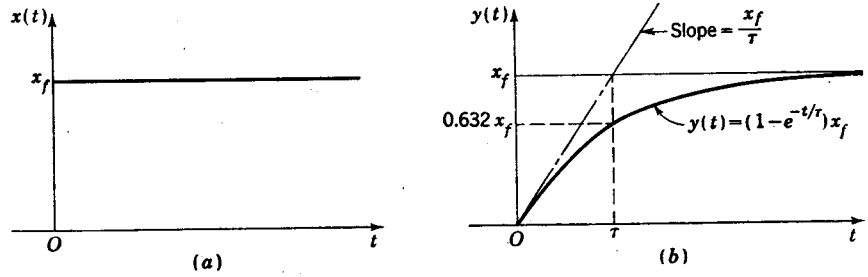


FIG. 3.4. Step-function response.

to a new or final position at which $x = x_f$, then a step change, as is illustrated graphically in Fig. 3.4*a*, has occurred.

The solution of the differential equation given by Eq. (3.9) is

$$y(t) = (1 - e^{-t/\tau})x_f \tag{3.11}$$

A graph of the response $y(t)$ is shown in Fig. 3.4*b*. A characteristic feature of such an exponential response curve is that, when $t = \tau$, then $y(t)$ has undergone 63.2 per cent of its total change. This is proved by letting $t = \tau$ in Eq. (3.11); thus

$$y(\tau) = (1 - e^{-1})x_f = (1 - 0.368)x_f = 0.632x_f \tag{3.12}$$

Another unique feature of such an exponential response is that the tangent to the curve at $t = 0$ intersects the final value at time $t = \tau$. This is proved as follows:

$$\left.\frac{dy}{dt}\right|_{t=0} = \left.\frac{1}{\tau} e^{-t/\tau} x_f\right|_{t=0} = \frac{x_f}{\tau} \tag{3.13}$$

Thus, as shown in Fig. 3.4*b*, the slope of the tangent line is such that it intersects the final value $y(t)_f = x_f$ at $t = \tau$. The term τ is called the

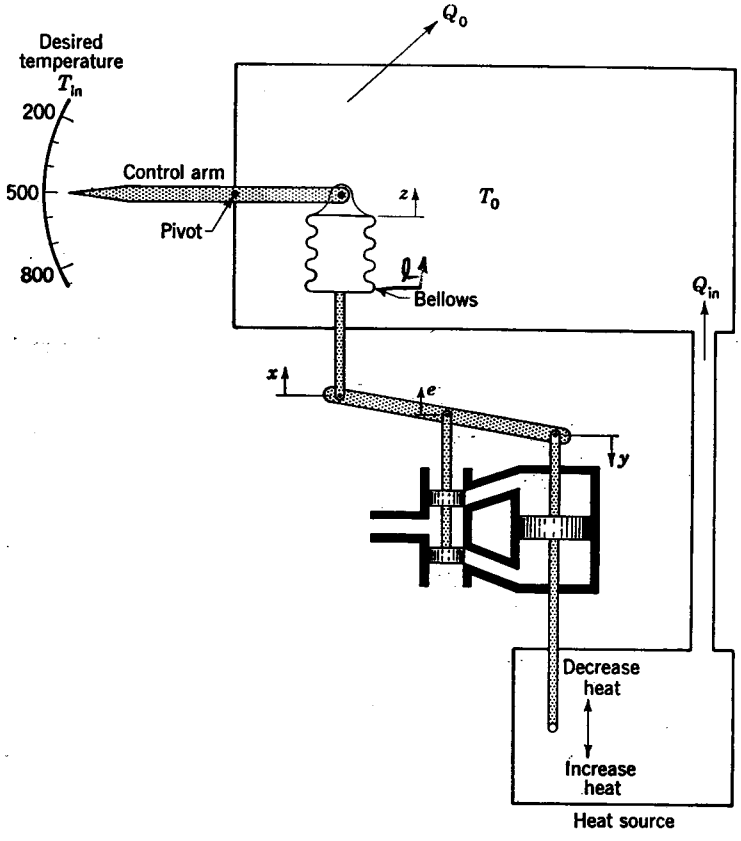


FIG. 3.5. Temperature control system.

time constant and is a measure of the speed of response. When τ is small, the system approaches its new operating condition very fast, and when τ is large, more time is required for the change to occur.

3.3. Temperature Control System. In Fig. 3.5 is shown a system for controlling the output temperature T_o of a chamber, such as an industrial oven, a heat-treating furnace, etc. The desired temperature T_{in} is

indicated by the pointer on the control arm, which pivots about its center. The other end of the control arm determines the position of the top of the temperature-sensitive bellows. When the pointer is moved to a higher temperature setting, the control arm raises the input position x of the linkage through the liquid-filled bellows. The bellows acts as a rigid connecting link in transmitting this motion because its length is fixed by the temperature of its liquid. This then raises the valve, which admits fluid to the top of the piston to move the piston extension, which actuates the heat source to increase the heat supply.

For the case in which the control arm is set for a desired temperature T_{in}, if the oven temperature T_o decreases, the bellows contracts. (The bellows is filled with a liquid which expands as T_o increases and contracts as T_o decreases.) Thus, as T_o decreases, so does the length of the bellows, which in turn increases x. As just described, when x increases, the piston moves in such a direction as to increase the heat Q_{in} to the oven. This in turn tends to bring the oven temperature T_o back to the desired value.

The over-all block-diagram representation of this system for operation in the neighborhood of any reference temperature T_i is obtained by working from the input to the output as follows: The variation z of the position of the top of the bellows from its reference position (i.e., the position when $T_{in} = T_i$) is a function only of the desired temperature T_{in}. That is, $Z = F(T_{in})$, and thus

$$z = \left.\frac{\partial Z}{\partial T_{in}}\right|_i t_{in} = C_2 t_{in} \tag{3.14}$$

where $t_{in} = T_{in} - T_i$ is the variation in the desired temperature and $C_2 = \left.\partial Z/\partial T_{in}\right|_i$ is the change in position Z of the top of the bellows per change in T_{in}. Thus, C_2 is equal to the slope of the curve of position Z versus T_{in} evaluated at the reference temperature.

The length L of the temperature-sensitive bellows is a function of the oven temperature T_o, or $L = F(T_o)$. Thus

$$l = \left.\frac{\partial L}{\partial T_o}\right|_i t_o = C_3 t_o \tag{3.15}$$

where $l = L - L_i$ is the change in the length of the bellows (L_i is the length of the bellows at the reference temperature), $t_o = T_o - T_i$ is the variation in the oven temperature (note that $T_o|_i = T_{in}|_i = T_i$), and $C_3 = \left.\partial L/\partial T_o\right|_i$ is the change in the length of the bellows per change in temperature of the oven. This is the slope of the curve of L versus T_o evaluated at the reference temperature.

From the geometry of Fig. 3.5, it follows that the change in length l of the temperature-sensitive bellows is equal to the change in the position z of the top of the bellows minus the change in the position x of the bottom. Thus, $l = z - x$, whence

$$x = z - l \tag{3.16}$$

The block-diagram representation for Eqs. (3.14) to (3.16) is shown in Fig. 3.6. It should be noted from Fig. 3.6 that this is, in effect, the comparator for this temperature controller.

The block diagram for the walking-beam, valve-and-piston combination is the same as that shown in Fig. 3.3*c*.

The rate of heat flow Q_{in} into the oven is a function of the position of the piston Y[that is, $Q_{in} = F(Y)$]; thus

$$q_{in} = \left.\frac{\partial Q_{in}}{\partial Y}\right|_i y = C_4 y \tag{3.17}$$

where $C_4 = \partial Q_{in}/\partial Y\big|_i$ is the change in heat supplied per unit change in piston position at the reference condition. It is necessary to obtain experimentally the curve of Q_{in} versus position Y in order to evaluate C_4. The net rate of heat flow into or out of the oven is $q_{in} - q_o$. The total heat accumulated is the time integral $\int (q_{in} - q_o)\, dt$, which is equal to the product $C_5 W t_o$; that is,

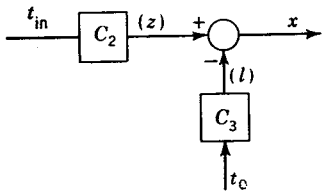


FIG. 3.6. Block diagram for comparator.

$$\int (q_{in} - q_o)\, dt = \frac{q_{in} - q_o}{p} = C_5 W t_o \tag{3.18}$$

where C_5 is the average specific heat of the substance in the oven, W is the total weight, and $t_o = T_o - T_i$ is the corresponding temperature change.

The rate of heat loss Q_o is a function of the temperature difference $T_o - T_a$,

$$Q_o = F(T_o - T_a) \tag{3.19}$$

where T_a is the ambient temperature. Linearization gives

$$q_o = \left.\frac{\partial Q_o}{\partial (T_o - T_a)}\right|_i (t_o - t_a) = C_6 (t_o - t_a) \tag{3.20}$$

where $$C_6 = \left.\frac{\partial Q_o}{\partial (T_o - T_a)}\right|_i$$

For a given oven, it is possible to plot a curve of Q_o versus $T_o - T_a$. The constant C_6 is the slope of this curve at the given reference point. Eliminating q_o between Eqs. (3.18) and (3.20) gives

$$\frac{q_{in} - C_6(t_o - t_a)}{p} = C_5 W t_o$$

or

$$\frac{q_{in} - C_6 t_o + C_6 t_a}{C_5 W p} = t_o \tag{3.21}$$

The block-diagram representation for Eqs. (3.17) and (3.21) is shown in Fig. 3.7. Combining the block diagrams of Figs. 3.3*c*, 3.6, and 3.7 yields the completed over-all block diagram shown in Fig. 3.8.

Usually, systems such as this temperature controller are subjected to an external disturbance. By an external disturbance is meant something

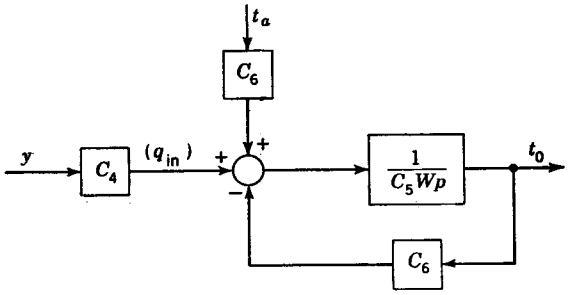


FIG. 3.7. Block diagram for heat source and oven.

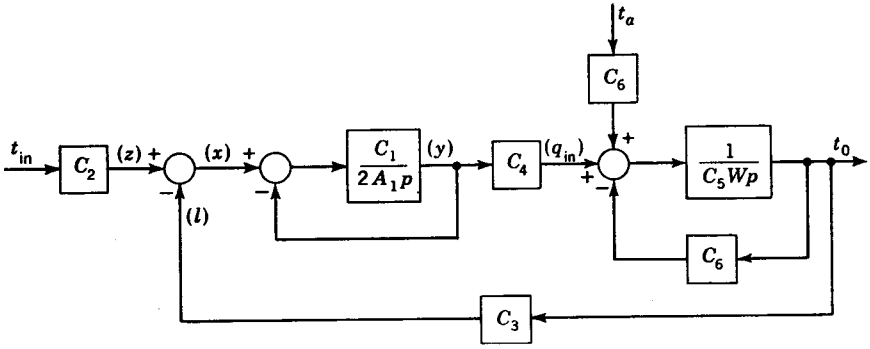


FIG. 3.8. Over-all block diagram for temperature control system.

which acts independently and usually undesirably to affect the operation of the system. In this case, the external disturbance is t_a. An extraneous or external disturbance may be regarded as an unwanted input to the system which tends to affect the value of the controlled variable.

In determining the dynamic behavior of a system, one is interested in the variation of the system parameters from some reference condition. This is the type of information which is obtainable from the block-diagram representation shown in Fig. 3.8. If absolute values are desired, it is an easy matter to convert from t_{in} to T_{in} or from t_o to T_o by merely adding the reference value.

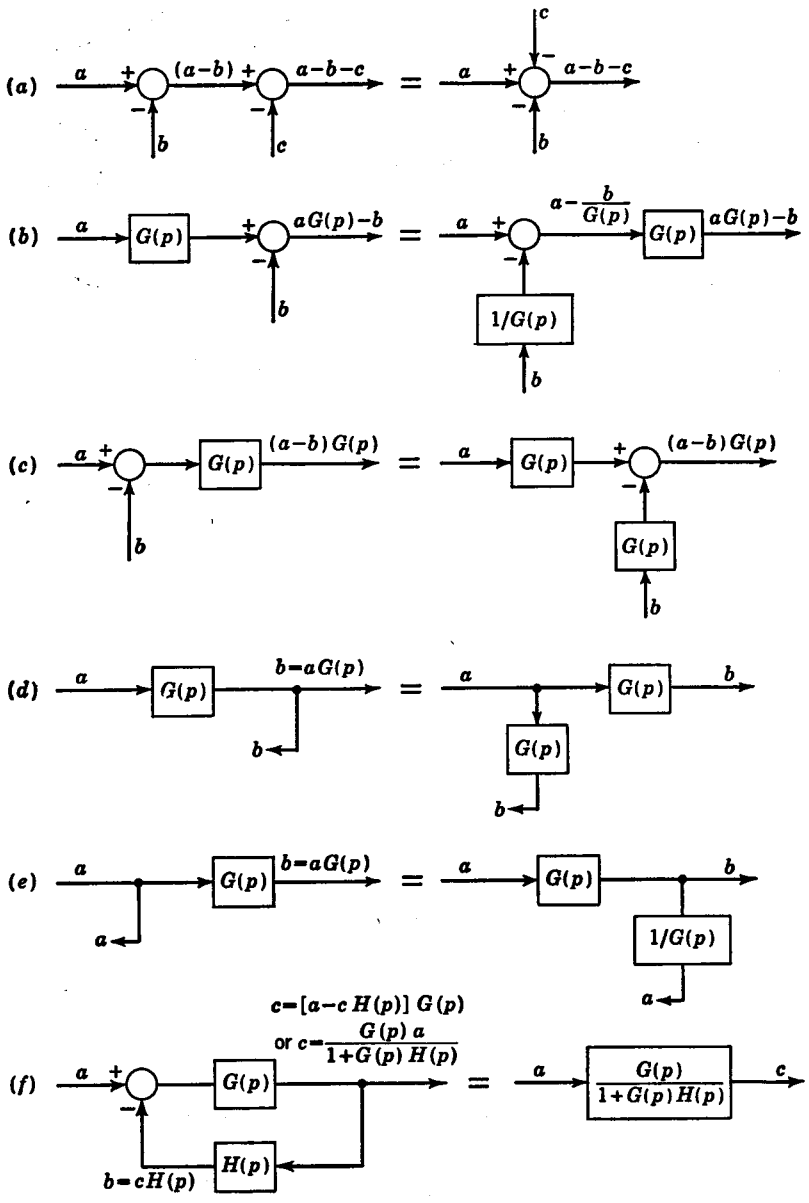


FIG. 3.9. Equivalent block diagrams. (*a*) Combining interconnected summing points; (*b*) moving a summing point behind an element; (*c*) moving a summing point ahead of an element; (*d*) moving a take-off point behind an element; (*e*) moving a take-off point ahead of an element; (*f*) eliminating a minor feedback loop.

In summary, the comparator for this temperature control is the temperature-sensitive bellows assembly. The desired input is the position z of the top of the bellows, and the variation of the length of the bellows is a measure of the output temperature. The error $x = z - l$ determines the input position of the walking-beam linkage. The servomotor serves as an amplifier which produces large hydraulic forces for moving the piston position y to actuate the heat supply. Thus, the major elements in a feedback control system consist of a comparator and a power amplifier which actuates the system to be controlled.

3.4. Block-diagram Algebra. It is often desirable to rearrange the form of a block diagram. In Fig. 3.9 are shown a number of rearrangements which are commonly employed. It is to be noticed that in all cases the rearrangement does not affect the over-all relationship between the input elements (i.e., elements with arrowheads pointing into the

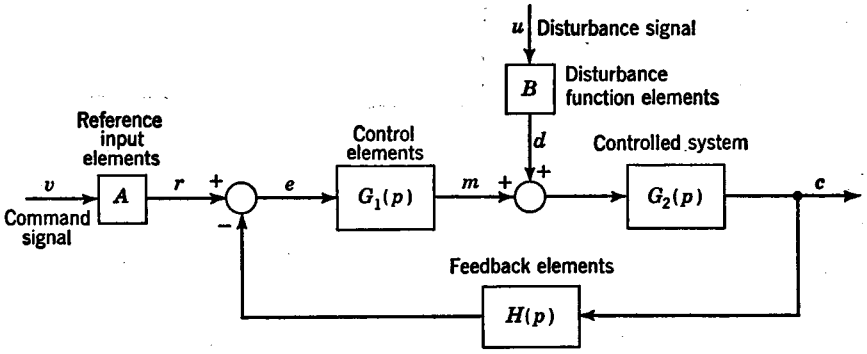


FIG. 3.10. General block-diagram representation for a control system.

diagram) and the output elements (i.e., elements with arrowheads pointing away from the diagram). There are many possible rearrangements for systems. However, it is usually desirable to make the ultimate form of the block diagram the same as that shown in Fig. 3.10. The reason for this is that the methods to be presented later for evaluating the performance of systems are based on systems which are represented in this general form.

Application of the technique shown in Fig. 3.9*f* to Fig. 3.8 yields the following for $G_1(p)$ and $G_2(p)$:

$$G_1(p) = \frac{C_1/2A_1p}{1 + C_1/2A_1p} C_4 = \frac{C_4}{1 + 2A_1p/C_1} = \frac{C_4}{1 + \tau_1 p} \tag{3.22}$$

$$G_2(p) = \frac{1/C_5Wp}{1 + C_6/C_5Wp} = \frac{1/C_6}{1 + C_5Wp/C_6} = \frac{1/C_6}{1 + \tau_2 p} \tag{3.23}$$

where $\tau_1 = \dfrac{2A_1}{C_1}$ and $\tau_2 = \dfrac{C_5W}{C_6}$

The resulting block diagram is shown in Fig. 3.11. In effect, block-diagram algebra was used to eliminate the minor feedback loops. A minor feedback loop is an internal feedback loop which takes place within the main loop.

The operational form of the differential equation relating the output t_o of this temperature controller to the input t_{in} and the external disturbance t_a is obtained from Fig. 3.11 as follows: Start with the reference input C_2t_{in}, subtract the feedback signal C_3t_o, and then continue to write the

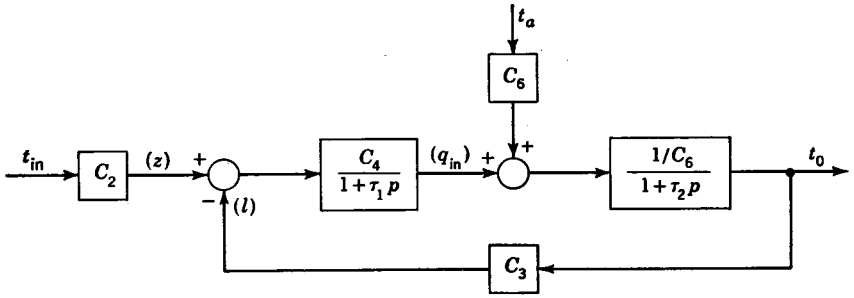


FIG. 3.11. Final block diagram for temperature control system.

mathematical operations indicated by the feedforward portion of the loop until the output t_o is obtained. That is,

$$\left\{\left[(C_2t_{in} - C_3t_o)\frac{C_4}{1+\tau_1 p}\right] + C_6t_a\right\}\frac{1}{C_6(1+\tau_2 p)} = t_o \qquad (3.24)$$

Solving for t_o yields

$$\frac{C_2C_4t_{in} - C_3C_4t_o + C_6(1+\tau_1 p)t_a}{C_6(1+\tau_1 p)(1+\tau_2 p)} = t_o$$

or

$$t_o = \frac{C_2C_4t_{in} + C_6(1+\tau_1 p)t_a}{C_6(1+\tau_1 p)(1+\tau_2 p) + C_3C_4} \qquad (3.25)$$

Equation (3.25) is the operational representation of a linear differential equation with constant coefficients. In Chap. 5, it is shown how linear differential equations with constant coefficients may be solved by use of Laplace transformations.

3.5. Speed Control System. Figure 3.12 shows a typical speed control system for gas turbines, steam turbines, or diesel engines. The position of the throttle lever sets the desired speed of the engine. The speed control is drawn in some reference operating position so that the values of all the lower-case parameters are zero. The positive direction of motion of these parameters is indicated by the arrowhead on each.

The center of gravity of the flyweights is at a distance $R = R_i + r$ from the center of rotation. The flyweights are geared directly to the

output shaft, so that the speed ω of the flyweights is proportional to the output speed. A lever which is pivoted as indicated in Fig. 3.12 transmits the centrifugal force from the flyweights to the bottom of the lower spring seat. The pivot and lever rotate with the flyweights as a unit. If the speed of the engine should drop below its reference value, then the centrifugal force of the flyweights decreases, thus decreasing the force

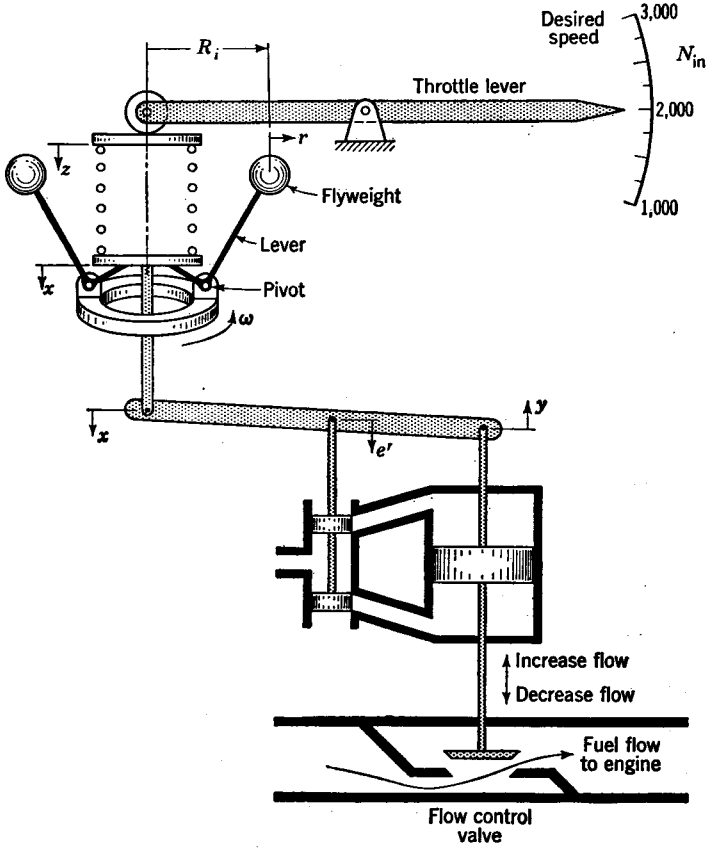


FIG. 3.12. Speed control system.

exerted on the bottom of the spring. This causes x to move downward, which in turn moves e' downward. Fluid then flows to the bottom of the big piston to increase y and thus open wider the flow control valve. By supplying more fuel, the speed of the engine will increase until equilibrium is again reached. For steam turbines, the flow control valve controls the flow of steam rather than fuel as is the case with gas turbines and diesels

Suppose that the throttle lever is moved to a higher speed setting, which in turn causes z to move downward. This in turn causes x to move down-

ward. As just discussed, moving x downward opens the fuel flow valve, which increases the speed.

The over-all block-diagram representation for this system is obtained as follows: The position of the top of the spring is a function of the desired speed only. Thus, the variation of the top of the spring z from its reference position is

$$z = C_2 n_{in} \tag{3.26}$$

where $n_{in} = N_{in} - N_i$ is the change in desired speed and $C_2 = \partial Z/\partial N_{in}\Big|_i$ is the slope of the curve of Z versus N_{in} evaluated at the reference point.

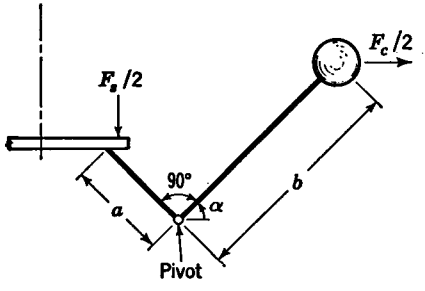


FIG. 3.13. Flyweight lever.

The centrifugal force F_c acting on the flyweights is

$$F_c = 2MR\omega^2 \tag{3.27}$$

where M = mass of each of the two flyweights
R = distance from center of rotation to center of gravity of each flyweight
ω = angular velocity of flyweights

Usually, a governor is geared directly to the output shaft such that ω is equal to the gear ratio times the output speed; i.e.,

$$\omega = C_g \frac{2\pi}{60} N_o$$

where C_g is the gear ratio and the constant $2\pi/60$ converts the output speed N_o from rpm to radians per second. Substitution of this value for ω into the preceding expression for F_c gives

$$F_c = 2\left(\frac{2\pi C_g}{60}\right)^2 MRN_o^2 = C_f MRN_o^2 \tag{3.28}$$

where $C_f = 2(2\pi C_g/60)^2$ is a force-conversion constant.

Effects due to inertia, friction, backlash, etc., are generally classified as secondary effects and do not have a marked influence upon the basic

operation of a system. In the initial design phases, the designer is chiefly interested in evaluating the primary factors which affect system performance, and thus he disregards secondary effects, as is illustrated in this analysis.

In Fig. 3.13 is shown an enlarged view of the flyweight lever. The distance from the pivot to the center of gravity of the flyweight is b, and the distance from the pivot to the spring seat is a. The angle of inclination of link b is designated as α. Taking moments about the pivot gives

$$\frac{F_c}{2} b \sin \alpha = \frac{F_s}{2} a \sin \alpha$$

or

$$F_s = \frac{b}{a} F_c = C_r F_c \tag{3.29}$$

where $C_r = b/a$ is a lever-ratio constant. Because of the large centrifugal force developed by a flyweight, gravitational effects are negligible. Substitution of F_c from Eq. (3.28) into the preceding expression gives

$$F_s = C_f C_r M R N_o^2 \tag{3.30}$$

The two independent variables in Eq. (3.30) are R and N_o, so that linearization gives

$$f_s = C_3 r + C_4 n_o \tag{3.31}$$

where

$$C_3 = \left.\frac{\partial F_s}{\partial R}\right|_i = C_f C_r M N_i^2$$

$$C_4 = \left.\frac{\partial F_s}{\partial N_o}\right|_i = 2 C_f C_r M R_i N_i$$

Note that

$$\left. N_o \right|_i = \left. N_{\text{in}} \right|_i = N_i$$

The compression of the spring from its reference length is $z - x$. Thus, the variation in force exerted by the spring is

$$f_s = K_s(z - x) \tag{3.32}$$

where K_s is the spring constant. Setting Eqs. (3.31) and (3.32) equal,

$$K_s(z - x) = C_3 r + C_4 n_o \tag{3.33}$$

The geometry of Fig. 3.12 shows that the motions of r and x are related by a lever so that $r = -C_r x$. The reason for the minus sign is that, as r increases, x decreases. Eliminating r from the preceding equation yields

$$K_s z - K_s x = -C_r C_3 x + C_4 n_o$$

or

$$x = \frac{K_s z - C_4 n_o}{K_s - C_r C_3} \tag{3.34}$$

The block-diagram representation for Eqs. (3.26) and (3.34) is given in Fig. 3.14, which shows the comparator for the speed control system.

The operation of the servomotor was discussed in Sec. 3-2, and the block diagram was given in Fig. 3.3c.

The flow through the flow control valve is a function of the position Y [that is, $Q = F(Y)$]. Linearization gives

$$q = \frac{\partial Q}{\partial Y}\bigg|_i y = C_5 y \tag{3.35}$$

where C_5 is the slope of the curve of Q versus Y evaluated at the reference condition.

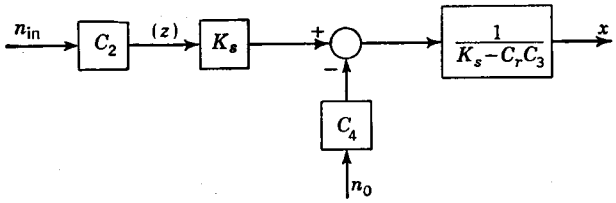


FIG. 3.14. Block diagram for comparator.

In general, the operating speed N_o of an engine is a function of the fuel flow Q supplied to the engine and the torque T exerted on the engine as it rotates. Thus

$$N_o = F(Q,T)$$

The linearized form of this expression is

$$n_o = C_6 q - C_7 t \tag{3.36}$$

where $C_6 = \partial N_o/\partial Q\big|_i$ is the change in speed per change in fuel and $C_7 = -\partial N_o/\partial T\big|_i$ is the change in speed per change in torque. For a constant flow Q, the speed decreases as the torque on the engine is increased, so that $\partial N_o/\partial T\big|_i$ is a negative number. However, the minus sign is seen to make C_7 a positive number. For convenience in using block diagrams, it is desirable that all constants are positive numbers. The variation in the torque t is

$$t = (T_L - T_{L_i}) + J\alpha = t_L + \frac{2\pi}{60} J p n_o \tag{3.37}$$

The variation t of the torque exerted on the engine is composed of the change in torque t_L due to a change in load plus the inertia torque $(2\pi/60)\ Jpn_o$ required to accelerate or decelerate the engine. (The inertia of a large engine is obviously a prime consideration.) The values

of the preceding partial derivatives are obtained from the curves of the operating characteristics for the particular engine under consideration.

Substitution of Eq. (3.37) into Eq. (3.36) yields

$$n_o = C_6 q - C_7 t_L - C_7 \frac{2\pi}{60} J p n_o$$

or
$$n_o = \frac{C_6 q - C_7 t_L}{1 + C_7 \frac{2\pi}{60} J p} = \frac{C_6 q - C_7 t_L}{1 + \tau_2 p} = \frac{C_6}{1 + \tau_2 p}(q - C_8 t_L) \qquad (3.38)$$

where $\tau_2 = C_7 \frac{2\pi}{60} J$ and $C_8 = \frac{C_7}{C_6} = \frac{-\partial N_o/\partial T\big|_i}{\partial N_o/\partial Q\big|_i} = \frac{\partial Q}{\partial T}\Big|_i$

For an implicit function, the product of the partial derivatives is -1 Thus, by writing $N_o = F(Q,T)$ in the implicit form $G(Q,T,N_o) = 0$, it follows that

$$\frac{\partial Q}{\partial T}\frac{\partial T}{\partial N_o}\frac{\partial N_o}{\partial Q} = -1 \qquad \text{or} \qquad \frac{-\partial N_o/\partial T}{\partial N_o/\partial Q} = \frac{\partial Q}{\partial T}$$

The block-diagram representation for Eqs. (3.35) and (3.38) is shown in Fig. 3.15.

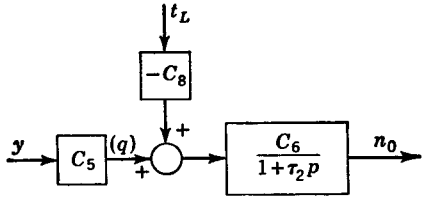


FIG. 3.15. Block diagram for engine.

The over-all block-diagram representation for this speed control system is obtained by combining Figs. 3.14, 3.3c, and 3.15, as is shown in Fig. 3.16, in which $\tau_1 = 2A_1/C_1$.

By letting $K_1 = C_5/(K_s - C_7C_3)$ and eliminating the minor feedback loop, then Fig. 3.16 may be represented as shown in Fig. 3.17. The operational form of the differential equation relating the output n_o to the input n_{in} and external disturbance t_L for the speed control system represented by the block diagram of Fig. 3.17 is obtained as follows: Subtract the feedback signal $C_4 n_o$ from the reference input $C_2 K_s n_{in}$, and perform the mathematical operations indicated by the feedforward portion of the block diagram until the output n_o is obtained. That is,

$$\left\{\left[(C_2 K_s n_{in} - C_4 n_o)\frac{K_1}{1 + \tau_1 p}\right] - C_8 t_L\right\}\frac{C_6}{1 + \tau_2 p} = n_o \qquad (3.39)$$

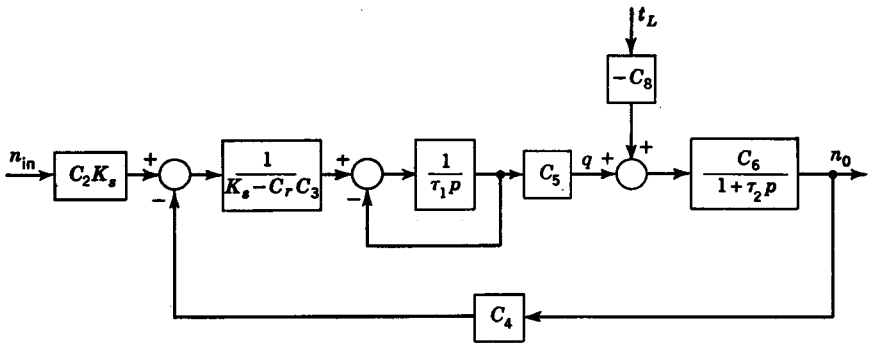


FIG. 3.16. Combined block diagrams.

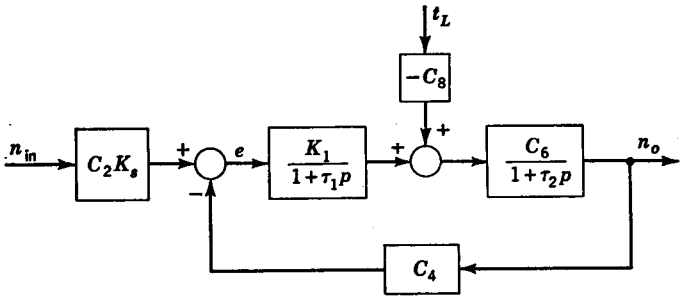


FIG. 3.17. Over-all block diagram for speed control system.

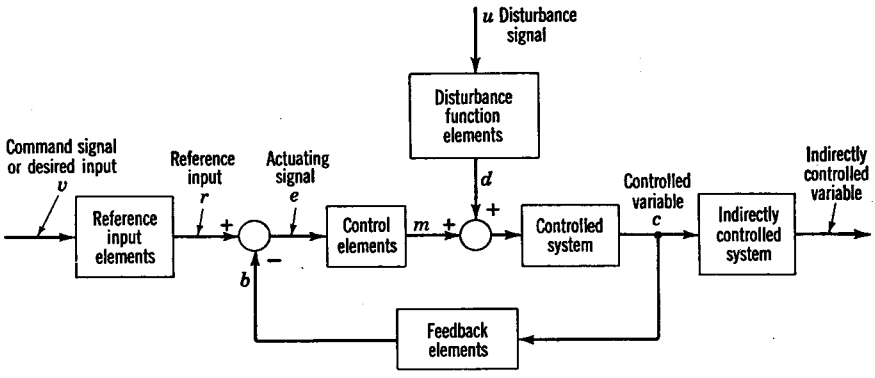


FIG. 3.18. Generalized feedback control system.

Solving for n_o yields

$$[(1 + \tau_1 p)(1 + \tau_2 p) + C_4 C_6 K_1] n_o = C_2 C_6 K_1 K_s n_{in} - C_6 C_8 (1 + \tau_1 p) t_L \tag{3.40}$$

or

$$n_o = \frac{C_2 C_6 K_1 K_s n_{in} - C_6 C_8 (1 + \tau_1 p) t_L}{(1 + \tau_1 p)(1 + \tau_2 p) + C_4 C_6 K_1} \tag{3.41}$$

3.6. Generalized Feedback Control System. A general representation for a feedback control system is shown in Fig. 3.18. It is to be noticed that the command signal, or desired input, does not usually go directly to the comparator but must be converted to a suitable input for this device. Similarly, the controlled variable, or output, in the general case must also be changed by the feedback elements $H(p)$ before it can be measured by the comparator. The actuating signal e is amplified by the control elements $G_1(p)$ before entering the system $G_2(p)$ being controlled. An external disturbance, as shown in Fig. 3.18, is a disturbance which acts independently to affect the operation of the system. Although in Fig. 3.18 the external disturbance is shown entering the system between the control elements and the controlled system, in general, the external disturbance may enter the system at any point.

It is also to be noticed from this generalized representation of a control system that the controlled variable is not necessarily the quantity which it is desired to control. For example, a household thermostat controls the temperature of the air around the thermostat, and depending upon the circulation of air in the house, the temperature of other areas may vary considerably. In addition, the idealized purpose of this control is to maintain the comfort of the persons of the household, which depends upon humidity, their clothing, their amount of physical activity, etc. Thus, it is apparent that the controlled variable is not necessarily the ultimate quantity which it is desired to control.

CHAPTER 4

STEADY-STATE OPERATION

4.1. Introduction. By steady-state operation is meant the equilibrium state attained such that there is no change with respect to time of any of the system variables. The system remains at this equilibrium state of operation until it is excited by a change in the desired input or in the external disturbance. A transient condition is said to exist as long as any of the variables of the system is changing with time. In this chapter, it is shown that considerable information about the basic character of a system may be obtained from an analysis of its steady-state operation.

4.2. Steady-state Analysis of a Control System. The differential equation relating the output n_o to the input n_{in} and external disturbance t_L for the speed control system represented by the block diagram of Fig. 3.17 is given by Eq. (3.40). For steady-state operation n_o, n_{in}, and t_L will have constant values, and therefore terms resulting from powers of p operating on these constant quantities will be zero. That is, $pn_o\Big|_{ss} = d(n_o)/dt\Big|_{ss} = 0$, etc. Thus, the equation describing the steady-state operation of this speed control system is

$$n_o = \frac{C_2C_6K_1K_sn_{in} - C_6C_8t_L}{1 + C_4C_6K_1} \tag{4.1}$$

It should be noticed that Eq. (4.1) could also have been obtained by letting $p = 0$ in the over-all block diagram of Fig. 3.17. Doing this yields the block diagram for steady-state operation shown in Fig. 4.1*a*.

In general, the block diagram describing the steady-state operation of a system may be represented as illustrated by Fig. 4.1*b*, in which

$$K_{G_1} = [G_1(p)]_{p=0} \qquad K_{G_2} = [G_2(p)]_{p=0} \qquad K_H = [H(p)]_{p=0} \tag{4.2}$$

From Fig. 4.1*b*, the equation for steady-state operation is found to be

$$[(Av - K_Hc)K_{G_1} + Bu]K_{G_2} = c$$

or

$$c = \frac{AK_{G_1}K_{G_2}}{1 + K_{G_1}K_{G_2}K_H}\,v + \frac{BK_{G_2}}{1 + K_{G_1}K_{G_2}K_H}\,u \tag{4.3}$$

Corresponding quantities for the speed control system are $c = n_o$, $v = n_{in}$, $u = t_L$, $A = C_2K_s$, $B = -C_8$, $K_{G_1} = K_1$, $K_{G_2} = C_6$, and $K_H = C_4$. Substitution of these results into Eq. (4.3) verifies Eq. (4.1).

The constant A which appears in Eq. (4.3) is, in effect, the scale factor for the input dial. To have the coefficient of the v term equal to unity, A must be selected such that

$$\frac{AK_{G_1}K_{G_2}}{1 + K_{G_1}K_{G_2}K_H} = 1$$

or

$$A = \frac{1 + K_{G_1}K_{G_2}K_H}{K_{G_1}K_{G_2}} = \frac{1}{K_{G_1}K_{G_2}} + K_H \tag{4.4}$$

When A is chosen in accordance with Eq. (4.4), the coefficient of the v term is unity and thus Eq. (4.3) becomes

$$c = v + \frac{BK_{G_2}}{1 + K_{G_1}K_{G_2}K_H}u = v + \frac{B}{1/K_{G_2} + K_{G_1}K_H}u \tag{4.5}$$

To have the controlled variable c equal to the command signal v (that is, $c = v$), it is necessary that the coefficient of the u term be zero. This

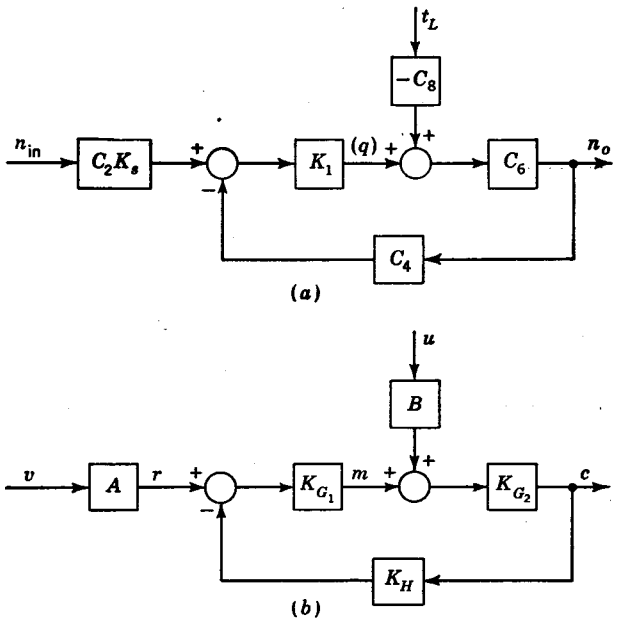


FIG. 4.1. Block diagram for steady-state operation.

coefficient is zero if either K_{G_1} or K_H is infinite. However, from Eq. (4.4), it follows that an infinite value of K_H would necessitate A being infinite, which is physically impossible. Thus, only K_{G_1} can be made infinite. This is accomplished by having *an integrator in the control elements* to yield a $1/p$ term which gives the effect of an infinite constant during

steady-state operation. This type of system is called an integral-type system.

Satisfactory performance may often be achieved by making the coefficient of the u term sufficiently small so that variations in the external disturbance cause only slight errors. Control systems for which the coefficient of the u term is finite are proportional-type systems. The next two sections treat proportional- and integral-type control systems, respectively.

4.3. Proportional-type Systems. The speed control system discussed in the latter part of Chap. 3 is a proportional-type control system. Substitution of the corresponding values for this speed control system into Eq. (4.4) gives

$$A = C_2K_s = \frac{1}{C_6K_1} + C_4$$

or

$$C_2 = \frac{1}{K_s}\left(\frac{1}{C_6K_1} + C_4\right) \tag{4.6}$$

The term $C_2 = \partial Z/\partial N_{in}\Big|_i$ is the scale factor for the speed-setting dial.

Because some of the terms in Eq. (4.6) are partial derivatives evaluated at the reference operating condition, the value of the scale factor C_2 is seen to vary for different reference points. This would result in a nonlinear scale for the input speed dial. The use of a nonlinear scale may be avoided by having the input speed-setting position a cam, which in turn sets the desired input position of the top of the spring, as is shown in Fig. 4.2. It is a relatively easy matter then to set up the speed control system so that Eq. (4.6) is satisfied for any reference condition.

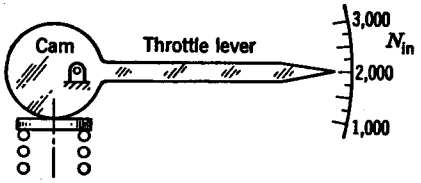


FIG. 4.2. Cam to avoid nonlinear input scale.

When this is so, Eq. (4.1) becomes

$$n_o = n_{in} - \frac{C_6C_8}{1 + C_4C_6K_1}t_L \tag{4.7}$$

Equation (4.7) is the typical form of the steady-state relationship which exists between the input, output, and external disturbance for a proportional-type control system. When the load torque T_L is not equal to the reference value T_{L_i} (that is, $t_L \neq 0$), then n_o is not equal to n_{in}. For example, suppose that this is the speed control system for the gas turbine of a jet airplane and that T_{L_i} is the torque required for the airplane in level flight. When the airplane is inclined to gain altitude,

a greater load torque T_L is required than that for level flight (that is, $t_L > 0$). Thus, the output speed is slightly less than the desired value for this flight condition.

The physical reason for this can be seen by looking at the actual speed control system shown in Fig. 3.12. For level flight, the system is set up so that $N_o = N_{in}$. When the airplane is gaining altitude, the load torque is increased. This increased torque results in a decreased speed, which in turn causes a lower position for x. Because of the lower position for x, there is a greater flow of fuel. To have the airplane continue to gain altitude, more flow is required than for level flight. To maintain this increased flow, the engine speed must be slightly less than for level flight.

Another method of understanding the operation of a proportional-type control system is obtained by considering individually the steady-state

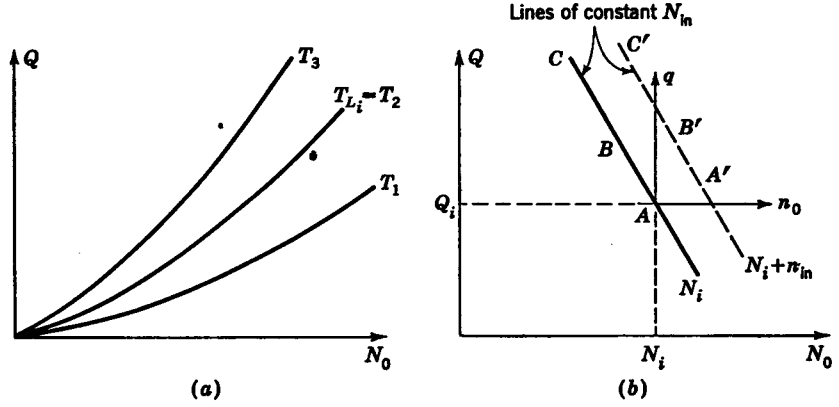


FIG. 4.3. (*a*) Steady-state engine characteristics; (*b*) steady-state control-element characteristics.

operating characteristics of the system to be controlled and those of the control elements. For an airplane in level flight, it is possible to plot a curve of fuel flow Q required to maintain various speeds N_o, as shown in Fig. 4.3*a* by the curve $T_{L_i} = T_2$. The curves for different load torques are also shown in Fig. 4.3*a*. The curve for T_3 would correspond to operation of the airplane at a certain angle of inclination. Similarly, the curve T_1 would correspond to the airplane losing altitude at a certain angle of declination. Thus, Fig. 4.3*a* represents the operating characteristics of the system to be controlled.

To determine the operating line ABC for the control elements as shown in Fig. 4.3*b*, first fix the desired speed setting at some value N_{in}. Then for various speeds of rotation N_o plot the corresponding flow Q coming from the controller. In this manner, the family of curves of N_o versus Q for various constant values of N_{in} is obtained as shown in Fig. 4.3*b*.

By superimposing the characteristics of the control elements and the controlled system as shown in Fig. 4.4*a*, much information about the operation of the system may be obtained. For a given engine speed N_o the fuel flow being supplied by the controller is obtained from Fig. 4.4*a*

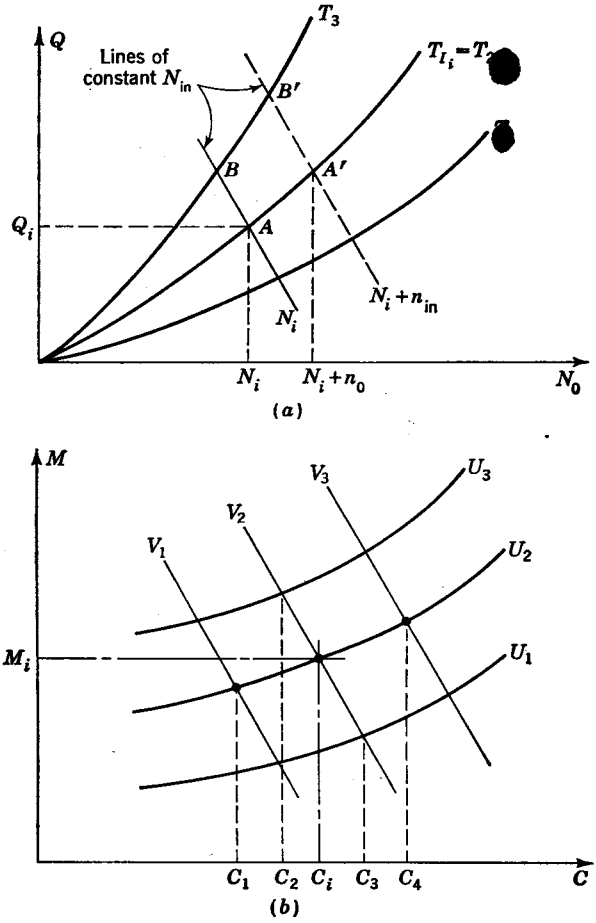


FIG. 4.4. (*a*) Engine and controller steady-state characteristics; (*b*) general representation of steady-state characteristics.

by proceeding vertically up from N_o to the line of operation of the controller and then proceeding horizontally to obtain the corresponding value of Q. Similarly, the fuel flow required to maintain the given flight condition ($T_L = T_1$, T_2, or T_3) is obtained by proceeding vertically up from N_o to the torque curve and then proceeding horizontally to obtain the corresponding value of Q. Steady-state operation exists at the

intersection of the line of operation of the controller and the torque line for the given flight condition, because at this intersection just enough flow is being supplied as is required to maintain the flight condition. For example, if the airplane is in level flight, then the intersection at point A of the line T_2 and the operating line for the desired speed setting N_i represents the steady-state operating point for the system. If the load is increased to T_3 while the desired speed is unchanged, then the new operating point must be on the line of T_3 at point B. Because AB is not a vertical line, variations in the load are seen to cause variations in the output speed. A proportional-type controller is sometimes called a droop-type controller, and the line AB is referred to as the droop line.

Steady-state Constants. The characteristics of the system to be controlled are specified in Fig. 4.4*a* by the family of curves of Q (the manipulated variable) versus N_o (the controlled variable) with lines of constant T_L (the disturbance). The characteristics of the controller are represented by the lines of constant N_{in} (the command signal) which are superimposed upon this plot.

In general, one can plot M (the manipulated variable) versus C (the controlled variable) with lines of constant U (the disturbance) for the controlled system, as shown is Fig. 4.4*b*. The characteristics of the controller may then be superimposed upon this plot, as illustrated by the lines of constant V (the command signal) in Fig. 4.4*b*. Thus, Fig. 4.4*b* is a general representation of the steady-state operating characteristics for a typical control system.

Equation (4.3) describes the steady-state operation about any reference point of operation. The constants in Eq. (4.3) may be evaluated from the operating curves of Fig. 4.4*b*. For example, the coefficient of the v term may be evaluated by maintaining U fixed so that $u = 0$. Thus, the u term in Eq. (4.3) vanishes. Solving the resultant expression for the coefficient of the v term gives

$$\frac{AK_{G_1}K_{G_2}}{1 + K_{G_1}K_{G_2}K_H} = \frac{c}{v}\bigg|_{u=0} = \frac{\Delta C}{\Delta V}\bigg|_{U=\text{constant}} = \frac{\partial C}{\partial V}\bigg|_{U=\text{constant}} \tag{4.8}$$

From Fig. 4.4*b* it is to be noticed that for operation about the point (C_i,M_i) the value of this coefficient is

$$\frac{AK_{G_1}K_{G_2}}{1 + K_{G_1}K_{G_2}K_H} = \frac{\Delta C}{\Delta V}\bigg|_{U_2} = \frac{C_4 - C_1}{V_3 - V_1} \tag{4.9}$$

Similarly, by maintaining a constant V so that $v = 0$, the coefficient of the u term is found, from Eq. 4.3, to be

$$\frac{BK_{G_2}}{1 + K_{G_1}K_{G_2}K_H} = \frac{c}{u}\bigg|_{v=0} = \frac{\Delta C}{\Delta U}\bigg|_{V=\text{constant}} = \frac{\partial C}{\partial U}\bigg|_{V=\text{constant}} \tag{4.10}$$

For operation in the vicinity of the point (C_i,M_i) this coefficient is

$$\frac{BK_{G_2}}{1 + K_{G_1}K_{G_2}K_H} = \frac{\Delta C}{\Delta U}\bigg|_{V_2} = \frac{C_3 - C_2}{U_1 - U_3} \tag{4.11}$$

When the droop line is vertical, $\Delta C = 0$ and the coefficient of the u term vanishes. A vertical droop line is characteristic of an integral-type control, as discussed in the next section.

When the operating curves are all parallel and equidistant, then the value of each of the preceding steady-state constants remains the same over the entire range of operation.

Illustrative Example. A typical family of steady-state operating curves for a speed control system for a diesel or turbine is shown in Fig. 4.5. Determine the steady-state equation for operation in the vicinity

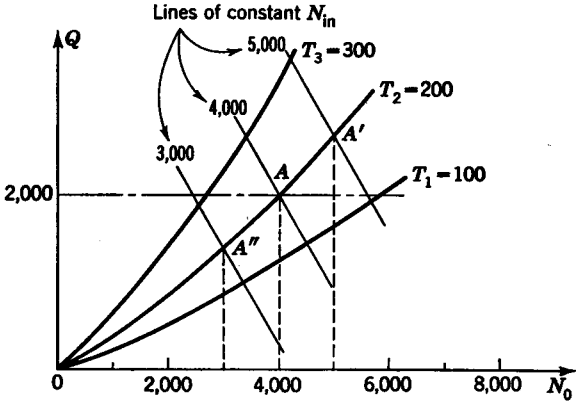


FIG. 4.5. Operating curves for a speed control.

of point A. If this were an open-loop rather than a closed-loop system, what would be the equation for steady-state operation about point A?

SOLUTION. From Eq. (4.8) it follows that

$$\frac{AK_{G_1}K_{G_2}}{1 + K_{G_1}K_{G_2}K_H} = \frac{\Delta N_o}{\Delta N_{in}}\bigg|_{T_2} = \frac{5{,}000 - 3{,}000}{5{,}000 - 3{,}000} = 1$$

Similarly from Eq. (4.10)

$$\frac{BK_{G_2}}{1 + K_{G_1}K_{G_2}K_H} = \frac{\Delta N_o}{\Delta T_L}\bigg|_{N_{in} = 4{,}000} = \frac{4{,}500 - 3{,}400}{100 - 300} = -5.5$$

Thus the equation for steady-state operation in the vicinity of point A is

$$n_o = n_{in} - 5.5\,t_L$$

For an open-loop control system there is but one set value of the flow Q for

each desired speed. This value of Q must be the flow required to make N_o equal to N_{in} at the reference load torque $T_{L_1} = T_2$. For a desired speed of 4,000 rpm, it is to be noted from Fig. 4.5 that the value of Q is 2,000 lb/hr. Thus the horizontal line of Q equal to 2,000 is the open-loop operating line of the controller when $N_{in} = 4{,}000$. Similarly, when N_{in} is 5,000 the operating line is a horizontal line through A', and when N_{in} is 3,000 the operating line is a horizontal line through A''. Thus

$$\frac{AK_{G_1}K_{G_2}}{1 + K_{G_1}K_{G_2}K_H} = \left.\frac{\Delta N_o}{\Delta N_{in}}\right|_{T_2} = \frac{5{,}000 - 3{,}000}{5{,}000 - 3{,}000} = 1$$

and

$$\frac{BK_{G_2}}{1 + K_{G_1}K_{G_2}K_H} = \left.\frac{\Delta N_o}{\Delta T_L}\right|_{N_{in}=4{,}000} = \frac{5{,}800 - 2{,}700}{100 - 300} = -15.5$$

The resulting steady-state equation is

$$n_o = n_{in} - 15.5t_L$$

In effect, an open-loop system is a proportional control in which the droop lines are horizontal.

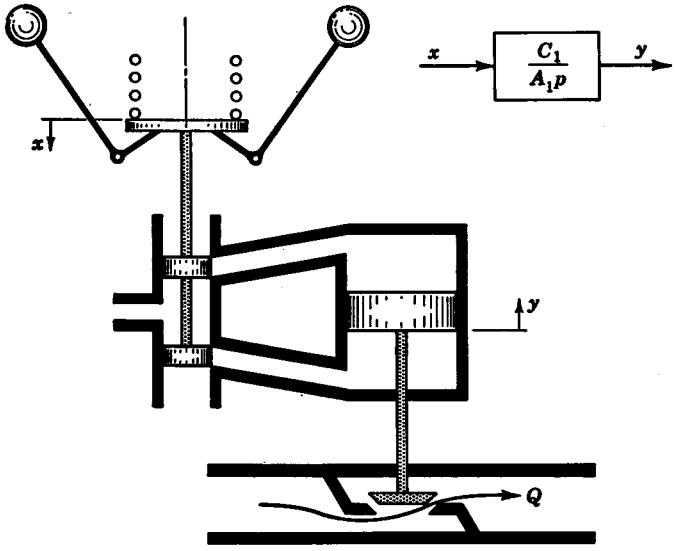


FIG. 4.6. Integral-type control system.

4.4. Integral-type Controller. By eliminating the linkage between x and y of Fig. 3.12 and using the hydraulic integrator shown in Fig. 4.6, the proportional-type speed controller is converted to an integral-type controller. The block-diagram representation for this integrator is also shown in Fig. 4.6. The substitution of this diagram for that of the servomotor which it replaces in Fig. 3.16 yields the block-diagram

representation shown in Fig. 4.7*a*. The value of K_{G_1} is computed as follows,

$$K_{G_1} = \left[\frac{C_1C_5}{(K_s - C_rC_3)A_1p}\right]_{p=0} = \left(\frac{K_I}{p}\right)_{p=0} = \infty \tag{4.12}$$

where $K_I = C_1C_5/(K_s - C_rC_3)A_1$ is the constant associated with the integrating portion of the system, as is shown in Fig. 4.7*b*.

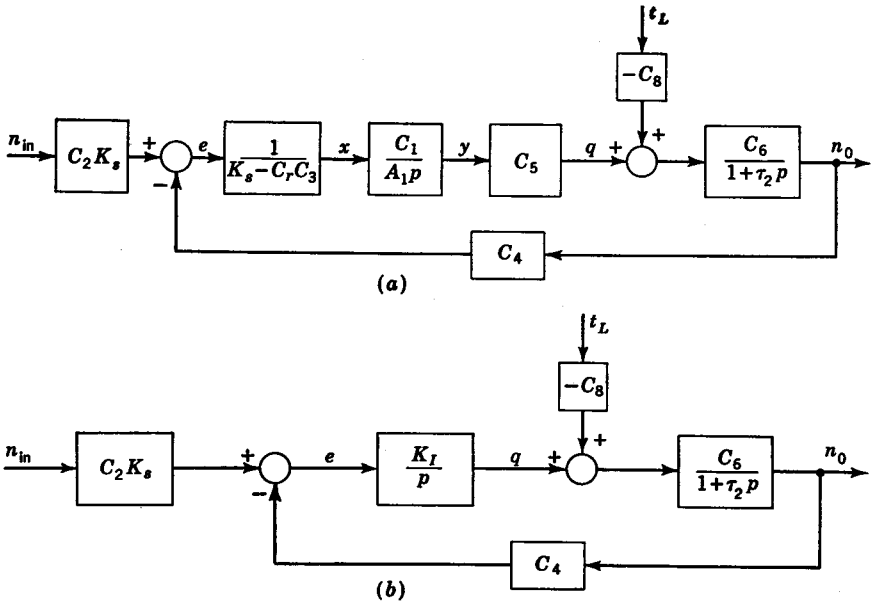


FIG. 4.7. Over-all block diagram for integral-type control.

Because K_{G_1} is infinite, e must be zero for steady-state operation. Thus, subtracting the feedback signal from the reference input in Fig. 4.7*b* gives

$$C_2K_sn_{in} - C_4n_o = e = 0$$

or

$$n_o = \frac{C_2K_s}{C_4}n_{in} \tag{4.13}$$

The preceding expression shows that the speed is independent of the load torque for an integral-type control system. It is an easy matter to adjust the scale factor C_2 for the input speed dial so that $C_2K_s/C_4 = 1$, in which case

$$n_o = n_{in} \tag{4.14}$$

The operation of an integral-type control system may be visualized as follows: From Fig. 4.6, it is to be seen that, if x momentarily changes and then returns to its line-on-line position, the position of y has been

changed permanently and so has the amount of flow going to the engine. Therefore, changing the amount of flow to account for a new operating torque does not change the steady-state position of x, which must be line on line. Because neither x nor the spring compression changes, the output speed must always be equal to the desired value in order that the flyweight force balances the spring force. (Note that, for the proportional-type control, changing the fuel flow requires a permanent change in the position x.)

An integral-type controller is easily recognized because *there must be an integrating component yielding a* $1/p$ *term in the operational expression between the comparator and the point where the external disturbance enters the system.* An integral-type control system has no speed droop, and so the line of operation of an integral-type controller is a vertical line. The operating characteristics of an integral-type control are shown in Fig. 4.8.

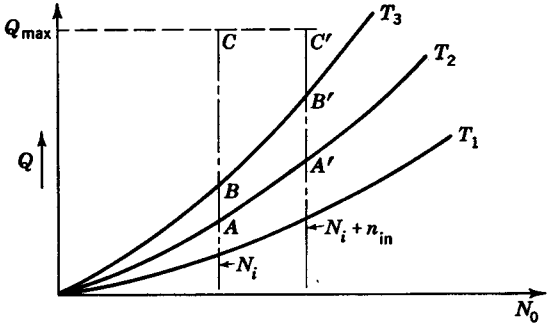


FIG. 4.8. Operating characteristics for integral-type control.

An integral-type controller is also called a floating-type controller because of the floating action of the position y of the flow-setting valve. Two other terms used for an integral-type controller are reset-type controller and isochronous-type controller.

4.5. Proportional- plus Integral-type Controller. From a consideration of steady-state operation only, integral-type systems seem preferable to proportional systems. However, it is generally easier to achieve good transient behavior with a proportional system rather than an integral system (techniques for determining the transient behavior of systems are presented in Chaps. 5 through 10). It is possible to combine the basic features of a proportional-type controller and an integral-type controller to form a proportional- plus integral-type controller, as is shown in Fig. 4.9.

The action of a proportional- plus integral-type controller to a change in the input or external disturbance is initially similar to that of a pro-

portional-type controller, but as the new equilibrium point is reached, the control action becomes the same as that of an integral-type controller. (In effect, the slope of the droop line continually increases.)

A proportional- plus integral-type controller combines the desirable transient characteristics of a proportional-type control and the feature of no steady-state error of the integral-type control.

A proportional- plus integral-type controller is shown in Fig. 4.9. The proportional action is provided by unit 1, which is the same as that

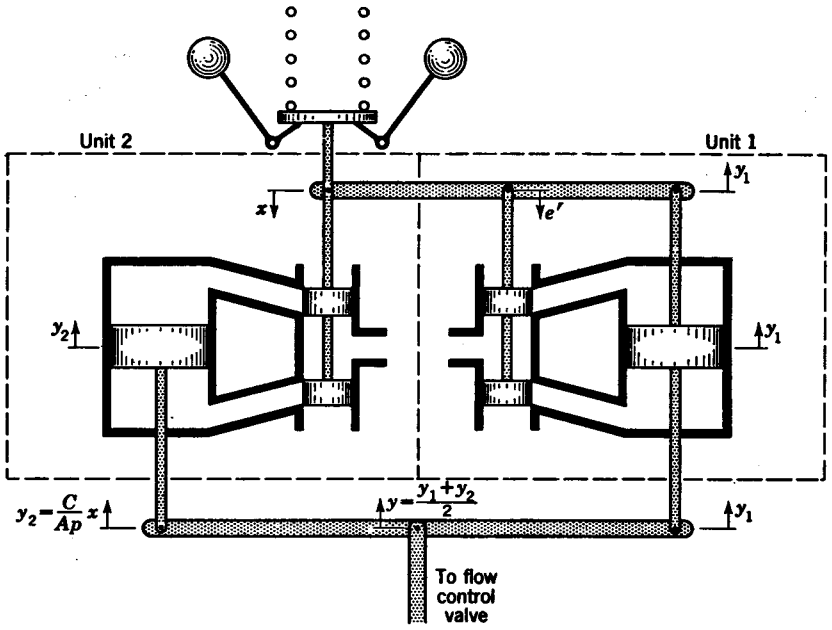


FIG. 4.9. Proportional- plus integral-type control.

for the proportional controller shown in Fig. 3.12. The equation for the proportional action is

$$y_1 = \frac{1}{1 + \tau_1 p} x \tag{4.15}$$

The integral action is provided by unit 2, which is the same as that for the integral controller shown in Fig. 4.6. The equation for this integral action is

$$y_2 = \frac{C}{Ap} x \tag{4.16}$$

The proportional and integral actions are added by a walking-beam linkage such that

$$y = \frac{y_1 + y_2}{2} \tag{4.17}$$

The substitution of y_1 and y_2 into the preceding expression gives

$$y = \frac{1}{2}\left(\frac{1}{1 + \tau_1 p} + \frac{C}{Ap}\right) x \tag{4.18}$$

The individual block diagrams for Eqs. (4.15) to (4.17) are shown in Fig. 4.10*a*. The combined diagram is shown in Fig. 4.10*b*. The substitution of this combined diagram into its corresponding position between x and

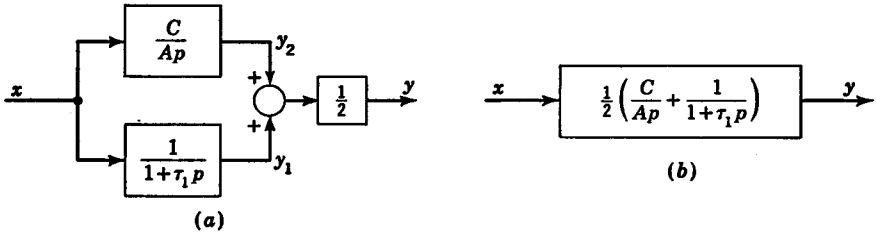


FIG. 4.10. Block diagrams for proportional plus integral action.

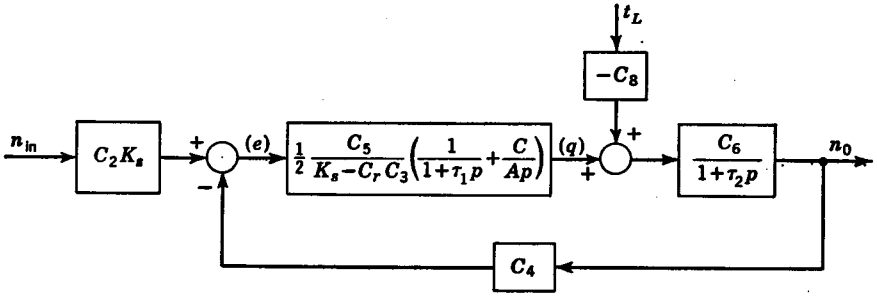


FIG. 4.11. Over-all block diagram for proportional- plus integral-type control.

y of Fig. 4.7*a* yields the resulting representation for this proportional plus integral control, as is shown in Fig. 4.11.

The value of K_{G_1} for this proportional plus integral controller is

$$K_{G_1} = \frac{C_5}{2(K_s - C_r C_3)}\left(\frac{1}{1 + \tau_1 p} + \frac{C}{Ap}\right)_{p=0} = \infty \tag{4.19}$$

The proportional plus integral actions are clearly evidenced by Eq. (4.19). Because K_{G_1} is infinite for steady-state operation, it follows that e is zero during steady-state operation. Thus, from Fig. 4.11

$$C_2 K_s n_{in} - C_4 n_o = e = 0$$

or

$$n_o = \frac{C_2 K_s}{C_4} n_{in} \tag{4.20}$$

Comparison of Eqs. (4.20) and (4.13) shows that the steady-state operation of a proportional- plus integral-type control system is the same as that of an integral-type control system alone. A proportional plus integral control is sometimes referred to as a compensated isochronous control. To better understand the action of this control, suppose that the throttle lever is moved to increase the speed. This causes the position x to move down as does e'. The time constant τ_1 of the proportional unit is small so that y_1 changes rapidly to increase the flow setting. The resulting motion of y_1 returns e' to its line-on-line position.

For the integrating unit, the quantity C/A is small so that y_2 continues to move at a slower rate to provide corrective action. As the speed increases, the position x moves up. The integrating unit continues to provide corrective action until x is returned to its line-on-line position (that is, $x = 0$). In summary, for proportional plus integral control, the initial effect is provided primarily by the proportional action, and the final effect is provided by the integrator.

4.6. Modes of Control.[1-3] In addition to proportional and integral, another mode of control is derivative, or rate, action. For a derivative-type controller, the steady-state expression for the control elements is

$$K_{G_1} = (K'p)_{p=0} = 0 \tag{4.21}$$

The output of a derivative controller is proportional to the rate of change of error. For any constant value of the actuating signal e, the output of the control elements is zero. Thus, steady state may exist in a derivative-type control system with any constant value of error signal. Because a derivative-type controller operates on the rate of change of error and not the error itself, the derivative mode of control is never used alone, but rather in combination with a proportional, or integral, or proportional- plus integral-type controller. The advantage of using derivative action is that the derivative is a measure of how fast the signal is changing and thus tends to give the effect of anticipation. The addition of derivative action is limited primarily to systems which respond very slowly, such as large industrial processes.

4.7. Summary. The selection of the control elements $G_1(p)$ was seen to have a predominant effect upon the steady-state operation of a system. For more complex control systems, it becomes increasingly difficult, if not impossible, to distinguish the individual modes of control. However,

[1] G. J. Murphy, "Basic Automatic Control Theory," chap. 6, D. Van Nostrand Company, Inc., Princeton, N.J., 1957.

[2] D. P. Eckman, "Automatic Process Control," John Wiley & Sons, Inc., New York, 1958.

[3] G. K. Tucker and D. M. Wills, "A Simplified Technique of Control System Engineering," Minneapolis-Honeywell Regulator Co., Philadelphia, 1958.

regardless of the various modes that may be present, it is a relatively simple matter to determine whether K_{G_1} is finite or infinite. For an infinite value, the integral action predominates and there is no steady-state error due to variations in the external disturbance. For a finite value, the system behaves as a proportional-type controller.

A major problem in the design of control systems is the determination of the system parameters to obtain satisfactory transient performance. The transient behavior of a system is prescribed by the differential equation of operation for the system. In the next chapter, it is shown how such differential equations may be solved algebraically by the use of Laplace transforms. In Chap. 6, it is shown that the transient behavior is governed primarily by the roots of the characteristic equation for the system. Thus, the transient characteristics of a system may be ascertained directly from a knowledge of the roots of the characteristic equation.

CHAPTER 5

LAPLACE TRANSFORMS

5.1. General. By transient response is meant the manner in which a system changes from some initial operating condition to some final condition. For example, in Fig. 5.1 it is to be seen that at some arbitrary time $t = 0$ the output is $y(0)$. The curve marked (a) represents the transient response of a system in which the output $y(t)$ slowly approaches its new operating condition. The curve marked (b) shows a system which successively overshoots and undershoots, but these oscillations gradually die out as the new operating condition is obtained.

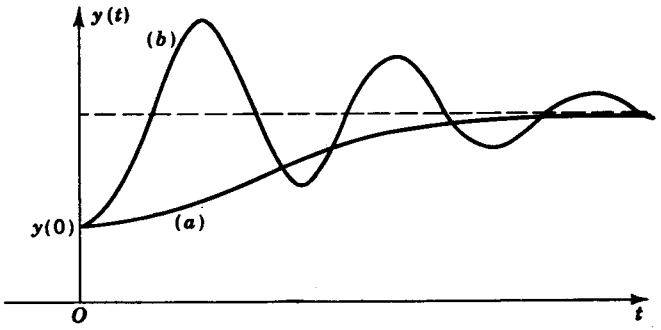


FIG. 5.1. Transient response.

A linear control system or component is one in which the operation is described by a linear differential equation, usually with constant coefficients. For a known input, classical methods could be used for determining the output. However, considerable time is saved by using the Laplace transform method of solving linear differential equations. In addition, Laplace transform analysis is closely related to other methods for evaluating system performance, as is explained in later chapters.

A brief review of classical methods for solving differential equations with constant coefficients is first presented. This is done so that, when the Laplace transform method is explained, a clearer understanding of the similarities and the differences between the two methods results.

5.2. Classical Methods. The transient response of a control system may be obtained by solving differential equations of the general form

$$(p^n + b_{n-1}p^{n-1} + \cdots + b_1p + b_0)y(t) = (a_mp^m + a_{m-1}p^{m-1} + \cdots + a_1p + a_0)x(t)$$

If the coefficient b_n of the $p^ny(t)$ term is not unity, it is an easy matter to divide each term by b_n to yield the preceding form. Solving for $y(t)$ gives

$$y(t) = \frac{a_mp^m + a_{m-1}p^{m-1} + \cdots + a_1p + a_0}{p^n + b_{n-1}p^{n-1} + \cdots + b_1p + b_0}\,x(t) \tag{5.1}$$

where $a_0, a_1, \ldots, a_m$ and $b_0, b_1, \ldots, b_n$ are constants. The term $x(t)$ represents the excitation to the system. This is called the forcing function because it forces or excites the system. The output $y(t)$ is called the response function because it responds to the forcing function $x(t)$. The denominator of Eq. (5.1) is the *characteristic function* of the differential equation. The equation which results by setting the characteristic function equal to zero is called the *characteristic equation.* The value of the exponent n, the highest power of p in the characteristic function, is the order of the differential equation. It is now shown how Eq. (5.1) may be written as the sum of first-order differential equations. First rewrite Eq. (5.1) in the form

$$y(t) = \frac{L_m(p)}{L_n(p)}\,x(t) \tag{5.2}$$

where
$$L_n(p) = p^n + b_{n-1}p^{n-1} + \cdots + b_1p + b_0$$
$$L_m(p) = a_mp^m + a_{m-1}p^{m-1} + \cdots + a_1p + a_0$$

The polynominal $L_n(p)$ may be factored into the form

$$L_n(p) = (p - r_1)(p - r_2) \cdots (p - r_n) \tag{5.3}$$

where $r_1, r_2, \ldots, r_n$ are the roots of the equation $L_n(p) = 0$. In factoring the polynomial $L_n(p)$ as shown in Eq. (5.3), it is to be noted that p may be treated as an algebraic quantity. For example, consider the function

$$L_n(p) = p^2 + 3p + 2 = (p - r_1)(p - r_2) \tag{5.4}$$

Setting $L_n(p)$ equal to zero yields

$$p^2 + 3p + 2 = 0 \tag{5.5}$$

The roots of the preceding equation are

$$r_{1,2} = \frac{-3 \pm \sqrt{9 - 8}}{2} = -1, -2 \tag{5.6}$$

Therefore, $L_n(p) = [p - (-1)][p - (-2)] = (p + 1)(p + 2)$.

Because $r_1, r_2, \ldots, r_n$ are the values of p for which $L_n(p) = 0$, then $r_1, r_2, \ldots, r_n$ are also called the zeros of the function $L_n(p)$. Thus, the roots of the equation $L_n(p) = 0$ are the zeros of the function $L_n(p)$.

The zeros of $L_n(p)$ are said to be distinct if each zero has a different value (that is, $r_1 \neq r_2 \neq r_3 \neq \cdots \neq r_n$). When two or more zeros are equal, the characteristic function is said to have repeated zeros. The case in which $L_n(p)$ has distinct zeros is considered first.

From the theory of partial-fraction expansion, it follows that for distinct zeros $L_m(p)/L_n(p)$ in Eq. (5.2) may be written in the form

$$\frac{L_m(p)}{L_n(p)} = \frac{K_1}{p - r_1} + \frac{K_2}{p - r_2} + \cdots + \frac{K_i}{p - r_i} + \cdots + \frac{K_n}{p - r_n} \tag{5.7}$$

The procedure for obtaining any constant K_i is as follows: First multiply both sides of Eq. (5.7) by $p - r_i$, that is,

$$(p - r_i)\frac{L_m(p)}{L_n(p)} = \frac{p - r_i}{p - r_1}K_1 + \frac{p - r_i}{p - r_2}K_2 + \cdots + K_i + \cdots + \frac{p - r_i}{p - r_n}K_n \tag{5.8}$$

The multiplication of the K_i term in Eq. (5.7) by $p - r_i$ is seen to cancel the denominator, thus leaving K_i alone as shown in Eq. (5.8). By letting $p = r_i$ in Eq. (5.8), each term of the right-hand side of Eq. (5.8) becomes zero except for K_i, which remains. Thus

$$K_i = \lim_{p \to r_i} \left[(p - r_i)\frac{L_m(p)}{L_n(p)} \right] \tag{5.9}$$

Successive application of Eq. (5.9), in which $i = 1, 2, \ldots, n$, yields each of the constants $K_1, K_2, \ldots, K_n$, respectively, in Eq. (5.7). As an example of the use of this partial-fraction-expansion technique, let it be desired to expand $L_m(p)/L_n(p)$, where

$$\frac{L_m(p)}{L_n(p)} = \frac{5p + 8}{p^2 + 3p + 2} = \frac{K_1}{p + 2} + \frac{K_2}{p + 1} \tag{5.10}$$

Application of Eq. (5.9) yields

$$K_1 = \lim_{p \to -2} \left[(p + 2)\frac{5p + 8}{(p + 2)(p + 1)} \right] = \lim_{p \to -2} \frac{5p + 8}{p + 1} = \frac{-10 + 8}{-2 + 1} = 2$$

and

$$K_2 = \lim_{p \to -1} \frac{5p + 8}{p + 2} = \frac{-5 + 8}{-1 + 2} = 3$$

Thus

$$\frac{L_m(p)}{L_n(p)} = \frac{2}{p + 2} + \frac{3}{p + 1} \tag{5.11}$$

The general form for expressing a differential equation as a sum of first-order equations is obtained by substitution of $L_m(p)/L_n(p)$ from

Eq. (5.7) into Eq. (5.2). That is,

$$y(t) = \frac{K_1}{p - r_1} x(t) + \cdots + \frac{K_i}{p - r_i} x(t) + \cdots + \frac{K_n}{p - r_n} x(t)$$

$$= \sum_{i=1}^{n} \frac{K_i}{p - r_i} x(t) = \sum_{i=1}^{n} K_i y_i(t) \quad (5.12)$$

where

$$y_i(t) = \frac{1}{p - r_i} x(t) \quad (5.13)$$

Equation (5.13) is a linear differential equation of the first order. Its solution is

$$y_i(t) = e^{r_i t}[\int e^{-r_i t} x(t)\, dt + c_i]$$

or

$$y_i(t) = c_i e^{r_i t} + e^{r_i t} \int e^{-r_i t} x(t)\, dt \quad (5.14)$$

After performing the integration indicated in Eq. (5.14), the resulting constant of integration will be multiplied by the term $e^{r_i t}$ which is in front of the integral sign. Thus, the effect of the constant of integration has already been included in the first term $c_i e^{r_i t}$ of Eq. (5.14). Substitution of Eq. (5.14) into Eq. (5.12) yields the following general solution,

$$y(t) = \sum_{i=1}^{n} k_i e^{r_i t} + \sum_{i=1}^{n} K_i e^{r_i t} \int e^{-r_i t} x(t)\, dt \quad (5.15)$$

where $k_i = c_i K_i$ is a constant. The terms in the first summation of the right-hand side of Eq. (5.15) comprise the complementary solution, and the terms in the second summation comprise the particular solution. That is,

$$y_c(t) = \sum_{i=1}^{n} k_i e^{r_i t} \quad (5.16)$$

and

$$y_p(t) = \sum_{i=1}^{n} K_i e^{r_i t} \int e^{-r_i t} x(t)\, dt \quad (5.17)$$

where $y_c(t)$ is the complementary solution and $y_p(t)$ is the particular solution.

Illustrative Example 1. Determine the solution for the following differential equation

$$(p^2 + 2p - 3) y(t) = t^2 \quad (5.18)$$

or

$$y(t) = \frac{1}{(p - 1)(p + 3)} t^2$$

Expanding the operator $L_m(p)/L_n(p)$ in a partial-fraction expansion gives

$$y(t) = \left[\frac{K_1}{p - 1} + \frac{K_2}{p + 3}\right] t^2 \quad (5.19)$$

The constants K_1 and K_2 which arise from the partial-fraction expansion are

$$K_1 = \lim_{p \to 1} \left[(p - 1) \frac{1}{(p - 1)(p + 3)} \right] = \frac{1}{4}$$

and

$$K_2 = \lim_{p \to -3} \frac{1}{p - 1} = -\frac{1}{4}$$

In Eq. (5.19), it is to be noted that $r_1 = 1$ and $r_2 = -3$; thus the complementary solution may be written directly from Eq. (5.16).

$$y_c(t) = k_1 e^{r_1 t} + k_2 e^{r_2 t} = k_1 e^{t} + k_2 e^{-3t} \tag{5.20}$$

The particular solution is evaluated from Eq. (5.17) as follows:

$$\begin{aligned} y_p(t) &= K_1 e^t \int e^{-t} t^2 \, dt + K_2 e^{-3t} \int e^{3t} t^2 \, dt \\ &= \tfrac{1}{4} e^t [-e^{-t}(t^2 + 2t + 2)] - \tfrac{1}{4} e^{-3t} \left[\frac{e^{3t}}{27} (9t^2 - 6t + 2) \right] \\ &= -\frac{t^2}{3} - \frac{4t}{9} - \frac{14}{27} \end{aligned} \tag{5.21}$$

The total solution is the sum of Eqs. (5.20) and (5.21). To evaluate the constants k_1 and k_2, two initial conditions are needed.

Repeated Zeros. Suppose that the characteristic function $L_n(p)$ has a multiple or repeated zero r which occurs q times, i.e.,

$$L_n(p) = (p - r)^q (p - r_1)(p - r_2) \cdots (p - r_{n-q}) \tag{5.22}$$

For repeated zeros the partial-fraction expansion has the general form

$$\begin{aligned} y(t) = \frac{L_m(p)}{L_n(p)} x(t) = \frac{C_q x(t)}{(p - r)^q} + \frac{C_{q-1} x(t)}{(p - r)^{q-1}} + \cdots + \frac{C_1 x(t)}{p - r} \\ + \frac{K_1 x(t)}{p - r_1} + \frac{K_2 x(t)}{p - r_2} + \cdots + \frac{K_{n-q} x(t)}{p - r_{n-q}} \end{aligned} \tag{5.23}$$

The constants $K_1, K_2, \ldots, K_{n-q}$ are evaluated as before by application of Eq. (5.9); however, the constants $C_q, C_{q-1}, \ldots, C_1$, which arise from the partial-fraction expansion of the repeated zero, are evaluated as follows:

$$\begin{aligned} C_q &= \lim_{p \to r} \left[(p - r)^q \frac{L_m(p)}{L_n(p)} \right] \\ C_{q-1} &= \lim_{p \to r} \left\{ \frac{1}{1!} \frac{d}{dp} \left[(p - r)^q \frac{L_m(p)}{L_n(p)} \right] \right\} \\ C_{q-k} &= \lim_{p \to r} \left\{ \frac{1}{k!} \frac{d^k}{dp^k} \left[(p - r)^q \frac{L_m(p)}{L_n(p)} \right] \right\} \end{aligned} \tag{5.24}$$

Illustrative Example 2. Determine the partial-fraction expansion for

$$\frac{L_m(p)}{L_n(p)} = \frac{2p + 7}{(p + 3)^2} = \frac{C_2}{(p + 3)^2} + \frac{C_1}{p + 3} \tag{5.25}$$

For this example, q is equal to 2, so that

$$C_2 = \lim_{p \to -3} \left[(p+3)^2 \frac{2p+7}{(p+3)^2} \right] = \lim_{p \to -3} (2p+7) = 1$$

and $$C_1 = \lim_{p \to -3} \left[\frac{d}{dp} (2p+7) \right] = \lim_{p \to -3} 2 = 2$$

Thus $$\frac{L_m(p)}{L_n(p)} = \frac{1}{(p+3)^2} + \frac{2}{p+3} \tag{5.26}$$

The portion of the response due to the term $C_i x(t)/(p-r)^i$ in Eq. (5.23) is

$$y_i(t) = \frac{C_i x(t)}{(p-r)^i} = (c_0 + c_1 t + \cdots + c_{i-1} t^{i-1}) e^{rt} + C_i e^{rt} \int \cdots \int e^{-rt} x(t) \, (dt)^i \qquad i = 2, 3, \ldots, q \tag{5.27}$$

where $c_0, c_1, \ldots, c_{i-1}$ are constants which must be evaluated from the initial conditions. The first term containing the c constants is the complementary solution, whereas the second term on the right-hand side of the preceding expression is the particular solution. The response due to the distinct zeros in Eq. (5.23) and also the response due to the term $C_1 x(t)/(p-r)$ may be evaluated by application of Eq. (5.15).

Illustrative Example 3. Let it be desired to solve the following equation, in which $x(t) = e^{-t}$, that is,

$$y(t) = \frac{2p+7}{(p+3)^2} e^{-t} = \frac{1}{(p+3)^2} e^{-t} + \frac{2}{p+3} e^{-t} \tag{5.28}$$

Application of Eq. (5.27) to the first term on the right-hand side of the preceding expression gives

$$\begin{aligned} y_2(t) &= (c_0 + c_1 t) e^{-3t} + e^{-3t} \int\int e^{3t} e^{-t} \, dt \, dt \\ &= (c_0 + c_1 t) e^{-3t} + \frac{e^{-t}}{4} \end{aligned}$$

The response due to the second term is

$$\begin{aligned} y_1(t) &= k_1 e^{-3t} + 2e^{-3t} \int e^{3t} e^{-t} \, dt \\ &= k_1 e^{-3t} + e^{-t} \end{aligned}$$

The total response $y(t)$ is the sum of the two preceding results, or

$$y(t) = [(c_0 + k_1) + c_1 t] e^{-3t} + \tfrac{5}{4} e^{-t} \tag{5.29}$$

Two initial conditions are required to evaluate the constant $c_0 + k_1$ and the constant c_1.

Numerous techniques such as the method of undetermined coefficients, variation of parameters, etc., have been developed for solving linear differential equations with constant coefficients. However, the method of

Laplace transforms,[1–5] which is next described, is best suited for solving the type of problems which are of interest to control engineers. In many ways, the Laplace transform method is similar to the preceding method of using the partial-fraction expansion to reduce an nth order equation to the sum of n lower-order equations. A major difference is that, in the Laplace transform method, the response due to each term in the partial-fraction expansion is determined directly from the transform table. Thus, there is no need to perform the integrations indicated by either Eq. (5.15) or Eq. (5.27). Because initial conditions are automatically incorporated into the Laplace transforms, the resulting response expression yields directly the total solution (i.e., complementary plus particular solution). Thus, the constants arising from the initial conditions are automatically evaluated, so that the final desired result is obtained directly.

5.3. Laplace Transformation Method. This method of solving differential equations is somewhat analogous to the process of multiplying or dividing by use of logarithms. In the well-known transformation of logarithms, numbers are transformed into powers of the base 10 or some other base. This process in effect makes it possible to multiply and divide by use of the simpler operations of addition and subtraction. After obtaining the desired answer in logarithms, the transformation back to the real-number system is accomplished by finding antilogarithms.

In the method of Laplace transforms, transformation of the terms of the differential equation yields an algebraic equation in another variable s. Thereafter, the solution of the differential equation is effected by simple algebraic manipulations in the s domain (the new variable is s rather than time t). To obtain the desired time solution, it is necessary to invert the transform of the solution from the s domain back to the time domain. Actually, for much control work, information obtained in the s domain suffices so that it may be unnecessary to invert back to the time domain.

The Laplace transformation $F(s)$ of a function of time $f(t)$ is defined as follows,

$$F(s) = \mathfrak{L}[f(t)] = \int_0^\infty f(t)e^{-st}\,dt \tag{5.30}$$

where $\mathfrak{L}$ is the symbol for taking the Laplace transform. The symbol $\mathfrak{L}$

[1] M. F. Gardner, and J. L. Barnes, "Transients in Linear Systems," John Wiley & Sons, Inc., New York, 1942.

[2] J. A. Aseltine, "Transform Method in Linear System Analysis," McGraw-Hill Book Company, Inc., New York, 1958.

[3] R. V. Churchill, "Operational Mathematics," 2d ed., McGraw-Hill Book Company, Inc., New York, 1958.

[4] W. T. Thomson, "Laplace Transformation," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1950.

[5] E. J. Scott, "Transform Calculus," Harper & Brothers, New York, 1955.

is read as "transform of" so that $\mathcal{L}[f(t)]$ means "transform of $f(t)$." For the integral on the right side of Eq. (5.30) the variable t vanishes after evaluation between the limits of integration. Thus, the resulting expression is a function of s only [that is, $F(s)$]. A verification for Eq. (5.30) is presented in Appendix I.

For most problems in control engineering, one is interested in solving for the response of the system only after some arbitrary time $t = 0$. To determine the response for $t > 0$, it is necessary to know only the initial conditions at $t = 0$ and the input or forcing function for $t > 0$. The limits of integration for the transform of Eq. (5.30) show that only positive values of time are considered.

Transforming Functions from the Time Domain to the s Domain. Some input functions which are frequently used for investigating the characteristics of a control system are the step function, pulse function, impulse function, exponentially decaying function, and sinusoid. The manner in which these time functions may be transformed to the s domain is shown

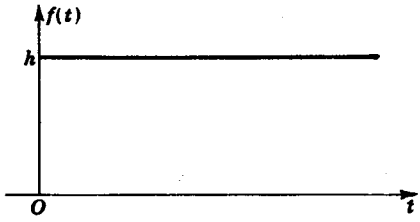


FIG. 5.2. Step function.

in this section. In addition, in the transformation of a linear differential equation with constant coefficients, it is necessary to transform the various derivatives to the s domain. In solving integrodifferential equations such as Eq. (2.25), it is necessary to transform integral terms. Thus, general expressions are also developed for transforming derivative and integral terms.

Step Function. A graphical representation of a step function is shown in Fig. 5.2. For positive values of time, the value of the function is h [from Eq. (5.30); it is to be noted that only positive values of time are of interest in obtaining the Laplace transform].

A step function is designated by the symbol $hu(t)$, where h is the height and $u(t)$ is the symbol for a unit step function whose height is 1. Application of Eq. (5.30) to a step function in which $f(t) = h$ for $t > 0$ gives

$$F(s) = \mathcal{L}[hu(t)] = \int_0^\infty he^{-st}\,dt = -\frac{he^{-st}}{s}\bigg|_0^\infty = \frac{h(-e^{-s(\infty)} + e^{-s(0)})}{s} = \frac{h}{s} \tag{5.31}$$

In evaluating a transform, the term s is regarded as any constant which makes $F(s)$ convergent. As illustrated by Eq. (5.31), if s is any positive constant ($s > 0$), then $e^{-s\infty} = 0$ and $e^{-s0} = e^{-0} = 1$ so that the result follows. However, it should be noted, for negative values of s, that $e^{-s\infty} = e^{\infty} = \infty$, in which case $F(s)$ would be divergent. As is discussed in Appendix I, the operator s must be taken as any constant such that $F(s)$ is convergent. Although there is a range of values of s over

TABLE 5.1. LAPLACE TRANSFORM PAIRS

$f(t)$	$F(s)$
$u(t) = 1$	$\frac{1}{s}$
$u_1(t)$	1
$e^{-\alpha t}$	$\frac{1}{s + \alpha}$
$e^{\alpha t}$	$\frac{1}{s - \alpha}$
$\sin \omega t$	$\frac{\omega}{s^2 + \omega^2}$
$\cos \omega t$	$\frac{s}{s^2 + \omega^2}$
$\frac{t^n}{n!}$	$\frac{1}{s^{n+1}}$
$\frac{t^n e^{\alpha t}}{n!}$	$\frac{1}{(s - \alpha)^{n+1}}$
$kf(t)$	$kF(s)$
$f_1(t) \pm f_2(t)$	$F_1(s) \pm F_2(s)$
$f'(t)$	$sF(s) - f(0)$
$f''(t)$	$s^2F(s) - sf(0) - f'(0)$
$f'''(t)$	$s^3F(s) - s^2f(0) - sf'(0) - f''(0)$
$f^n(t)$	$s^nF(s) - s^{n-1}f(0) - \cdots - f^{n-1}(0)$
$f^{(-1)}(t)$	$\frac{F(s)}{s} + \frac{f^{(-1)}(0)}{s}$
$f^{(-n)}(t)$	$\frac{F(s)}{s^n} + \frac{f^{(-1)}(0)}{s^n} + \frac{f^{(-2)}(0)}{s^{n-1}} + \cdots + \frac{f^{(-n)}(0)}{s}$
$f(\tau) = f(t - t_0)$	$e^{-st_0}F(s)$

which $F(s)$ is convergent, there is but one transform $F(s)$ corresponding to each time function $f(t)$. In Table 5.1 is shown a list of time functions $f(t)$ and their corresponding transforms $F(s)$. In solving problems by Laplace transforms, the term s acts as a dummy operator, and thus there is no need for knowing the range of values over which $F(s)$ exists.

The listing of transform pairs [i.e., corresponding values of $F(s)$ and $f(t)$] given in Table 5.1 is adequate for the solution of most problems which arise in control engineering. The derivation of most of these transform pairs is now explained.

Pulse Function. A pulse function is shown in Fig. 5.3. The height of the function is h and the width t_0 so that its area is ht_0. The Laplace transform is obtained by applying Eq. (5.30), in which $f(t) = h$ for $0 < t \leq t_0$ and $f(t) = 0$ for $t > t_0$.

$$F(s) = \int_0^{t_0} he^{-st}\,dt = h\left[\frac{-e^{-st}}{s}\right]_0^{t_0} = \frac{h}{s}(1 - e^{-st_0}) \qquad (5.32)$$

A special case of a pulse function is an impulse function. By designating the height as $h = k/t_0$, it follows that the area is always equal to k. Now, as the width t_0 approaches zero, the height becomes infinite but the area remains equal to k. This limiting case of a pulse function is called an impulse. The symbol $ku_1(t)$ represents an impulse function whose area is k. Substitution of $h = k/t_0$ into Eq. (5.32) and taking

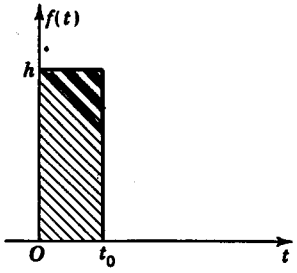


FIG. 5.3. Pulse function.

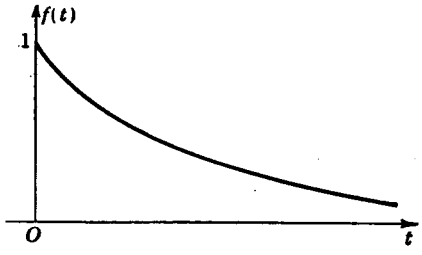


FIG. 5.4. Exponentially decaying function.

the limit as t_0 approaches zero gives the following transform for an impulse:

$$F(s) = \mathcal{L}[ku_1(t)] = \lim_{t_0 \to 0}\left[\frac{k}{st_0}(1 - e^{-st_0})\right] = \frac{0}{0}$$

Application of Lhopital's rule for evaluating the preceding indeterminant gives

$$F(s) = \lim_{t_0 \to 0}\frac{d[k(1 - e^{-st_0})]/dt_0}{d(st_0)/dt_0} = \lim_{t_0 \to 0}\frac{kse^{-st_0}}{s} = k \qquad (5.33)$$

The transform of an impulse function is thus seen to be equal to the area of the function. The impulse function whose area is unity, $u_1(t)$, is called a unit impulse. Much information as to the transient behavior of a system may be obtained by determining the manner in which a system returns to its equilibrium state after the system has been excited by a momentary disturbance such as a pulse or an impulse.

Exponentially Decaying Function. The function $f(t) = e^{-at}$ is shown in Fig. 5.4. Applying Eq. (5.30) gives the transform of this exponentially decaying function.

$$F(s) = \mathcal{L}(e^{-at}) = \int_0^\infty e^{-(a+s)t}\,dt = -\frac{e^{-(a+s)t}}{a+s}\bigg|_0^\infty = \frac{1}{a+s} \tag{5.34}$$

Sinusoidal Function. A sinusoidal time function is shown in Fig. 5.5. The equation for this sinusoidal is $f(t) = \sin \omega t$. Thus

$$F(s) = \mathcal{L}(\sin \omega t) = \int_0^\infty e^{-st} \sin \omega t\,dt \tag{5.35}$$

The preceding integration is simplified by making use of Euler's equations.

$$\begin{aligned} e^{j\theta} &= \cos\theta + j \sin\theta \\ e^{-j\theta} &= \cos\theta - j \sin\theta \end{aligned} \tag{5.36}$$

Addition of Euler's equations and division by 2 yields

$$\cos\theta = \frac{e^{j\theta} + e^{-j\theta}}{2} \tag{5.37}$$

Subtracting the second of Euler's equations from the first and dividing by $2j$ yields

$$\sin\theta = \frac{e^{j\theta} - e^{-j\theta}}{2j} \tag{5.38}$$

The validity of Euler's equations is proved by expanding $e^{j\theta}$, $\sin\theta$, and

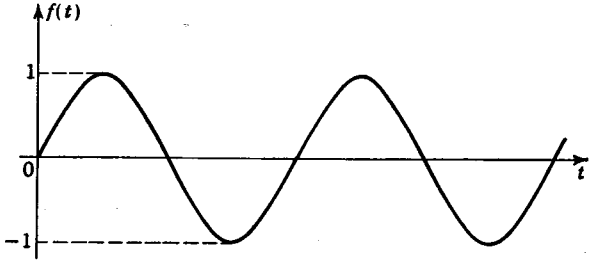


FIG. 5.5. Sinusoidal function.

$\cos\theta$ by use of Maclaurin's series. Thus

$$\begin{aligned} e^{j\theta} &= 1 + j\theta + \frac{(j\theta)^2}{2!} + \frac{(j\theta)^3}{3!} + \frac{(j\theta)^4}{4!} + \frac{(j\theta)^5}{5!} + \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) + j\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right) \\ \cos\theta &= 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4} - \cdots \\ \sin\theta &= \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots \end{aligned}$$

The results of Eq. (5.36) follow directly from the preceding expansions. Because the vector $e^{j\theta}$ is the vector sum of $\cos\theta + j \sin\theta$, the magni-

tude of $e^{j\theta}$ is always $\sqrt{\cos^2\theta + \sin^2\theta} = 1$. The physical significance of the unit vector $e^{j\theta}$ is shown in Fig. 5.6. Here it is to be noticed that $e^{j\theta}$ is a unit vector which is rotated counterclockwise an angle θ from the real axis. Similarly, $e^{-j\theta}$ is a unit vector rotated an angle $-\theta$ from the real axis. For θ equal to 90°, it is seen from the unit circle that $e^{j90^\circ} = j$. Squaring the preceding unit vector yields

$$j^2 = e^{j90^\circ}e^{j90^\circ} = e^{j180^\circ}$$

From the unit circle, this vector is seen to be equal to -1. Thus, $j^2 = -1$, which establishes the well-known result that $j = \sqrt{-1}$.

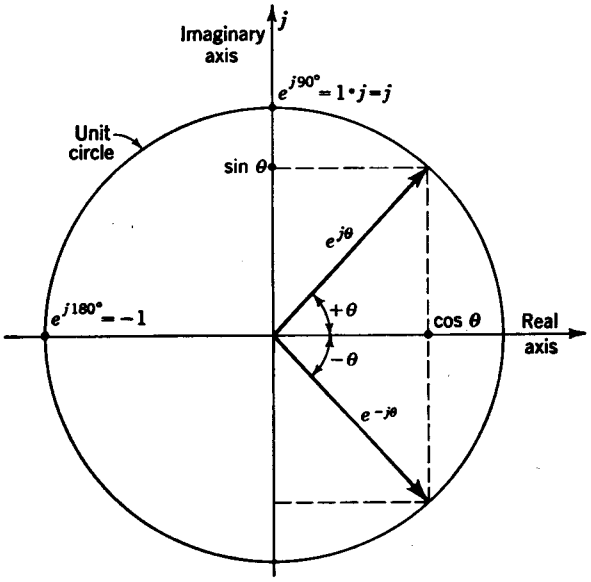


FIG. 5.6. Unit circle.

Substitution of the exponential form for sin θ as given by Eq. (5.38) into Eq. (5.35) gives

$$F(s) = \int_0^\infty \frac{e^{j\omega t} - e^{-j\omega t}}{2j} e^{-st}\, dt$$

$$= \int_0^\infty \frac{e^{-(s-j\omega)t} - e^{-(s+j\omega)t}}{2j}\, dt$$

$$F(s) = \frac{1}{2j}\left(\frac{1}{s - j\omega} - \frac{1}{s + j\omega}\right) = \frac{\omega}{s^2 + \omega^2} \tag{5.39}$$

The response of a system to a sinusoidal forcing function forms the basis for appraising the performance of systems by frequency-response techniques, as discussed in Chaps. 9 and 10.

Real Translation. The Laplace transform of a function $f(\tau)$ which begins at some time $t = t_0$ as shown in Fig. 5.7 rather than at $t = 0$ may be obtained directly by application of the real-translation theorem. From Eq. (5.30), the Laplace transform for $f(\tau)$ is

$$\mathcal{L}[f(\tau)] = \int_0^\infty f(\tau)e^{-st}\,dt = \int_{t_0}^\infty f(\tau)e^{-st}\,dt \tag{5.40}$$

By noting that $t = t_0 + \tau$, $dt = d\tau$, and the lower limit of integration

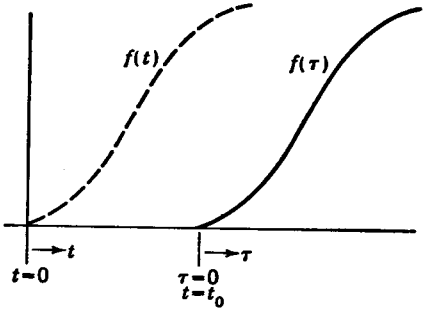


FIG. 5.7. Delayed time function.

$t = t_0$ corresponds to $\tau = 0$, the preceding integral becomes

$$\mathcal{L}[f(\tau)] = \int_0^\infty f(\tau)e^{-s(t_0+\tau)}\,d\tau = e^{-st_0}\int_0^\infty f(\tau)e^{-s\tau}\,d\tau \tag{5.41}$$

It is to be noted that

$$\int_0^\infty f(\tau)e^{-s\tau}\,d\tau = \int_0^\infty f(t)e^{-st}\,dt = F(s)$$

where $F(s)$ would be the usual transform of the function if it were not delayed. Substitution of the preceding result into Eq. (5.41) gives

$$\mathcal{L}[f(\tau)] = e^{-st_0}F(s) \tag{5.42}$$

An application of Eq. (5.42) is immediately evident by noting that the pulse function shown in Fig. 5.3 may be regarded as a step function of height h which begins at $t = 0$ minus a step function of height h which begins at $t = t_0$. The transform for the first step function is h/s, while that for the delayed step function is $(h/s)e^{-st_0}$. Subtracting the delayed step function from the first yields the following transform for the pulse:

$$\frac{h}{s}(1 - e^{-st_0})$$

This is the result given by Eq. (5.32).

Transform of a Derivative. Any linear differential equation will of course have derivatives of various orders. The order of each derivative

is the same as the exponent of the operator p in the operational representation of the term. The general expression for transforming derivatives is obtained as follows:

Let $u = f(t)$ and $v = -e^{-st}/s$; then $du = (d/dt)[f(t)]\,dt$ and $dv = e^{-st}\,dt$. Integration by parts yields

$$\int u\,dv = uv - \int v\,du$$

Substitution of the preceding values for u and v and integration between the limits of zero and infinity gives

$$\int_0^\infty f(t)e^{-st}\,dt = -f(t)\frac{e^{-st}}{s}\Big|_0^\infty + \frac{1}{s}\int_0^\infty \frac{d}{dt}f(t)e^{-st}\,dt \tag{5.43}$$

The left-hand member of the preceding expression is seen to be $F(s)$. For t equal to infinity, the upper limit of the first term on the right-hand side goes to zero. Thus

$$\begin{aligned} F(s) &= \frac{f(0)}{s} + \frac{1}{s}\int_0^\infty \frac{d[f(t)]}{dt}\,e^{-st}\,dt \\ &= \frac{f(0)}{s} + \frac{1}{s}\,\mathcal{L}\left[\frac{df(t)}{dt}\right] \end{aligned} \tag{5.44}$$

where the initial condition $f(0)$ is the value of $f(t)$ for $t = 0$. Therefore

$$\mathcal{L}\left[\frac{df(t)}{dt}\right] = \mathcal{L}[f'(t)] = sF(s) - f(0) \tag{5.45}$$

By the extension of the preceding techniques to higher derivatives, the following equations for transforms of higher derivatives are obtained,

$$\begin{aligned} \mathcal{L}[f''(t)] &= s^2F(s) - sf(0) - f'(0) \\ \mathcal{L}[f'''(t)] &= s^3F(s) - s^2f(0) - sf'(0) - f''(0) \\ \mathcal{L}[f^n(t)] &= s^nF(s) - s^{n-1}f(0) - \cdots - f^{n-1}(0) \end{aligned} \tag{5.46}$$

where $f'(t) = df(t)/dt$, $f''(t) = d^2f(t)/dt^2$, . . . , $f^n(t) = d^nf(t)/dt^n$, and $f'(0)$ is the initial value of $f'(t)$ at $t = 0$, etc. The initial conditions $f(0)$, $f'(0)$, $f''(0)$, . . . associated with a particular differential equation must, of course, be given.

An interesting result is obtained when Eq. (5.45) is applied to the function shown in Fig. 5.8. The initial value of this function is $f(0)$. Because of the step change of height h_c, the value of the function for $t > 0$ is $h = h_c + f(0)$. The transform of this function is

$$F(s) = \mathcal{L}[f(t)] = \mathcal{L}\{[h_c + f(0)]u(t)\} = \frac{h_c + f(0)}{s}$$

Application of Eq. (5.45) to obtain the transform of the derivative of this function gives

$$\mathcal{L}\left[\frac{d}{dt}f(t)\right] = sF(s) - f(0) = \frac{s(h_c + f(0))}{s} - f(0) = h_c \qquad (5.47)$$

The preceding transform is the same as that obtained for an impulse function. Thus, the derivative of a step change is an impulse function whose area is equal to the change in height h_c of the step.

Transform of an Integral. In using the Laplace transform method to solve integrodifferential equations, it is necessary to obtain the transform of an integral. The procedure for obtaining the equation for the transform of an integral is similar to that for a differential. In the

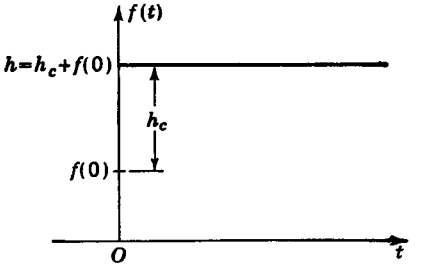


FIG. 5.8. Step function.

general expression for integration by parts let $u = \int f(t)\,dt = f^{(-1)}(t)$ and $dv = e^{-st}\,dt$. Then application of the equation for integration by parts yields

$$\mathcal{L}[f^{(-1)}(t)] = \frac{F(s)}{s} + \frac{f^{(-1)}(0)}{s} \qquad (5.48)$$

where $f^{(-1)}(t) = \int f(t)\,dt$ and $f^{(-1)}(0) = \int f(t)\,dt\Big|_{t=0}$ is the constant of integration which results from integrating $f(t)$.

By the repeated application of this procedure, it is found that

$$\begin{aligned}\mathcal{L}[f^{(-2)}(t)] &= \frac{F(s)}{s^2} + \frac{f^{(-1)}(0)}{s^2} + \frac{f^{(-2)}(0)}{s} \\ \mathcal{L}[f^{(-n)}(t)] &= \frac{F(s)}{s^n} + \frac{f^{(-1)}(0)}{s^n} + \frac{f^{(-2)}(0)}{s^{n-1}} + \cdots + \frac{f^{(-n)}(0)}{s}\end{aligned} \qquad (5.49)$$

where $f^{(-n)}(t) = \int \cdots \int f(t)\,dt^n$, $f^{(-2)}(0) = \int\int f(t)\,dt^2\Big|_{t=0}$ is the second constant of integration which results from the double integration of $f(t)$, etc.

Linearity Theorem. The linearity characteristic of Laplace transformations is a very useful property. If k is a constant or a variable which is independent of both t and s, then it follows from Eq. (5.30) that

$$\mathcal{L}[kf(t)] = k\mathcal{L}[f(t)] = kF(s) \qquad (5.50)$$

Another important linearity property is

$$\mathcal{L}[f_1(t) \pm f_2(t)] = F_1(s) \pm F_2(s) \tag{5.51}$$

5.4. Application of Laplace Transforms. Let it be desired to transform from the time domain to the s domain the following linear differential equation. (When $K = 1$, then this equation describes the operation of the servomotor shown in Fig. 3.1.)

$$(\tau p + 1)y(t) = Kx(t) \tag{5.52}$$

For convenience in using the transform tables, it is desirable to have the coefficient of the highest power of p in $L_n(p)$ unity, thus,

$$\left(p + \frac{1}{\tau}\right) y(t) = \frac{K}{\tau} x(t) \tag{5.53}$$

The Laplace transform of each term in this differential equation is

$$\mathcal{L}[py(t)] = [sY(s) - y(0)]$$

$$\mathcal{L}\left[\frac{1}{\tau} y(t)\right] = \frac{1}{\tau} Y(s)$$

$$\mathcal{L}\left[\frac{K}{\tau} x(t)\right] = \frac{K}{\tau} X(s)$$

The capital letter with s in parentheses indicates the value of the parameter in the s domain. Transforming each term of the differential equation for the system yields

$$\left(s + \frac{1}{\tau}\right) Y(s) = \frac{K}{\tau} X(s) + y(0)$$

or

$$Y(s) = \frac{(K/\tau)X(s) + y(0)}{s + 1/\tau} \tag{5.54}$$

For the case in which all the initial conditions are zero, from Eq. (5.46) $\mathcal{L}[p^n f(t)] = s^n F(s)$, and thus it follows that the substitution of s for p, $Y(s)$ for $y(t)$, and $X(s)$ for $x(t)$ yields directly the transformed equation.

The symbol $X(s)$ is the transform of the input. The nature of the expression $X(s)$ depends upon the particular input to the system, such as a step function, exponential, sinusoidal, etc. The value of $y(0)$ which is $y(t)$ evaluated at $t = 0$ depends upon the initial condition of the system.

Let it be desired to determine the response $y(t)$ to the input function $x(t)$ shown in Fig. 5.9*a*. The initial value is $x(0) = 0$, and then a step change h_c occurs so that the height of this function is $h = h_c$.

Application of the linearity theorem to the step input gives

$$X(s) = \mathcal{L}[hu(t)] = \frac{h}{s} = \frac{h_c}{s} \tag{5.55}$$

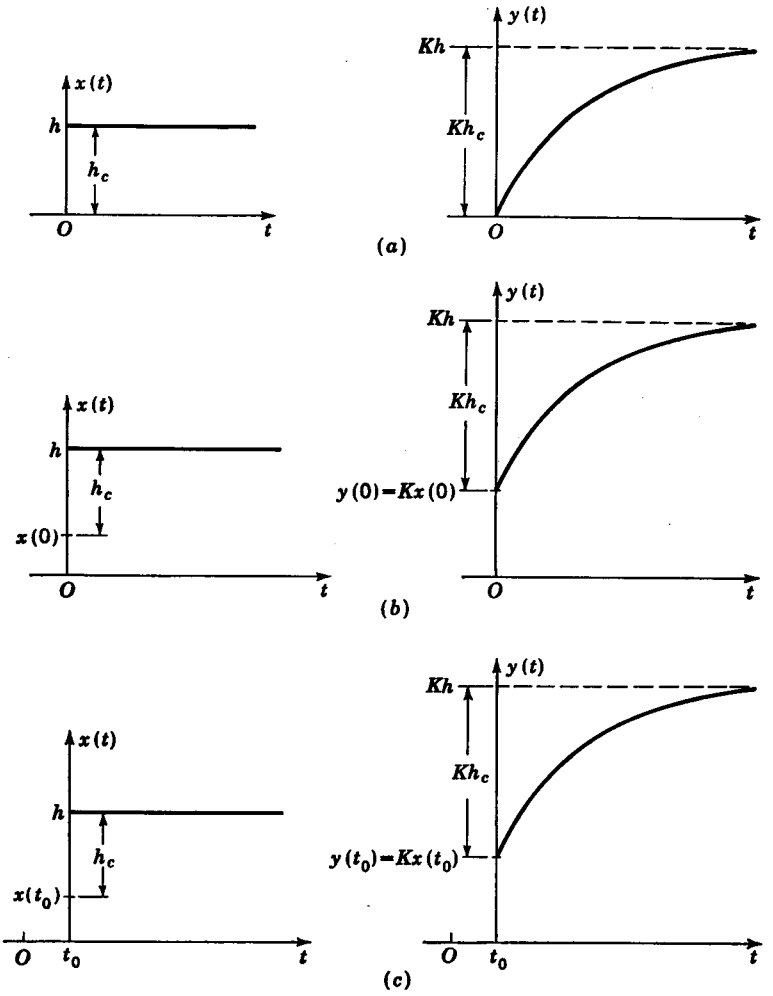


FIG. 5.9. Response of a first-order system to a step input.

Substitution of the preceding result into Eq. (5.54) yields

$$Y(s) = \frac{Kh_c/\tau s + y(0)}{s + 1/\tau} = \frac{Kh_c/\tau + sy(0)}{s(s + 1/\tau)} \tag{5.56}$$

The partial-fraction expansion for $Y(s)$ is

$$Y(s) = \frac{Kh_c}{s} - \frac{Kh_c - y(0)}{s + 1/\tau} \tag{5.57}$$

To invert Eq. (5.57) from the s domain back to the time domain, it is to be noted that

$$\mathfrak{L}[y(t)] = Y(s) \qquad \mathfrak{L}[u(t)] = \frac{1}{s} \qquad \mathfrak{L}(e^{-\alpha t}) = \frac{1}{s + \alpha}$$

Thus, the inverse transform of each term designated by $\mathfrak{L}^{-1}$ is

$$\mathfrak{L}^{-1}[Y(s)] = y(t) \qquad \mathfrak{L}^{-1}\left(\frac{1}{s}\right) = u(t) \qquad \mathfrak{L}^{-1}\left(\frac{1}{s + \alpha}\right) = e^{-\alpha t} \tag{5.58}$$

The inverse transform of Eq. (5.57) is

$$y(t) = Kh_c - [Kh_c - y(0)]e^{-t/\tau} \qquad t > 0$$

This result has meaning for positive values of time only. It may be written in the form

$$y(t) = Kh_c(1 - e^{-t/\tau}) + y(0)e^{-t/\tau} \qquad t > 0 \tag{5.59}$$

Steady state exists when all of the time derivatives are zero (that is, $p^n y = p^m x = 0$, $n = 1, 2, 3, \ldots$ and $m = 1, 2, 3, \ldots$). Thus, if the system was initially at a steady-state operating condition so that $py(0) = 0$, then from Eq. (5.53) it follows that $y(0) = Kx(0)$. Because $x(0) = 0$ and $y(0) = Kx(0) = 0$, all the initial conditions are seen to be zero. Hence, the preceding expression becomes

$$y(t) = Kh_c(1 - e^{-t/\tau}) \qquad t > 0 \tag{5.60}$$

A graph of this result is shown in Fig. 5.9*a*.

Let it now be desired to determine the response for the case in which the initial value of the input $x(0)$ is not zero, as is shown in Fig. 5.9*b*. The total height h of the function is $x(0)$ plus the step change h_c. The transform of this input is

$$X(s) = \mathfrak{L}[hu(t)] = \mathfrak{L}[h_c + x(0)]u(t) = \frac{h_c + x(0)}{s} \tag{5.61}$$

Substitution of the preceding value for $X(s)$ into Eq. (5.54) gives the following transformed equation for $Y(s)$:

$$Y(s) = \frac{K[h_c + x(0)]/\tau s + y(0)}{s + 1/\tau} = \frac{K[h_c + x(0)]/\tau + sy(0)}{s(s + 1/\tau)} \tag{5.62}$$

The partial-fraction expansion gives

$$Y(s) = \frac{K[h_c + x(0)]}{s} + \frac{y(0) - K[h_c + x(0)]}{s + 1/\tau}$$

Inverting the preceding expression yields the desired response

$$y(t) = K[h_c + x(0)] + [y(0) - Kh_c - Kx(0)]e^{-t/\tau} \tag{5.63}$$

If the system is initially in a steady-state condition at $t = 0$, then

$Kx(0) = y(0)$ so that the response is

$$y(t) = Kh_c(1 - e^{-t/\tau}) + y(0) \qquad t > 0 \tag{5.64}$$

The first term on the right-hand side of the preceding expression is seen to be the response due to the step change h_c alone, and the last term is the initial value $y(0)$. *For a system which is initially at a steady-state operating condition, the response may be obtained by merely adding the initial value $y(0)$ to the response for the case in which all the initial conditions are zero.*

The response to a delayed input as shown in Fig. 5.9*c* is obtained as follows: Application of Eqs. (5.42) and (5.45) for obtaining the transform of the derivative for a delayed time function gives

$$\mathfrak{L}\left[\frac{df(\tau)}{dt}\right] = [sF(s) - f(t_0)]e^{-t_0 s}$$

and in general

$$\mathfrak{L}[f^n(\tau)] = [s^nF(s) - s^{n-1}f(t_0) - \cdots - f^{n-1}(t_0)]e^{-t_0 s} \tag{5.65}$$

Thus, to convert a transform to a delayed transform, multiply each term by $e^{-t_0 s}$, and evaluate the initial conditions at $t = t_0$ rather than $t = 0$. The delayed transformed expression corresponding to Eq. (5.62) is

$$Y(s)e^{-t_0 s} = \frac{\{K[h_c + x(t_0)]/\tau + sy(t_0)\}e^{-t_0 s}}{s(s + 1/\tau)} \tag{5.66}$$

Performing a partial-fraction expansion yields

$$Y(s)e^{-t_0 s} = \frac{K[h_c + x(t_0)]e^{-t_0 s}}{s} + \frac{\{y(t_0) - K[h_c + x(t_0)]\}e^{-t_0 s}}{s + 1/\tau}$$

Thus, the time solution is

$$y(t - t_0) = K[h_c + x(t_0)] + [y(t_0) - Kh_c - Kx(t_0)]e^{-(t-t_0)/\tau} \tag{5.67}$$

If the system had initially been at a steady-state operating condition so that $y(t_0) = Kx(t_0)$, then Eq. (5.67) would reduce to

$$y(t - t_0) = Kh_c(1 - e^{-(t-t_0)/\tau}) + y(t_0) \qquad t > t_0 \tag{5.68}$$

By comparison of Eq. (5.67) and Eq. (5.63), it should be noted that the substitution of $t - t_0$ for t gives the effect of translation by an amount t_0. It is not necessary that the system be initially at a steady-state operating condition to effect a time shift by substituting $t - t_0$ for t. For most problems in which a time shift occurs, it is more convenient to work the problem initially as though there were no time shift and then substitute $t - t_0$ for t to obtain the desired result.

The general procedure used to solve differential equations by Laplace transforms may be summarized as follows:

1. Transform each term of the differential equation from the time domain to the s domain, and solve for $Y(s)$.
2. Substitute the value of the initial conditions and the transform of the input into the expression obtained in step 1.
3. Perform a partial-fraction expansion.
4. Invert each term back to the time domain to obtain the desired time response.

Much simplification in carrying out the algebraic manipulations of a Laplace transform solution is afforded for the following special cases:

1. If the system is initially at steady state [that is, $p^n y(0) = p^m x(0) = 0$, $n,m = 1, 2, \ldots$; and thus $y(0) = Kx(0)$] and in addition if

$$y(0) = Kx(0) = 0$$

then all the initial conditions are zero. For this case, the transform is obtained by merely substituting s for p and $Y(s)$ for $y(t)$ and $X(s)$ for $x(t)$ in the original differential equation.

2. If the system is initially at steady state but $y(0) = Kx(0) \neq 0$, then the response is obtained by adding the initial value $y(0)$ to the result obtained for case 1 above.

3. A time shift is effected by substituting $t - t_0$ for t. It is not necessary that the system be initially at a steady-state operating condition to effect a time shift.

Illustrative Example. Determine the response of the following differential equation to a unit step function when all the initial conditions are zero:

$$(p^2 + 7p + 12)y(t) = (p + 2)x(t) \tag{5.69}$$

Because the initial conditions are zero, substitution of s for p, $Y(s)$ for $y(t)$, and $X(s)$ for $x(t)$ yields

$$Y(s) = \frac{(s + 2)X(s)}{s^2 + 7s + 12} = \frac{s + 2}{s(s^2 + 7s + 12)} \tag{5.70}$$

where $X(s) = 1/s$ for a unit step function. Performing a partial-fraction expansion gives

$$Y(s) = \frac{K_1}{s} + \frac{K_2}{s + 3} + \frac{K_3}{s + 4} \tag{5.71}$$

where

$$K_1 = \lim_{s \to 0} \frac{s + 2}{(s + 3)(s + 4)} = \frac{1}{6}$$

$$K_2 = \lim_{s \to -3} \frac{s + 2}{s(s + 4)} = \frac{1}{3}$$

$$K_3 = \lim_{s \to -4} \frac{s + 2}{s(s + 3)} = -\frac{1}{2}$$

Thus

$$Y(s) = \frac{1}{6s} + \frac{1}{3(s + 3)} - \frac{1}{2(s + 4)} \tag{5.72}$$

Inverting the preceding expression gives the desired time response

$$y(t) = \tfrac{1}{6} + \tfrac{1}{3}e^{-3t} - \tfrac{1}{2}e^{-4t} \tag{5.73}$$

If initially $x(0)$ is not zero but the system is at a steady-state operating condition so that $p^2y(0) = py(0) = px(0) = 0$, then from Eq. (5.69) $12y(0) = 2x(0)$. The response is

$$y(t) = \tfrac{1}{6} + \tfrac{1}{3}e^{-3t} - \tfrac{1}{2}e^{-4t} + y(0) \tag{5.74}$$

A time shift by an amount t_0 can be effected by merely substituting $t - t_0$ for t.

5.5. Final-value Theorem. This theorem enables one to obtain the value $f(t)$ of a time function at $t = \infty$ directly from the Laplace transform $F(s)$. This is in effect the same type of information which is obtained from a steady-state analysis.

To develop the final-value theorem, first write the equation for the transform of a derivative in the form

$$\int_0^\infty f'(t)e^{-st}\,dt = sF(s) - f(0)$$

For s equal to zero, $e^{-st} = 1$, thus,

$$\int_0^\infty f'(t)\,dt = [sF(s)]_{s=0} - f(0) \tag{5.75}$$

The preceding expression may be written as

$$\int_0^\infty f'(t)\,dt = f(t)\Big|_0^\infty = \lim_{t\to\infty}[f(t)] - f(0) \tag{5.76}$$

The desired final-value theorem is obtained directly from Eqs. (5.75) and (5.76), i.e.,

$$\lim_{t\to\infty} f(t) = [sF(s)]_{s=0} \tag{5.77}$$

Application of the final-value theorem to Eq. (5.56) gives

$$\lim_{t\to\infty} y(t) = \frac{Kh_c/\tau + sy(0)}{s + 1/\tau}\bigg|_{s=0} = Kh_c \tag{5.78}$$

This is the result that is obtained by substitution of $t = \infty$ into Eq. (5.59). The value of $y(t)$ at $t = \infty$ has no significance when $y(t)$ is a pure sinusoid or when $y(t)$ becomes infinite, and so in these cases the final-value theorem is meaningless.

Initial-value Theorem. With the aid of the initial-value theorem, the value $f(t)$ of a time function at $t = (0+)$ may be computed directly from the transform $F(s)$ for the function. It is to be noted that $f(0+)$ is not the initial value but rather the value of the function at a time slightly greater than zero. Response equations obtained by Laplace

transforms are not valid for $t = 0$, but only for $t > 0$. For most response functions $f(0+)$ is equal to $f(0)$; however, when a step change occurs at $t = 0$, then $f(0) \neq f(0+)$.

The derivation of the initial-value theorem follows: For $t \approx 0$, then $e^{-st} \approx 1$; thus the equation for the transform of a derivative can be written as

$$\mathfrak{L}[f'(t)] = \int_0^{0+} f'(t)(1)\,dt + \int_{0+}^{\infty} f'(t)e^{-st}\,dt = sF(s) - f(0) \quad (5.79)$$

As s approaches infinity, $e^{-st} \approx 0$, which makes the second integral vanish. Letting $s \to \infty$ in Eq. (5.79) gives

$$f(0+) - f(0) = \lim_{s\to\infty} sF(s) - f(0)$$

or

$$f(0+) = \lim_{s\to\infty} sF(s) \quad (5.80)$$

Equation (5.80) is the initial-value theorem. Application of this initial-value theorem to Eq. (5.56) gives

$$y(0+) = \left[\frac{Kh_c/\tau s + y(0)}{1 + 1/\tau s}\right]_{s=\infty} = \frac{0 + y(0)}{1 + 0} = y(0)$$

This merely indicates that there is not a step change in $y(t)$ at $t = 0$. The step function shown in Fig. 5.8 illustrates the case in which a step change does occur at $t = 0$. Application of the initial-value theorem gives

$$x(0+) = [sX(s)]_{s=\infty} = \frac{s[h_c + x(0)]}{s}\bigg|_{s=\infty} = h_c + x(0) \quad (5.81)$$

Thus, a step change of height h_c is seen to occur at $t = 0$.

5.6. Input Functions Which Are Piecewise Continuous. A piecewise continuous function is characterized by the fact that the equation for the function changes from interval to interval as is illustrated by the functions in Fig. 5.10. For example, in Fig. 5.10*a* for $0 \leq t \leq t_0$, then $x(t) = at$; and for $t \geq t_0$, then $x(t) = at_0$. For the first interval $0 \leq t \leq t_0$, the input function is inclined at a slope a. Such an inclined straight line is called a ramp function. For $t \geq t_0$, the input is seen to be a step function. The solution of such problems is effected by starting with the first interval and successively solving for the response in each interval.

Let it be desired to solve Eq. (5.52) for the case in which the input is that shown in Fig. 5.10*a*. From the transform table, the transform for $t^n/n!$ is $1/s^{n+1}$; so for $n = 1$, it follows that $\mathfrak{L}(at) = a/s^2$. Thus, the transform $X(s)$ for the input is

$$X(s) = \frac{a}{s^2} \quad (5.82)$$

Substitution of the preceding result into Eq. (5.54) gives

$$Y(s) = \frac{Ka/\tau s^2 + y(0)}{s + 1/\tau} = \frac{Ka/\tau + s^2 y(0)}{s^2(s + 1/\tau)} \tag{5.83}$$

A partial-fraction expansion of Eq. (5.83) gives

$$Y(s) = \frac{C_2}{s^2} + \frac{C_1}{s} + \frac{K_1}{s + 1/\tau} \tag{5.84}$$

where from Eq. (5.24)

$$C_2 = \lim_{s \to 0} s^2 Y(s) = \lim_{s \to 0} \frac{Ka/\tau + s^2 y(0)}{s + 1/\tau} = Ka$$

$$C_1 = \lim_{s \to 0} \frac{d}{ds} [s^2 Y(s)] = \lim_{s \to 0} \frac{(s + 1/\tau)2sy(0) - [Ka/\tau + s^2 y(0)]}{(s + 1/\tau)^2}$$

$$= \frac{-Ka/\tau}{1/\tau^2} = -Ka\tau$$

$$K_1 = \lim_{s \to -1/\tau} \left[\left(s + \frac{1}{\tau} \right) Y(s) \right] = Ka\tau + y(0)$$

Substitution of these constants into Eq. (5.84) gives

$$Y(s) = \frac{Ka}{s^2} - \frac{Ka\tau}{s} + \frac{Ka\tau + y(0)}{s + 1/\tau} \tag{5.85}$$

Inverting the preceding expression gives the desired response for $0 < t \leq t_0$,

$$\begin{aligned} y(t) &= Kat - Ka\tau + [Ka\tau + y(0)]e^{-t/\tau} \\ &= Kat - Ka\tau(1 - e^{-t/\tau}) + y(0)e^{-t/\tau} \qquad 0 < t \leq t_0 \end{aligned} \tag{5.86}$$

For the second interval the input is a step function. From Fig. 5.10*a*, it is to be seen that the initial value $x(t_0)$ is at_0. Because there is no change in the height of this function from its initial value, h_c equals zero. The step-function response for this system has previously been determined and is given by Eq. (5.67). The substitution of $x(t_0) = at_0$ and $h_c = 0$ into Eq. (5.67) gives

$$y(t - t_0) = Kat_0 + [y(t_0) - Kat_0]e^{-(t-t_0)/\tau} \qquad t > t_0 \tag{5.87}$$

The value of $y(t_0)$ is obtained by letting $t = t_0$ in Eq. (5.86).

$$y(t_0) = Kat_0 - Ka\tau(1 - e^{-t_0/\tau}) + y(0)e^{-t_0/\tau} \tag{5.88}$$

Substitution of this result for $y(t_0)$ into Eq. (5.87) gives

$$\begin{aligned} y(t - t_0) &= Kat_0 + [-Ka\tau(1 - e^{-t_0/\tau}) + y(0)e^{-t_0/\tau}]e^{-(t-t_0)/\tau} \\ &= Kat_0 - Ka\tau e^{-t/\tau}(e^{t_0/\tau} - 1) + y(0)e^{-t/\tau} \qquad t > t_0 \end{aligned} \tag{5.89}$$

In this method, the initial conditions for a new interval are obtained by evaluating the equation for the preceding interval at the value of time when the preceding interval ceases and the new one begins.

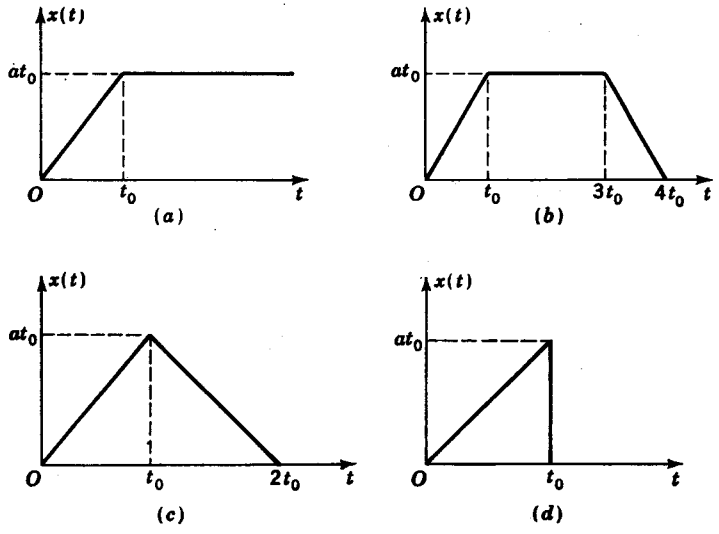


FIG. 5.10. Piecewise continuous functions.

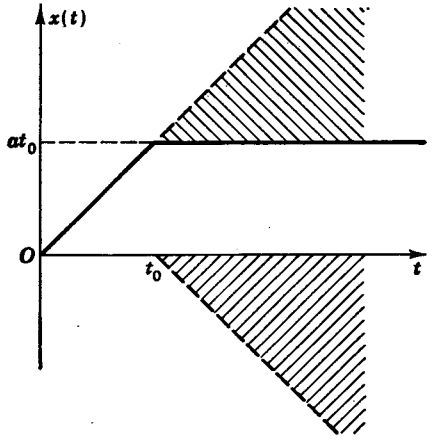


FIG. 5.11. Sum of two ramp functions.

An alternate method for solving piecewise continuous problems is to regard the input as being the sum of separate functions as is illustrated in Fig. 5.11. That is, the sum of the ramp function which begins at $t = 0$ and the equal but opposite ramp function which begins at $t = t_0$ is seen to yield the function of Fig. 5.10*a*. Similarly, it is possible to

represent any piecewise continuous function as the sum of other functions which are continuous.

The transform for the first ramp function of Fig. 5.11 is a/s^2, and that for the delayed ramp function is $-(a/s^2)e^{-t_0 s}$. Thus, the transform $X(s)$ for the input is

$$X(s) = \frac{a}{s^2} - \frac{a}{s^2} e^{-t_0 s} \tag{5.90}$$

Substitution of the preceding result into Eq. (5.54) gives

$$Y(s) = \frac{Ka/\tau s^2 + y(0)}{s + 1/\tau} - \frac{(Ka/\tau s^2)e^{-t_0 s}}{s + 1/\tau} \tag{5.91}$$

Because of the delaying factor $e^{-t_0 s}$ the second term on the right-hand side of Eq. (5.91) should be ignored for $t \leq t_0$. Thus, for $0 < t \leq t_0$, the transform is the same as that given by Eq. (5.83). The corresponding response for $0 < t \leq t_0$ is given by Eq. (5.86). For $t > t_0$, the response is the sum of that due to the second term of Eq. (5.91) plus that already obtained for the first term [Eq. (5.86)]. By noting that the second term of Eq. (5.91) is the same as the first with the exception of no initial-condition term $y(0)$, the response due to the second term is obtained by letting $y(0) = 0$ in Eq. (5.86) and substituting $t - t_0$ for t to take into account the time shift. Thus, for $t > t_0$ the total response is

$$y(t) = Kat - Ka\tau(1 - e^{-t/\tau}) + y(0)e^{-t/\tau} - [Ka(t - t_0) - Ka\tau(1 - e^{-(t-t_0)/\tau})]$$

or

$$y(t) = Kat_0 - Ka\tau e^{-t/\tau}(e^{t_0/\tau} - 1) + y(0)e^{-t/\tau} \tag{5.92}$$

An advantage of this second method of solution is that initial conditions appear only in the transformed expression for the first interval. However, in this latter method, the amount of computational effort increases with the number of separate functions required to make up the over-all piecewise continuous function. Thus, the choice of the first or second method depends on the particular problem to be solved. (In Appendix II, it is shown how the convolution integral may be used to determine the response to any arbitrary input function.)

CHAPTER 6

THE CHARACTERISTIC FUNCTION

6.1. General. Because an actual system may be subjected to all types and varieties of input excitations $x(t)$, it becomes impractical to calculate the system response for every possible excitation. In this chapter, it is shown that a very good measure of the transient behavior may be obtained directly from the zeros of the characteristic function (i.e., roots of the characteristic equation). This criterion for evaluating transient performance is obtained by considering the essential characteristics of a general system of order n.

The general operational representation for a differential equation of order n is

$$y(t) = \frac{a_m p^m + a_{m-1}p^{m-1} + \cdots + a_1 p + a_0}{p^n + b_{n-1}p^{n-1} + \cdots + b_1 p + b_0} x(t) \tag{6.1}$$

The transform of each term is

$$\begin{aligned}
\mathcal{L}[p^n y(t)] &= s^n Y(s) - I(s)_n \\
b_{n-1}\mathcal{L}[p^{n-1}y(t)] &= b_{n-1}s^{n-1}Y(s) - I(s)_{n-1} \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
a_m\mathcal{L}[p^m x(t)] &= a_m s^m X(s) - I(s)_m \\
a_{m-1}\mathcal{L}[p^{m-1}x(t)] &= a_{m-1}s^{m-1}X(s) - I(s)_{m-1} \\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

where I_n, I_{n-1}, . . . represent the initial conditions associated with each transform. Transforming each term of Eq. (6.1) accordingly and collecting terms yields

$$\begin{aligned}
Y(s) &= \frac{(a_m s^m + a_{m-1}s^{m-1} + \cdots + a_1 s + a_0)X(s) + I(s)}{s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0} \\
&= \frac{L_m(s)X(s) + I(s)}{L_n(s)}
\end{aligned} \tag{6.2}$$

where $I(s) = I(s)_n + I(s)_{n-1} + \cdots - I(s)_m - I(s)_{m-1} \cdots$ is the sum of all the initial conditions. By comparison of Eqs. (6.1) and (6.2), it is to be noted that the form of the characteristic function in the s domain, $L_n(s)$, remains the same as that in the p domain, $L_n(p)$. The numerator also has the same form with the exception that the initial conditions $I(s)$

are added in the s domain. Comparison of Eqs. (6.1) and (6.2) shows that, *when all the initial conditions are zero, the transform is obtained by merely substituting s for p, $Y(s)$ for $y(t)$, and $X(s)$ for $x(t)$ in the operational form of the differential equation.*

For this case

$$Y(s) = \frac{L_m(s)}{L_n(s)} X(s) \tag{6.3}$$

where $L_m(s)/L_n(s)$ is called the *transfer function.* It is to be noted that $L_m(s)$ and $L_n(s)$ are obtained directly from the differential equation of operation for the system. Thus, *the transfer function contains basic information concerning the essential characteristics of a system* without regard to initial conditions or excitation.

The term $X(s)$ in Eq. (6.2) is the general representation for the transform of the input signal or forcing function. This term may be written as

$$X(s) = \frac{N_{X(s)}}{D_{X(s)}}$$

where $N_{X(s)}$ is the numerator of $X(s)$ and $D_{X(s)}$ is the denominator of $X(s)$. For example, for a unit step function, $X(s) = 1/s$ so that $N_{X(s)} = 1$ and $D_{X(s)} = s$. Substitution of the preceding representation $X(s) = N_{X(s)}/D_{X(s)}$ into Eq. (6.2) yields the following general transformed form for $Y(s)$,

$$Y(s) = \frac{L_m(s)N_{X(s)} + I(s)D_{X(s)}}{L_n(s)D_{X(s)}} = \frac{A(s)}{B(s)} \tag{6.4}$$

where $A(s)$ and $B(s)$ are polynomials in s.

6.2. Transforming from the s Domain Back to the Time Domain—Inverse Transformations. By an inverse transformation is meant the process of inverting a function from the s domain back to the time domain. The inverse transform $\mathcal{L}^{-1}$ of a function $F(s)$ is defined by the equation

$$\mathcal{L}^{-1}[F(s)] = \frac{1}{2\pi j} \int_C F(s)e^{ts}\, ds = f(t) \tag{6.5}$$

where C is a suitably chosen contour in the s domain. This integral method of evaluating the inverse transform is not employed when the much simpler process of entering a transform table with the given $F(s)$ and reading directly the desired $f(t)$ can be utilized, as is the case for ordinary control analysis. A partial listing of commonly used transforms is given in Table 5.1.

The transform table may be used to obtain the Laplace transform $F(s)$ of a given function of time or to obtain the inverse transform $f(t)$ for a given function of s. This process is analogous to the use of a logarithmic

table for obtaining the logarithm of a number or to the use of the same table for the opposite process of obtaining antilogarithms.

At first it would appear that the listing of transforms given in Table 5.1 would have to be extended considerably so that it would be applicable to the wide range of problems encountered in the design of systems. However, this is not the case. The listing given in Table 5.1 is adequate for the solution of most ordinary problems that arise in control engineering. The reason for this is that there are relatively few different types of terms which appear in the differential equation after it has been expanded by a partial-fraction expansion. In particular, the zeros of $B(s)$ are either distinct or repeated.

Distinct Zeros. The transformed function $B(s)$ is the denominator of Eq. (6.4). When the zeros of $B(s)$ are distinct, the denominator $B(s)$ can be factored in the form

$$B(s) = (s - r_1)(s - r_2) \cdots (s - r_n) \tag{6.6}$$

where $r_1, r_2, \ldots, r_n$ are n distinct zeros of $B(s)$.

The partial-fraction expansion of Eq. (6.4) is of the form

$$Y(s) = \frac{K_1}{s - r_1} + \frac{K_2}{s - r_2} + \cdots + \frac{K_i}{s - r_i} + \cdots + \frac{K_n}{s - r_n} \tag{6.7}$$

where $K_1, K_2, \ldots, K_n$ are n constants. Each constant K_i may be evaluated by the method used to obtain Eq. (5.9). That is, first multiply both sides of Eq. (6.7) by $s - r_i$; then take the limit as s approaches r_i. After performing these operations, the only term remaining on the right-hand side of Eq. (6.7) is K_i. Thus

$$K_i = \lim_{s \to r_i} [(s - r_i) Y(s)] \tag{6.8}$$

The inverse transform of Eq. (6.7) is obtained directly from the transform table and is

$$y(t) = K_1 e^{r_1 t} + K_2 e^{r_2 t} + \cdots + K_n e^{r_n t} \tag{6.9}$$

Equation (6.9) shows that *each distinct zero of* $B(s) = L_n(s) D_{X(s)}$ *yields an exponential-type term* $K_i e^{r_i t}$ *in the response function.* The exponent r_i is the corresponding zero of $B(s)$. Each zero $r_1, r_2, \ldots, r_n$ must be negative in order that each term $K_i e^{r_i t}$ in $y(t)$ be a decaying function. If any zero of $B(s)$ is positive, $y(t)$ will increase without bound as t increases to infinity. A constant term results if $r_i = 0$, because $K_i e^{(0)t} = K_i$.

Repeated Zeros. For the case in which $B(s)$ has a multiple or repeated zero r which occurs q times, $B(s)$ may be factored in the form

$$B(s) = (s - r)^q (s - r_1)(s - r_2) \cdots (s - r_{n-q}) \tag{6.10}$$

The corresponding partial-fraction expansion for $Y(s)$ is

$$Y(s) = \frac{C_q}{(s-r)^q} + \frac{C_{q-1}}{(s-r)^{q-1}} + \cdots + \frac{C_1}{s-r} + \frac{K_1}{s-r_1} + \frac{K_2}{s-r_2} + \cdots + \frac{K_{n-q}}{s-r_{n-q}} \quad (6.11)$$

The constant coefficients for the multiple terms are evaluated as follows:

$$\begin{aligned} C_q &= \lim_{s \to r} [(s-r)^q Y(s)] \\ C_{q-1} &= \lim_{s \to r} \left\{ \frac{d}{ds} [(s-r)^q Y(s)] \right\} \\ C_{q-k} &= \lim_{s \to r} \left\{ \frac{1}{k!} \frac{d^k}{ds^k} [(s-r)^q Y(s)] \right\} \end{aligned} \quad (6.12)$$

From the transform table, the inverse transform of Eq. (6.11) is found to be

$$y(t) = \frac{C_q t^{q-1} e^{rt}}{(q-1)!} + \frac{C_{q-1} t^{q-2} e^{rt}}{(q-2)!} + \cdots + \frac{C_2 t e^{rt}}{1!} + C_1 e^{rt} + K_1 e^{r_1 t} + K_2 e^{r_2 t} + \cdots + K_{n-q} e^{r_{n-q} t} \quad (6.13)$$

Each response term associated with the repeated zero $(s-r)^q$ is seen to be multiplied by the exponential factor e^{rt}. If the value of r is positive, $y(t)$ will become infinite as time increases. For negative values of r, a decreasing exponential results, and thus the response term due to the repeated zero eventually vanishes.

Illustrative Example. Let it be desired to determine the time response $y(t)$ for the transformed equation

$$Y(s) = \frac{11s + 28}{(s+2)^2(s+5)} = \frac{C_2}{(s+2)^2} + \frac{C_1}{s+2} + \frac{K_1}{s+5} \quad (6.14)$$

The constants are evaluated as follows:

$$C_2 = \lim_{s \to -2} \frac{11s + 28}{s+5} = 2$$

$$C_1 = \lim_{s \to -2} \left[\frac{d}{ds} \left(\frac{11s + 28}{s+5} \right) \right] = \lim_{s \to -2} \frac{(s+5)11 - (11s+28)}{(s+5)^2} = 3$$

$$K_1 = \lim_{s \to -5} \frac{11s + 28}{(s+2)^2} = -3$$

Thus,
$$Y(s) = \frac{2}{(s+2)^2} + \frac{3}{s+2} - \frac{3}{s+5}$$

By use of Table 5.1, the inverse transform of the preceding equation is found to be

$$y(t) = 2te^{-2t} + 3e^{-2t} - 3e^{-5t}$$

or
$$y(t) = (2t+3)e^{-2t} - 3e^{-5t} \quad (6.15)$$

6.3. Complex Conjugate Zeros. Complex zeros of $B(s)$ always occur in pairs, and furthermore these zeros are always conjugates of one another. That is, they have the same real part but equal and opposite imaginary parts. Thus, if the polynomial $B(s)$ has a complex zero $a + jb$, the complex conjugate $a - jb$ will also be a zero of $B(s)$. Although the preceding discussion of distinct zeros is also applicable to complex conjugate zeros, the following analysis brings out more clearly the fact that a pair of complex conjugate zeros in $B(s)$ combine to introduce an exponentially damped sinusoidal term in $y(t)$.

A pair of complex conjugate zeros when multiplied together yield the following quadratic:

$$(s - a - jb)(s - a + jb) = s^2 - 2as + (a^2 + b^2) \tag{6.16}$$

For any given quadratic term, the values of a and b may be computed by equating coefficients of like terms as follows: Consider the expression $s^2 + 4s + 9$. The coefficient 4 of the s term is equal to $-2a$ so that $-2a = 4$ or $a = -2$. Similarly equating the constant terms gives $a^2 + b^2 = 9$ or $b = \sqrt{9 - 4} = \sqrt{5}$. Because of the square root, b may be either positive or negative. For convenience in working numerical problems, only the positive value is used. If in the determination of b it is found that b is an imaginary number, the two zeros are real and unequal rather than complex conjugates. For example, in the expression $s^2 + 8s + 12$ the value of a is equal to -4 so that $b = \sqrt{12 - 16} = j\sqrt{4} = j2$. For this case, the zeros are $a \pm jb = -4 \pm (j^2 2) = -4 \mp 2 = -6, -2$. Because the case of real zeros has been previously discussed, in the following analysis it is assumed that b is real so that the zeros are complex conjugates.

Suppose that $B(s)$ may be factored in the form

$$B(s) = (s - a - jb)(s - a + jb)(s - r_1) \cdots (s - r_{n-2}) \tag{6.17}$$

The partial-fraction expansion for $Y(s) = A(s)/B(s)$ is of the form

$$Y(s) = \frac{K_c}{s - a - jb} + \frac{K_{-c}}{s - a + jb} + \frac{K_1}{s - r_1} + \cdots + \frac{K_{n-2}}{s - r_{n-2}} \tag{6.18}$$

The inverse transform of Eq. (6.18) is

$$y(t) = K_c e^{(a+jb)t} + K_{-c} e^{(a-jb)t} + K_1 e^{r_1 t} + \cdots + K_{n-2} e^{r_{n-2} t} \tag{6.19}$$

The constants K_c and K_{-c} associated with the complex conjugate zeros are evaluated as usual for distinct zeros by the application of Eq. (6.8). That is,

$$K_c = \lim_{s \to a+jb} \left[(s - a - jb) \frac{A(s)}{(s - a - jb)(s - a + jb)(s - r_1) \cdots (s - r_{n-2})} \right]$$

$$= \lim_{s \to a+jb} \left[\frac{1}{2jb} \frac{A(s)}{(s - r_1) \cdots (s - r_{n-2})} \right] = \frac{1}{2jb} K(a + jb) \tag{6.20}$$

where

$$K(a+jb) = \lim_{s\to a+jb} \frac{A(s)}{(s-r_1)\cdots(s-r_{n-2})} = \left[(s^2-2as+a^2+b^2)\frac{A(s)}{B(s)}\right]_{s=a+jb}$$

Similarly, the constant K_{-c} is obtained as follows:

$$K_{-c} = \lim_{s\to a-jb}\left[(s-a+jb)\frac{A(s)}{(s-a-jb)(s-a+jb)(s-r_1)\cdots(s-r_{n-2})}\right]$$
$$= \lim_{s\to a-jb}\frac{1}{(-2jb)}\frac{A(s)}{(s-r_1)\cdots(s-r_{n-2})} = -\frac{1}{2jb}K(a-jb) \quad (6.21)$$

where $K(a-jb) = \lim_{s\to a-jb}\frac{A(s)}{(s-r_1)\cdots(s-r_{n-2})} = \left[(s^2-2as+a^2+b^2)\frac{A(s)}{B(s)}\right]_{s=a-jb}$

The constants $K(a+jb)$ and $K(a-jb)$ are complex conjugate numbers. These complex numbers may be represented as shown graphically in Fig. 6.1, whence

$$\begin{aligned} K(a+jb) &= |K(a+jb)|e^{j\alpha} \\ K(a-jb) &= |K(a+jb)|e^{-j\alpha} \end{aligned} \quad (6.22)$$

where $|K(a+jb)| = |K(a-jb)|$ is the length of either vector, α is the angle of the vector $K(a+jb)$, and $-\alpha$ is the angle of the vector $K(a-jb)$.

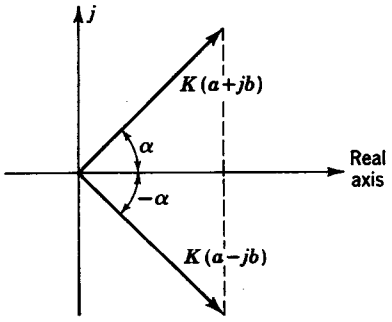


FIG. 6.1. Vector representation for $K(a+jb)$ and $K(a-jb)$.

The constants K_c and K_{-c}, which are also complex conjugate numbers, may be written in the form

$$\begin{aligned} K_c &= \frac{1}{2jb}|K(a+jb)|e^{j\alpha} \\ K_{-c} &= -\frac{1}{2jb}|K(a+jb)|e^{-j\alpha} \end{aligned} \quad (6.23)$$

Substitution of K_c and K_{-c} from Eq. (6.23) into Eq. (6.19) gives

$$y(t) = \frac{1}{b}|K(a+jb)|e^{at}\frac{e^{j(bt+\alpha)} - e^{-j(bt+\alpha)}}{2j} + K_1e^{r_1t} + \cdots + K_{n-2}e^{r_{n-2}t}$$

or

$$y(t) = \frac{1}{b}|K(a+jb)|e^{at}\sin(bt+\alpha) + K_1e^{r_1t} + \cdots + K_{n-2}e^{r_{n-2}t} \quad (6.24)$$

Illustrative Example. Determine the inverse transformation of the following transformed equation:

$$Y(s) = \frac{20}{(s^2+4s+13)(s+6)}$$

SOLUTION. Equating coefficients to obtain the value of a and b for the quadratic yields $-2a = 4$, or $a = -2$, and $a^2 + b^2 = 13$, or

$$b = \sqrt{13-4} = 3$$

Evaluation of $K(a+jb)$ gives

$$K(a+jb) = \left[(s^2 - 2as + a^2 + b^2)\frac{A(s)}{B(s)}\right]_{s=a+jb}$$

$$= \left(\frac{20}{s+6}\right)_{s=-2+j3} = \frac{20}{4+j3} \quad (6.25)$$

The magnitude is

$$|K(a+jb)| = \frac{20}{\sqrt{16+9}} = 4.0$$

Similarly, the angle α is

$$\alpha = \measuredangle K(a+jb) = \tan^{-1}\frac{\text{imaginary part of } K(a+jb)}{\text{real part of } K(a+jb)}$$

$$= \measuredangle \frac{20}{4+j3} = \measuredangle \frac{20}{4+j3}\frac{4-j3}{4-j3}$$

$$= \measuredangle 20/25(4-j3) = \tan^{-1} -3/4 = -36.8° \quad (6.26)$$

The general form of the inverse transformation is

$$y(t) = \frac{1}{b}|K(a+jb)|e^{at}\sin(bt+\alpha) + K_1e^{r_1t}$$

Evaluation of K_1 gives

$$K_1 = \lim_{s\to -6}\frac{20}{s^2+4s+13} = \frac{20}{25} = 0.8$$

Thus the desired result is

$$y(t) = 4/3e^{-2t}\sin(3t - 36.8°) + 0.8e^{-6t}$$

Equation (6.24) shows the exponentially damped sinusoidal term which results from complex conjugate zeros of $B(s)$. The exponential factor a is the real part of the complex conjugate zeros. The imaginary part b is the frequency of oscillation of the exponentially damped sinusoid. Thus, b is referred to as the damped frequency of oscillation. The period of each oscillation is $2\pi/b$. The envelope of this sinusoid is $(1/b)|K(a + jb)|e^{at}$. To have the exponential term decreasing with time, it is necessary that a be negative. For the case in which $a = 0$, a

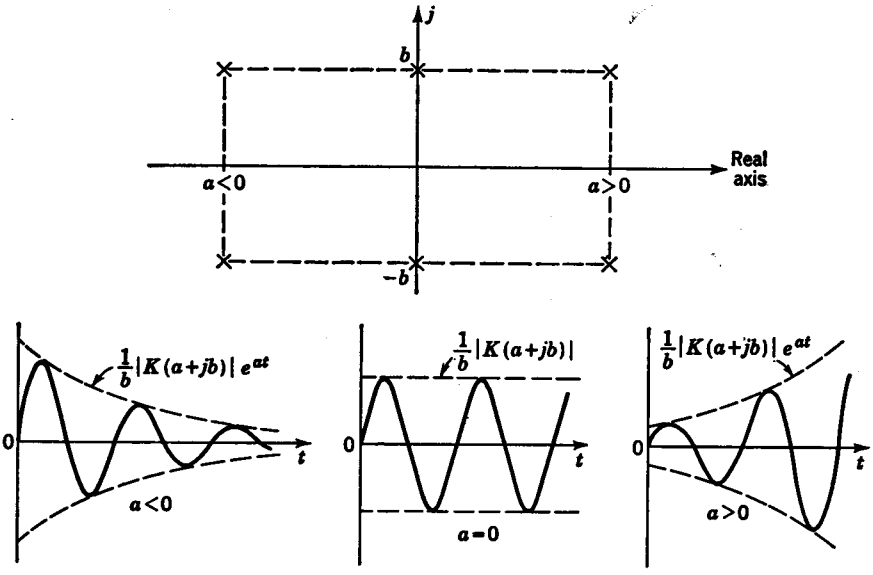


FIG. 6.2. Response terms that result from complex conjugate zeros.

sinusoid of constant amplitude $(1/b)|K(a + jb)|$ results. For $a = 0$, Eq. (6.24) becomes

$$y(t) = \frac{1}{b}|K(a + jb)| \sin(bt + \alpha) + K_1 e^{r_1 t} + \cdots + K_{n-2} e^{r_{n-2} t} \quad (6.27)$$

In Fig. 6.2 is graphically illustrated the type of time-response terms that result from complex conjugate zeros. When the zeros lie to the left of the imaginary axis ($a < 0$), a decreasing sinusoid results; when the zeros are on the imaginary axis ($a = 0$), a sinusoid of constant amplitude results; and when the zeros are to the right of the imaginary axis ($a > 0$), an increasing sinusoid results.

In Fig. 6.3 is shown a plot of the type of response terms that result from real zeros. These results follow directly from Eq. (6.9). Because the exponential factor is the value of the zero, a negative zero ($r_i < 0$) yields

an exponentially decreasing term, while a positive zero ($r_i > 0$) yields an exponentially increasing term. A zero at the origin ($r_i = 0$) results in a constant term. Even though a zero may be repeated, the exponential term dominates the response.

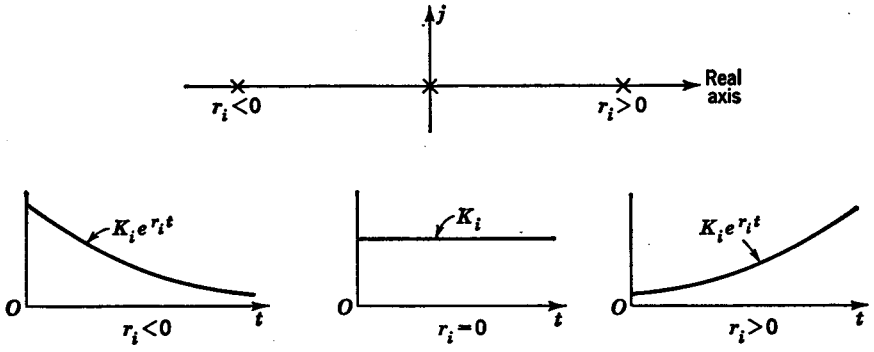


FIG. 6.3. Response terms that result from real zeros.

6.4. Damping Ratio and Natural Frequency. A pair of complex conjugate zeros may be specified by the damping ratio ζ (zeta) and the natural frequency ω_n rather than by the real part a and the imaginary part b. As is later shown, the damping ratio and natural frequency give a better general indication of the transient behavior than do a and b. In Fig. 6.4 is shown a pair of complex conjugate zeros. The distance from the origin to each zero is ω_n such that $\omega_n^2 = a^2 + b^2$. The angle β which is measured from the negative real axis to the radial vector ω_n is such that $\zeta = \cos\beta$ or $\beta = \cos^{-1}\zeta$. Thus, the damping ratio determines the angular location of the zeros and ω_n the distance of the zeros from the origin. Because

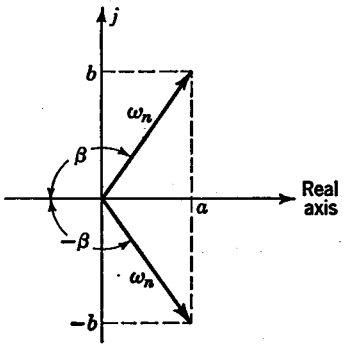


FIG. 6.4. Complex conjugate zeros.

$$a = -\omega_n \cos\beta = -\zeta\omega_n$$

and $b = \omega_n \sin\beta = \omega_n\sqrt{1 - \cos^2\beta} = \omega_n\sqrt{1 - \zeta^2}$, the quadratic form given by Eq. (6.16) may be written as

$$s^2 - 2as + (a^2 + b^2) = s^2 + 2\zeta\omega_n s + \omega_n^2 \tag{6.28}$$

For any given quadratic expression, the numerical values of ζ and ω_n are computed by equating coefficients. For example, consider the expression $s^2 + 4s + 9$. The constant terms are $\omega_n^2 = 9$ or $\omega_n = 3$. The

reason for using only the positive value when mathematically

$$\omega_n = \sqrt{9} = \pm 3$$

is that ω_n is the distance from the origin to the zero and must be positive. The coefficients for the s term give $2\zeta\omega_n = 4$, or $\zeta = 4/2\omega_n = \frac{2}{3}$. If so desired, the values of a and b are readily computed; that is,

$$a = -\omega_n\zeta = -3(\tfrac{2}{3}) = -2$$

and $b = \omega_n\sqrt{1-\zeta^2} = 3\sqrt{1-\frac{4}{9}} = 3\sqrt{\frac{5}{9}} = \sqrt{5}$.

The time response due to complex conjugate zeros of $B(s)$ [Eqs. (6.24)] may be written in terms of the damping ratio and natural frequency as follows:

$$y(t) = \frac{1}{b}|K(a+jb)|e^{-\zeta\omega_n t}\sin[(\omega_n\sqrt{1-\zeta^2})t + \alpha] + K_1e^{r_1t} + \cdots + K_{n-2}e^{r_{n-2}t} \qquad -1 < \zeta < 1 \quad (6.29)$$

For the case in which $\zeta = 0$, Eq. (6.29) becomes

$$y(t) = \frac{1}{b}|K(a+jb)|\sin(\omega_n t + \alpha) + K_1e^{r_1t} + \cdots + K_{n-2}e^{r_{n-2}t} \quad (6.30)$$

Thus, a sinusoidal term of constant amplitude $(1/b)|K(a+jb)|$ and natural frequency of oscillation ω_n is seen to result.

In Fig. 6.5 is shown a more general plot of a pair of complex conjugate zeros. When a is negative so that a decreasing exponential results, the zeros are to the left of the imaginary axis so that $\beta < 90°$, in which case $1 > \zeta > 0$. For positive values of a, the zeros are to the right of the imaginary axis so that $\beta > 90°$ and $0 > \zeta > -1$. A positive value of

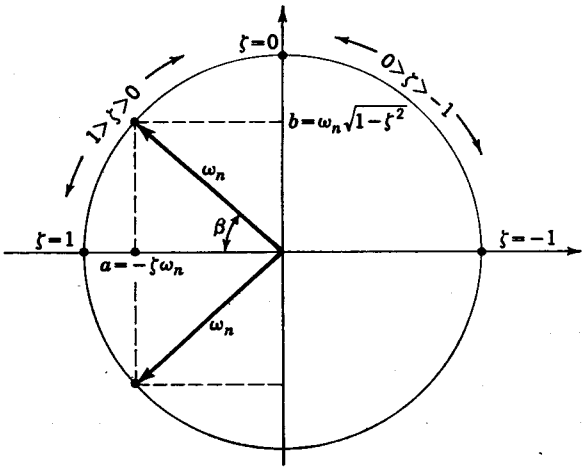


FIG. 6.5. General plot of complex conjugate zeros.

ζ yields a decreasing sinusoid, while a negative value results in an increasing sinusoid. For ζ equal to zero ($\beta = 90°$) a sinusoid of constant amplitude is obtained. When the magnitude of ζ is greater than 1, two real zeros result rather than complex conjugate zeros. The case of real zeros has previously been discussed.

Logarithmic Decrement. For an exponentially damped sinusoid as shown in Fig. 6.6, the amplitude of the sinusoid after each oscillation changes in a geometric series. At time t_1, the amplitude is $(1/b)|K(a + jb)|e^{at_1}$. The time required to complete one period is $T = 2\pi/b$, and thus the time

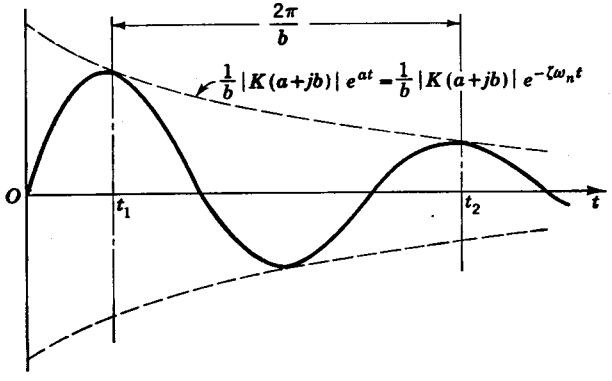


FIG. 6.6. Logarithmic decrement.

t_2 after one oscillation is completed is $t_2 = t_1 + T = t_1 + 2\pi/b$. The new amplitude is $(1/b)|K(a + jb)|e^{a(t_1+2\pi/b)}$, and the ratio of amplitudes is

$$\frac{(1/b)|K(a + jb)|e^{at_1}}{(1/b)|K(a + jb)|e^{at_1}e^{2\pi a/b}} = e^{-2\pi a/b} = e^{2\pi\zeta/\sqrt{(1-\zeta^2)}} \tag{6.31}$$

The log e of this amplitude ratio is $-2\pi a/b = 2\pi\zeta/\sqrt{1 - \zeta^2}$, which is called the logarithmic decrement. The amplitude ratio after one oscillation is thus seen to be a function of the damping ratio only.

Illustrative Example. Let it be desired to determine the general equation for the transient response of a second-order system to a unit-step-function change which occurs at $t = 0$. The operational form of the differential equation is

$$y(t) = \frac{\omega_n^2}{p^2 + 2\zeta\omega_n p + \omega_n^2}\, x(t) \tag{6.32}$$

Assume that all the initial conditions are zero.

SOLUTION. The transform for this differential equation is

$$Y(s) = \frac{\omega_n^2 X(s)}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{\omega_n^2}{(s^2 + 2\zeta\omega_n s + \omega_n^2)s} \tag{6.33}$$

where $X(s) = 1/s$ is the transform for the step input.

For complex conjugate zeros (that is, $-1 < \zeta < 1$), the response has the general form

$$y(t) = \frac{1}{b}|K(a+jb)|e^{at}\sin(bt+\alpha) + K_1 \tag{6.34}$$

Evaluation of $|K(a+jb)|$ gives

$$|K(a+jb)| = \left|\frac{\omega_n^2}{s}\right|_{s=a+jb} = \left|\frac{\omega_n^2}{a+jb}\right| = \frac{\omega_n^2}{\sqrt{a^2+b^2}}$$

Substitution of $a = -\zeta\omega_n$ and $b = \omega_n\sqrt{1-\zeta^2}$ gives

$$|K(a+jb)| = \frac{\omega_n^2}{\sqrt{\zeta^2\omega_n^2 + \omega_n^2(1-\zeta^2)}} = \omega_n \tag{6.35}$$

Similarly, it follows that

$$\alpha = \measuredangle K(a+jb) = \measuredangle \frac{\omega_n^2}{a+jb}\frac{a-jb}{a-jb} = \measuredangle(a-jb) = \tan^{-1}\frac{-b}{a}$$

$$= \tan^{-1}\frac{-\sqrt{1-\zeta^2}}{-\zeta} \tag{6.36}$$

The constant K_1 is evaluated as follows:

$$K_1 = \lim_{s\to 0}\frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} = 1$$

Thus, the desired transient response is

$$y(t) = \frac{1}{\sqrt{1-\zeta^2}}e^{-\zeta\omega_n t}\sin[(\omega_n\sqrt{1-\zeta^2})t + \alpha] + 1 \qquad -1 < \zeta < 1 \tag{6.37}$$

For the case in which $\zeta = 1$, the quadratic term in Eq. (6.33) is

$$s^2 + 2\omega_n s + \omega_n^2 = (s+\omega_n)^2$$

Thus the partial-fraction expansion for $Y(s)$ is

$$Y(s) = \frac{C_2}{(s+\omega_n)^2} + \frac{C_1}{s+\omega_n} + \frac{K_1}{s} \tag{6.38}$$

The constants C_1, C_2, and K_1 are evaluated as follows:

$$C_2 = \lim_{s\to-\omega_n}\frac{\omega_n^2}{s} = -\omega_n$$

$$C_1 = \lim_{s\to-\omega_n}\left[\frac{d}{ds}\left(\frac{\omega_n^2}{s}\right)\right] = \left.\frac{-\omega_n^2}{s^2}\right|_{s=-\omega_n} = -1$$

$$K_1 = \lim_{s\to 0}\frac{\omega_n^2}{s^2 + 2\omega_n s + \omega_n^2} = 1$$

Thus
$$y(t) = (C_2 t + C_1)e^{-\omega_n t} + K_1$$

or
$$y(t) = 1 - (\omega_n t + 1)e^{-\omega_n t} \tag{6.39}$$

In a similar manner, the equation of the time response for $\zeta > 1$ may be derived. For this case, the zeros are both negative and distinct so that the response function will include two decaying exponential terms.

In Fig. 6.7 is shown the response $y(t)$ to a step change in the input for various values of the damping ratio ζ. It is to be noted that, for $\zeta < 0.4$, there is an excessive amount of overshooting and oscillations. For $\zeta > 1$, an excessive amount of time is required to reach the new operating condition. Thus, for most control work, it is desired to have $0.4 < \zeta < 1$.

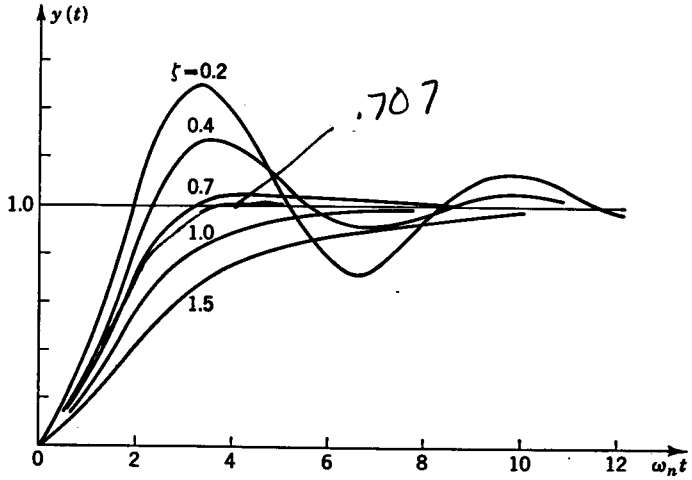


FIG. 6.7. Response of second-order system to a step input.

6.5. Predicting the Transient Response from the Zeros of $B(s)$. The general form of the transient response can be ascertained directly from the zeros of the transformed function $B(s)$. For example, suppose that the zeros of $B(s)$ are those plotted in Fig. 6.8. It follows that $B(s)$ may be factored in the form

$$B(s) = (s - a_1 - jb_1)(s - a_1 + jb_1)(s - jb_2) \\ (s + jb_2)(s - 0)(s - r_2)(s - r)^2 \quad (6.40)$$

Performing a partial-fraction expansion on $Y(s) = A(s)/B(s)$ yields

$$Y(s) = \frac{K_{c1}}{s - a_1 - jb_1} + \frac{K_{-c1}}{s - a_1 + jb_1} + \frac{K_{c2}}{s - jb_2} + \frac{K_{-c2}}{s + jb_2} + \frac{K_1}{s} \\ + \frac{K_2}{s - r_2} + \frac{C_2}{(s - r)^2} + \frac{C_1}{s - r} \quad (6.41)$$

Taking the inverse transform of the preceding expression yields

$$y(t) = \frac{1}{b_1} |K(a_1 + jb_1)| e^{a_1 t} \sin (b_1 t + \alpha_1) + \frac{1}{b_2} |K(jb_2)| \sin (b_2 t + \alpha_2) \\ + K_1 + K_2 e^{r_2 t} + (C_2 t + C_1) e^{rt} \quad (6.42)$$

It is to be noted that a pair of complex conjugate zeros yields an exponentially varying sinusoidal term. A pair of complex conjugate zeros on the imaginary axis yields a sinusoid with a constant amplitude. The zero at the origin contributes a constant term. Distinct or multiple zeros on the real axis yield exponential terms.

The term $B(s) = L_n(s)D_{X(s)}$ consists of the zeros of $L_n(s)$ of the characteristic function for the system plus the zeros $D_{X(s)}$ corresponding to the denominator of the transform of the input excitation. If any zero of the characteristic function for the system $L_n(s)$ lies to the right of the imaginary axis, the response contains an increasing time function and will increase without bound. *Thus, if any zero of $L_n(s)$ lies in the right half plane (i.e., to the right of the imaginary axis), then the system is basically unstable.* Whether a system is stable or unstable is a basic property of the system $L_n(s)$ itself and not the particular input, or excitation, to the system.

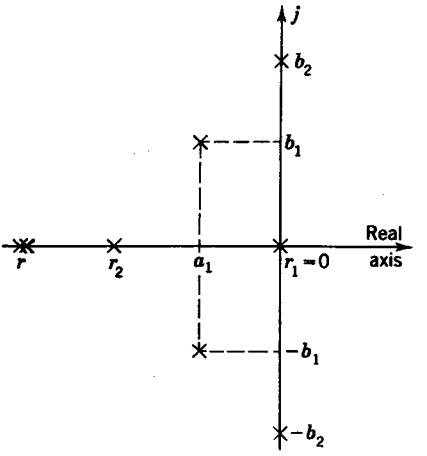


FIG. 6.8. Graphical representation of zeros of $B(s)$.

The zeros of $D_{X(s)}$ yield response terms associated with the particular excitation to the system. Take, for example, a ramp function $D_{X(s)} = s^2$ which gives a response term of the form C_1t. It should be noted that the input which is a ramp function eventually becomes infinite, and thus the output of the system has been forced to infinity because of the particular input. As is illustrated by this example, the *zeros of $D_{X(s)}$ do not affect the basic stability of a system* but merely yield response terms appropriate to the particular excitation.

Illustrative Example. The differential equation of operation for a control system is given by Eq. (6.43). Determine the general form of the response equation when the input excitation $x(t)$ is a unit step function.

$$y(t) = \frac{360(p^2 + p + 1)}{(p^2 + 2p + 5)(p^2 + 6p + 9)(p^2 + 6p + 8)}\,x(t) \qquad (6.43)$$

SOLUTION. Because of the unit-step-function input $N_{X(s)} = 1$, $D_{X(s)} = s$ so that $B(s) = L_n(s)D_{X(s)} = L_n(s)s$. Thus

$$Y(s) = \frac{360(s^2 + s + 1)(1) + sI(s)}{(s^2 + 2s + 5)(s^2 + 6s + 9)(s^2 + 6s + 8)s} = \frac{A(s)}{B(s)}$$

The partial-fraction expansion for $Y(s)$ gives

$$Y(s) = \frac{K_c}{s+1+j2} + \frac{K_{-c}}{s+1-j2} + \frac{C_2}{(s+3)^2} + \frac{C_1}{s+3} + \frac{K_1}{s+2} + \frac{K_2}{s+4} + \frac{K_3}{s}$$

Thus, the general form of the time solution is

$$y(t) = \frac{1}{2}|K(a+jb)|e^{-t}\sin(2t+\alpha) + (C_2t + C_1)e^{-3t} + K_1e^{-2t} + K_2e^{-4t} + K_3 \quad (6.44)$$

It should be noted in all cases that the exponent of each exponential term is equal to the horizontal distance from the imaginary axis to the zero of interest. That is, the exponential factor is equal to the numerical value of the real part of the zero. The terms due to zeros which are located far to the left of the imaginary axis have large exponential decaying factors and tend to decrease very rapidly to negligible quantities. Thus, zeros closer to the imaginary axis usually have a more predominant effect upon the transient behavior. Accordingly, the analysis of complicated control systems is often approximated by omitting from the characteristic function zeros which do not affect substantially the performance of the system.

6.6. Response of System to Change in External Disturbance. In this section, it is shown that the characteristic function for the differential

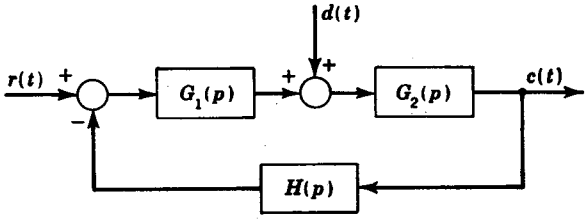


FIG. 6.9. General representation for a feedback control system.

equation which relates the output of a system to a change in the external disturbance is the same as that for the differential equation which relates the output to a change in the desired input. In Fig. 6.9 is shown the general representation for a feedback control system in which $d(t)$ represents the external disturbance.

As previously discussed, the effect of the input $r(t)$ and external disturbance $d(t)$ on the output or controlled variable $c(t)$ may be considered individually and then each result added by superposition to obtain

the total variation in $c(t)$. The block diagram which relates the input $r(t)$ to the output $c(t)$ without regard to the external disturbance is shown in Fig. 6.10. The equation relating $r(t)$ and $c(t)$ is obtained as follows:

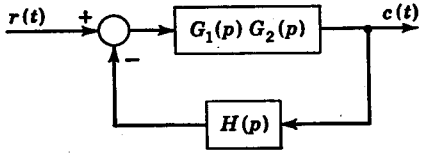


FIG. 6.10. Block diagram for consideration of the input $r(t)$.

$$[r(t) - H(p)c(t)]G_1(p)G_2(p) = c(t) \tag{6.45}$$

Solving for $c(t)$ gives

$$c(t) = \frac{G_1(p)G_2(p)}{1 + G_1(p)G_2(p)H(p)}\, r(t)$$

By using N_{G_1} to designate the numerator of $G_1(p)$ and D_{G_1} to designate the denominator of $G_1(p)$, etc., the preceding expression may be written in the form

$$\begin{aligned} c(t) &= \frac{N_{G_1}N_{G_2}/D_{G_1}D_{G_2}}{1 + N_{G_1}N_{G_2}N_H/D_{G_1}D_{G_2}D_H}\, r(t) \\ &= \frac{N_{G_1}N_{G_2}D_H}{D_{G_1}D_{G_2}D_H + N_{G_1}N_{G_2}N_H}\, r(t) \end{aligned} \tag{6.46}$$

The characteristic function $L_n(p)$ for the system is $D_{G_1}D_{G_2}D_H + N_{G_1}N_{G_2}N_H$.

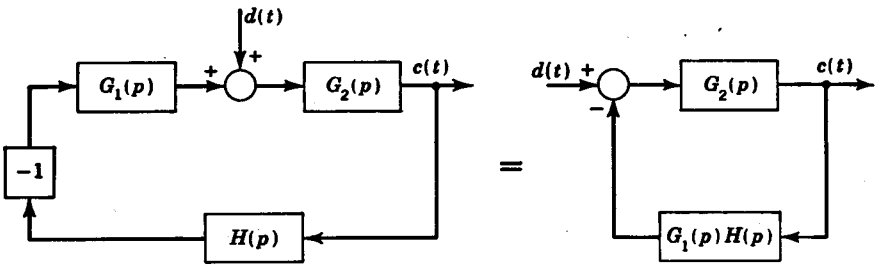


FIG. 6.11. Block diagram for consideration of effect of an external disturbance $d(t)$.

The block diagram which relates the external disturbance $d(t)$ to the output when $r(t)$ is considered zero is shown in Fig. 6.11. The equation relating $d(t)$ and $c(t)$ is

$$[d(t) - G_1(p)H(p)c(t)]G_2(p) = c(t) \tag{6.47}$$

or

$$\begin{aligned} c(t) &= \frac{G_2(p)}{1 + G_1(p)G_2(p)H(p)}\, d(t) = \frac{N_{G_2}/D_{G_2}}{1 + N_{G_1}N_{G_2}N_H/D_{G_1}D_{G_2}D_H} \\ &= \frac{N_{G_2}D_{G_1}D_H}{D_{G_1}D_{G_2}D_H + N_{G_1}N_{G_2}N_H} \end{aligned} \tag{6.48}$$

where $G_1(p) = \dfrac{N_{G_1}}{D_{G_1}}$ $G_2(p) = \dfrac{N_{G_2}}{D_{G_2}}$ $H(p) = \dfrac{N_H}{D_H}$

It can be shown that, if an excitation enters any place in the loop, the differential equation relating the disturbance and the output will always

have the same characteristic function $L_n(p)$ which is the product of all the denominator terms plus the product of all the numerators. However, the constant terms which appear in the partial-fraction expansion depend upon where the disturbance enters the system. Thus, there is but one characteristic function for a system, and this function gives basic information as to the transient behavior of a system.

6.7. System Stability and Transient Behavior Based on Impulse Response. The response of a system to a unit impulse excitation provides a good indication or measure of the general transient behavior of the system. The unit impulse is in effect a momentary disturbance which

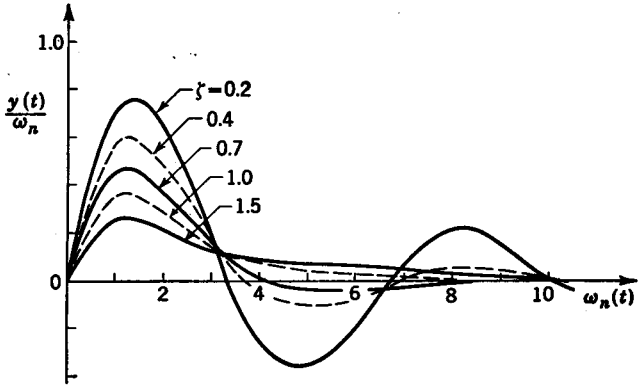


FIG. 6.12. Response of a second-order system to a unit impulse.

upsets the initial state of equilibrium of the system. In time, a stable system will return again to its equilibrium position.

Substitution of $X(s) = 1$ into Eq. (6.2) yields the transformed equation for a unit impulse excitation.

$$Y(s) = \frac{L_m(s) + I(s)}{L_n(s)} \tag{6.49}$$

It is to be noted that the basic form of the response of a system to a unit impulse is determined entirely by the zeros of the characteristic function of the original differential equation. If any zeros lie to the right of the imaginary axis the output increases without bound so that the system is basically unstable. For an unstable system, it would be impossible to achieve any initial equilibrium state, for as soon as the power was turned on, the output would continually increase with time. As mentioned in the preceding section, the characteristic function is an inherent property of the system, and thus the basic form of the response to an input is independent of where the excitation enters the loop.

In Fig. 6.12 is shown the response to a unit impulse of a second-order system whose operational equation is

$$(p^2 + 2\zeta\omega_n p + \omega_n^2)y(t) = \omega_n^2 x(t) \tag{6.50}$$

It is to be seen from Fig. 6.12 that for $\zeta < 0.4$ there is a considerable amount of oscillation before the system again reaches equilibrium operation. Also, for $\zeta > 1.0$ a considerable amount of time is required for the system to return to its initial state. Thus, it seems desirable to have $0.4 < \zeta < 1.0$. This is the result that was obtained from a consideration of the step-function response.

6.8. Stability Criteria. A major difficulty in using the Laplace transform method for determining the transient response of a feedback control system is that it necessitates determining the zeros of the characteristic function. The general form for the characteristic function has previously been shown to be

$$D_{G_1}D_{G_2}D_H + N_{G_1}N_{G_2}N_H \tag{6.51}$$

In determining the block diagram for a system, the terms N_{G_1}, N_{G_2}, N_H, D_{G_1}, D_{G_2}, and D_H are usually obtained in factored form. Because of the plus sign in Eq. (6.51), the zeros of N_{G_1}, N_{G_2}, etc., are not the zeros of this characteristic function. Thus, it becomes necessary to determine the zeros of the general polynomial represented by Eq. (6.51). This presents no difficulty for first- and second-order systems. The zero of a first-order system is immediately obvious, and the two zeros of a quadratic equation are readily solved. However, a third-order system requires determining the three zeros of a cubic, and a fourth-order system necessitates solving a quartic, etc.

Routh's Criterion.[1] Routh's criterion is a method for determining whether or not any of the zeros of the characteristic function are in the right half plane. The application of this criterion is as follows:

First write the characteristic function in the general form

$$b_n s^n + b_{n-1}s^{n-1} + b_{n-2}s^{n-2} + \cdots + b_2 s^2 + b_1 s + b_0 \tag{6.52}$$

Next arrange the coefficients of the characteristic function according to the following schedule

$$\begin{array}{cccc}
b_n & b_{n-2} & b_{n-4} & b_{n-6} \quad \cdots \\
b_{n-1} & b_{n-3} & b_{n-5} & b_{n-7} \quad \cdots \\
c_1 & c_2 & c_3 & c_4 \quad \cdots \\
d_1 & d_2 & d_3 & \cdots \\
\cdots & \cdots & \cdots & \cdots \\
e_1 & e_2 & 0 & \\
f_1 & f_2 & 0 & \\
g_1 & 0 & & \\
h_1 & 0 & &
\end{array} \tag{6.53}$$

[1] E. J. Routh, "Dynamics of a System of Rigid Bodies," 3d ed., The Macmillan Company, New York, 1877.

After first arranging the b coefficients as shown, the row of c terms is evaluated as follows:

$$c_1 = \frac{b_{n-1}b_{n-2} - b_n b_{n-3}}{b_{n-1}}$$
$$c_2 = \frac{b_{n-1}b_{n-4} - b_n b_{n-5}}{b_{n-1}} \tag{6.54}$$
$$c_3 = \frac{b_{n-1}b_{n-6} - b_n b_{n-7}}{b_{n-1}}$$

The arrows in Eq. (6.53) show the schedule of cross multiplication that is used to evaluate c_2. This same general schedule is used for each successive term. By dropping down a row, the same schedule is used for evaluating each d term. That is,

$$d_1 = \frac{c_1 b_{n-3} - b_{n-1}c_2}{c_1}$$
$$d_2 = \frac{c_1 b_{n-5} - b_{n-1}c_3}{c_1} \tag{6.55}$$

This process is continued until only zeros would result in all successive rows. To illustrate, consider the function

$$s^4 + 3s^3 + s^2 + 6s + 2 \tag{6.56}$$

The first two rows of the following array are obtained directly from the coefficients of Eq. (6.56), and the remaining rows are computed as just described:

$$\begin{array}{cccc} 1 & 1 & 2 & 0 \\ 3 & 6 & 0 & \\ -1 & 2 & 0 & \\ 12 & 0 & & \\ 2 & 0 & & \end{array} \tag{6.57}$$

Routh's criterion states that the number of changes of sign of the coefficients in the left-hand column is equal to the number of zeros of the characteristic function that are located to the right of the imaginary axis.

For the preceding example, the signs of the numbers in the left-hand column are seen to go from plus to minus and then back to plus again so that there are two changes of sign. Thus, there are two zeros located to the right of the imaginary axis.

It is not necessary to use Routh's criterion if any of the coefficients of the characteristic function are zero or negative, because when this is so, it can be shown that there is at least one zero located on, or to the right of, the imaginary axis.

To determine how many zeros of the characteristic function lie to the right of some vertical line, a distance σ from the imaginary axis (i.e., the number of zeros that have a real part greater than σ), transform the characteristic function by substituting $s + \sigma$ for s, and apply Routh's criterion as just described. The number of changes of sign in the first column for this new function is equal to the number of zeros which are located to the right of the vertical line through σ.

If a 0 appears in the left-hand column and the sign of the term above the 0 is opposite of that below the 0, then there is but one sign change. However, when the sign of the term above the 0 is the same as that below it, a pair of zeros on the imaginary axis is indicated. For example, consider the function

$$s^3 + 3s^2 + 4s + 12 = (s + 3)(s^2 + 4) \tag{6.58}$$

The array of coefficients is

$$\begin{array}{lll} 1 & 4 & 0 \\ 3 & 12 & 0 \\ 0 & 0 & \\ \underline{0 - 0} & & \\ 0 & & \end{array} \tag{6.59}$$

To evaluate the fourth coefficient in the left-hand column, replace the 0 obtained for the third coefficient by ϵ, which is a very small number. Thus

$$\begin{array}{lll} 1 & 4 & 0 \\ 3 & 12 & 0 \\ \epsilon \approx 0 & 0 & \\ 12 & 0 & \end{array} \tag{6.60}$$

The value of the fourth coefficient is $(12\epsilon - 0)/\epsilon = 12$. The factored form of Eq. (6.58) shows the imaginary zeros indicated by the preceding array. The number of coefficients in the left-hand column is $n + 1$, where n is the order of the characteristic function. For the case in which a zero of the characteristic function is located at the origin, then the coefficient of the $(n + 1)$st row is 0.

Hurwitz Criterion.[1] This criterion determines the conditions which must be satisfied by the coefficients of the characteristic function so that the system is stable (i.e., all the zeros lie in the left half plane). In the Hurwitz criterion, the coefficients of the characteristic function [Eq. (6.52)] are arranged in the following array:

[1] E. A. Guillemin, "The Mathematics of Circuit Analysis," pp. 395–409, John Wiley & Sons, Inc., New York, 1949.

$$\begin{array}{llll|lll} D_1 & b_1 & b_0 & 0 & 0 & 0 & \cdots \\ D_2 & b_3 & b_2 & b_1 & b_0 & 0 & \cdots \\ D_3 & b_5 & b_4 & b_3 & b_2 & b_1 & \cdots \\ D_4 & b_7 & b_6 & b_5 & b_4 & b_3 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{array} \tag{6.61}$$

whence the determinants $D_1, D_2, D_3, \ldots$ are found to be

$$D_1 = b_1 \qquad D_2 = \begin{vmatrix} b_1 & b_0 \\ b_3 & b_2 \end{vmatrix}$$
$$D_3 = \begin{vmatrix} b_1 & b_0 & 0 \\ b_3 & b_2 & b_1 \\ b_5 & b_4 & b_3 \end{vmatrix} \tag{6.62}$$

For stability, it is necessary that all the determinants be positive. That is, $D_1 > 0$, $D_2 > 0$, $D_3 > 0$, etc.

For a general cubic expression

$$b_3 s^3 + b_2 s^2 + b_1 s + b_0 \tag{6.63}$$

the application of the Hurwitz criterion gives

$$\begin{gathered} D_1 = b_1 > 0 \\ D_2 = \begin{vmatrix} b_1 & b_0 \\ b_3 & b_2 \end{vmatrix} = b_1 b_2 - b_0 b_3 > 0 \\ D_3 = b_3 D_2 > 0 \text{ or } b_3 > 0 \end{gathered} \tag{6.64}$$

Thus for stability

$$b_1 b_2 > b_0 b_3 \tag{6.65}$$

With the Hurwitz criterion, similar conditions necessary for the stability of higher-order systems may also be derived. The preceding techniques of Routh and Hurwitz can be extended to yield information regarding the location of the zeros as the coefficients are varied.[1–4]

6.9. Summary. In this chapter it is shown that the transient response of a system is governed primarily by the location of the zeros of $B(s) = L_n(s)D_{X(s)}$. The zeros of $D_{X(s)}$ yield response terms appropriate to the particular excitation to the system. The function $L_n(s)$ is a basic property of the system itself. When all the zeros of $L_n(s)$ are located in

[1] E. Sponder, On the Representation of the Stability Region in Oscillation Problems with the Aid of Hurwitz Determinants, *NACA Tech. Mem.* 1348, August, 1952.

[2] J. F. Koenig, On the Zeros of Polynomials and the Degree of Stability of Linear Systems, *J. Appl. Phys.*, vol. 24, p. 476, 1953.

[3] T. J. Higgins and J. G. Levinthal, Stability Limits for Third Order Servomechanisms, *Trans. AIEE*, vol. 71, pt. 2, p. 459, 1952.

[4] H. A. Hogan, and T. J. Higgins, Stability Boundaries for Fifth Order Servomechanisms, *Proc. Natl. Electronics Conf.*, vol. 11, pp. 1001–1011, 1955.

the left half plane, the system is stable (i.e., for any bounded input the response is also bounded). If any zero of $L_n(s)$ is located in the right half plane, the system is unstable (i.e., the response is always unbounded). The imaginary axis is the border line between stable and unstable systems.

Complex imaginary zeros are undesirable because they yield constant sinusoids. A zero of $L_n(s)$ at the origin is also undesirable because it indicates an integration of the input. Note that an integrator in the feedforward elements (which integrates the error signal) does not yield a zero of $L_n(s)$ at the origin because the zeros of $D_{G(s)}$ are not the zeros of $L_n(s)$. When the characteristic function $L_n(s)$ has a zero at the origin, a constant input (i.e., a step function) yields an unbounded time term $C_1 t$ in the output which is the integral of the input. The output is bounded only if the integral of the input is bounded.

The basic form of the response due to repeated zeros is the same as that for distinct zeros, with the exception that repeated zeros on the imaginary axis yield increasing time terms rather than time terms with constant amplitudes. Because $B(s) = L_n(s)D_{X(s)}$, the zeros of $L_n(s)$ and $D_{X(s)}$ act independently to yield the time response unless one or more of the zeros of $L_n(s)$ and $D_{X(s)}$ are the same. This introduces repeated zeros in $B(s)$, which affects the basic form of the response equation only if the zeros are on the imaginary axis. For example, a repeated complex imaginary zero results in an increasing sinusoid rather than a constant sinusoid. Similarly, a repeated zero at the origin yields an increasing time function rather than a constant. As previously discussed in this section, a step input ($D_{X(s)} = s$) applied to a system in which $L_n(s)$ has a zero at the origin introduces an s^2 term in $B(s)$ that yields an increasing time function $C_1 t$ in the response. To ensure stability, zeros of $L_n(s)$ should be excluded not only from the right half plane but also from the imaginary axis.

Table 6.1 summarizes the type of response terms associated with the zeros of $B(s)$.

TABLE 6.1. LOCATION OF ZEROS AND CORRESPONDING RESPONSE FUNCTIONS

Zeros of $B(s) = L_n(s)D_{X(s)}$	*Type of response*
Left half plane (distinct or repeated)	Decaying exponential and/or decaying sinusoid
Right half plane (distinct or repeated)	Increasing exponential and/or increasing sinusoid
Imaginary axis:	
Distinct	Constant and/or constant sinusoid
Repeated	Increasing time function and/or increasing sinusoid

CHAPTER 7

THE ROOT-LOCUS METHOD

7.1. Significance of Root Loci. The root-locus method was developed by W. R. Evans.[1–4] This method enables one to determine the roots of the characteristic equation (i.e., the zeros of the characteristic function) by knowing the factored form of the feedforward and feedback elements of a control system.

As is discussed in Chap. 6, the transient behavior of a system is governed primarily by the roots of the characteristic equation for the system. Neither the initial conditions nor the particular excitation affects the basic operation of a system.

To illustrate the significance of the root loci, consider the control system represented by Fig. 7.1*a*. The general form for the characteristic equation is

$$D_G D_H + N_G N_H = 0 \tag{7.1}$$

From Fig. 7.1*a* it follows that $N_G = K$, $D_G = s(s + 4)$, $N_H = 1$, and $D_H = 1$. Thus, the characteristic function for this system is

$$s(s + 4) + K = s^2 + 4s + K \tag{7.2}$$

The roots of the characteristic equation depend upon the value of K, which is the *static loop sensitivity*. As is shown in Fig. 7.1*a*, the static loop sensitivity K is the product of all the constant terms in the control loop when the coefficient of each s term is unity.

Because Eq. (7.2) is a quadratic, the roots are $r_{1,2} = a \pm jb$, in which $-2a = 4$ or $a = -2$ and $b = \sqrt{K - a^2} = \sqrt{K - 4}$. Thus

$$\begin{aligned} &\text{For } K > 4 \quad r_{1,2} = -2 \pm j\sqrt{K - 4} \\ &\text{For } K = 4 \quad r_1 = r_2 = -2 \\ &\text{For } K < 4 \quad r_{1,2} = -2 \pm j\sqrt{K - 4} = -2 \mp \sqrt{4 - K} \end{aligned} \tag{7.3}$$

[1] W. R. Evans, Graphical Analysis of Control Systems, *Trans. AIEE*, vol. 67, pp. 547–551, 1948.

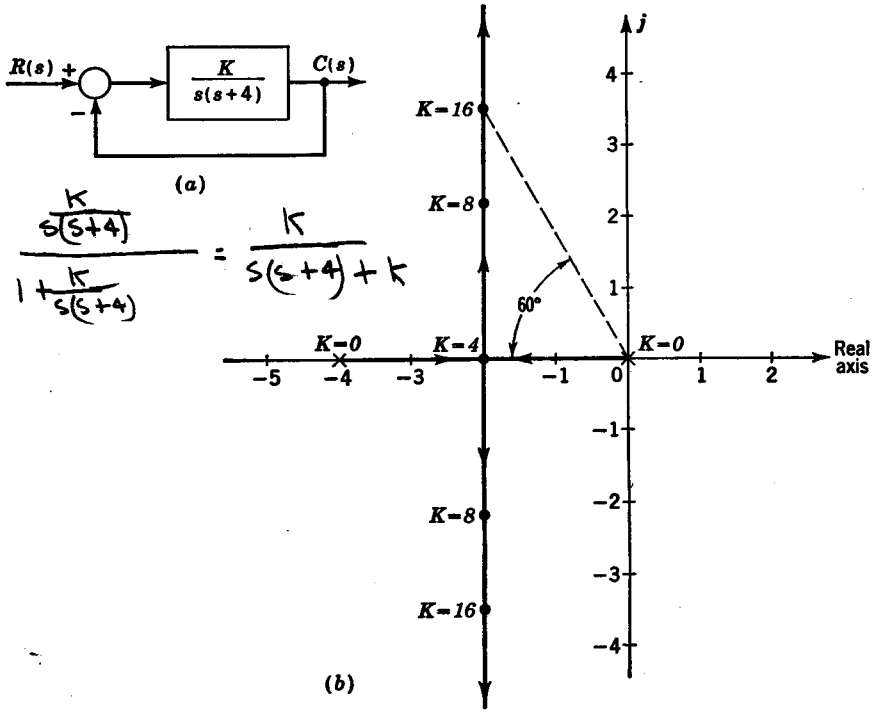
[2] W. R. Evans, "Control-system Dynamics," McGraw-Hill Book Company, Inc., New York, 1954.

[3] J. J. D'azzo, and C. H. Houpis, "Control System Analysis and Synthesis," McGraw-Hill Book Company, Inc., New York, 1960.

[4] C. J. Savant, Jr., "Basic Feedback Control System Design," McGraw-Hill Book Company, Inc., New York, 1958.

In this last case, the roots are real and unequal. The heavy lines in Fig. 7.1*b* are a plot of the roots of this characteristic equation for various values of K. When $K = 0$, the roots are $r_{1,2} = 0, -4$; when $K = 4$, the roots are $r_1 = r_2 = -2$; when $K = 16$, the roots are $r_{1,2} = -2 \pm j\sqrt{12}$; etc.

Such a plot of the roots of the characteristic equation for each value of K as K varies from 0 to ∞ is a root-locus plot. From such a root-locus



$$\frac{C}{R} = \frac{\frac{K}{s(s+4)}}{1 + \frac{K}{s(s+4)}} = \frac{K}{s(s+4) + K}$$

FIG. 7.1. (*a*) Second-order system; (*b*) root-locus plot for $s(s + 4) + K = 0$.

plot, it is an easy matter to select the value of K to yield the desired roots of the characteristic equation. For example, let it be desired to have a damping ratio $\zeta = 0.5$. As discussed in the preceding chapter, $\beta = \cos^{-1}\zeta = \cos^{-1} 0.5 = 60°$. As is shown in Fig. 7.1*b*, the line inclined at the angle $\beta = 60°$ intersects the root-locus plot at a value of $K = 16$. From this plot, the corresponding roots are

$$r_{1,2} = -2 \pm j\sqrt{12}$$

in which case the factored form of the characteristic function is

$$(s - r_1)(s - r_2) = (s + 2 - j2\sqrt{3})(s + 2 + j2\sqrt{3})$$
$$= s^2 + 4s + 16 \quad (7.4)$$

From the preceding expression, it follows that $\omega_n^2 = 16$ or $\omega_n = 4$, and equating $2\zeta\omega_n$ to the coefficient 4 of the s term gives $\zeta = 4/2\omega_n = \frac{1}{2}$, which verifies the preceding results.

A plot of the roots of the characteristic equation as K varies from 0 to ∞ yields very valuable information. For example, consider the feedback control system shown in Fig. 7.2*a*. It is to be noted that

$$N_G = K \qquad D_G = s(s + 4)(s + 6) \qquad N_H = 1 \qquad D_H = 1 \tag{7.5}$$

Thus the characteristic function is

$$s(s + 4)(s + 6) + K = (s - r_1)(s - r_2)(s - r_3) \tag{7.6}$$

The right-hand side of Eq. (7.6) is the factored form of the characteristic function. Because the number of roots $r_1, r_2, \ldots, r_n$ is equal to the order of the equation, the number of loci is also equal to the order of the equation, which in this case is 3. For each value of K, there corresponds a particular value of r_1, r_2, and r_3. Thus, for each value of K from 0 to ∞ one may plot the loci of corresponding values of r_1, r_2, and r_3. The three loci for Eq. (7.6) are drawn in Fig. 7.2*b*. One locus starts at $r_1 = -6$ for $K = 0$ and proceeds out the negative real axis as K increases. Another locus starts at $r_2 = -4$ and goes to the right along the real axis to the point $r_2 = -1.57$ and then leaves the real axis and proceeds out along a 60° asymptote toward infinity. The third locus starts at $r_3 = 0$ and moves along the negative real axis to $r_3 = -1.57$. This locus then leaves the real axis and proceeds toward infinity along a $-60°$ asymptote. In the construction of the loci, the three loci are determined without regard to which is considered the r_1, the r_2, or the r_3 locus. That is, from the similarity of the terms on the right-hand side of Eq. (7.6), it is seen that the particular subscripts 1, 2, and 3 may be used interchangeably.

From Fig. 7.2*b* it is to be noted that, if K is 240, then the three roots of the characteristic equation are

$$r_1 = -10 \qquad r_2 = j4.9 \qquad r_3 = -j4.9$$

Thus, the characteristic function may be written in the following factored form:

$$s(s + 4)(s + 6) + 240 = (s + 10)(s + j4.9)(s - j4.9) \tag{7.7}$$

For $K = 240$, a pair of roots lie on the imaginary axis which would yield a sinusoidal response term of constant amplitude. From the root-locus plot, it is to be noted that, for $K > 240$, a pair of roots will always lie in the right half plane, thus making the system unstable.

For each value of K the corresponding roots of the characteristic equation may be determined directly from the root-locus plot. These roots

in turn govern the transient behavior. From the root-locus plot, the designer may select the value of K such that the system will have a desired transient behavior. For example, let it be desired to have a damping ratio of 0.5. The intersection of the line drawn at the angle $\beta = 60°$ with the loci determines the value of $K = 44$. It is to be noted that the loci are always symmetrical with respect to the real axis because complex roots always occur as conjugate pairs.

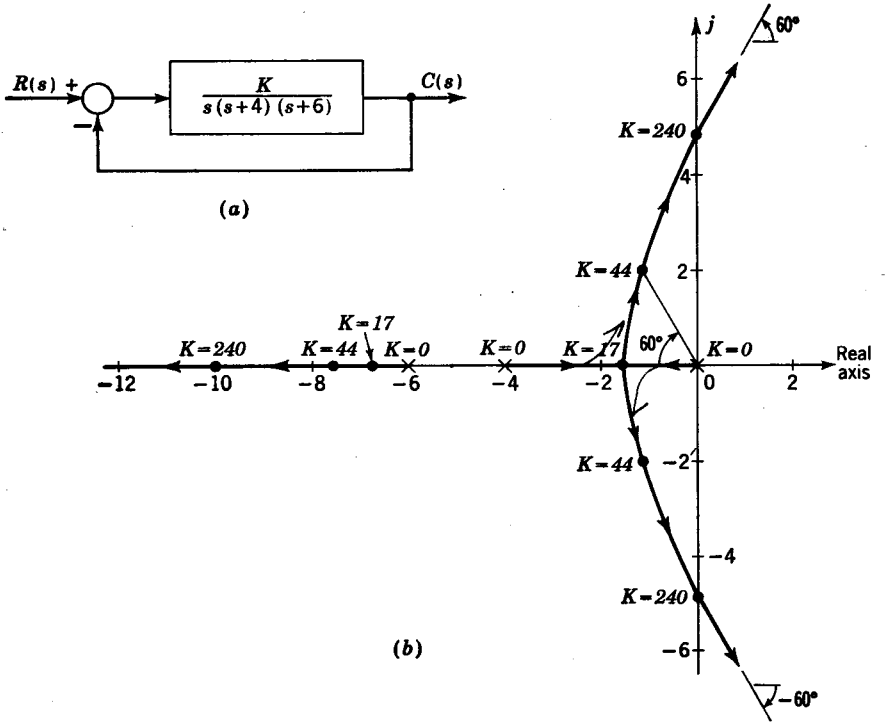


FIG. 7.2. (*a*) Third-order system; (*b*) root-locus plot for $s(s + 4)(s + 6) + K = 0$.

7.2. Construction of Loci. By determining certain points and asymptotes, the loci may be sketched in quite readily. The loci always originate, or begin, at $K = 0$, in which case the value of the roots is obtained directly from the characteristic equation. For example, from Eq. (7.6) when $K = 0$, the roots are 0, -4, and -6.

The zeros $r_1, r_2, \ldots, r_n$ of a characteristic function are those values of s which make the characteristic function equal to zero, i.e.,

$$s(s + 4)(s + 6) + K = (s - r_1)(s - r_2)(s - r_3) = 0 \qquad (7.8)$$

Thus, a root $r_1, r_2, \ldots, r_n$ of a characteristic equation is a value of s

such that

$$s(s + 4)(s + 6) + K = 0$$

or

$$s(s + 4)(s + 6) = -K \tag{7.9}$$

In Fig. 7.3, it is to be noted that the term s in Eq. (7.9) may be represented as a vector drawn from the origin to any point $s = a + jb$ in the s plane. Similarly, the term $s + 4$ is the vector sum of the vector drawn from the -4 point to the origin plus the vector from the origin to the point s. This vector sum is equal to the vector drawn directly from the -4 point to s. Similarly, $s + 6$ is the vector drawn from -6

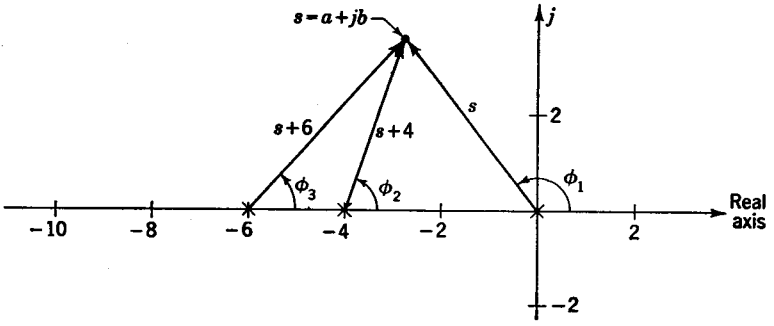


FIG. 7.3. Application of the angle condition to a trial point.

to s. Thus, Eq. (7.9) may be regarded as a vector equation. The length or magnitude of each vector satisfies the magnitude condition

$$|s|\,|s + 4|\,|s + 6| = |-K| = K \tag{7.10}$$

where $|s|$ is the length of the s vector in Fig. 7.3, $|s + 4|$ is the length of the $s + 4$ vector, etc. In addition, the angle of each vector must satisfy the angle condition

$$\measuredangle s + \measuredangle(s + 4) + \measuredangle(s + 6) = \measuredangle(-K) = 180° \pm k360° \tag{7.11}$$

where $k = 0, 1, 2, 3, \ldots$

$\measuredangle s = \phi_1$

$\measuredangle(s + 4) = \phi_2$

$\measuredangle(s + 6) = \phi_3$

The angles ϕ_1, ϕ_2, and ϕ_3 are illustrated in Fig. 7.3.

The term $-K$ is a negative number which may be represented as being at the point $-K$ on the negative real axis so that $\measuredangle(-K)$ is $180° \pm k360°$.

In order that a point in the s plane be a root of the characteristic equation and thus lie on a locus of the roots, it is necessary that the point be located so that Eq. (7.11) is satisfied. Thus, the paths of the

loci are determined from Eq. (7.11). After the root-locus plot has been obtained by application of Eq. (7.11), then the value of K at any point on a locus may be computed by means of Eq. (7.10).

The first place to start investigating the location of loci is along the real axis. For a trial point s on the positive real axis,

$$\measuredangle s = \measuredangle (s + 4) = \measuredangle (s + 6) = 0° \tag{7.12}$$

Thus, $\phi_1 + \phi_2 + \phi_3 = 0°$, in which case the angle condition is not satisfied. Therefore, there is no locus on the positive real axis.

Next, consider a trial point s that lies on the real axis between 0 and -4; then

$$\measuredangle s = 180° \qquad \measuredangle (s + 4) = 0° \qquad \measuredangle (s + 6) = 0° \tag{7.13}$$

Because $\phi_1 + \phi_2 + \phi_3 = 180°$ satisfies the angle condition, there is a locus on the real axis between 0 and -4.

For the region from -4 to -6, it follows that

$$\measuredangle s = 180° \qquad \measuredangle (s + 4) = 180° \qquad \measuredangle (s + 6) = 0° \tag{7.14}$$

Thus, $\phi_1 + \phi_2 + \phi_3 = 360°$ does not satisfy the angle condition. Finally for the region from -6 to $-\infty$ on the negative real axis,

$$\phi_1 + \phi_2 + \phi_3 = 180° + 180° + 180° = 540°$$

so that Eq. (7.11) is again satisfied, which signifies a locus in this region.

The next step in the construction of the loci is to determine the asymptotes as s approaches infinity. For very large values of s, the terms 0, 4, 6 in Eq. (7.9) become negligible compared with the value of s, so that Eq. (7.9) becomes

$$s(s + 4)(s + 6) \approx s(s)(s) = s^3 = -K \tag{7.15}$$

The angle condition is

$$\begin{aligned} \measuredangle s^3 &= 3 \measuredangle s = 180° \pm k360° \\ \text{or} \qquad \measuredangle s &= \frac{180° \pm k360°}{3} = 60° \pm k120° \end{aligned} \tag{7.16}$$

Only three distinct angles, 60°, $-60°$, 180°, are obtained from Eq. (7.16).

In order to locate the asymptotes, it is necessary to know where they intersect the real axis. The point σ_c where the asymptotes cross the real axis is determined from the general equation.

$$\sigma_c = \frac{\Sigma \text{ zeros of } D_G D_H - \Sigma \text{ zeros of } N_G N_H}{(\text{no. of zeros of } D_G D_H) - (\text{no. of zeros of } N_G N_H)} \tag{7.17}$$

For this system, Eq. (7.17) becomes

$$\sigma_c = \frac{[(0) + (-4) + (-6)] - 0}{3 - 0} = \frac{-10}{3} = -3\frac{1}{3} \tag{7.18}$$

The information thus far obtained concerning the location of the roots is shown graphically in Fig. 7.4. By knowing the point at which the locus between 0 and -4 breaks away from the real axis, one can construct a reasonably good sketch of the locus. This breakaway point σ_b

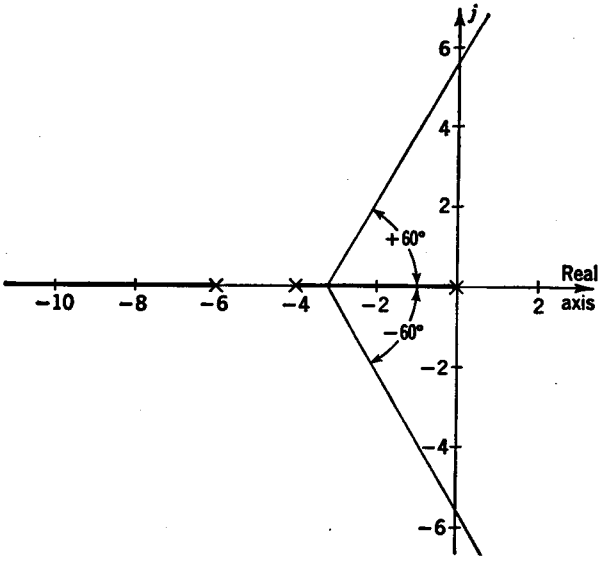


FIG. 7.4. Location of loci on real axis, and the asymptotes.

is determined as follows: Consider in Fig. 7.5 a point s which is a small vertical distance Δ above the real axis. The equation for each angle is

$$\begin{aligned} \phi_1 &= \pi - \tan^{-1}\frac{\Delta}{0 - \sigma_b} \\ \phi_2 &= \tan^{-1}\frac{\Delta}{\sigma_b - (-4)} \\ \phi_3 &= \tan^{-1}\frac{\Delta}{\sigma_b - (-6)} \end{aligned} \tag{7.19}$$

For small values of Δ, $\tan^{-1}(\Delta/-\sigma_b) \approx -\Delta/\sigma_b$, etc. From Fig. 7.5, it is to be seen that, as Δ approaches zero, $\phi_1 + \phi_2 + \phi_3 \approx \pi + 0 + 0 = \pi$. Thus

$$\phi_1 + \phi_2 + \phi_3 = \pi + \frac{\Delta}{\sigma_b} + \frac{\Delta}{4 + \sigma_b} + \frac{\Delta}{6 + \sigma_b} = \pi$$

or

$$\frac{\Delta}{\sigma_b} + \frac{\Delta}{\sigma_b + 4} + \frac{\Delta}{\sigma_b + 6} = 0 \tag{7.20}$$

After eliminating Δ and obtaining a common denominator, the preceding expression becomes

$$3\sigma_b^2 + 20\sigma_b + 24 = 0$$

whence

$$\sigma_b = -1.57 \tag{7.21}$$

The value of K at the point where the locus breaks away from the real axis is the maximum value that K attains between the points 0 and -4. Thus, an alternative method for determining σ_b is first to solve the characteristic equation for K:

$$K = -s(s+4)(s+6) = -(s^3 + 10s^2 + 24s) \tag{7.22}$$

Differentiation to obtain the value of $s = \sigma_b$ at which K is a maximum gives

$$\frac{dK}{ds} = -\frac{d}{ds}(s^3 + 10s^2 + 24s) = -(3s^2 + 20s + 24) = 0 \tag{7.23}$$

Equation (7.23) is the result that was obtained by the preceding geometric evaluation. When dK/ds is a cubic or higher, the point $s = \sigma_b$ at which

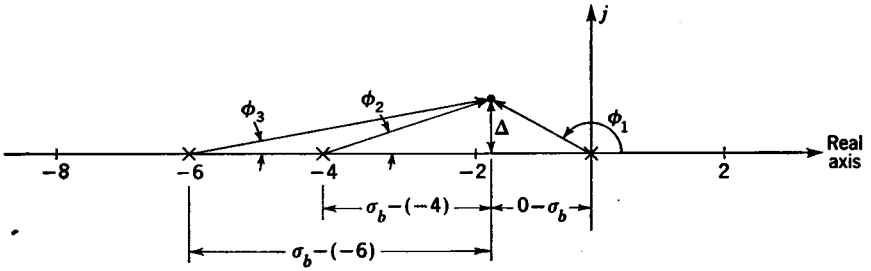


FIG. 7.5. Determination of the breakaway point.

K is a maximum may be obtained by plotting K as a function of s. The occurrence of a breakaway point is ascertained by the fact that *there is always a breakaway point between any two adjacent* ×*'s on the real axis which are connected by a locus.* An × is used to designate a zero of $D_G D_H$.

With the information obtained thus far, the root-locus plot may be sketched in quite accurately, as is shown in Fig. 7.6. For better accuracy in constructing the path of the loci from the breakaway point to the asymptote, it is necessary by trial and error to select trial points in this region and apply the angle condition. Trial points which satisfy the angle condition are points on the locus. It is a good idea to start by choosing all trial points on the same vertical line until the point on the loci is obtained. For a trial point on one side of the loci, the angle condition will yield an angle greater than 180°, and for a trial point on the other side, an angle less than 180° is indicated. This information indicates in what direction a new trial point should be taken. The

use of a spirule, which is a commercially available device, saves much effort in finding the angle of a trial point.

The basis for determining the value of the gain K at various places along the loci is the magnitude condition, that is, Eq. (7.10).

Referring to Fig. 7.6, it is to be noted that the product of the length of each vector $|s + 6|$, $|s + 4|$, and $|s|$ yields directly the value of K at any point on the loci. To illustrate the application of the magnitude

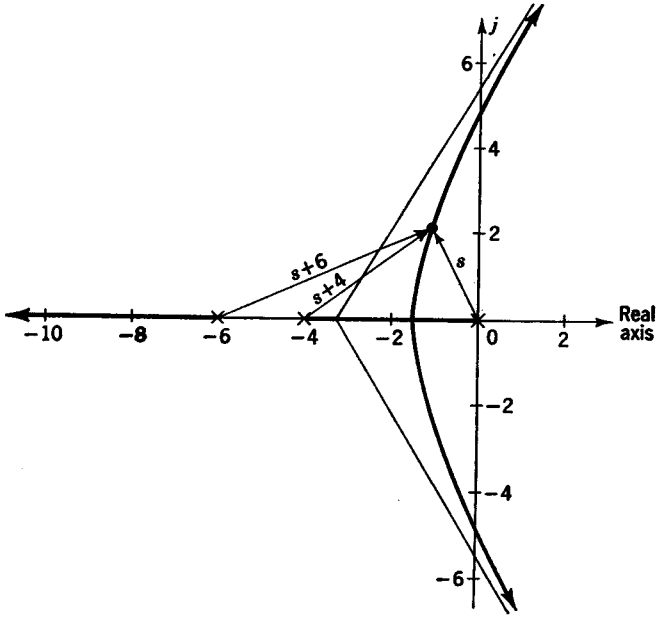


FIG. 7.6. Application of magnitude condition.

condition, consider the point shown on the locus in Fig. 7.6. The value of s at this point is

$$s = -1.2 + j2.1$$

From Eq. (7.10), it follows that

$$\begin{aligned} K &= |s|\,|s + 4|\,|s + 6| \\ &= \sqrt{-1.2^2 + 2.1^2}\,\sqrt{2.8^2 + 2.1^2}\,\sqrt{4.8^2 + 2.1^2} = 44 \end{aligned}$$

In summary, then, the path of the locus plots is obtained entirely from the angle condition [Eq. (7.11)], and the values of K along the loci are determined from the magnitude condition [Eq. (7.10)].

The value of K at which the locus crosses the imaginary axis may be computed by Routh's method if so desired. That is, application of

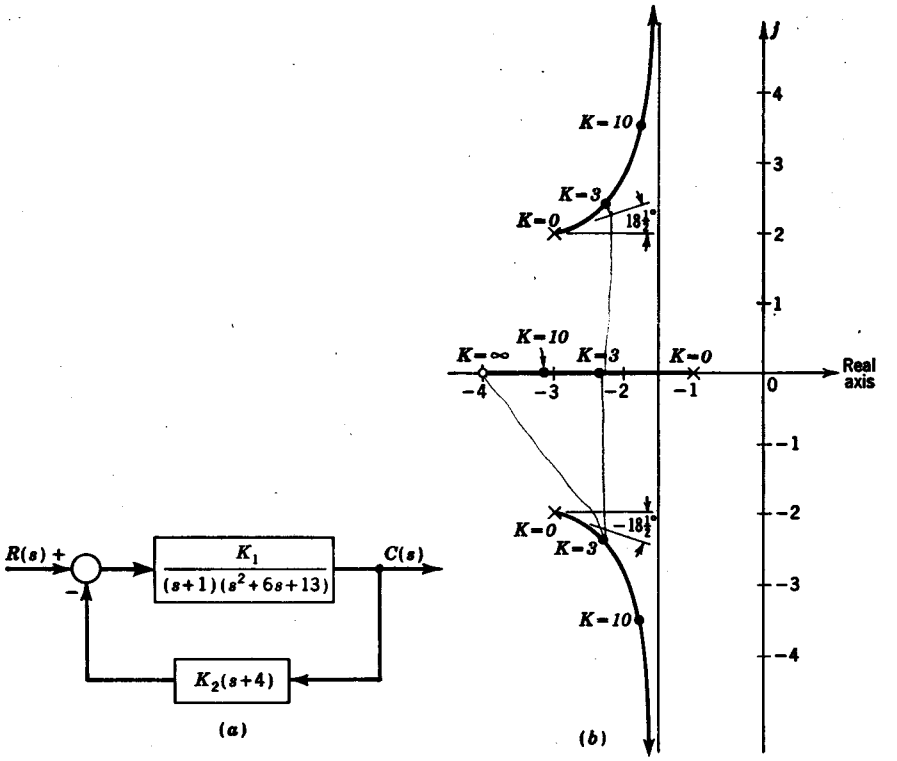


FIG. 7.7. (*a*) Feedback control system; (*b*) root-locus plot for

$$(s + 1)(s^2 + 6s + 13) + K(s + 4) = 0$$

Routh's criterion to the characteristic function $s^3 + 10s^2 + 24s + K$ gives the following array:

$$\begin{array}{lll} 1 & 24 & 0 \\ 10 & K & 0 \\ \dfrac{240 - K}{10} = \epsilon & 0 & \\ \dfrac{K\epsilon}{\epsilon} = K & & \end{array}$$

The value of K which makes the third coefficient vanish is $K = 240$.

Complex Conjugate Terms. The application of the root-locus method is now illustrated for the system shown in Fig. 7.7*a*, in which it is noticed that D_G has complex conjugate zeros. The characteristic equation for this system is

$$(s + 1)(s^2 + 6s + 13) + K(s + 4) = (s - r_1)(s - r_2)(s - r_3) = 0 \tag{7.24}$$

where $K = K_1K_2$. Because the highest power of s in the characteristic equation is 3, there are three loci. Factoring the quadratic term in Eq. (7.24) gives

$$(s + 1)(s + 3 + j2)(s + 3 - j2) + K(s + 4) = 0$$

or

$$\frac{(s + 1)(s + 3 + j2)(s + 3 - j2)}{s + 4} = -K \qquad (7.25)$$

The angle condition is

$$\measuredangle(s + 1) + \measuredangle(s + 3 + j2) + \measuredangle(s + 3 - j2) - \measuredangle(s + 4) = 180° \pm k360° \qquad (7.26)$$

In general, the angle condition is written

$$\measuredangle D_GD_H - \measuredangle N_GN_H = 180° \pm k360° \qquad (7.27)$$

To begin to construct a locus, first plot the zeros of D_GD_H and the zeros of N_GN_H. As shown in Fig. 7.8, the vector drawn from each zero to

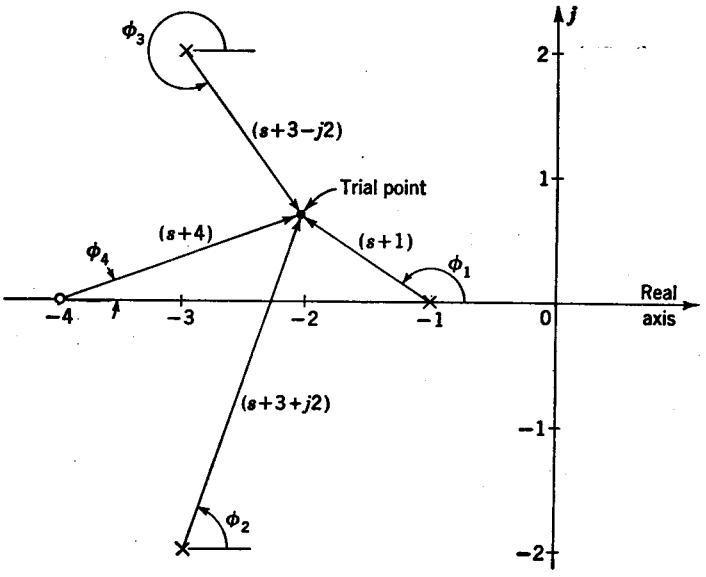


FIG. 7.8. Application of angle condition to a trial point.

any point s represents the corresponding term in the characteristic equation. The location of the zeros of D_GD_H is indicated by crosses (×), and the zeros of N_GN_H are always represented by small circles (○).

In order that a trial point s be a root of the characteristic equation (i.e., be a point on a locus of the roots), from the angle condition it follows that

$$\phi_1 + \phi_2 + \phi_3 - \phi_4 = 180° \pm k360° \qquad (7.28)$$

where

$$\phi_1 = \measuredangle(s + 1)$$
$$\phi_2 = \measuredangle(s + 3 + j2)$$
$$\phi_3 = \measuredangle(s + 3 - j2)$$
$$\phi_4 = \measuredangle(s + 4)$$

As usual, the first place to investigate the possible location of loci is along the real axis. As is illustrated in Fig. 7.9, the value of $\phi_2 + \phi_3$ will always be 360° when s is on the real axis. For any trial point s on the real axis, the sum of the angular contribution $\phi_2 + \phi_3$ for any complex conjugate ×'s (zeros of $D_G D_H$) or O's (zeros of $N_G N_H$) will always

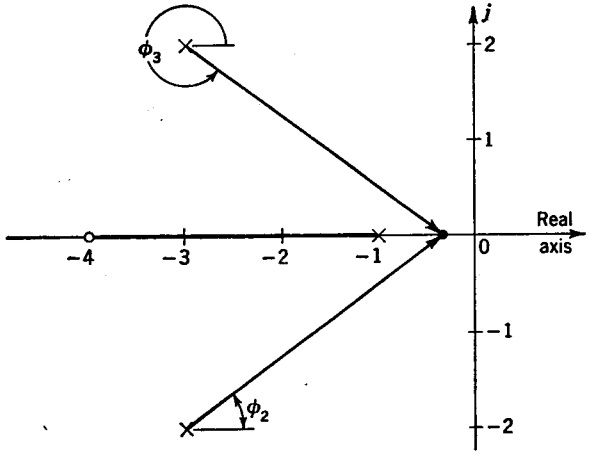


FIG. 7.9. Location of loci on the real axis.

be 360°. Thus, *the location of the loci along the real axis is determined only by the ×'s and O's which lie on the real axis.* When s is to the right of the -1 point, $\phi_1 = \phi_4 = 0°$. Thus, there is no locus located to the right of the -1 point. When s is between -1 and -4,

$$\phi_1 - \phi_4 = 180° - 0° = 180°$$

so that a locus is located in this region. Similarly, it can be established that there is no locus to the left of the -4 point.

The next step in the construction of the loci is to determine the asymptotes for very large values of s. In Eq. (7.26), the terms 1, $3 + j2$, $3 - j2$, and 4 become negligible compared with the value of s, so that Eq. (7.26) becomes

$$\measuredangle s + \measuredangle s + \measuredangle s - \measuredangle s = 180° \pm k360° \qquad k = 0, 1, 2, \ldots \tag{7.29}$$

or

$$3\measuredangle s - \measuredangle s = 2\measuredangle s = 180° \pm k360°$$

$$\measuredangle s = \frac{180° \pm k360°}{2} = 90° \pm k180° = \pm 90° \tag{7.30}$$

In general, it can be shown that the angle of the asymptotes is given

by the following equation,

$$\angle s = \frac{180° \pm k360°}{n - m} \qquad k = 0, 1, 2, \ldots \tag{7.31}$$

where n = highest power of s in $D_G D_H$ = no. of ×'s
m = highest power of s in $N_G N_H$ = no. of ○'s

The number of distinct asymptotes is equal to $n - m$. Although it would appear from Eq. (7.31) that there are more asymptotes, the angles

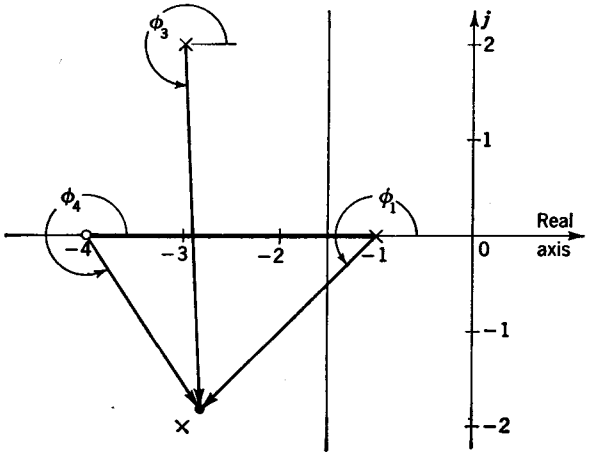


FIG. 7.10. Determination of the angle of departure.

repeat for other values of k after $n - m$ distinct angles have been determined.

The intersection of these asymptotes with the real axis is found as follows by application of Eq. (7.17), which is the general expression for obtaining the point of crossing σ_c:

$$\sigma_c = \frac{[(-1) + (-3 - j2) + (-3 + j2)] - (-4)}{3 - 1} = -1.5 \tag{7.32}$$

The information thus far obtained is sketched in Fig. 7.10. The next step is to determine the angle of departure of the loci from the complex conjugate zeros of $D_G D_H$. To do this, a trial point s is taken which is located very close to the point $(-3 - j2)$. From Fig. 7.10, the following values for ϕ_1, ϕ_3, and ϕ_4 are obtained:

$$\begin{aligned}
\phi_1 &= \tan^{-1} \frac{(-2) - 0}{(-3) - (-1)} = \tan^{-1} \frac{-2}{-2} = 225° \\
\phi_3 &= \tan^{-1} \frac{(-2) - 2}{(-3) - (-3)} = \tan^{-1} \frac{-4}{0} = 270° \\
\phi_4 &= \tan^{-1} \frac{(-2) - 0}{(-3) - (-4)} = \tan^{-1} \frac{-2}{1} = 296.5°
\end{aligned} \tag{7.33}$$

The angle ϕ_2 of departure of the loci from the point $(-3 - j2)$ is determined from the angle condition as follows:

$$\phi_1 + \phi_2 + \phi_3 - \phi_4 = 225° + \phi_2 + 270° - 296.5° = 180° \pm k360° \tag{7.34}$$

Thus $$\phi_2 + 198.5° = 180° \pm k360°$$

or $$\phi_2 = -18.5° \pm k360° = -18.5° \tag{7.35}$$

Because the loci are symmetrical about the real axis, the angle of departure from the other conjugate zero will be $+18.5°$. The complete root-locus plot may now be sketched in as shown in Fig. 7.7*b*. The value of K at any place along the loci is determined by application of the magnitude condition.

From Eq. (7.25), it is to be noted that for $s = -1$, $-3 - j2$, or $-3 + j2$, the value of K is 0. These values of s are the zeros of $D_G D_H$. Thus, in general, one locus begins at each zero of $D_G D_H$ (that is, each ×). Similarly, for $s = -4$, $K = \infty$. The value $s = -4$ is a zero of $N_G N_H$, and thus one locus will terminate at each zero of $N_G N_H$ (that is, each ○). If n is the number of ×'s and m the number of ○'s, then the remaining $n - m$ loci terminate along asymptotes at infinity.

7.3. General Procedure for Determining Root Loci. In the preceding section, it was shown that the loci could be sketched in quite accurately by knowing a few critical points, asymptotes, and the angle of departure from complex zeros, etc. The general procedure for constructing root loci is summarized as follows.

1. *Origin.* When K is zero, the zeros of the characteristic function are the zeros of $D_G D_H$. Thus, *each locus originates at a zero of* $D_G D_H$ (*designated by* ×'*s*) *and the number of individual loci is equal to n, the number of zeros of* $D_G D_H$.

2. *Terminus.* As K becomes very large, m loci (m is the number of zeros of $N_G N_H$) will approach the m zeros of $N_G N_H$. That is, *one locus will terminate at each of the m zeros of* $N_G N_H$. *The remaining* $n - m$ *loci will approach infinity along asymptotes.*

3. *Asymptotes.* The angles at which each of the $n - m$ loci approaches infinity is determined from Eq. (7.31); i.e.,

$$\measuredangle s = \frac{180° \pm k360°}{n - m} \tag{7.31}$$

The point σ_c at which the asymptotes intersect or cross the real axis is computed by Eq. (7.17); i.e.,

$$\sigma_c = \frac{\Sigma \text{ (zeros of } D_G D_H) - \Sigma \text{ (zeros of } N_G N_H)}{n - m} \tag{7.17}$$

4. *Loci on Real Axis.* Complex zeros of $D_G D_H$ or $N_G N_H$ have no effect on the location of loci on the real axis. The place at which the loci are located along the real axis is determined by considering only zeros of $D_G D_H$ and $N_G N_H$ which lie on the real axis. As is illustrated in Fig. 7.11, there is never a locus to the right of the first ○ or × on the real axis, but there is always a locus to the left of the first ○ or ×, there is never a locus to the left of the second ○ or ×, there is always a locus to the left of the third ○ or ×, never left of fourth, always left of fifth, and so on, alternating.

5. *Breakaway from Real Axis.* The point σ_b at which the locus breaks away from the real axis is obtained by applying the angle condition to

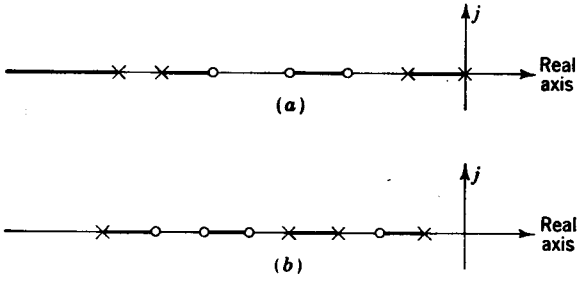


FIG. 7.11. Loci on real axis.

an arbitrarily chosen point which is a small vertical distance Δ from the real axis and then solving this resultant equation for the only unknown term σ_b. This may also be determined by finding the real value $s = \sigma_b$ at which K is a maximum (that is, $dK/ds = 0$).

6. *Angle of Departure.* The angle of departure of a locus from a complex zero of $D_G D_H$ is obtained by selecting a trial point very close to this zero and applying the angle condition.

7. *Angle of Arrival.* The angle at which a locus will terminate at a complex zero of $N_G N_H$ is determined by taking a trial point which is close to this zero and applying the angle condition. This process is similar to that used to obtain the angle of departure from a zero of $D_G D_H$.

The following technique for evaluating a break-in point completes the preceding list of rules for constructing root-locus plots. A break-in point is similar to a breakaway point with the exception that a pair of loci comes into the real axis rather than leaving it. A break-in point is illustrated in Fig. 7.12. This is the system shown in Fig. 7.1, with the addition of the $s + 6$ term in N_G. The characteristic equation for the system of Fig. 7.12 is

$$\frac{s(s+4)}{s+6} = -K \tag{7.36}$$

Application of rule 4 to determine the loci on the real axis yields the location of loci along the real axis as shown in Fig. 7.12.

The breakaway and break-in points are evaluated as follows:

$$\frac{dK}{ds} = -\frac{d}{ds}\frac{s(s+4)}{s+6} = -\left[\frac{(s+6)(2s+4) - s(s+4)}{(s+6)^2}\right] = 0 \quad (7.37)$$

Thus, $$s^2 + 12s + 24 = 0$$

whence $$s = -6 \pm 2\sqrt{3} = -2.54, -9.46 \quad (7.38)$$

Further investigation shows that the value -2.54 makes K a maximum in the region $-4 < s < 0$; thus this is the breakaway point. The value

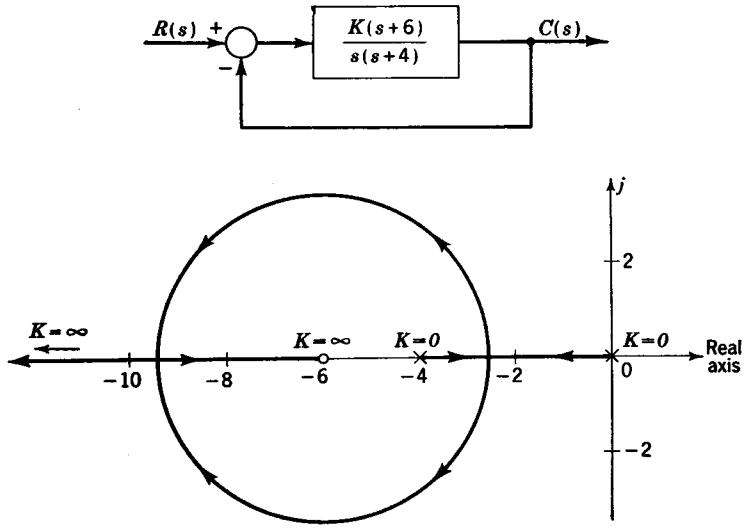


FIG. 7.12. Root-locus plot for $s(s + 4) + K(s + 6) = 0$.

-9.46 makes K a minimum in the range $-\infty < s < -6$; thus this is a break-in point as illustrated in Fig. 7.12. These breakaway and break-in points could also be computed by the geometrical technique previously discussed of assuming a trial point s which is a small distance Δ above the real axis.

The occurrence of a breakaway or break-in point can be recognized from a consideration of the ×'s and ○'s which lie on the real axis. Every locus begins at an × and terminates at a ○ or along an asymptote at ∞. Thus, *there must be a breakaway point between any two adjacent ×'s on the real axis which are connected by a locus. Similarly, a break-in point is required if a ○ on the real axis is not connected to an adjacent × on the real axis by a locus.* Thus, if the locus is not located entirely on the

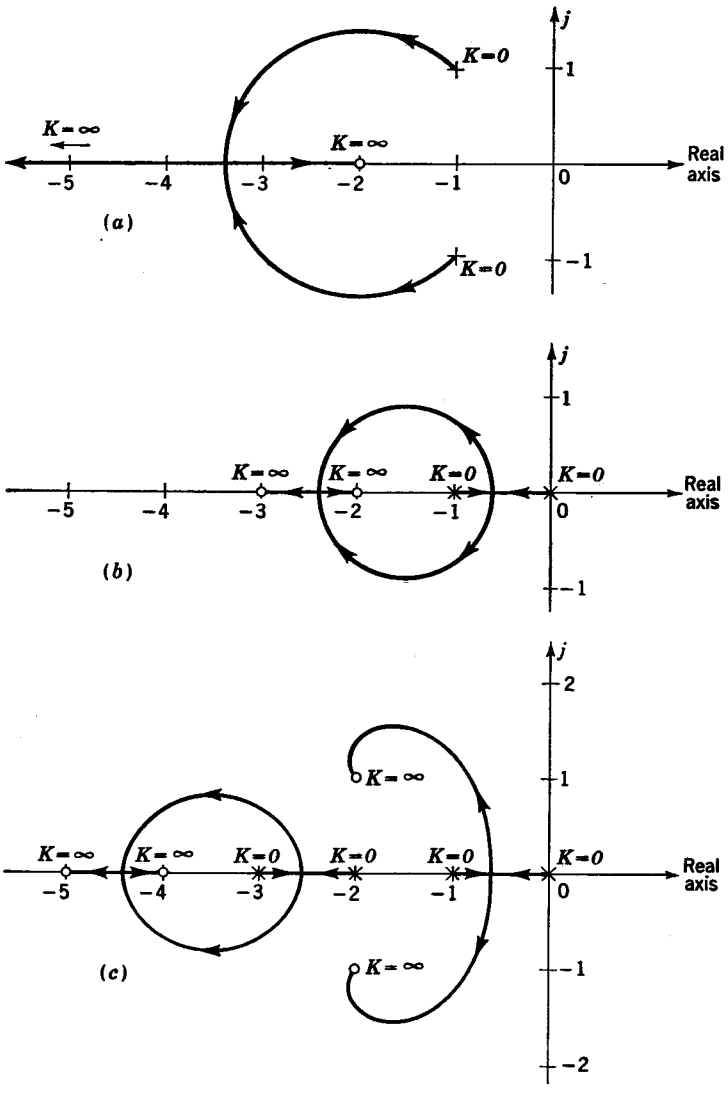


FIG. 7.13. Root-locus plots.

real axis between an adjacent ○ and ×, it is necessary that it come into the real axis from elsewhere. The preceding rules may be verified for the root-locus plots shown in Fig. 7.13.

A comparison of the root-locus plot given in Fig. 7.12 with that of Fig. 7.1 shows that a zero of $N_G N_H$ (that is, a ○) has the effect of pulling the locus to the left. For large values of K the system of Fig. 7.1 would have a small damping ratio and would thus exhibit a very oscillatory

type of response. For the system of Fig. 7.12, large values of K would result in negative real roots and thus a damped exponential type of response. In general, the addition of a $\circ$ in the left half plane tends to make the system more stable. On the other hand, adding an $\times$ in the left half plane tends to push the root-locus plot to the right and thus decrease stability.

CHAPTER 8

ANALOG COMPUTERS

8.1. Introduction. The use of computers has played a major role in the recent advances in the design of automatic control systems. These computers may be divided into two types, analog computers[1–3] and digital computers.[4,5]

An analog computer is one in which the equation describing the operation of the computer is analogous to that for the actual system. The most commonly used analog computer is the electronic analog computer, in which voltages at various places within the computer are proportional to the variable terms in the actual system. As is shown in this chapter, the operation of a control system can be simulated by the use of an analog computer.

Basically a digital computer can only add and subtract. Thus it is necessary to reduce all problems to rather elementary mathematical manipulations. This process is called programming. The programming of a problem for solution on a digital computer makes extensive use of the methods of "numerical analysis" to convert the problem to the numerical operations which the computer can perform. It may require weeks or even months to program a problem for a computer, which in turn completes the solution in a few minutes or seconds. A digital computer has been referred to as an "energetic moron" in that it is capable of performing thousands of simple additions and subtractions in a second. A digital computer must store information for use in later computations. This is usually done by means of a magnetic drum, which acts as a memory device.

[1] G. A. Korn and T. M. Korn, "Electronic Analog Computers," 2d ed., McGraw-Hill Book Company, Inc., New York, 1956.

[2] A. S. Jackson, "Analog Computation," McGraw-Hill Book Company, Inc., New York, 1960.

[3] G. W. Smith and R. C. Wood, "Principles of Analog Computation," McGraw-Hill Book Company, Inc., 1959.

[4] N. R. Scott, "Analog and Digital Computer Technology," McGraw-Hill Book Company, Inc., 1960.

[5] R. S. Ledley, "Digital Computer and Control Engineering," McGraw-Hill Book Company, Inc., 1960.

The input to a digital computer consists of numbers and instructions for the operation of the machine on these numbers. These numbers and instructions may be fed into the machine by various methods such as punched cards, tape, typewriter, etc.

Because of the ability of both digital and analog computers to solve complicated mathematical equations almost instantaneously, they are often incorporated as part of control systems to compute desired information. This information may then be used immediately to improve the control of the particular system. For example, in an inertial guidance system, the output of three mutually perpendicular accelerometers is fed into a computer, which in turn calculates the position of the vehicle. Thus, the output of this computer is the actual position of the vehicle, which is compared with the desired position to yield an error signal for actuating the steering mechanism.

8.2. Analog Computer. The electronic analog computer is a very powerful tool for investigating the performance of control systems. For more complex systems, the advantages of the analog computer become more apparent. Analog computers are used for many purposes besides that of investigating linear and nonlinear control systems. For example, they are used to solve nonlinear differential equations, partial differential equations, systems of differential or partial differential equations, matrix and eigenvalue problems, operational research problems, etc. New applications and uses for this versatile computing device are continually being discovered.

This chapter is primarily concerned with the use of these computers for simulating control systems. For this purpose, the equation describing the operation of the analog computer is analogous to that which represents the actual physical system. The variable quantities of the actual system such as the output, input, error, etc., are represented by voltages at various places within the analog computer. Permanent records of these voltages may be obtained by using recording equipment. By using potentiometers or variable capacitors to vary the resistance or capacitance at various places within the computer, the effect upon system performance of changing the corresponding parameters in the actual system (e.g., gain, time constants, damping ratio, etc.) may be determined immediately.

To solve any linear differential equation with constant coefficients, it is necessary only to make use of the processes of integration, summation, and multiplication by a constant. This is illustrated by the block diagram of Fig. 8.1 for the equation

$$M\ddot{y} + C\dot{y} + Ky = x(t) \tag{8.1}$$

To set up the block-diagram representation for a differential equation,

first solve for the highest-order differential appearing in the original equation. The highest-order term appearing in Eq. (8.1) is $M\ddot{y}$. Solving Eq. (8.1) for $M\ddot{y}$ yields

$$M\ddot{y} = x(t) - C\dot{y} - Ky \tag{8.2}$$

Successive integration of the highest-order differential and multiplication by appropriate constants yields the other lower-order terms. That is,

$$\begin{aligned} \frac{C/M}{p}(M\ddot{y}) &= C\dot{y} \\ \frac{K/C}{p}(C\dot{y}) &= Ky \end{aligned} \tag{8.3}$$

Each term on the right-hand side of Eq. (8.2) goes into the summer of Fig. 8.1, so that the output of the summer is proportional to the acceleration. Successive integration of this acceleration yields the velocity and displacement.

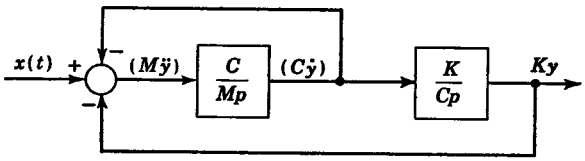


FIG. 8.1. Block diagram for $M\ddot{y} + C\dot{y} + Ky = x(t)$.

The heart of the electronic analog computer is the operational amplifier, which is a very-high-gain d-c amplifier. This device may be used as an integrator, summer, or multiplier. The particular mathematical operation depends upon the particular network of resistors and capacitors which are placed around it.

8.3. Computer Operations. In Fig. 8.2 is shown the schematic representation of an operational amplifier. The input voltage is e_i, the output is e_2, and the amplification is $-A$. Thus

$$e_2 = -Ae_i \tag{8.4}$$

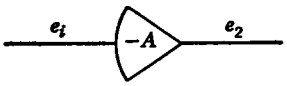


FIG. 8.2. Schematic representation for an operational amplifier.

The reason for the minus sign is that the amplifier reverses the phase of the input. For most operational amplifiers, the value of A is very large. Values of A may range from 50,000 to 100×10^6.

Multiplication by a Constant. By feeding the input voltage through a resistor R_1 and by putting a resistor R_2 in parallel with the amplifier as shown in Fig. 8.3, a circuit for multiplication by a constant is obtained. Because the input grid of the amplifier draws little or no current (a

typical value is 10^{-9} amp),

$$i_1 \approx i_2 \tag{8.5}$$

The values of i_1 and i_2 are

$$i_1 = \frac{e_1 - e_i}{R_1} \qquad i_2 = \frac{e_i - e_2}{R_2} \tag{8.6}$$

Equating i_1 and i_2 yields

$$\frac{e_1 - e_i}{R_1} = \frac{e_i - e_2}{R_2} \tag{8.7}$$

Usually e_2 is less than 100 volts; so for very large values of A, it follows from Eq. (8.4) that $e_i = -e_2/A \approx 0$. The substitution of $e_i \approx 0$ into Eq. (8.7) gives

$$e_2 = -\frac{R_2}{R_1} e_1 \tag{8.8}$$

The value of the multiplication constant is $-R_2/R_1$. When $R_2 = R_1$, the constant is -1, which means simply that the phase of the input signal has been inverted.

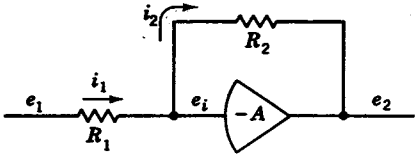


FIG. 8.3. Operational amplifier circuit for multiplying by a constant.

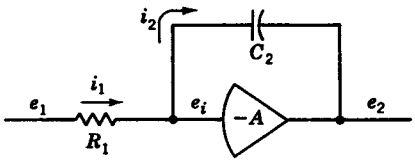


FIG. 8.4. Operational amplifier circuit for integrating.

Integration. By replacing the resistor R_2 of Fig. 8.3 by a capacitor as shown in Fig. 8.4, then an integrating circuit is obtained. The current i_2 flowing through this capacitor is

$$i_2 = C_2 p(e_i - e_2) \tag{8.9}$$

By equating i_1 and i_2 as before, it follows that

$$\frac{e_1 - e_i}{R_1} = C_2 p(e_i - e_2) \tag{8.10}$$

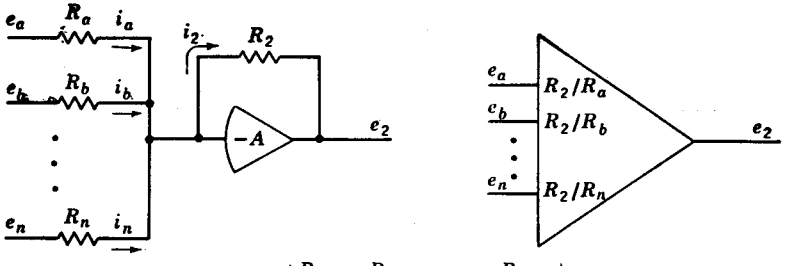
Because $e_i \approx 0$, the preceding expression reduces to

$$e_2 = \frac{-e_1}{R_1 C_2 p} = \frac{-1}{R_1 C_2} \int_0^t e_1 \, dt + e_2(0) \tag{8.11}$$

where $e_2(0)$ is the value of e_2 when $t = 0$. In addition to integration, the circuit of Fig. 8.4 also multiplies by the constant $-1/R_1C_2$.

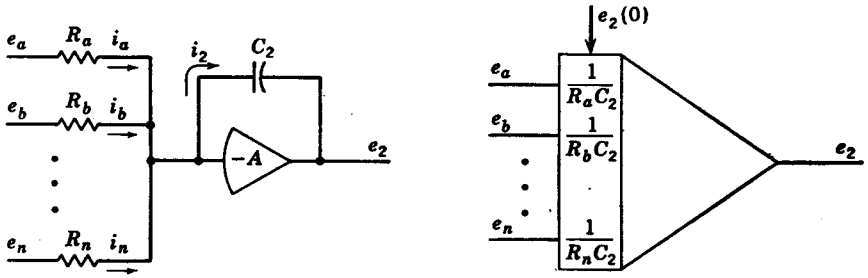
Summation. The effect of summation is obtained by placing the desired quantities to be added in parallel at the input to the computer

circuit. The output will then be the summation of the effect due to each individual input. A general summing circuit is shown in Fig. 8.5*a*, and a circuit to integrate more than one quantity is shown in Fig. 8.5*b*. A simplified notation for these various computer elements is also shown in Fig. 8.5*a* and *b*. This schematic notation saves much effort in drawing the computer diagram for a circuit.



$$e_2 = -\left(\frac{R_2}{R_a}e_a + \frac{R_2}{R_b}e_b + \cdots + \frac{R_2}{R_n}e_n\right)$$

(*a*)



$$e_2 = -\left(\frac{1}{R_aC_2}\int_0^t e_a\,dt + \frac{1}{R_bC_2}\int_0^t e_b\,dt + \cdots + \frac{1}{R_nC_2}\int_0^t e_n\,dt\right) + e_2(0)$$

(*b*)

FIG. 8.5. General summing circuits.

8.4. Computer Diagrams. Let it be desired to determine the computer diagram for solving the following equation:

$$\ddot{y} + 8\dot{y} + 2y = x(t) \tag{8.12}$$

The initial conditions are $y(0) = 2$ ft and $\dot{y}(0) = 5$ ft/sec, and the expected maximum values are $y_m = 5$ ft, $\dot{y}_m = 10$ ft/sec, $\ddot{y}_m = 20$ ft/sec^2, and $x(t)_m = 20$ lb. Solving Eq. (8.12) for the highest-order derivative yields

$$\ddot{y} = x(t) - 8\dot{y} - 2y \tag{8.13}$$

The general computer diagram for Eq. (8.13) is shown in Fig. 8.6*a*.

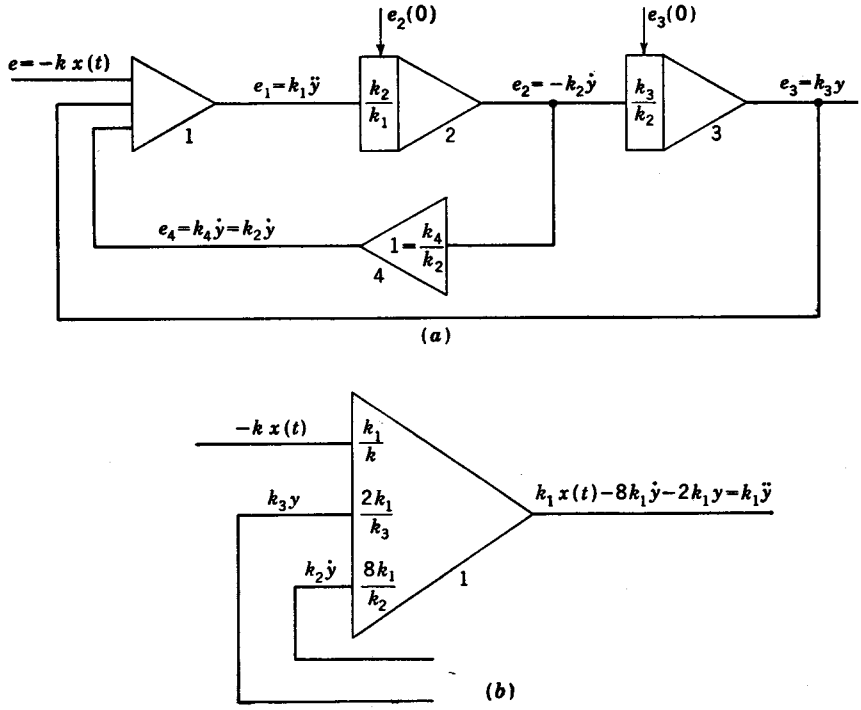


FIG. 8.6. Computer diagram for $\ddot{y} = x(t) - 8\dot{y} - 2y$.

In Fig. 8.6*b*, it is to be noted that the first amplifier, a summer, solves Eq. (8.13). Other amplifiers merely integrate to yield desired terms.

The output voltage e_1 of the first amplifier is proportional to the acceleration $\ddot{y}$ so that $e_1 = k_1\ddot{y}$. The voltage relationship for the second amplifier, which is an integrator, is

$$e_2 = \frac{-1}{(R_1C_2)_2}\int_0^t e_1\,dt + e_2(0) \tag{8.14}$$

where $(R_1C_2)_2$ is the product of the resistance and capacitance for amplifier 2. Substitution of $e_1 = k_1\ddot{y}$ into the preceding expression gives

$$e_2 = \frac{-k_1}{(R_1C_2)_2}\int_0^t \ddot{y}\,dt + e_2(0) \tag{8.15}$$

Because the integral of acceleration is velocity, it follows that

$$\dot{y} = \int_0^t \ddot{y}\,dt + \dot{y}(0) \tag{8.16}$$

Multiplication of the preceding expression by $-k_2$ yields

$$-k_2\dot{y} = -k_2\int_0^t \ddot{y}\,dt - k_2\dot{y}(0) \tag{8.17}$$

Equation (8.15) is the voltage relationship for the amplifier, and Eq. (8.17) is the corresponding physical relationship for the system. It is desired to have e_2 equal to $-k_2\dot{y}$ so that the voltage is proportional to the velocity. This is accomplished by equating the right-hand sides of Eqs. (8.15) and (8.17); thus

$$\frac{-k_1}{(R_1C_2)_2}\int_0^t \ddot{y}\,dt + e_2(0) = -k_2\int_0^t \ddot{y}\,dt - k_2\dot{y}(0)$$

Because corresponding terms in the preceding expression are equal, the following relationships must hold:

$$\frac{1}{(R_1C_2)_2} = \frac{k_2}{k_1} \quad \text{and} \quad e_2(0) = -k_2\dot{y}(0) \tag{8.18}$$

Each integrator may be initially biased by a d-c voltage to give the effect of initial conditions, as is represented diagrammatically at the top of each integrator in Fig. 8.6*a*. In practice, the initial condition is obtained by placing a source of constant potential such as a battery of potential $e_2(0)$ in parallel with the capacitor, as is shown in Fig. 8.7. For $t < 0$, the switch is up so that the battery can charge the capacitor to the desired initial value. At $t = 0$, the switch moves down to disconnect the battery from the circuit so that the integrator circuit functions as previously discussed.

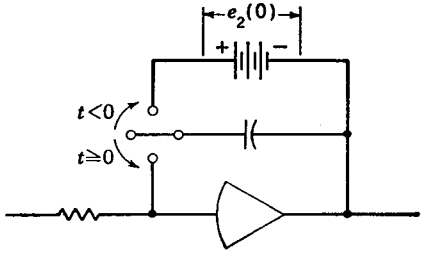


FIG. 8.7. Circuit for obtaining initial conditions.

The voltage equation for the third amplifier circuit is

$$e_3 = \frac{-1}{(R_1C_2)_3}\int_0^t e_2\,dt + e_3(0) = \frac{k_2}{(R_1C_2)_3}\int_0^t \dot{y}\,dt + e_3(0) \tag{8.19}$$

Because the integral of velocity is displacement,

$$y = \int_0^t \dot{y}\,dt + y(0) \tag{8.20}$$

Multiplication of the preceding expression by k_3 gives

$$k_3y = k_3\int_0^t \dot{y}\,dt + k_3y(0) \tag{8.21}$$

In order to have $e_3 = k_3y$, by comparison of Eqs. (8.19) and (8.21) it follows that

$$\frac{1}{(R_1C_2)_3} = \frac{k_3}{k_2} \quad \text{and} \quad e_3(0) = k_3y(0) \tag{8.22}$$

Similarly, the voltage equation for amplifier 4, which multiplies by a constant, is

$$e_4 = -\frac{R_2}{R_1} e_2 = \frac{R_2}{R_1} k_2 \dot{y} \tag{8.23}$$

The desired relationship for amplifier 4 is

$$e_4 = k_4 \dot{y} \tag{8.24}$$

Thus it follows from Eqs. (8.23) and (8.24) that

$$\frac{R_2}{R_1} = \frac{k_4}{k_2} \tag{8.25}$$

There is never any need to apply an initial voltage $e(0)$ to account for initial conditions when an amplifier is used to multiply by a constant. Only integration requires initial conditions.

Scale Factors. The quantities k_1, k_2, k_3, and k_4 are scale factors which relate voltages in the computer to corresponding physical quantities. In general, it can be shown that the scale factor for the output of an amplifier is equal to the scale factor of the input quantity times the value of $-1/R_1C_2$ for an integrating amplifier or the value $-R_2/R_1$ for a summing amplifier. This is apparent from Fig. 8.6*a*, in which

$$-k_2 = \frac{-1}{(R_1C_2)_2} k_1 \qquad k_3 = \frac{-1}{(R_1C_2)_3} (-k_2) \qquad k_4 = \frac{-R_2}{R_1} (-k_2) \tag{8.26}$$

An amplifier always reverses the sign from the input to the output. Thus, the sign of the scale factor changes in going from amplifier to amplifier, as shown in Fig. 8.6*a*.

$$e_1 = k_1 \ddot{y} \qquad e_2 = -k_2 \dot{y} \qquad e_3 = k_3 y \qquad e_4 = k_4 \dot{y} \tag{8.27}$$

The first amplifier is the one which actually solves the differential equation. To obtain $e_1 = k_1 \ddot{y}$, Eq. (8.13) is multiplied by k_1.

$$k_1 \ddot{y} = k_1 x(t) - 8k_1 \dot{y} - 2k_1 y \tag{8.28}$$

Because of the change of sign as a quantity goes through an amplifier, the negative of each term on the right-hand side of Eq. (8.28) must be fed into the first amplifier to obtain $k_1 \ddot{y}$, as is shown in Fig. 8.6*b*. That is, the input to the first amplifier must be

$$-k_1 x(t) + 8k_1 \dot{y} + 2k_1 y \tag{8.29}$$

Because $k_3 y$ is available as the output of the third amplifier, then multiplication by $2(k_1/k_3)$ yields one of the desired inputs, $2k_1 y$. In order to obtain the correct sign for the second term in Eq. (8.29), it is necessary to use amplifier 4 to change $-k_2 \dot{y}$ to $k_2 \dot{y}$. Multiplication of $k_2 \dot{y}$ by

$8(k_1/k_2)$ yields $8k_1\dot{y}$. The remaining term $-k_1x(t)$ is obtained by multiplying the input signal $-kx(t)$ by k_1/k. Each of these constant multiplication factors is shown in Fig. 8.6*b* at the input to the first amplifier.

A more direct method for obtaining these factors is to write Eq. (8.28) as a function of the available terms $kx(t)$, $k_2\dot{y}$, and k_3y. That is,

$$k_1\ddot{y} = -\frac{k_1}{k}[-kx(t)] - \frac{8k_1}{k_2}[k_2\dot{y}] - \frac{2k_1}{k_3}[k_3y] \tag{8.30}$$

The voltage equation for the first amplifier is

$$e_1 = -\frac{R_2}{R_a}e - \frac{R_2}{R_b}e_4 - \frac{R_2}{R_c}e_3 \tag{8.31}$$

In order that $e_1 = k_1\ddot{y}$, it follows from Eqs. (8.30) and (8.31) that

$$\frac{R_2}{R_a} = \frac{k_1}{k} \qquad \frac{R_2}{R_b} = \frac{8k_1}{k_2} \qquad \frac{R_2}{R_c} = \frac{2k_1}{k_3} \tag{8.32}$$

Thus, the coefficients in front of the brackets of Eq. (8.30) represent the ratio of the resistors to be used at the input to the first amplifier. Similarly, the terms within the brackets of Eq. (8.30) are the input quantities to the first amplifier. Because the term $-k_2\dot{y}$ which exists in the computer has the opposite sign of $k_2\dot{y}$ in the brackets, it is necessary to multiply by -1 in the feedback path.

Because the voltages e, e_1, e_2, and e_3 correspond to values of quantities in the physical system [$e = -kx(t)$, $e_1 = k_1\ddot{y}$, $e_2 = -k_2\dot{y}$, $e_3 = k_3y$], it is necessary to select the value of the scale factors k, k_1, k_2, and k_3 in order to interpret the results. For most computers, the voltages e, e_1, e_2, and e_3 should be limited to within 100 volts. Thus, the scale factor may be obtained from the equation

$$\text{Scale factor} = \frac{100}{\text{maximum value of parameter}} \tag{8.33}$$

From the originally given information, the scale factors are computed as follows:

$$\begin{aligned} k &= \frac{100}{x(t)_m} = \frac{100}{20} = 5 \text{ volts/lb} \\ k_1 &= \frac{100}{\ddot{y}_m} = \frac{100}{20} = 5 \text{ volts/ft/sec}^2 \\ k_2 &= \frac{100}{\dot{y}_m} = \frac{100}{10} = 10 \text{ volts/ft/sec} \\ k_3 &= \frac{100}{y_m} = \frac{100}{5} = 20 \text{ volts/ft} \end{aligned} \tag{8.34}$$

Substitution of the preceding results into Fig. 8.6*a* and *b* yields the final scaled computer diagram shown in Fig. 8.8, where the values of the initial voltages $e_2(0)$ and $e_3(0)$ are computed as follows:

$$e_2(0) = -k_2\dot{y}(0) = -10(5) = -50 \text{ volts} \tag{8.35}$$

$$e_3(0) = k_3y(0) = 20(2) = 40 \text{ volts} \tag{8.36}$$

The procedure for setting up an electronic analog computer to solve a differential equation of order n may be summarized as follows.

1. Solve the differential equation for the highest-order derivative. For example, $\dddot{y} = x(t) - C_2\ddot{y} - C_3\dot{y} - C_4y$.

2. Write each term of the right-hand side of the preceding expression as a function of the voltage $e_1 = k_1\dddot{y}$ and the input voltage $e = -kx(t)$

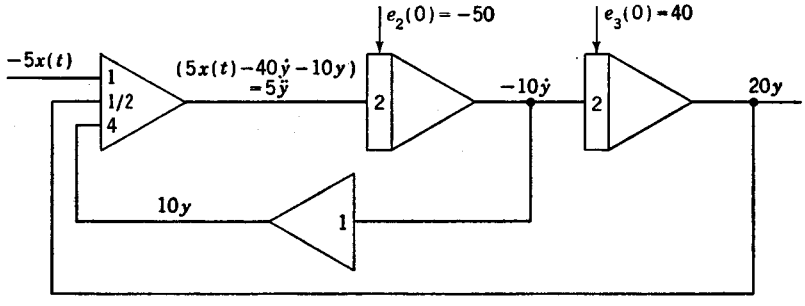


FIG. 8.8. Final computer diagram.

and each successive voltage that appears after integration (that is, $e_2 = -k_2\ddot{y}$, $e_3 = k_3\dot{y}$, $e_4 = -k_4y$). Multiplying the expression in step 1 by k_1 yields

$$k_1\dddot{y} = \frac{k_1}{k}[kx(t)] - C_2\frac{k_1}{k_2}[k_2\ddot{y}] - C_3\frac{k_1}{k_3}[k_3\dot{y}] - C_4\frac{k_1}{k_4}[k_4y] \tag{8.37}$$

The input to the first amplifier should be the negative of the right-hand side of the preceding expression, i.e.,

$$\frac{k_1}{k}[-kx(t)] + C_2\frac{k_1}{k_2}[k_2\ddot{y}] + C_3\frac{k_1}{k_3}[k_3\dot{y}] + C_4\frac{k_1}{k_4}[k_4y] \tag{8.38}$$

The coefficients in front of the brackets are the ratio of the resistors (that is, R_2/R_a, R_2/R_b, . . . , R_2/R_n) to be used at the input to the first amplifier. When the sign of the term in the computer is opposite to that of the corresponding term in the brackets (for example, $e_2 = -k_2\ddot{y}$ and $e_4 = -k_4y$), use an amplifier to multiply the feedback signal by -1.

3. The value of $1/R_1C_2$ for each integrator is equal to the ratio of the scale factor for the output voltage term to that for the input [for example $(1/R_1C_2)_2 = k_2/k_1$, $(1/R_1C_2)_3 = k_3/k_2$, . . .].

In accordance with Eq. (8.33), the value of each scale factor is seen to

depend on the maximum value of the corresponding parameter. For most systems being studied, there is usually sufficient information available to make a reasonable estimate of the maximum value of each term. If an error is made in predicting the maximum value of a term, then the maximum value of the voltage corresponding to that term will not be 100 volts. Such a situation is easily detected and is corrected as follows: Suppose that the maximum value of the voltage e_1 is found to be 50 volts rather than 100 volts. The original scale factor was $k_1 = 5$, so that the maximum value of $\ddot{y}$ is now found to be $\ddot{y}_m = e_{1_m}/k_1 = {}^{50}\!/_{5} = 10$ ft/sec^2 rather than 20 ft/sec^2. Thus, the value of k_1 should be $k_1 = {}^{100}\!/_{10} = 10$ rather than 5. It is an easy matter to revise the computer diagram using $k_1 = 10$ so that the maximum value of the voltage e_1 will be 100 volts.

8.5. Time Scale. For many problems, it is desired that the speed at which the analog computer solves the problem be different from the speed at which the phenomena actually occur. For example, various phenomena of astronomy require years; so obviously it is desirable to increase the speed at which such problems are solved on the computer. For other phenomena which take place very rapidly, it is necessary to slow down the speed at which such problems are simulated by the computer. By letting t represent the time at which a phenomenon actually occurs and the term τ represent the time required for this phenomenon to occur on the computer, $\tau = at$ relates the actual time t to the computer or machine time τ. For $a < 1$ the phenomenon occurs faster in the computer than it does in nature. For example, if $a = 0.1$, something which requires 10 sec actually to complete is completed by the computer in $\tau = 0.1t = 0.1 \times 10 = 1$ sec. Similarly, if $a > 1$, the phenomenon is slowed down by the computer.

Illustrative Example. Let it be desired to slow down the computer solution of the preceding problem by a factor of 5. The first step in the solution of this problem is to transform the original equation from a function of actual time t to a function of machine time τ. This is accomplished by noting that

$$\tau = at \tag{8.39}$$

and

$$\frac{d\tau}{dt} = a \tag{8.40}$$

Thus

$$\dot{y} = \frac{dy}{dt} = \frac{d\tau}{dt}\frac{dy}{d\tau} = a\frac{dy}{d\tau}$$
$$\ddot{y} = \frac{d^2y}{dt^2} = \frac{d}{dt}\left(\frac{dy}{dt}\right) = \frac{d\tau}{dt}\frac{d}{d\tau}\left(a\frac{dy}{d\tau}\right) = a^2\frac{d^2y}{d\tau^2} \tag{8.41}$$

Similarly, it may be shown that in general

$$\frac{d^ny}{dt^n} = a^n\frac{d^ny}{d\tau^n} \tag{8.42}$$

Application of the preceding rules to convert the original time expression given by Eq. (8.12) from a function of t to a function of τ gives

$$a^2 \frac{d^2y}{d\tau^2} + 8a \frac{dy}{d\tau} + 2y = x\left(\frac{\tau}{a}\right) \tag{8.43}$$

where $x(\tau/a)$ is obtained by substitution of τ/a for t in the original function $x(t)$. Because of the change of variable, the term y in Eq. (8.43) is now

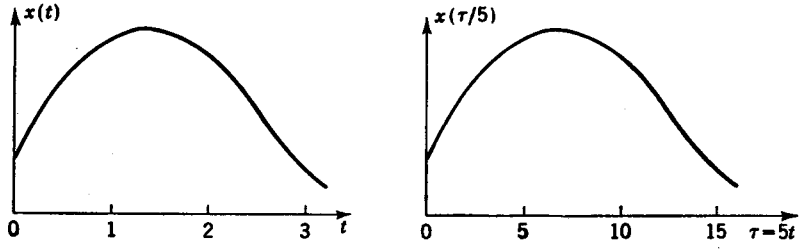


FIG. 8.9. Graphs of $x(t)$ versus t and $x\left(\frac{\tau}{5}\right)$ versus τ.

a function of τ rather than t. To slow down the solution by a factor of 5, the value of a is 5 so that Eq. (8.43) becomes

$$25 \frac{d^2y}{d\tau^2} + 40 \frac{dy}{d\tau} + 2y = x\left(\frac{\tau}{5}\right) \tag{8.44}$$

The transformed initial conditions are

$$\left.\frac{dy}{d\tau}\right|_{\tau=0} = \frac{1}{a}\left.\frac{dy}{dt}\right|_{t=0} = \frac{1}{5}(5) = 1 \text{ ft/sec} \tag{8.45}$$

Similarly, the transformed maximum values are

$$\begin{aligned} \left(\frac{d^2y}{d\tau^2}\right)_m &= \frac{1}{a^2}\left(\frac{d^2y}{dt^2}\right)_m = \frac{20}{25} = 0.8 \text{ ft/sec}^2 \\ \left(\frac{dy}{d\tau}\right)_m &= \frac{1}{a}\left(\frac{dy}{dt}\right)_m = \frac{10}{5} = 2.0 \text{ ft/sec} \end{aligned} \tag{8.46}$$

As is shown in Fig. 8.9, the function $x(\tau/5)$ is obtained, in effect, by multiplying the original time scale by the factor $a = 5$. Thus, a time-scale change does not affect the initial and maximum values for $x(\tau/a)$. Similarly, the initial and maximum values for y are unaffected by a change in the time scale.

The new scale factors are computed as follows:

$$\begin{aligned} k &= \frac{100}{x(\tau/a)_m} = \frac{100}{20} = 5 \text{ volts/lb} \\ k_1 &= \frac{100}{(d^2y/d\tau^2)_m} = \frac{100}{0.8} = 125 \text{ volts/ft/sec}^2 \\ k_2 &= \frac{100}{(dy/d\tau)_m} = \frac{100}{2} = 50 \text{ volts/ft/sec} \\ k_3 &= \frac{100}{y_m} = \frac{100}{5} = 20 \text{ volts/ft} \end{aligned} \tag{8.47}$$

The computer diagram is now obtained by application of the general procedure given in the preceding section. Solving Eq. (8.44) for $d^2y/d\tau^2$ and multiplying by the scale factor $k_1 = 125$ gives

$$\begin{aligned} 125\frac{d^2y}{d\tau^2} &= \frac{125}{25}x\left(\frac{\tau}{5}\right) - \frac{125 \times 40}{25}\frac{dy}{d\tau} - \frac{125 \times 2y}{25} \\ &= 5x\left(\frac{\tau}{5}\right) - 200\frac{dy}{d\tau} - 10y \end{aligned} \tag{8.48}$$

The right-hand side of Eq. (8.48) is next written in terms of the voltages. That is

$$125\frac{d^2y}{d\tau^2} = \frac{5}{k}\left[kx\left(\frac{\tau}{5}\right)\right] - \frac{200}{k_2}\left[k_2\frac{dy}{d\tau}\right] - \frac{10}{k_3}[k_3y] \tag{8.49}$$

The negative of the right-hand side of Eq. (8.49) is

$$\frac{5}{k}\left[-kx\left(\frac{\tau}{5}\right)\right] + \frac{200}{k_2}\left[k_2\frac{dy}{d\tau}\right] + \frac{10}{k_3}[k_3y] \tag{8.50}$$

Substitution of the numerical values for the scale factors in the preceding expression gives

$$1\left[-5x\left(\frac{\tau}{5}\right)\right] + 4\left[50\frac{dy}{d\tau}\right] + 0.5[20y] \tag{8.51}$$

The resulting computer diagram is shown in Fig. 8.10*a*.

The coefficients in front of the bracketed terms of Eq. (8.51) are the effective input multipliers at the first amplifier. The multiplication factors for the second and third amplifiers are obtained as follows:

$$\begin{aligned} \left(\frac{1}{R_1C_2}\right)_2 &= \frac{k_2}{k_1} = \frac{50}{125} = 0.4 \\ \left(\frac{1}{R_1C_2}\right)_3 &= \frac{k_3}{k_2} = \frac{20}{50} = 0.4 \end{aligned} \tag{8.52}$$

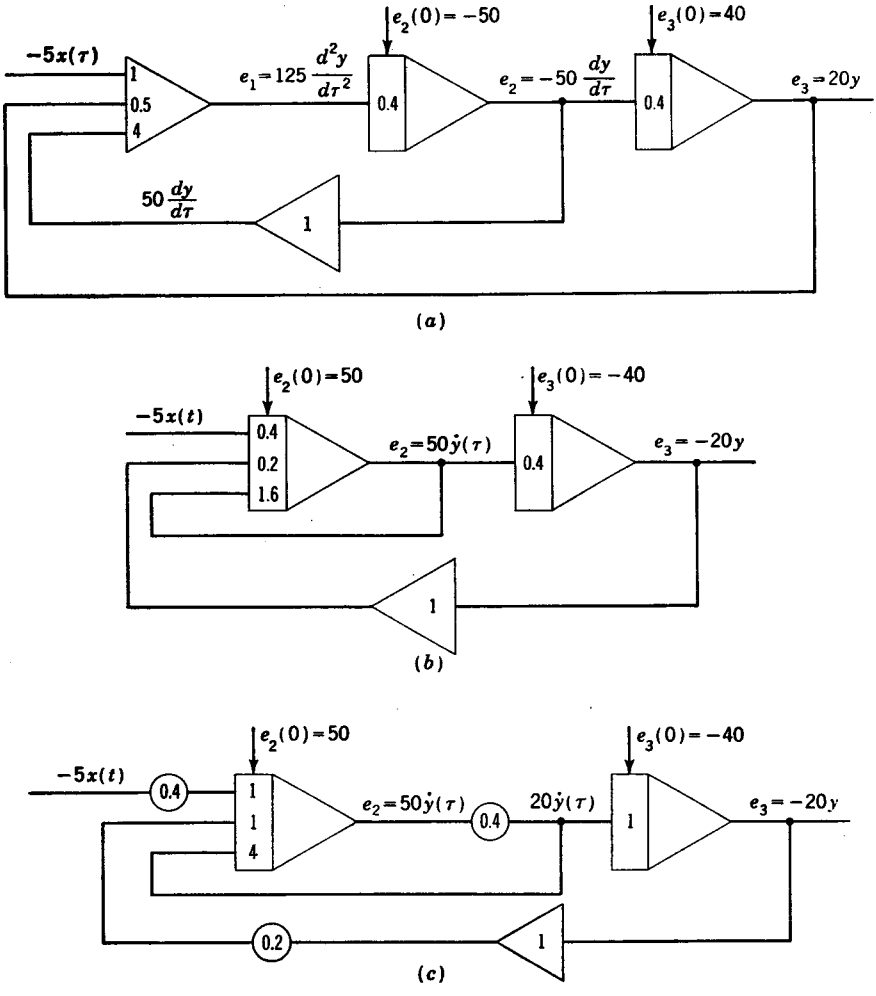


FIG. 8.10. Computer diagram for $25 \frac{d^2y}{d\tau^2} = x\left(\frac{\tau}{5}\right) - 40 \frac{dy}{d\tau} - 2y$.

The values of the initial voltages $e_2(0)$ and $e_3(0)$ are computed in the following manner:

$$\begin{aligned} e_2(0) &= -k_2 \frac{dy}{d\tau}\bigg|_{\tau=0} = -50(1) = -50 \text{ volts} \\ e_3(0) &= k_3 y\bigg|_{\tau=0} = 20(2) = 40 \text{ volts} \end{aligned} \tag{8.53}$$

If it is not necessary to measure $d^2y/d\tau^2$, the first amplifier may be eliminated by multiplying the input constants of the first summer by the factor

0.4 for the next amplifier [that is, (1)(0.4) = 0.4, (0.5)(0.4) = 0.2, (4)(0.4) = 1.6], as shown in Fig. 8.10*b*. The elimination of this amplifier changes the signs of the resulting voltages, and thus the sign of each feedback quantity must be reversed. Similarly the sign of each initial condition voltage must be changed.

Some operational amplifiers such as the REAC (Reeves electronic analog computer) use standard resistors of 1×10^6, 0.25×10^6, and 0.10×10^6 ohms and a standard capacitor of 1×10^{-6} farad. Thus, only gains of $1/RC$ equal to 1, 4, or 10 are readily available for each integrator. It is possible to put two resistors in parallel or series at the

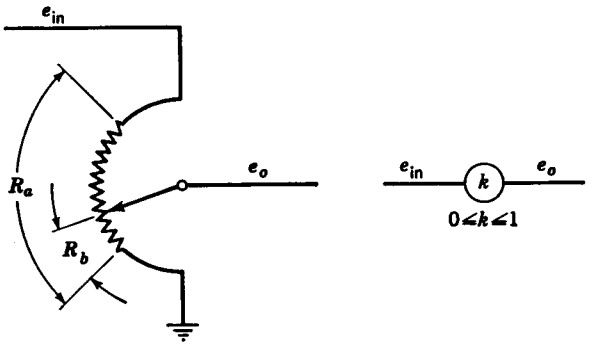


FIG. 8.11. Potentiometer.

input to obtain some other effective value of resistance. For example, two 1×10^6 ohm resistors in series yield a 2×10^6 ohm resistance, while two 1×10^6 ohm resistors in parallel yield a 0.5×10^6 ohm resistance.

In general, it is necessary to use a potentiometer to obtain the desired effective resistance. Figure 8.11 shows a schematic diagram of a potentiometer. The voltage relationship is

$$e_o = \frac{R_b}{R_a} e_{in} = k e_{in} \tag{8.54}$$

where

$$k = \frac{R_b}{R_a} \leq 1$$

The computer diagram of Fig. 8.10*b* may be modified by the addition of three potentiometers as shown in Fig. 8.10*c* such that only gains of 1, 4, or 10 are necessary at each amplifier. By comparison of Fig. 8.10*b* and *c*, it is to be noted that the effective gain of an amplifier is the product of the gain of the amplifier and the value of k for the potentiometer in front of the amplifier.

8.6. Simulation. A very important application of the analog computer is the simulation of automatic control systems. One method that

could be used to simulate a control system would be to determine the over-all differential equation and solve this on the computer. Usually though, one is interested in determining the effect on the system performance when certain parameters are varied. Using the preceding technique would mean solving a new differential equation for each change.

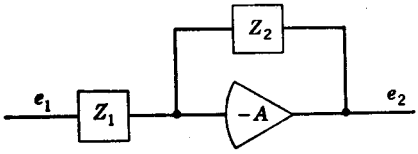
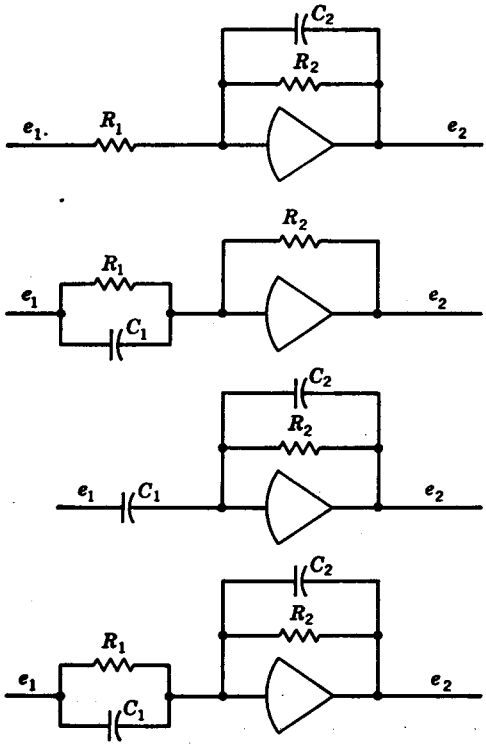


FIG. 8.12. General schematic representation of an operational amplifier.

Because of the similarity between a block diagram and a computer diagram, it is customary to simulate each portion of the system and then interconnect each of these elements. Thus, the effect of changing one of the terms in the original block diagram may be accomplished by changing the corresponding quantity in the computer diagram.

In Fig. 8.12 is shown a schematic representation of an operational amplifier in which the input impedance is Z_1 and the parallel impedance is

TABLE 8.1



$$e_2 = \frac{-R_2}{R_1(1+R_2C_2p)}e_1$$

$$e_2 = -\frac{R_2}{R_1}(1+R_1C_1p)e_1$$

$$e_2 = \frac{-R_2C_1p}{1+R_2C_2p}e_1$$

$$e_2 = -\frac{R_2}{R_1}\frac{1+R_1C_1p}{1+R_2C_2p}e_1$$

Z_2. The equation of operation for this amplifier is

$$\frac{e_2}{e_1} = -\frac{Z_2}{Z_1} \tag{8.55}$$

When the input impedance is a resistor R_1 and the parallel impedance is a resistor R_2, the preceding expression reduces to the result given by Eq. (8.8). For the case in which $Z_1 = R_1$ and $Z_2 = 1/C_2p$ the result given by Eq. (8.11) is verified.

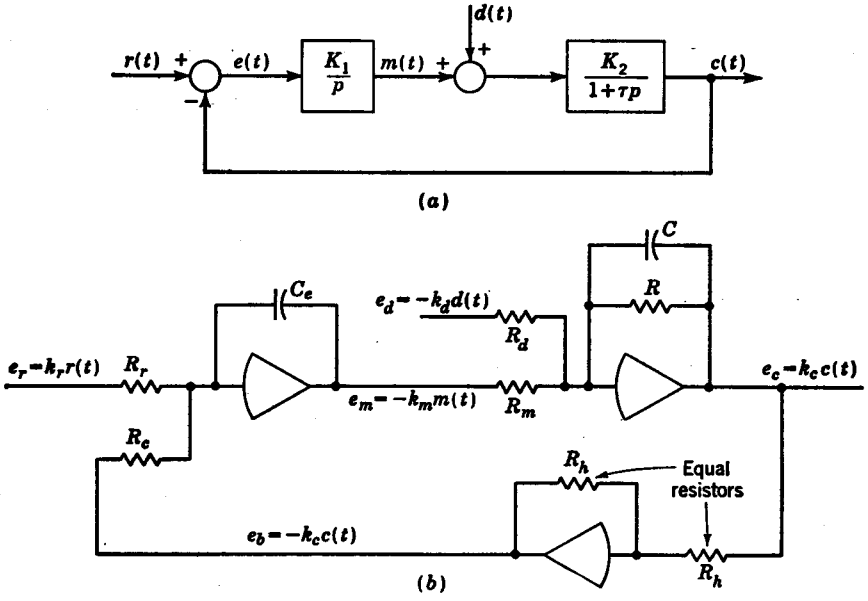


FIG. 8.13. Block diagram and corresponding computer diagram.

In Table 8.1 is shown a number of computer circuits for simulating various transfer functions. For the first circuit $Z_1 = R_1$ and $Z_2 = \dfrac{1}{1/R_2 + C_2p} = \dfrac{R_2}{1 + R_2C_2p}$. Substitution of the values of these impedances into Eq. (8.55) gives the equation of operation for this computer circuit, i.e.,

$$e_2 = -\frac{R_2}{R_1(1 + R_2C_2p)}\, e_1 \tag{8.56}$$

In Fig. 8.13*a* is shown a typical block diagram, and the corresponding computer diagram is shown in Fig. 8.13*b*.

Illustrative Example. Suppose that the system shown in Fig. 8.13*a* is used to control the angular position of a shaft. For this system, it is known that $K_1 = 10$, $K_2 = 5$, and $\tau = 1.0$. The maximum values have

been estimated to be $c(t)_m = 20$ radians, $r(t)_m = 10$ radians, $m_m = 50$ in.-lb, and $d(t)_m = 100$ in.-lb. Determine the values of the resistors and capacitors for the computer diagram of Fig. 8.13*b*.

SOLUTION. The voltage equation for the first amplifier is

$$e_m = \frac{-1}{R_r C_e} \int_0^t e_r \, dt - \frac{1}{R_c C_e} \int_0^t e_b \, dt + e_m(0) \tag{8.57}$$

The substitution of $e_r = k_r r(t)$ and $e_b = -k_c c(t)$ into the preceding expression gives

$$e_m = \frac{-k_r}{R_r C_e} \int_0^t r(t) \, dt + \frac{k_c}{R_c C_e} \int_0^t c(t) \, dt + e_m(0) \tag{8.58}$$

The equation which describes the operation of the corresponding portion of the actual system is

$$\begin{aligned} m(t) &= K_1 \int_0^t [r(t) - c(t)] \, dt + m(0) \\ &= K_1 \int_0^t r(t) \, dt - K_1 \int_0^t c(t) \, dt + m(0) \end{aligned} \tag{8.59}$$

Multiplication of Eq. (8.59) by $-k_m$ gives

$$-k_m m(t) = -k_m K_1 \int_0^t r(t) \, dt + k_m K_1 \int_0^t c(t) \, dt - k_m m(0) \tag{8.60}$$

In order that $e_m = -k_m m(t)$, by comparison of corresponding terms in Eqs. (8.58) and (8.60) it follows that

$$\frac{1}{R_r C_e} = \frac{k_m K_1}{k_r} \qquad \frac{1}{R_c C_e} = \frac{k_m K_1}{k_c} \qquad e_m(0) = -k_m m(0) \tag{8.61}$$

For any system k_m, k_r, k_c, R_r, and R_c are constant, and thus the gain K_1 may be changed independently by varying the capacitance C_e.

Numerical values are obtained by first computing the scale factors, thus,

$$\begin{aligned} k_r &= {}^{100}\!/_{10} = 10 \text{ volts/radian} \\ k_c &= {}^{100}\!/_{20} = 5 \text{ volts/radian} \\ k_m &= {}^{100}\!/_{50} = 2 \text{ volts/in.-lb} \\ k_d &= {}^{100}\!/_{100} = 1 \text{ volt/in.-lb} \end{aligned} \tag{8.62}$$

From Eqs. (8.61), it follows that

$$\frac{1}{R_r C_e} = \frac{(2)K_1}{10} = 2 \qquad \frac{1}{R_c C_e} = \frac{(2)K_1}{5} = 4 \qquad e_m(0) = -2m(0) \tag{8.63}$$

If a 1-μf capacitor is used for C_e,

$$\begin{aligned} R_r &= \frac{1}{2C_e} = \frac{1}{2 \times 10^{-6}} = 500{,}000 \text{ ohms} \\ R_c &= \frac{1}{4C_e} = 250{,}000 \text{ ohms} \end{aligned} \tag{8.64}$$

From Table 8.1, it follows that the voltage equation for the second amplifier is

$$e_c = -\frac{R}{R_m(1 + RCp)} e_m - \frac{R}{R_d(1 + RCp)} e_d \tag{8.65}$$

Substitution of $e_m = -k_m m(t)$ and $e_d = -k_d d(t)$ into the preceding expression yields

$$e_c = \frac{R}{1 + RCp}\left[\frac{k_m m(t)}{R_m} + \frac{k_d d(t)}{R_d}\right] \tag{8.66}$$

The equation for the corresponding portion of the actual system is

$$c(t) = \frac{K_2}{1 + \tau p}[m(t) + d(t)] \tag{8.67}$$

Multiplication by k_c gives

$$k_c c(t) = \frac{k_c K_2}{1 + \tau p}[m(t) + d(t)] \tag{8.68}$$

By comparison of corresponding terms in Eqs. (8.66) and (8.68), it follows that, to have $e_c = k_c c(t)$, the following relationships must hold:

$$\frac{Rk_m}{R_m} = k_c K_2 \qquad \frac{Rk_d}{R_d} = k_c K_2 \qquad \tau = RC \tag{8.69}$$

The substitution of numerical values gives

$$\frac{R}{R_m} = \frac{k_c K_2}{k_m} = \frac{(5)(5)}{2} = 12.5 \tag{8.70}$$

$$\frac{R}{R_d} = \frac{k_c K_2}{k_d} = \frac{(5)(5)}{1} = 25 \tag{8.71}$$

$$RC = \tau = 1 \tag{8.72}$$

If a 1-μf capacitor is used for C,

$$\begin{aligned} R &= \frac{1}{C} = 1{,}000{,}000 \text{ ohms} \\ R_m &= \frac{R}{12.5} = 80{,}000 \text{ ohms} \\ R_d &= \frac{R}{25} = 40{,}000 \text{ ohms} \end{aligned} \tag{8.73}$$

Because C appears only in Eq. (8.72), τ may be varied independently by varying C. A variable capacitor provides a convenient means for varying C. From Eqs. (8.70) and (8.71), it follows that to change the value of K_2 both R_m and R_d must be changed accordingly. A variable resistance is provided by a potentiometer.

Because the output e_c is to be subtracted from the input e_r, it is neces-

sary to multiply the output by -1, as shown in the feedback path of Fig. 8.13*b*.

If, in testing, it is found that $e_{c_m} \neq 100$ volts, this is evidence that the originally estimated value for $c(t)_m$ is not correct. Because $e_c = k_c c(t)$, the actual value of $c(t)_m$ is now easily found to be

$$c(t)_m = \frac{e_{c_m}}{k_c} \tag{8.74}$$

When the maximum value of $c(t)$ is near 100 volts, there is no need to change the value of k_c. However, if the maximum voltage is quite small, then it is desirable to increase the accuracy with which the output voltage can be read by using a new value of k_c based on the value of $c(t)_m$ obtained from Eq. (8.74). Similarly, if e_{c_m} was sufficiently greater than 100 volts to cause overloading, then it would be necessary to decrease the value of k_c. For most amplifiers, an overload light turns on when the maximum voltage is high enough to overload an amplifier.

A time-scale change $\tau = at$ is readily effected by substituting ap for p in the block diagram of Fig. 8.13*a*.

A major application of analog computers is in the design of systems with nonlinear components. Standard electronic circuits are available for simulating commonly encountered nonlinear effects such as coulomb friction, backlash, dead zone, saturation, continuous nonlinear functions, etc.

8.7. Digital Computers. As is discussed in Sec. 8.1, a problem must be reduced to elementary mathematical operations before being submitted to a digital computer. For example, consider the problem discussed in Sec. 8.3, that is,

$$\ddot{y} + 8\dot{y} + 2y = x(t) \tag{8.12}$$

where the initial values $y(0)$ and $\dot{y}(0)$ are known, as is the function $x(t)$. The initial value of $\ddot{y}(0)$ is obtained by solving Eq. (8.12) for $\ddot{y}$ and substituting the given initial conditions into this expression, i.e.,

$$\ddot{y}(0) = x(0) - 8\dot{y}(0) - 2y(0) \tag{8.75}$$

By choosing a small increment Δt such that the slope $\dot{y}(0)$ of the curve of y versus t may be considered to remain constant in the interval $0 \leq t \leq \Delta t$, it follows that

$$y(\Delta t) = y(0) + \dot{y}(0)\,\Delta t \tag{8.76}$$

Similarly, if Δt is small enough so that the slope $\ddot{y}(0)$ of the graph of velocity vs. time does not vary appreciably,

$$\dot{y}(\Delta t) = \dot{y}(0) + \ddot{y}(0)\,\Delta t \tag{8.77}$$

The value of $\ddot{y}(\Delta t)$ may now be computed from the original differential equation, i.e.,

$$\ddot{y}(\Delta t) = x(\Delta t) - 8\dot{y}(\Delta t) - 2y(\Delta t) \tag{8.78}$$

Similarly, the evaluation of each term at time $t = 2\ \Delta t$ is accomplished as follows:

$$\begin{aligned} y(2\ \Delta t) &= y(\Delta t) + \dot{y}(\Delta t)\ \Delta t \\ \dot{y}(2\ \Delta t) &= \dot{y}(\Delta t) + \ddot{y}(\Delta t)\ \Delta t \\ \ddot{y}(2\ \Delta t) &= x(2\ \Delta t) - 8\dot{y}(2\ \Delta t) - 2y(2\ \Delta t) \end{aligned} \tag{8.79}$$

Thus, in general, it follows that, at time $t = \eta\ \Delta t$,

$$\begin{aligned} y(\eta\ \Delta t) &= y[(\eta - 1)\ \Delta t] + \dot{y}[(\eta - 1)\ \Delta t]\ \Delta t \\ \dot{y}(\eta\ \Delta t) &= \dot{y}[(\eta - 1)\ \Delta t] + \ddot{y}[(\eta - 1)\ \Delta t]\ \Delta t \\ \ddot{y}(\eta\ \Delta t) &= x(\eta\ \Delta t) - 8\dot{y}(\eta\ \Delta t) - 2y(\eta\ \Delta t) \end{aligned} \tag{8.80}$$

By continuing this process, values for y, $\dot{y}$, and $\ddot{y}$ may be obtained for any time $\eta\ \Delta t$. Accuracy is increased by using smaller increments for Δt. The preceding method for reducing a linear differential equation to a sequence of simple mathematical operations is known as the straight-line approximation method. Although more accurate techniques of numerical analysis could have been employed, the preceding method suffices to illustrate the fundamental concepts. It should be noticed that it is necessary to store the initial conditions and values of $x(t)$ at $t = 0, \Delta t, 2\ \Delta t, 3\ \Delta t, \ldots, \eta\ \Delta t$ in the machine so that they are available for computation when needed. Similarly, the answers to Eqs. (8.76) and (8.77) must be stored in the machine so that $\ddot{y}(\Delta t)$ in Eq. (8.78) may be computed, etc. Such storage devices in a digital computer are referred to as memory units. In essence, then, numbers and directions are the input to a digital computer, and the output is in the form of numbers.

Usually an analog computer can be set up to simulate a control system much faster than it can be programmed for solution on a digital computer. In addition, an analog computer tends to be more versatile and convenient to use for most problems encountered in control work.

CHAPTER 9

FREQUENCY-RESPONSE METHODS

9.1. Introduction. The frequency-response methods to be discussed in this chapter provide a convenient means for investigating the dynamic behavior of a control system. By frequency response[1–3] is meant the response of a system to a sinusoidal input $x = x_0 \sin \omega t$. After the effect of the initial transients has "died out," the output becomes a sinusoid

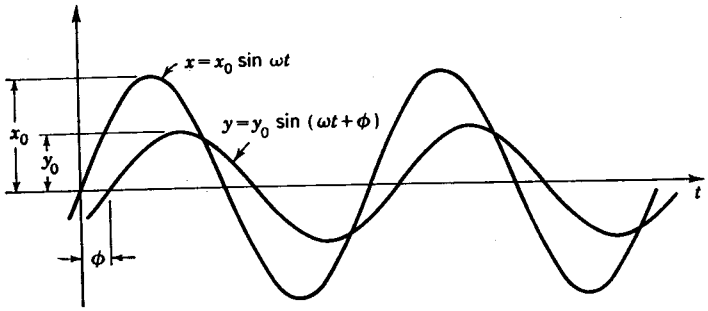


FIG. 9.1. Sinusoidal response.

with the same angular velocity ω of the input. In general, the output $y = y_0 \sin (\omega t + \phi)$ is displaced some phase angle ϕ from the input, and the amplitude of the output, y_0, is different from that of the input, x_0, as is illustrated in Fig. 9.1.

Both the phase angle ϕ and the ratio of the amplitude of the output to that of the input, y_0/x_0, are a function of the angular velocity ω of the input signal. Graphs of ϕ versus ω and of amplitude ratio y_0/x_0 versus ω form the basis for frequency-response methods.

Both the Laplace transform method and the frequency-response method are means for determining the dynamic behavior of control systems. With the Laplace transform method, emphasis is placed upon

[1] H. Harris, The Frequency Response of Automatic Control Systems, *Trans. AIEE*, vol. 65, pp. 534–545, 1946.

[2] G. S. Brown and D. P. Campbell, "Principles of Servomechanisms," John Wiley & Sons, Inc., New York, 1948.

[3] H. Chestnut and R. Mayer, "Servomechanisms and Regulating System Design," 2d ed., John Wiley & Sons, Inc., New York, 1959.

the characteristic equation. With frequency-response techniques, it is not necessary to determine the roots of the characteristic equation. A major advantage of frequency-response methods is that they provide a good basis for synthesizing systems. That is, they indicate what type of changes should be made in the system to yield the desired response.

9.2. Frequency Response. Because frequency-response methods are based on a knowledge of ϕ versus ω and y_0/x_0 versus ω, it is now shown how these quantities may be determined directly by substitution of $j\omega$ for p in the operational form of the differential equation for the system.

The general operational form of a differential equation is

$$y(t) = \frac{(a_m p^m + a_{m-1}p^{m-1} + \cdots + a_1 p + a_0)x(t)}{p^n + b_{n-1}p^{n-1} + \cdots + b_1 p + b_0} = \frac{L_m(p)x(t)}{L_n(p)} \tag{9.1}$$

The transform of the preceding expression is

$$Y(s) = \frac{L_m(s)X(s) + I(s)}{L_n(s)} = \frac{L_m(s)N_{X(s)}}{L_n(s)D_{X(s)}} + \frac{I(s)}{L_n(s)} \tag{9.2}$$

The transform for the input is

$$X(s) = \mathcal{L}(x_0 \sin \omega t) = \frac{\omega x_0}{s^2 + \omega^2} = \frac{N_{X(s)}}{D_{X(s)}} \tag{9.3}$$

Expanding Eq. (9.2) in a partial-fraction expansion and noting that $D_{X(s)} = s^2 + \omega^2 = (s - j\omega)(s + j\omega)$ gives

$$Y(s) = \frac{K_1}{s - r_1} + \cdots + \frac{K_n}{s - r_n} + \frac{K_C}{s - j\omega} + \frac{K_{-C}}{s + j\omega} + \frac{I_1}{s - r_1} + \cdots + \frac{I_n}{s - r_n} \tag{9.4}$$

where $r_1, r_2, \ldots, r_n$ are the zeros of $L_n(s)$ and $K_1, K_2, \ldots, K_n, K_C$ and K_{-C} are the constants which arise from the partial-fraction expansion of $L_m(s)N_{X(s)}/L_n(s)D_{X(s)}$ and $I_1, I_2, \ldots, I_n$ are the constants which arise from the partial-fraction expansion of $I(s)/L_n(s)$. Inverting Eq. (9.4) gives

$$y(t) = (K_1 + I_1)e^{r_1 t} + \cdots + (K_n + I_n)e^{r_n t} + \frac{1}{b}|K(a + jb)|e^{at} \sin (bt + \phi) \tag{9.5}$$

For a stable system $r_1, r_2, \ldots, r_n$ must have negative real parts, so that after sufficient time the effect of these terms becomes negligible. Thus, for stable systems, the steady-state sinusoidal response $y(t)_{ss}$ is determined by the last term of Eq. (9.5). Because $a = 0$ and $b = \omega$, it follows that

$$y(t)_{ss} = \frac{1}{\omega}|K(j\omega)| \sin (\omega t + \phi) \tag{9.6}$$

The terms $|K(j\omega)|$ and $\phi = \measuredangle K(j\omega)$ are evaluated as follows:

$$K(j\omega) = \lim_{s \to j\omega} [(s^2 + \omega^2)Y(s)] \tag{9.7}$$

By noting from Eq. (9.2) that

$$Y(s) = \frac{L_m(s)X(s)}{L_n(s)} + \frac{I(s)}{L_n(s)} \tag{9.8}$$

then
$$K(j\omega) = \lim_{s \to j\omega} \left[\frac{(s^2 + \omega^2)L_m(s)\omega x_0}{L_n(s)(s^2 + \omega^2)} + (s^2 + \omega^2) \frac{I(s)}{L_n(s)} \right]$$
$$= \frac{L_m(j\omega)\omega x_0}{L_n(j\omega)} \tag{9.9}$$

In the preceding development, the effect of the initial conditions is seen to drop out so that $y(t)_{ss}$ is independent of the initial behavior of the system. The terms $L_m(j\omega)$ and $L_n(j\omega)$ are obtained by substituting $j\omega$ for p in $L_m(p)$ and $L_n(p)$. From Eq. (9.9) it follows that

$$|K(j\omega)| = \left| \frac{L_m(j\omega)}{L_n(j\omega)} \right| \omega x_0 \tag{9.10}$$

and
$$\phi = \measuredangle \left[\frac{L_m(j\omega)}{L_n(j\omega)} \right] \omega x_0 = \measuredangle \frac{L_m(j\omega)}{L_n(j\omega)} = \phi \tag{9.11}$$

Substitution of Eq. (9.10) into Eq. (9.6) yields

$$y(t)_{ss} = \left| \frac{L_m(j\omega)}{L_n(j\omega)} \right| x_0 \sin(\omega t + \phi) = y_0 \sin(\omega t + \phi) \tag{9.12}$$

From Eq. (9.12), the amplitude y_0 of the steady-state response $y(t)_{ss}$ is seen to be

$$y_0 = \frac{|L_m(j\omega)|}{|L_n(j\omega)|} x_0$$

The amplitude ratio is

$$\frac{y_0}{x_0} = \left| \frac{L_m(j\omega)}{L_n(j\omega)} \right| \tag{9.13}$$

It is now shown how to obtain plots of the amplitude ratio y_0/x_0 versus ω and also phase angle ϕ versus ω. Consider the first-order linear differential equation

$$y(t) = \frac{1}{1 + \tau p} x(t) \tag{9.14}$$

The substitution of $j\omega$ for p in the preceding equation yields

$$y(j\omega) = \frac{1}{1 + j\tau\omega} x(j\omega)$$

or
$$\frac{y(j\omega)}{x(j\omega)} = \frac{L_m(j\omega)}{L_n(j\omega)} = \frac{1}{1 + j\tau\omega} \tag{9.15}$$

The ratio of the amplitude of the output to that of the input is

$$\frac{y_0}{x_0} = \left| \frac{L_m(j\omega)}{L_n(j\omega)} \right| = \frac{1}{\sqrt{1 + (\tau\omega)^2}} \tag{9.16}$$

The phase angle ϕ is

$$\phi = \measuredangle \frac{1}{1 + j\tau\omega} \tag{9.17}$$

The angle of the numerator is $\measuredangle(1 + j0) = \tan^{-1} 0 = 0$. Thus, subtracting the angle of the denominator from the angle of the numerator gives

$$\phi = 0 - \measuredangle(1 + j\tau\omega) = -\tan^{-1}\tau\omega \tag{9.18}$$

In Fig. 9.2 is shown a graph of the amplitude ratio y_0/x_0 versus ω and a graph of the phase angle ϕ as a function of the angular velocity ω of the

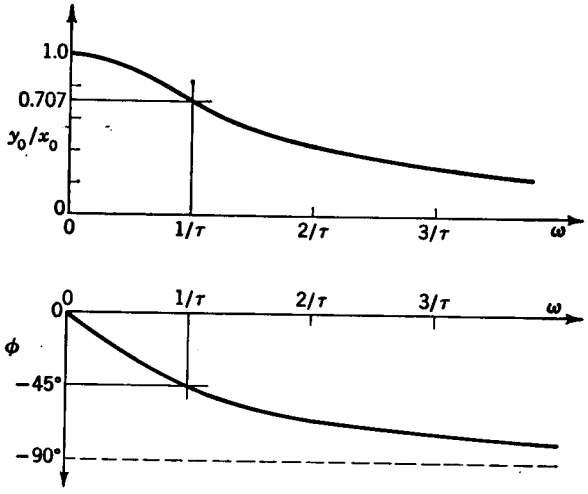


FIG. 9.2. Response curves for the function $\frac{1}{1 + j\tau\omega}$.

forcing function. When ω is small, the amplitude of the output is almost equal to that of the input and the phase angle ϕ is quite small. The output cannot keep up with the input at higher frequencies, and it begins to lag behind the input. For $\omega > 1/\tau$ this effect becomes very pronounced.

From Fig. 9.2, it is to be noticed that, for a given ω, the amplitude ratio has a given value and also there is a certain phase angle ϕ between the input and the output. This amplitude ratio $|L_m(j\omega)/L_n(j\omega)|$ and phase angle $\measuredangle L_m(j\omega)/L_n(j\omega)$ determine a vector $L_m(j\omega)/L_n(j\omega)$. The path of the tip of this vector (vector loci) for values of ω from zero to ∞ is shown in Fig. 9.3. This is called a polar plot. The polar plot shown in Fig. 9.3

is seen to convey the same information as the two separate curves shown in Fig. 9.2. The polar plot for this first order system is a semicircle as shown in Fig. 9.3. At $\omega = 0$, the length of the vector is 1, and the phase angle zero. At $\omega = 1/\tau$ the phase angle $\phi = -45°$, and the length of the vector is 0.707.

The operation of the feedforward part of a control system is given by the equation

$$c(t) = \frac{L_m(p)}{L_n(p)}\, e(t) = G(p)e(t) \tag{9.19}$$

When the actuating signal $e(t)$ is a sinusoid [$e(t) = e_0 \sin \omega t$], then the controlled variable $c(t)$ is also a sinusoid [$c(t) = c_0 \sin (\omega t + \phi)$]. The amplitude ratio c_0/e_0 and the phase angle ϕ are

$$\begin{aligned} \frac{c_0}{e_0} &= \left| \frac{c(j\omega)}{e(j\omega)} \right| = |G(j\omega)| \\ \phi &= \measuredangle \left[\frac{c(j\omega)}{e(j\omega)} \right] = \measuredangle G(j\omega) \end{aligned} \tag{9.20}$$

Thus the vector $G(j\omega)$ completely describes the frequency response of the feedforward elements. Similarly, the response of the feedback elements is determined by $H(j\omega)$.

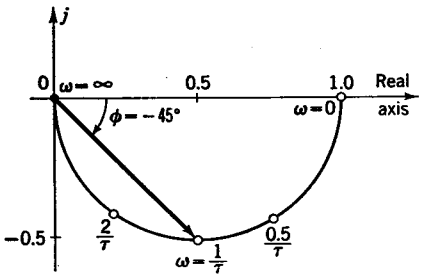


FIG. 9.3. Polar plot for the function $\frac{1}{1 + j\tau\omega}$.

In Chap. 6 it is shown that, *when the initial conditions are zero, the substitution of s for p yields directly the transform of the differential equation which is called the transfer function.* Thus, *either the substitution of $j\omega$ for p in the differential equation or the substitution of $j\omega$ for s in the transfer function gives the vector equation for evaluating the frequency response.*

9.3. Logarithmic Representation of Frequency Response. Frequency-response methods are based on the response $G(j\omega)$ of the feedforward elements and $H(j\omega)$ of the feedback elements. The transfer functions for these quantities are $G(s)$ and $H(s)$, respectively. These quantities are usually obtained in factored form and are composed of multiples or ratios of one or more of the following types of terms: s, $1 + \tau s$, $(s^2 + 2\zeta\omega_n s + \omega_n^2)/\omega_n^2$. The substitution of $j\omega$ for s means that $G(j\omega)$ or $H(j\omega)$ will be composed of terms such as $j\omega$, $1 + j\tau\omega$, or $(\omega_n^2 - \omega^2 + j2\zeta\omega_n\omega)/\omega_n^2$. To obtain the resulting polar plot, the multiplication of such terms is simplified by the use of logarithms. For

example, let the value of $G(s)$ be given by the equation

$$G(s) = \frac{K}{s(1 + \tau s)} \tag{9.21}$$

Substituting $j\omega$ for s yields

$$G(j\omega) = \frac{K}{j\omega(1 + j\tau\omega)} \tag{9.22}$$

The magnitude of Eq. (9.22) is

$$|G(j\omega)| = \frac{K}{|j\omega|\,|1 + j\tau\omega|} \tag{9.23}$$

Taking the logarithm of the preceding expression gives

$$\begin{aligned} \log |G(j\omega)| &= \log K - \log |j\omega| - \log |1 + j\tau\omega| \\ \log |G(j\omega)| - \log K &= -\log \omega - \tfrac{1}{2} \log (1 + \tau^2\omega^2) \end{aligned} \tag{9.24}$$

The angle ϕ is

$$\begin{aligned} \phi = \measuredangle G(j\omega) &= -\measuredangle j\omega - \measuredangle (1 + j\tau\omega) \\ &= -90^\circ - \tan^{-1} \tau\omega \end{aligned} \tag{9.25}$$

By using logarithms, the effect of each term is added separately if it appears in the numerator or is subtracted if it is in the denominator. The contribution due to the $1/j\omega$ term is shown in Fig. 9.4. The equation for the amplitude is $\log |1/j\omega| = -\log \omega$. When ω is 1, the value of $-\log \omega$ is 0, and when ω changes by multiples of 10, the value of $-\log \omega$

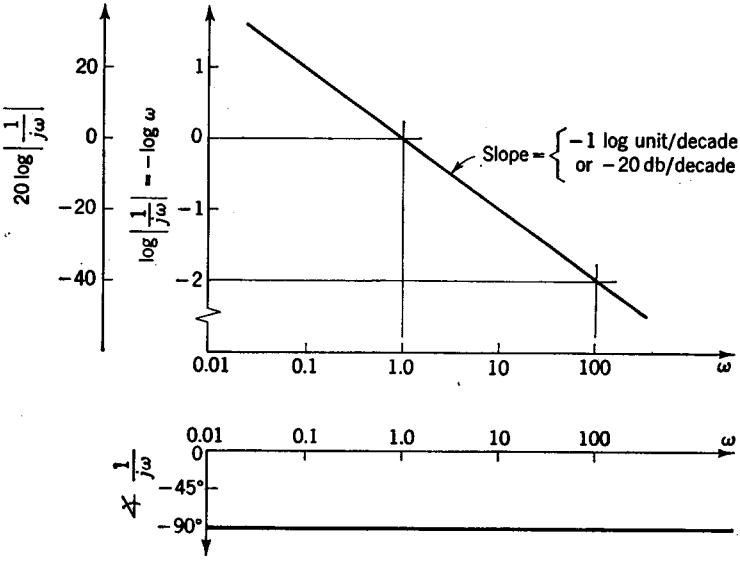


FIG. 9.4. log-magnitude plot for $\frac{1}{j\omega}$.

changes in increments of 1. The slope of this straight-line logarithmic response shown in Fig. 9.4 is -1 log unit/decade. A decade is the horizontal distance on the frequency scale from any value of ω to ten times ω. Thus, the distance from $\omega = 1$ to $\omega = 10$ or from $\omega = 3$ to $\omega = 30$, etc., is a decade.

The vertical log-magnitude scale is sometimes expressed in decibel units. To convert the vertical log-magnitude scale of Fig. 9.4 to decibel units, it is necessary only to multiply by a factor of 20. In using decibel

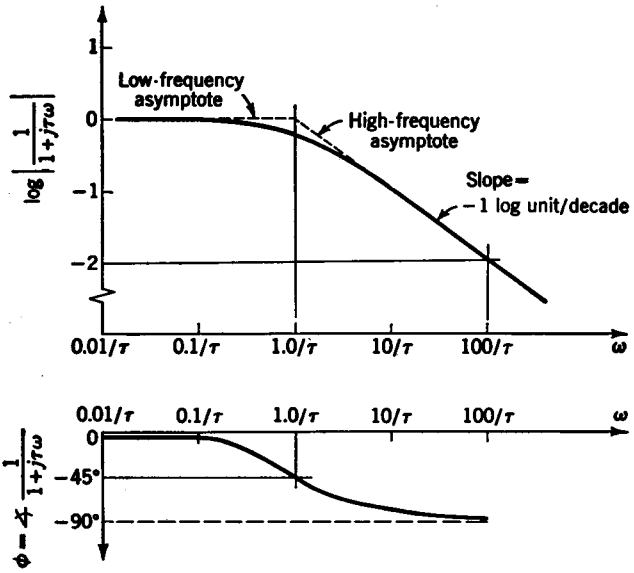


FIG. 9.5. log-magnitude plot for $\dfrac{1}{1 + j\tau\omega}$.

units the slope is -20 db/decade rather than -1 log unit/decade. In this text, the log-magnitude scale is expressed directly in logarithmic units rather than decibels.

The contribution due to $\log |1/(1 + j\tau\omega)|$ is shown in Fig. 9.5. For small values of ω such that $\omega \ll 1/\tau$,

$$-\tfrac{1}{2} \log (1 + \tau^2\omega^2) \approx -\tfrac{1}{2} \log 1 = 0 \tag{9.26}$$

This is the equation for the low-frequency asymptote to the exact curve, as is shown in Fig. 9.5. For $\omega \gg 1/\tau$,

$$-\tfrac{1}{2} \log (1 + \tau^2\omega^2) \approx -\tfrac{1}{2} \log \tau^2\omega^2 = -\log \tau\omega \tag{9.27}$$

This is the high-frequency asymptote. When $\omega = 1/\tau$, the value of $-\log \tau\omega$ is zero, and for $\omega = 10/\tau$, the function is -1, etc. The slope of this high-frequency asymptote is thus seen to be -1 log unit/decade.

For most preliminary design work, the asymptotes are sufficiently close to the exact curve so that the extra effort involved in using the exact curve is generally not warranted. The maximum error occurs at the "break frequency" ($\omega = 1/\tau$) and is $0 - \frac{1}{2}\log(1 + 1) = -0.303/2 = -0.1515$. It should be noticed that Fig. 9.5 is applicable for any value of τ. Thus, the same curve may be used to represent various functions of the form $(1 + \tau_1 s)(1 + \tau_2 s) \cdots$ by using a separate scale for each break frequency,

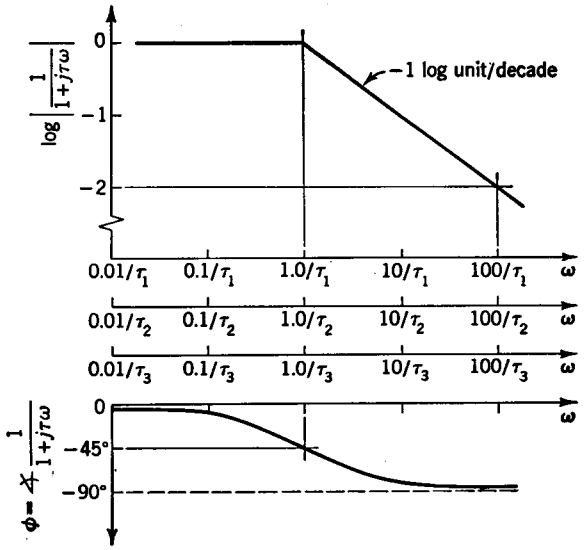


FIG. 9.6. General plot for $\dfrac{1}{1 + j\tau\omega}$.

as is shown in Fig. 9.6. The break frequency is located directly under the intersection of the two asymptotes.

The phase-angle curve is obtained by solving the equation

$$\phi = -\tan^{-1}\tau\omega \tag{9.28}$$

In Fig. 9.7, it is shown graphically how Figs. 9.4 and 9.5 may be added to solve Eqs. (9.24) and (9.25). For numerical purposes, it is assumed that the value of τ is 0.1 sec. For $\omega > 1/\tau = 10$, the slope is the sum of that due to the $1/j\omega$ term plus that due to the high-frequency asymptote of $1/(1 + j\tau\omega)$, that is, -2 log units/decade. Such graphs of the log magnitude vs. log frequency are called log-magnitude or Bode diagrams. H. W. Bode made many contributions to the development of frequency-response techniques.

A feature of the logarithmic method is that, if the term appears in the numerator rather than the denominator, it is necessary merely to change

the sign of the amplitude and phase-angle scales in tabulating the result. For example, if Eq. (9.22) were of the form

$$G(j\omega) = \frac{K(1 + j\tau\omega)}{j\omega} \tag{9.29}$$

then $$\log |G(j\omega)| - \log K = -\log \omega + \tfrac{1}{2} \log (1 + \tau^2\omega^2) \tag{9.30}$$

and $$\measuredangle G(j\omega) = -90^\circ + \tan^{-1} \tau\omega \tag{9.31}$$

By comparing Eqs. (9.30) and (9.31) with Eqs. (9.24) and (9.25), it is seen that only the sign of the term which went from the denominator to the numerator has changed. In Fig. 9.8 is shown the log-magnitude diagram for $G(j\omega)/K = (1 + j\tau\omega)/j\omega$, in which the value of τ is assumed to be 0.1 sec. Comparison of the log-magnitude diagram for $1 + j\tau\omega$ of Fig. 9.8 with that for $1/(1 + j\tau\omega)$ of Fig. 9.7 shows that the sign of the

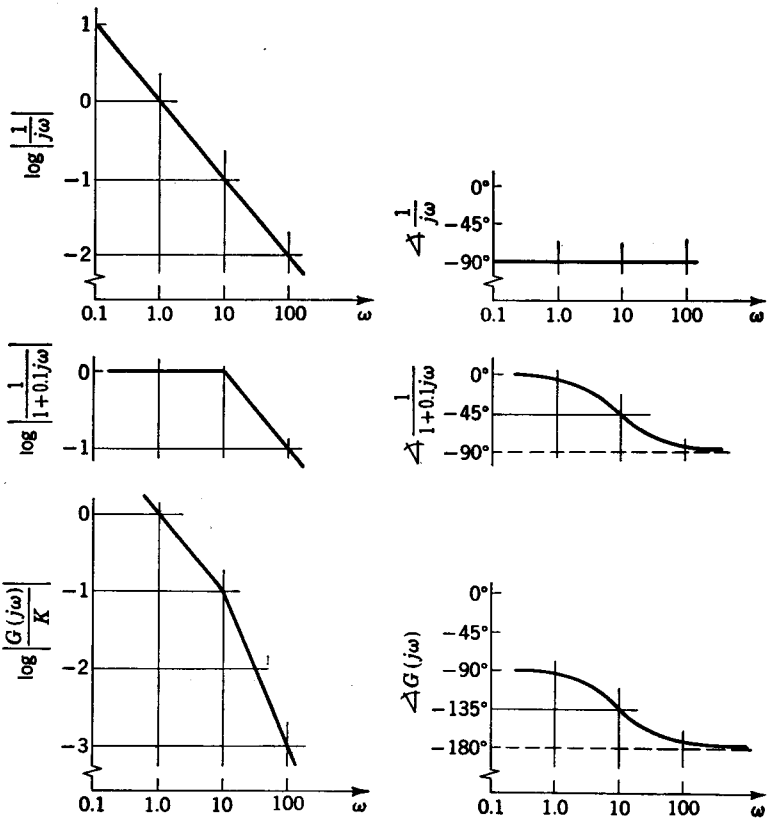


FIG. 9.7. log-magnitude plot for $\dfrac{G(j\omega)}{K} = \dfrac{1}{j\omega(1 + j0.1\omega)}$.

$j\omega - .1\omega^2$

$-\tan^{-1}\frac{10}{-\omega}$ $-\tan 135$

$-90^\circ - \tan$

logarithm of the amplitude ratio and the sign of the phase angle have been changed.

The third type of term which occurs is of the form $(\omega_n^2 - \omega^2 + j2\zeta\omega_n\omega)/\omega_n^2$. By using a generalized graph, this term may be treated in a manner analogous to that described for using Fig. 9.6 to evaluate terms of the form $1 + j\tau\omega$. Consider the function

$$G(j\omega) = \frac{K_1}{(1 + j\tau\omega)(\omega_n^2 - \omega^2 + j2\zeta\omega_n\omega)} \tag{9.32}$$

This may be rewritten in the form

$$G(j\omega) = \frac{K}{(1 + j\tau\omega)\left[\dfrac{(\omega_n^2 - \omega^2) + j2\zeta\omega_n\omega)}{\omega_n^2}\right]} \tag{9.33}$$

where $K = K_1/\omega_n^2$.

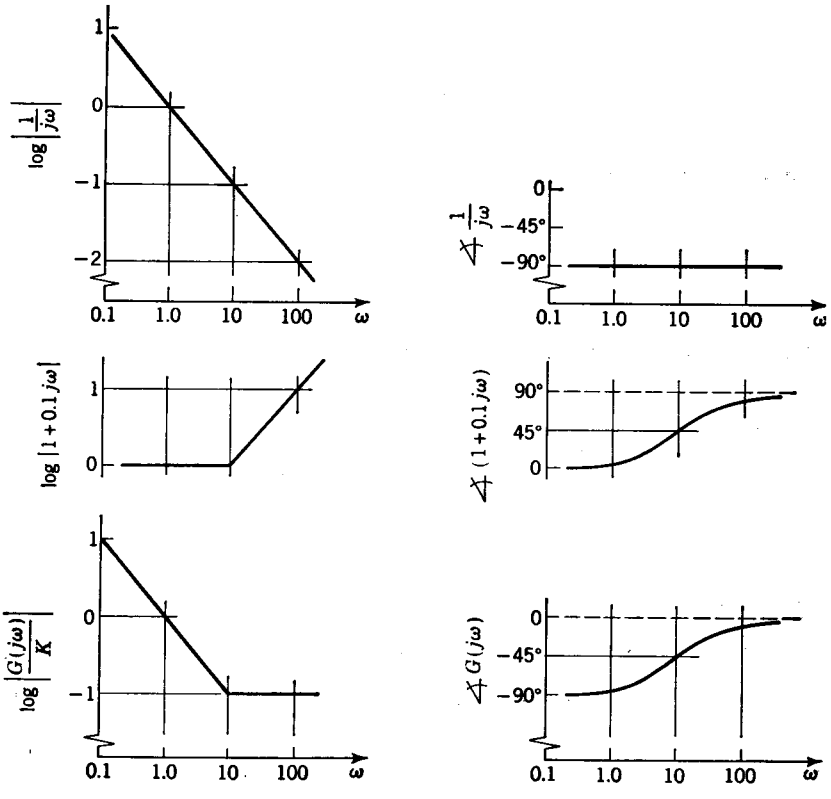


FIG. 9.8. log-magnitude plot for $\dfrac{1 + j0.1\omega}{j\omega}$.

The logarithm of the magnitude of the quadratic term is

$$\log \left| \frac{1}{[1 - (\omega/\omega_n)^2] + j2\zeta(\omega/\omega_n)} \right| = \log \frac{1}{\sqrt{[1 - (\omega/\omega_n)^2]^2 + [2\zeta(\omega/\omega_n)]^2}}$$
$$= -\frac{1}{2} \log \left\{ \left[1 - \left(\frac{\omega}{\omega_n} \right)^2 \right]^2 + \left(2\zeta \frac{\omega}{\omega_n} \right)^2 \right\} \quad (9.34)$$

For small values of ω/ω_n such that $\omega/\omega_n \ll 1$, Eq. (9.34) becomes

$$-\tfrac{1}{2} \log 1 = 0 \quad (9.35)$$

This is the equation for the low-frequency asymptote which is a horizontal straight line like that obtained for a $1 + j\tau\omega$ term.

The equation for the high-frequency asymptote is obtained by noting that, for $\omega/\omega_n \gg 1$, in Eq. (9.34) it follows that

$$\left[1 - \left(\frac{\omega}{\omega_n} \right)^2 \right]^2 \approx \left(\frac{\omega}{\omega_n} \right)^4 \gg \left(2\zeta \frac{\omega}{\omega_n} \right)^2$$

Thus, for $\omega/\omega_n \gg 1$, Eq. (9.34) becomes

$$-\frac{1}{2} \log \left(\frac{\omega}{\omega_n} \right)^4 = -2 \log \frac{\omega}{\omega_n} \quad (9.36)$$

The slope of the high-frequency asymptote is -2 log units/decade, and this asymptote intersects the low-frequency asymptote at $\omega/\omega_n = 1$.

The value of the phase angle is

$$\phi = -\tan^{-1} \frac{2\zeta(\omega/\omega_n)}{1 - (\omega/\omega_n)^2} \quad (9.37)$$

The nondimensional curves for the logarithm of the amplitude and the phase angle as given by Eqs. (9.34) and (9.37), respectively, are shown in Fig. 9.9. The curves for the reciprocal of this function are obtained by merely changing the sign of the amplitude and phase-angle scales.

In summary, then, for any function which is composed of multiples of terms such as $j\omega$, $1 + j\tau\omega$, and $(\omega_n^2 - \omega^2 + j2\zeta\omega_n\omega)/\omega_n^2$ the value $\log G(j\omega) - \log K$ and the $\measuredangle G(j\omega)$ at any given angular velocity ω of the driving sine wave is the sum of the contribution due to each term which is obtained from Fig. 9.4, 9.6, or 9.9. Thus, having the $|G(j\omega)|$ and the $\measuredangle G(j\omega)$, one may construct a polar plot of $|G(j\omega)|$, as was demonstrated by Fig. 9.3.

Experimental Determination of Frequency Response.[1] A feature of frequency-response methods is that the response $G(j\omega)$ and $H(j\omega)$ may be determined experimentally. For example, at a given frequency ω, the

[1] R. A. Bruns and R. M. Saunders, "Analysis of Feedback Control Systems," chap. 14, McGraw-Hill Book Company, Inc., New York, 1955.

value of $G(j\omega)$ is obtained by exciting the feedforward elements with a sinusoidal input of angular velocity ω and then measuring the ratio of the amplitude of the output to that of the input and also measuring the phase angle ϕ. By repeating this process for a wide range of values of ω

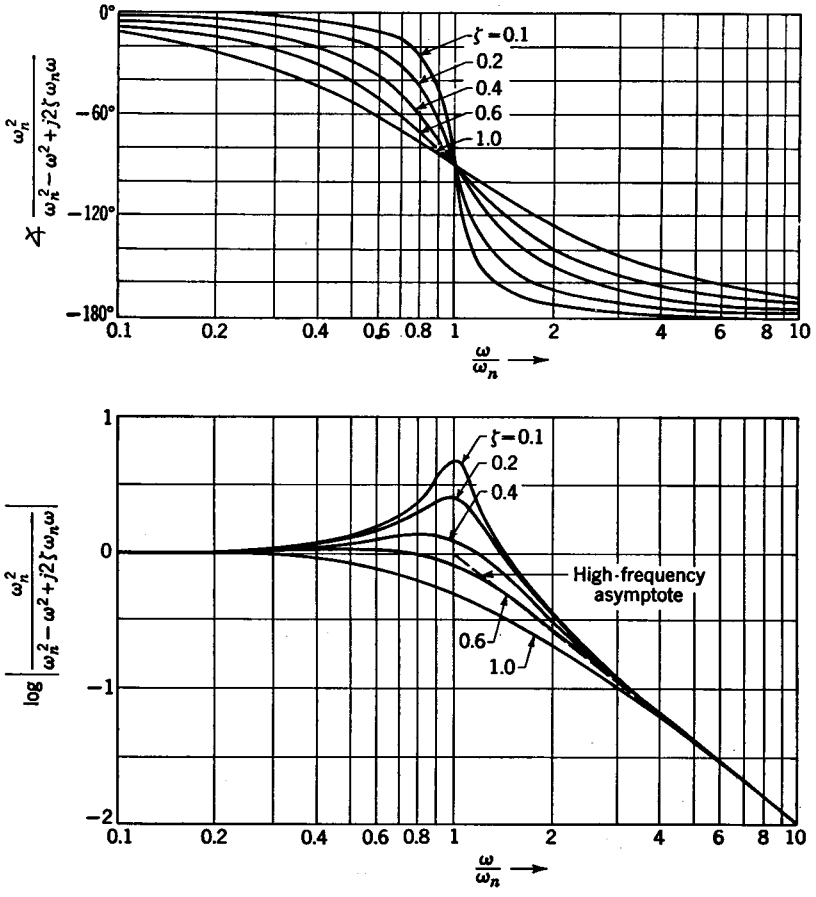


FIG. 9.9. log-magnitude plot for $\dfrac{\omega_n^2}{\omega_n^2 - \omega^2 + j2\zeta\omega_n\omega} = \dfrac{1}{1 - (\omega/\omega_n)^2 + j2\zeta(\omega/\omega_n)}$.

the frequency response is obtained. The response $H(j\omega)$ is similarly obtained by sinusoidally exciting the feedback elements.

In Fig. 9.10 are shown the asymptotes of an experimentally determined log-magnitude plot. Because of the change of slope at ω_1, there is a term $1/(1 + j\tau_1\omega)$ where $\tau_1 = 1/\omega_1$ in the frequency-response equation. At the angular velocity ω_2 there is a net increase of $+1$ log units/decade, and thus the term $1 + j\tau_2\omega$, where $\tau_2 = 1/\omega_2$ appears in the numerator of the

response expression. The change in slope at ω_3 is indicative of the term $1/(1 + j\tau_3\omega)$, where $\tau_3 = 1/\omega_3$. Because the slope changes by -2 log units/decade at ω_4, there is a quadratic term in the denominator. The value of the break frequency ω_4 is equal to the natural frequency ω_n for the quadratic. That is, from Fig. 9.9, it follows that at the break point for a quadratic $\omega/\omega_n = \omega_4/\omega_n = 1$ or $\omega_n = \omega_4$. To determine the damping ratio ζ, it is necessary to compare the exact response curve for the component which causes this quadratic term in $G(j\omega)$ to the general response curves of Fig. 9.9. From the preceding, it follows that the frequency response for $G(j\omega)$ is

$$G(j\omega) = \frac{K(1 + j\tau_2\omega)}{(1 + j\tau_1\omega)(1 + j\tau_3\omega)[(\omega_4^2 - \omega^2 + j2\zeta\omega_4\omega)/\omega_4^2]} \tag{9.38}$$

In the next section, it is shown how the value of K can be determined directly from the low-frequency portion of the log-magnitude diagram.

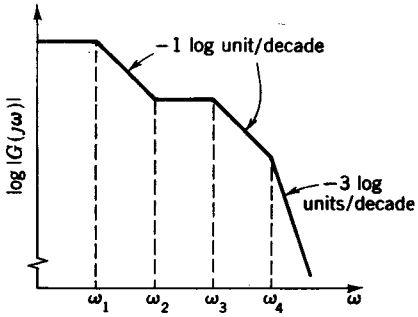


FIG. 9.10. Experimentally determined log-magnitude plot.

It is a rather lengthy process to excite control elements at various frequencies in order to obtain experimentally the frequency response. An alternative procedure is to determine the transient response to some known input and then use the method presented in Appendix III for calculating the frequency response from the transient response.

After the equation for the frequency response has been experimentally determined, it is a simple matter to substitute s for $j\omega$ to obtain the transfer function. The substitution of s for $j\omega$ in Eq. (9.38) gives

$$G(s) = \frac{K(1 + \tau_2 s)}{(1 + \tau_1 s)(1 + \tau_3 s)[(s^2 + 2\zeta\omega_4 s + \omega_4^2)/\omega_4^2]} \tag{9.39}$$

Similarly, the substitution of p for s gives the differential equation of operation.

The magnitude of $1 + j\tau\omega$ is the same as that of $1 - j\tau\omega$. However, as the phase angle for $1 + j\tau\omega$ goes from 0 to $+90°$, the phase angle for $1 - j\tau\omega$ goes from 0 to $-90°$. Most systems are minimum-phase systems;[1] i.e., all factors are of the form $j\omega$, $1 + j\tau\omega$, or $(\omega_n^2 - \omega^2 + j2\zeta\omega_n\omega)/\omega_n^2$. For minimum-phase systems, as ω becomes infinite, the phase angle is $\phi = -90°(n - m)$, where n is the order of the denominator and m that

[1] J. L. Bower and P. M. Schultheiss, "Introduction to the Design of Servomechanisms," John Wiley & Sons, Inc., New York, 1958.

of the numerator. Non-minimum-phase systems may be detected from the phase-angle plot, because as ω becomes infinite $\phi \neq -90°(n - m)$. For either minimum or non-minimum-phase systems the slope of the log-magnitude diagram at high frequencies is $-(n - m)$ log units/decade.

9.4. Evaluating the Gain K. In general a transfer function may be expressed in the form

$$G(s) = \frac{K_n\{(1 + \tau_a s) \cdot \cdot \cdot [(s^2 + 2\zeta_a\omega_{n_a} s + \omega_{n_a}{}^2)/\omega_{n_a}{}^2] \cdot \cdot \cdot\}}{s^n\{(1 + \tau_1 s) \cdot \cdot \cdot [(s^2 + 2\zeta_1\omega_{n_1} s + \omega_{n_1}{}^2)/\omega_{n_1}{}^2] \cdot \cdot \cdot\}} \quad (9.40)$$

where K_n is the over-all gain of the transfer function $G(s)$ and n is the power to which the s term in the denominator is raised. Usually the value of n is 0, 1, or 2. The first time constant in the numerator is τ_a, the second τ_b, etc. The natural frequency for the first quadratic term in the numerator is ω_{n_a}, and the damping ratio is ζ_a. Similarly, τ_1 is the first time constant which appears in the denominator, the second τ_2, etc. The natural frequency for the first quadratic term in the denominator is ω_{n_1}, and its damping ratio is ζ_1.

The substitution of $j\omega$ for s in Eq. (9.40) gives

$$G(j\omega) = \frac{K_n\{(1 + j\tau_a\omega) \cdot \cdot \cdot [(\omega_{n_a}{}^2 - \omega^2 + j2\zeta_a\omega_{n_a}\omega)/\omega_{n_a}{}^2] \cdot \cdot \cdot\}}{(j\omega)^n\{(1 + j\tau_1\omega) \cdot \cdot \cdot [(\omega_{n_1}{}^2 - \omega^2 + j2\zeta_1\omega_{n_1}\omega)/\omega_{n_1}{}^2] \cdot \cdot \cdot\}} \quad (9.41)$$

For small values of ω all the terms inside the braces of the preceding expression approach 1, so that

$$G(j\omega) = \frac{K_n}{(j\omega)^n} \qquad \omega \approx 0 \quad (9.42)$$

As is indicated from the preceding expression, the gain K_n can be determined from the low-frequency portion of a log-magnitude plot. Equation (9.42) is valid for minimum- as well as non-minimum-phase systems. All the frequency-response techniques to be discussed in this and the next chapter are equally valid for minimum- or non-minimum-phase systems. The techniques for evaluating K_n for $n = 0$, 1, or 2 are described in the following.

$n = 0$. When n is zero, there is no $j\omega$ term in $G(j\omega)$. Thus, Eq. (9.42) becomes

$$G(j\omega) = K_0 \qquad \omega \approx 0 \quad (9.43)$$

A typical log-magnitude plot for this case is shown in Fig. 9.11*a*. For small values of ω the low-frequency asymptote has a constant value $G(j\omega) = K_0$.

$n = 1$. For small values of ω and for n equal to 1, it follows from

Eq. (9.42) that

$$G(j\omega) = \frac{K_1}{j\omega} \qquad \omega \approx 0 \tag{9.44}$$

The logarithm of the magnitude of $G(j\omega)$ is

$$\log |G(j\omega)| = \log K_1 - \log \omega \qquad \omega \approx 0 \tag{9.45}$$

From Eq. (9.45), it follows that, as ω_1 changes by a factor of 10, then $\log |G(j\omega)|$ changes by -1. Thus for $\omega \approx 0$ the slope of the curve of

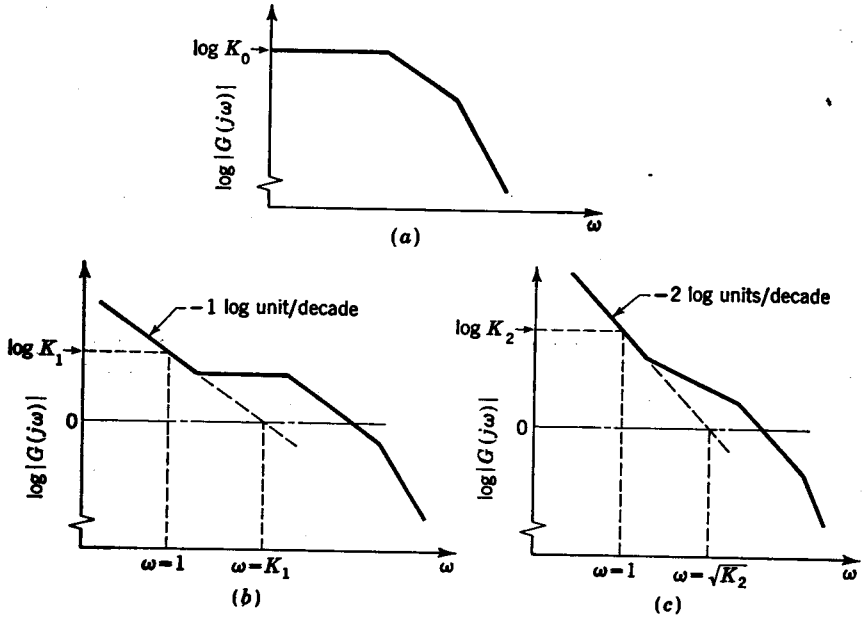


FIG. 9.11. log-magnitude plot. (*a*) $n = 0$; (*b*) $n = 1$; (*c*) $n = 2$.

$\log |G(j\omega)|$ versus ω is -1 log unit/decade. A typical log-magnitude diagram for the case in which $n = 1$ is shown in Fig. 9.11*b*. It is to be noted that the low-frequency slope of -1 log unit/decade or its extension intersects the horizontal axis $\log |G(j\omega)| = 0$ at the point where $\omega = K_1$. This fact follows directly from Eq. (9.45) by noting that, when $\log |G(j\omega)| = 0$, then $\log \omega = \log K_1$ or simply $\omega = K_1$. In addition, for ω equal to 1, Eq. (9.45) becomes $\log |G(j\omega)| = \log K$. Thus, as is shown in Fig. 9.11*b*, a vertical line through $\omega = 1$ intersects the low-frequency asymptote or its extension at the value $\log K_1$.

$n = 2$. The low-frequency equation for this case is

$$G(j\omega) = \frac{K_2}{(j\omega)^2} = -\frac{K_2}{\omega^2} \tag{9.46}$$

The logarithm of the magnitude of $G(j\omega)$ is

$$\log |G(j\omega)| = \log K_2 - 2 \log \omega \tag{9.47}$$

A typical log-magnitude diagram for $n = 2$ is illustrated in Fig. 9.11*c*. The slope at low frequencies is -2 log units/decade. From Eq. (9.47), it follows that, when $\log |G(j\omega)| = 0$, then $\log \omega = \frac{1}{2} \log K_2$ or $\omega = \sqrt{K_2}$. Thus, the low-frequency asymptote or its extension intersects the horizontal line $\log |G(j\omega)| = 0$ at the frequency $\omega = \sqrt{K_2}$. When $\omega = 1$, then Eq. (9.47) becomes $\log |G(j\omega)| = \log K_2$. Thus, a vertical line through $\omega = 1$ intersects the low-frequency asymptote or its extension at $\log K_2$.

9.5. Equivalent Unity-feedback System. Much simplification is afforded in the application of frequency-response methods to systems having unity feedback. A control system having feedback elements $H(p)$ can usually be represented by an equivalent unity-feedback system, as is illustrated in Fig. 9.12. For the case in which $H(p)$ is a constant, the equivalent unity-feedback system is readily obtained by moving the constant $H(p)$ to the input side of the main loop. The systems represented by Figs. 3.17, 4.7*b*, and 4.11 have a constant term $H(p) = C_4$ in the feedback path. Moving C_4 to the input side of the main loop yields the unity-feedback systems shown in Fig. 9.13*a*, *b*, and *c*, respectively.

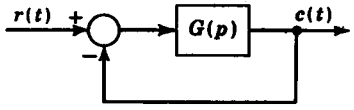


FIG. 9.12. Unity-feedback system.

To obtain the equivalent unity-feedback system when $H(p)$ is not a constant, first write $H(p)$ in the form $H(p) = C[1 + H_1(p)]$. The constant C may now be taken out of the feedback path, and the remaining term $1 + H_1(p)$ may be represented as shown in Fig. 9.14 by two separate paths. The design of such systems in which there is an inner, or minor, feedback path is often facilitated by the use of inverse polar plots, as is discussed in Sec. 10-5.

In obtaining the equivalent unity-feedback system, only constant terms are to be taken outside the main loop. The fact that $r(t)$ is equal to some constant times the command signal does not affect the basic dynamic behavior of the system.

System Type. When a system is represented in its equivalent unity-feedback form, the value of n in $G(s)$ as indicated by Eq. (9.40) has a predominant effect upon the behavior of the system. When $n = 0$, the system is designated as a type 0 system. A type 1 system is one for which $n = 1$, a type 2 system is one for which $n = 2$, etc.

A type 0 system results when there is no integration as in a proportional-type control. For an integral-type control in which there is one integrator in the feedforward elements, n is equal to 1. A type 2 system has two integrations in the feedforward elements, etc.

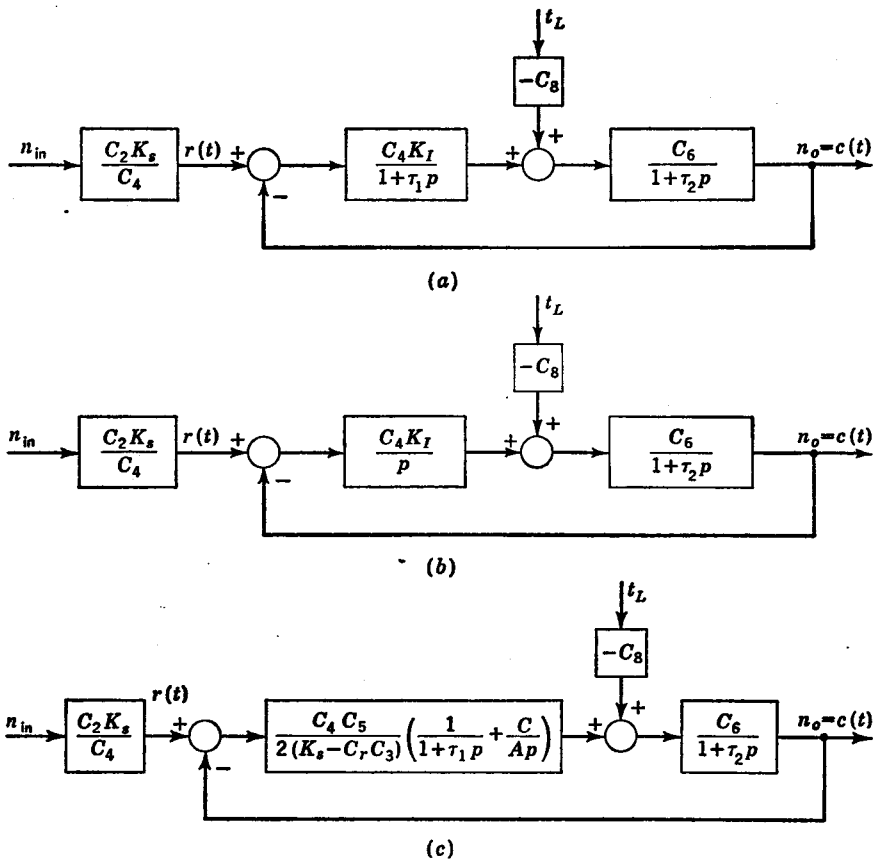


FIG. 9.13. Equivalent unity-feedback systems when $H(p)$ is a constant.

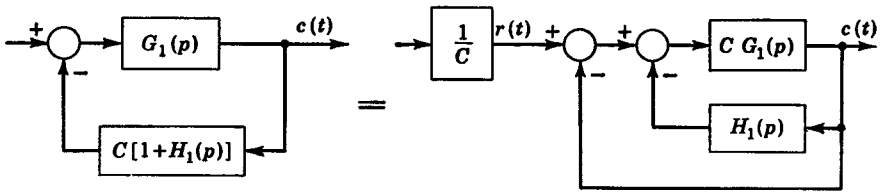


FIG. 9.14. Equivalent unity-feedback system when $H(p)$ is not a constant.

9.6. Polar Plots. Vector loci or polar plots are better suited for the solution of certain control problems than are log-magnitude diagrams, and vice versa. As is later explained, other methods of representing frequency-response information are the log-modulus, or Nichols, plot and the inverse polar plot. A control engineer must be familiar with all these means of

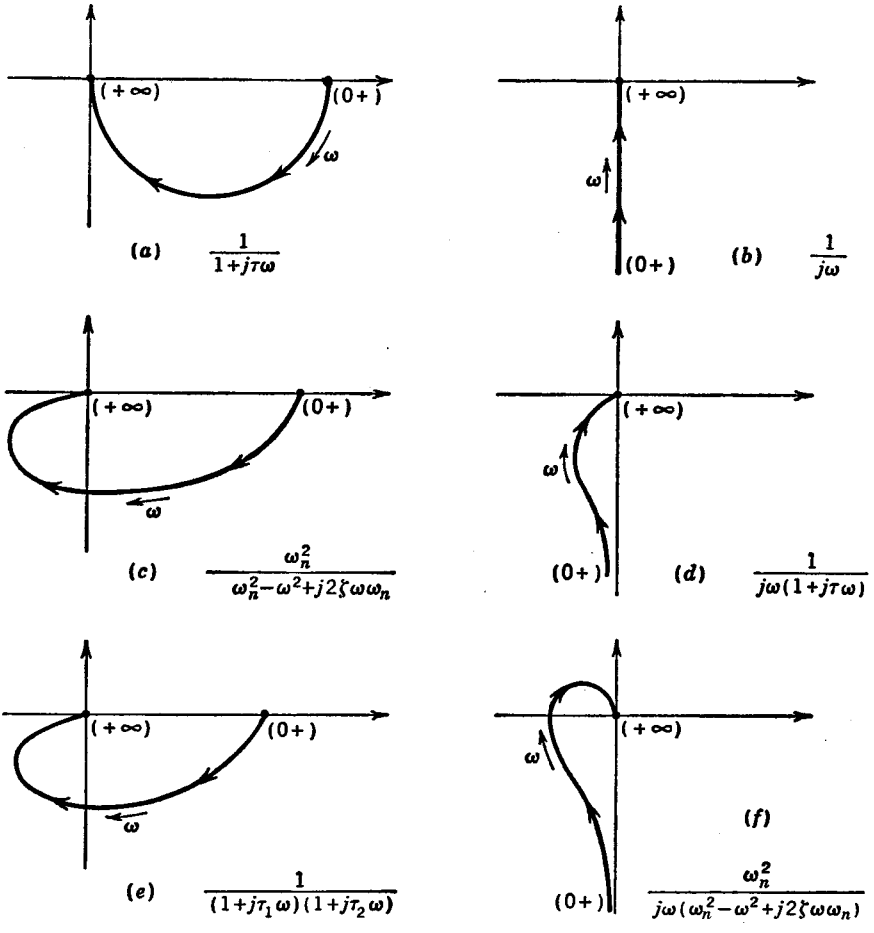


FIG. 9.15. Common polar plots.

plotting frequency-response data so that he can select the method which is best suited to his particular problem.

The polar plots for a number of commonly encountered functions are shown in Fig. 9.15.

Polar plots may often be roughly sketched by knowing the location at low frequencies ($\omega \to 0$) and at high frequencies ($\omega \to \infty$).

For example, in Fig. 9.15*a*

$$\frac{1}{j} = e^{-j90^\circ}$$

$$\left.\frac{1}{1+j\tau\omega}\right|_{\omega=0+} = 1$$

$$\left.\frac{1}{1+j\tau\omega}\right|_{\omega=+\infty} \approx \left.\frac{1}{j\tau\omega}\right|_{\omega=+\infty} = \left.\frac{e^{-j90^\circ}}{\tau\omega}\right|_{\omega=+\infty} = (0+)e^{-j90^\circ} \quad (9.48)$$

Thus, the locus of Fig. 9.15*a* begins at the +1 point on the positive real axis, and as ω approaches ∞, the locus approaches 0 along the $-90°$ axis.

The low- and high-frequency values for Fig. 9.15*b* are

$$\left.\frac{1}{j\omega}\right|_{\omega=0+} = \left.\frac{1}{\omega}e^{-j90°}\right|_{\omega=0+} = (+\infty)e^{-j90°}$$
$$\left.\frac{1}{j\omega}\right|_{\omega=+\infty} = \left.\frac{1}{\omega}e^{-j90°}\right|_{\omega=+\infty} = (0+)e^{-j90°} \tag{9.49}$$

For Fig. 9.15*c*

$$\left.\frac{1}{1-(\omega/\omega_n)^2+j2\zeta(\omega/\omega_n)}\right|_{\omega=0+} = 1$$
$$\left.\frac{1}{1-(\omega/\omega_n)^2+j2\zeta(\omega/\omega_n)}\right|_{\omega=+\infty} \approx \left.\frac{1}{-(\omega/\omega_n)^2}\right|_{\omega=+\infty} = -\left.\left(\frac{\omega_n}{\omega}\right)^2\right|_{\omega=+\infty}$$
$$= (0+)e^{-j180°} \tag{9.50}$$

The limiting value for a function composed of multiples of the preceding terms is equal to the product of the contribution of each term. For example, in Fig. 9.15*d*

$$\left.\frac{1}{j\omega(1+j\tau\omega)}\right|_{\omega=0+} = \left.\frac{1}{j\omega}\right|_{\omega=0+} \times \left.\frac{1}{1+j\tau\omega}\right|_{\omega=0+} = (+\infty)e^{-j90°}$$
$$\left.\frac{1}{j\omega(1+j\tau\omega)}\right|_{\omega=+\infty} = \left.\frac{1}{j\omega}\right|_{\omega=+\infty} \times \left.\frac{1}{1+j\tau\omega}\right|_{\omega=+\infty} = (0+)e^{-j90°}(0+)e^{-j90°}$$
$$= (0+)e^{-j180°} \tag{9.51}$$

Application of this technique will verify the results shown in Fig. 9.15*e* and *f*.

The low- and high-frequency locations are summarized in Table 9.1. The reciprocal of any term in Table 9.1 is obtained by changing the sign of

TABLE 9.1

ω	$\frac{1}{j\omega}$	$\frac{1}{1+j\tau\omega}$	$\frac{\omega_n^2}{\omega_n^2-\omega^2+j2\zeta\omega\omega_n}$
0+	$\infty/\underline{-90°}$	$1/\underline{0°}$	$1/\underline{0°}$
$+\infty$	$0/\underline{-90°}$	$0/\underline{-90°}$	$0/\underline{-180°}$

the phase angle and taking the reciprocal of the magnitude. At low frequencies, there is a phase shift with the $1/j\omega$ term only. Thus, for a type 0 system, the polar plot originates on the positive real axis, where ϕ is zero. Typical polar plots for type 0, 1, 2, and 3 systems are shown in

Fig. 9.16. A type 1 system begins at infinity on the negative 90° axis, a type 2 at infinity on the negative 180° axis, etc. For most systems, the polar plot terminates at the origin when $\omega = \infty$.

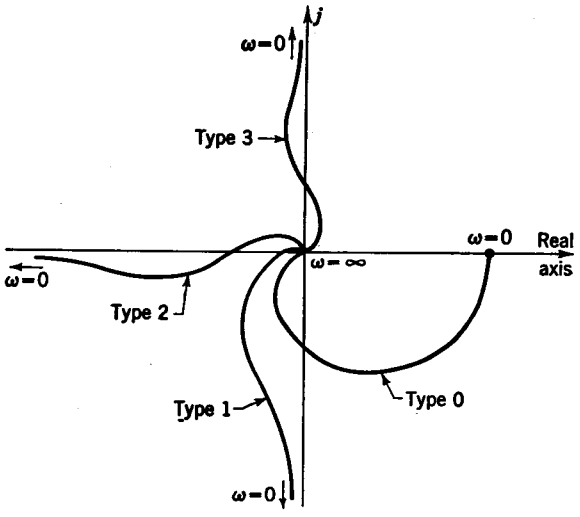


FIG. 9.16. Polar plots for type 0, 1, 2, and 3 systems.

9.7. Determining the Closed-loop Frequency Response from the Open-loop Response. Frequency-response methods make extensive use of the open-loop frequency response $G(j\omega)$. The open-loop response is the response that would be obtained if the feedback path were disconnected at the comparator (i.e., opened).

For unity-feedback systems, the closed-loop frequency response

$$c(j\omega)/r(j\omega)$$

is related to the open-loop response by the equation

$$\frac{c(j\omega)}{r(j\omega)} = \frac{G(j\omega)}{1 + G(j\omega)} \qquad (9.52)$$

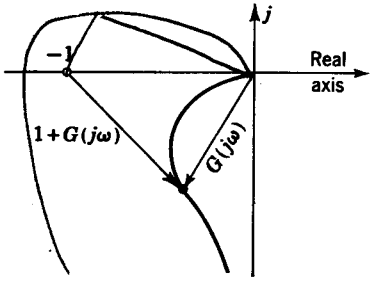


FIG. 9.17. Determination of the closed-loop frequency response from the open-loop response.

The preceding expression has meaning for a stable system only. For an unstable system $c(j\omega)/r(j\omega)$ becomes infinite. Whether or not a system is stable may be determined by the application of Routh's criterion or from the Nyquist stability criterion, which is discussed in the next chapter. In Fig. 9.17 is shown a typical $G(j\omega)$ plot. The vector from the origin to a point on the curve is $G(j\omega)$, and the vector from the point $-1 + j0$ to the

same point on this curve is $1 + G(j\omega)$. The ratio of these two vectors is the closed-loop frequency response for the value of ω at that point. This shows that every point on the $G(j\omega)$ plane corresponds to a certain value of $c(j\omega)/r(j\omega)$. The magnitude of the ratio of the amplitude of the output sinusoid to the input is designated by the symbol $M = |c(j\omega)/r(j\omega)|$.

In Fig. 9.18, it is to be seen that the locus of lines of constant M are circles on the $G(j\omega)$ plane. The proof of this follows: Consider any point

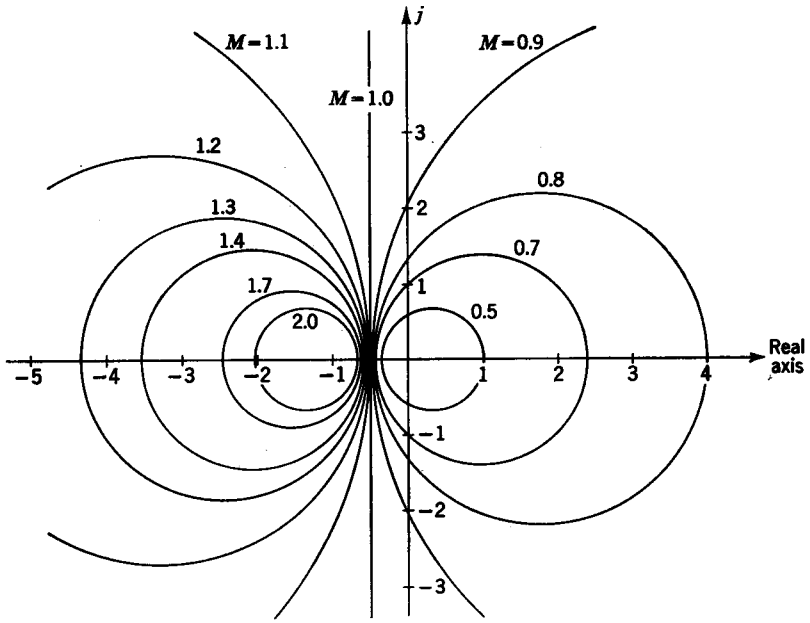


FIG. 9.18. Constant-M circles.

$G(j\omega) = x + jy$ in the $G(j\omega)$ plane of Fig. 9.18. The closed-loop frequency response is

$$\frac{c(j\omega)}{r(j\omega)} = \frac{x + jy}{1 + x + jy} \tag{9.53}$$

The magnitude of the preceding equation is

$$M = \left|\frac{c(j\omega)}{r(j\omega)}\right| = \left(\frac{x^2 + y^2}{1 + 2x + x^2 + y^2}\right)^{1/2} \tag{9.54}$$

Squaring and cross multiplying gives

$$x^2(M^2 - 1) + 2xM^2 + y^2(M^2 - 1) = -M^2 \tag{9.55}$$

Dividing by $M^2 - 1$ and completing the square by adding $M^4/(M^2 - 1)^2$

to both sides yields

$$x^2 + \frac{2xM^2}{M^2 - 1} + \frac{M^4}{(M^2 - 1)^2} + y^2 = \frac{M^4}{(M^2 - 1)^2} - \frac{M^2}{M^2 - 1} \tag{9.56}$$

Thus,

$$\left(x + \frac{M^2}{M^2 - 1}\right)^2 + y^2 = \frac{M^2}{(M^2 - 1)^2} \tag{9.57}$$

Equation (9.57) is the equation of a circle as shown in Fig. 9.19 with center at

$$x = \frac{M^2}{1 - M^2} \qquad y = 0 \tag{9.58}$$

and radius

$$r = \left| \frac{M}{M^2 - 1} \right| \tag{9.59}$$

The closed-loop frequency response may be expressed in the form

$$\frac{c(j\omega)}{r(j\omega)} = Me^{j\alpha} \tag{9.60}$$

where $M = |c(j\omega)/r(j\omega)|$ and $\alpha = \measuredangle c(j\omega)/r(j\omega)$.

The loci of lines of constant phase angle α for the closed loop response are also circles. The circles of constant α are shown in Fig. 9.20. The centers of these circles are located at the point

$$x = -\tfrac{1}{2} \qquad y = 1/(2N) \tag{9.61}$$

where $N = \tan \alpha$. The radius of each circle is

$$r = \frac{1}{2N}\sqrt{N^2 + 1} \tag{9.62}$$

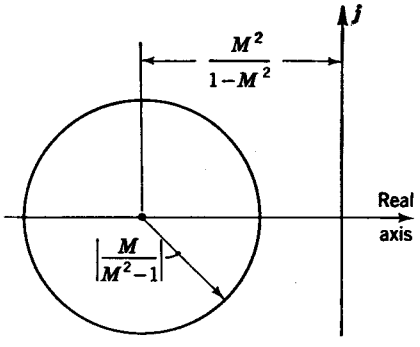


FIG. 9.19. Typical M circle.

9.8. Correlation between Transient and Frequency Response. The transient response of a system can be ascertained directly from the frequency response. For example, for a second-order system, the differential equation of operation is

$$c(t) = \frac{\omega_n^2}{p^2 + 2\zeta\omega_n p + \omega_n^2} r(t) \tag{9.63}$$

The substitution of $j\omega$ for p yields the following equation for the closed-loop frequency response of a second-order system:

$$\frac{c(j\omega)}{r(j\omega)} = \frac{\omega_n^2}{\omega_n^2 - \omega^2 + j2\zeta\omega\omega_n} = \frac{1}{1 - (\omega/\omega_n)^2 + j2\zeta(\omega/\omega_n)} \tag{9.64}$$

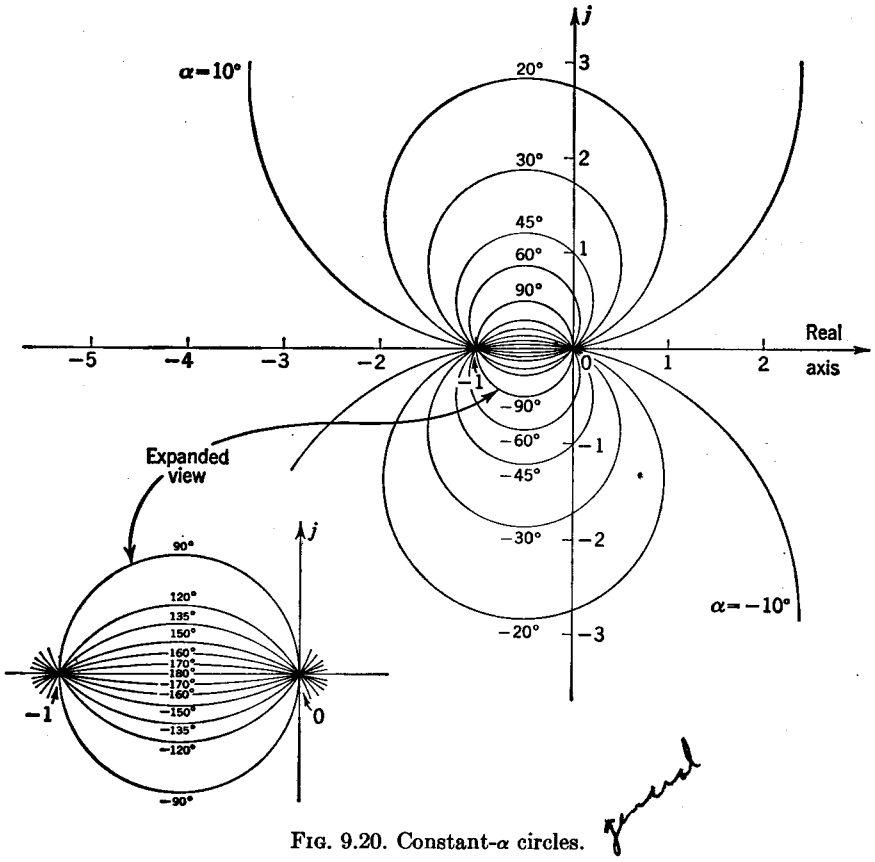


FIG. 9.20. Constant-α circles.

The ratio of the amplitude of the output to the input is

$$M = \left|\frac{c(j\omega)}{r(j\omega)}\right| = \frac{1}{\{[1 - (\omega/\omega_n)^2]^2 + [2\zeta(\omega/\omega_n)]^2\}^{\frac{1}{2}}} \tag{9.65}$$

The angular velocity $\omega = \omega_m$ at which the amplitude ratio M becomes a maximum is obtained by differentiating Eq. (9.65) with respect to ω and equating this result to zero. Performing this operation yields

$$\omega_m = \omega_n \sqrt{1 - 2\zeta^2} \qquad 0 \leq \zeta \leq 0.707 \tag{9.66}$$

where ω_m is the angular velocity at which the maximum value M_m occurs.

The substitution of ω_m from Eq. (9.66) for ω in Eq. (9.65) gives

$$M_m = \frac{1}{2\zeta\sqrt{1 - \zeta^2}} \qquad 0 \leq \zeta \leq 0.707 \tag{9.67}$$

The preceding result has significance only for $0 \leq \zeta \leq 0.707$, in which

case $M_m \geq 1$. In Fig. 9.21 is shown a plot of corresponding values of M_m and ζ.

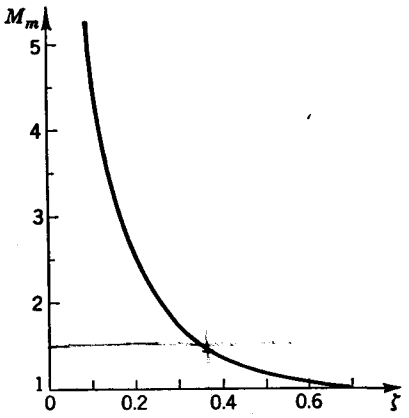


FIG. 9.21. M_m versus ζ for a second-order system.

As previously discussed, the zeros of the characteristic function which are located nearest the imaginary axis have a predominant effect upon the transient behavior of higher-order systems. Thus, the transient behavior of a higher-order system for which $M_m > 1$ (the value of M_m may be obtained from a polar plot for the system) may be *approximated* by a second-order system whose damping ratio ζ as obtained from Fig. 9.21 corresponds to the value of M_m for the system. For the case in which $M_m < 1$, the transient behavior is better *approximated* by a first-order system, as is soon to be discussed.

Illustrative Example. In Fig. 9.22 is shown the polar plot for a third-order system with unity feedback in which $G(s) = 80/[s(s^2 + 10s + 26)]$. By superimposing the polar plot for $G(j\omega)$ on a family of M circles as shown in Fig. 9.22, it is found that $M_m = 1.6$. From Fig. 9.21, it follows that, for $M_m = 1.6$, the corresponding value of ζ is 0.33. Thus, the transient behavior of this system may be approximated by a second-order system whose damping ratio ζ is 0.33.

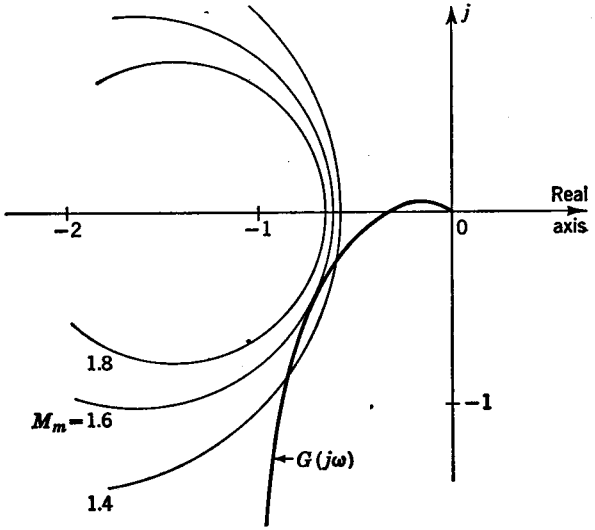


FIG. 9.22. Determination of M_m from polar plot.

The horizontal coordinate in Figs. 6.7 and 6.12 is $\omega_n t$, and thus, for a second-order system, the value of ω_n determines the speed of response. For a given damping ratio, it follows from Eq. (9.66) that ω_m is proportional to ω_n. Therefore, the value of ω_m may be used to provide a measure or indication of the speed of response.

When M_m is less than 1 the transient response can be approximated by an equivalent first-order system. It is to be noticed from Fig. 6.7 that, when $\zeta > 0.707$, there is no overshoot of the response to a step change in the input. This is similar to the type of response that is obtained from a first-order system. The correlation which exists between the frequency and transient response of a first-order system is obtained as follows: A first-order system is one for which

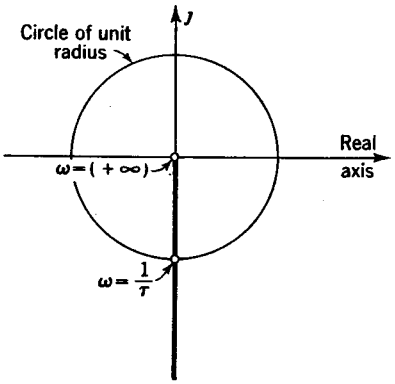


FIG. 9.23. Polar plot for $G(j\omega) = \frac{1}{j\tau\omega}$.

$$\frac{c(j\omega)}{r(j\omega)} = \frac{1}{1 + j\tau\omega} \tag{9.68}$$

Writing the preceding in terms of $G(j\omega)$ gives

$$\frac{c(j\omega)}{r(j\omega)} = \frac{G(j\omega)}{1 + G(j\omega)} = \frac{1}{1 + j\tau\omega}$$

Solving for $G(j\omega)$ yields

$$G(j\omega) = \frac{1}{j\tau\omega} \tag{9.69}$$

The polar plot for $G(j\omega) = 1/j\tau\omega$ is shown in Fig. 9.23. When the magnitude of $G(j\omega)$ is 1, then $\omega = 1/\tau$. For a first-order system, the value of ω at which $G(j\omega)$ has a length of 1 is the reciprocal of the time constant $\tau = 1/\omega$. Thus, for systems in which $M_m < 1$, an indication of the equivalent time constant τ may be obtained by taking the reciprocal of the angular velocity ω at which the magnitude of $G(j\omega)$ is 1. This value of τ completely determines the response for a first-order system. The preceding techniques for estimating the transient behavior are adequate for most design purposes. In Appendix IV is shown an exact method for obtaining the transient response from the frequency response.

9.9. Determining the Gain K so that a System Will Have a Desired Value of M_m. In Fig. 9.24 is shown a typical polar plot of $G(j\omega)$. If the gain K of the original function is doubled, the value of $G(j\omega)$ is doubled at every point. As shown in Fig. 9.24, it is not necessary to change the shape of the polar plot, but merely to change the scale by multiplying the old scale by the factor 2. Values of this new scale are shown in parentheses.

In the preceding section, it was shown that, when M_m is greater than 1, the transient behavior may be approximated by a second-order system in which ζ is obtained from Fig. 9.21. For M_m less than 1, the transient behavior is not governed by the value of M_m. It is now shown how the gain K can be adjusted so that the polar plot $G(j\omega)$ will be tangent to any desired $M_m > 1$ circle. This, in effect, is determining the gain K so that the system will have a desired M_m.

From Fig. 9.25, the line drawn from the origin, tangent to the desired M_m circle at the point P, has an included angle of ψ. The value of $\sin \psi$ is

$$\sin \psi = \frac{|M_m/(M_m^2 - 1)|}{|M_m^2/(1 - M_m^2)|} = \frac{1}{M_m} \tag{9.70}$$

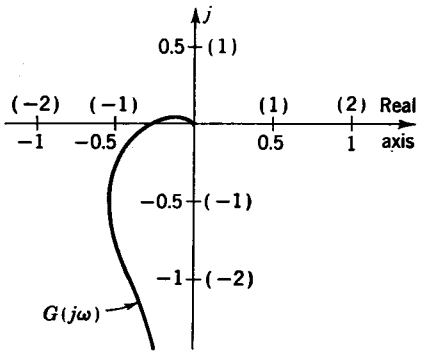


FIG. 9.24. Typical polar plot.

A characteristic feature of the point of tangency P is that a line drawn from the point P perpendicular to the negative real axis intersects this axis at the -1 point. This characteristic may be proved from the geometry of Fig. 9.25.

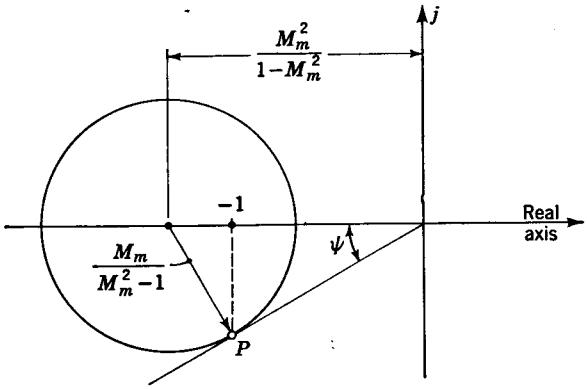


FIG. 9.25. Tangent to a M_m circle.

The procedure for determining the gain K so that $G(j\omega)$ will have a desired value of M_m is as follows:

1. Draw the polar plot for $G(j\omega)/K$.
2. Draw the tangent line to the desired M_m circle [Eq. (9.70)].
3. Draw the circle with center on the negative real axis that is tangent to both the $G(j\omega)/K$ plot and the tangent line, as is shown in Fig. 9.26.

4. Erect the perpendicular to the negative real axis from point P, the point of tangency of this circle and the tangent line. This perpendicular intersects the negative real axis at a value $-A + j0 = -A$.

5. In order that the circle drawn in step 3 correspond to the desired M_m circle, this point should be $-1 + j0 = -1$ rather than $-A$. The desired gain is that value of K which changes the scale so that this does become the -1 point; thus $K(-A) = -1$ or $K = 1/A$.

As is illustrated in Fig. 9.26, the perpendicular drawn from point P to the negative real axis intersects the negative real axis at a value of

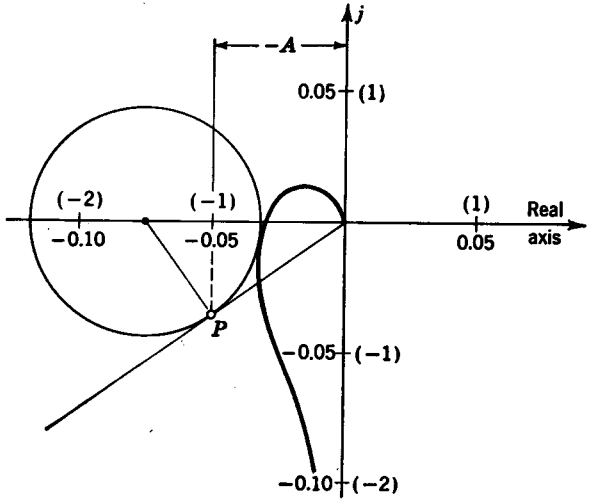


FIG. 9.26. Determination of K to yield a desired M_m.

-0.05. However, this value should be -1. Multiplication of the scale by a factor of 20 (that is, $-0.05 \times 20 = -1$), as is shown in Fig. 9.26 by the numbers in parentheses, converts this point to the -1 point. Thus, the original function should have a gain of 20 in order that all the necessary conditions be satisfied.

Illustrative Example. Let it be desired to determine the gain K such that a unity-feedback system for which $G(s) = K/[s(1 + 0.1s)]$ will have a maximum value $M_m = 1.4$.

SOLUTION. First construct the polar plot for

$$\frac{G(j\omega)}{K} = \frac{1}{j\omega(1 + 0.1j\omega)}$$

as is shown in Fig. 9.27. The value of ψ is obtained from Eq. (9.70).

$$\psi = \sin^{-1}\frac{1}{M_m} = \sin^{-1}\frac{1}{1.4} = 45.6° \tag{9.71}$$

By trial and error, the circle which is tangent to both the $G(j\omega)/K$ plot and

also to the line drawn at the angle $\psi = 45.6°$ is determined. The perpendicular drawn from point P intersects the negative real axis at the point $-A = -0.067$, or $A = 0.067$. Thus, the gain is $K = 1/0.067 = 16$.

9.10. log-modulus, or Nichols, Plots. In addition to log-magnitude and polar plots, another method of representing frequency-response

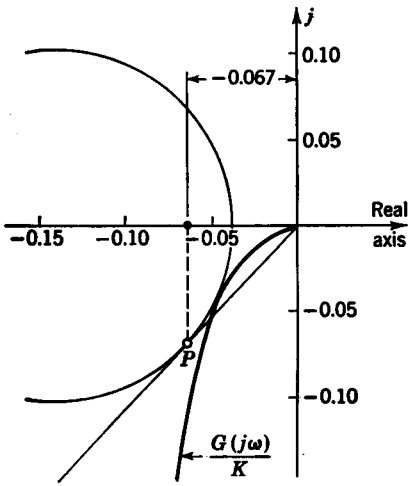


FIG. 9.27. Polar plot for $\frac{G(j\omega)}{K} = \frac{1}{j\omega(1 + 0.1j\omega)}$

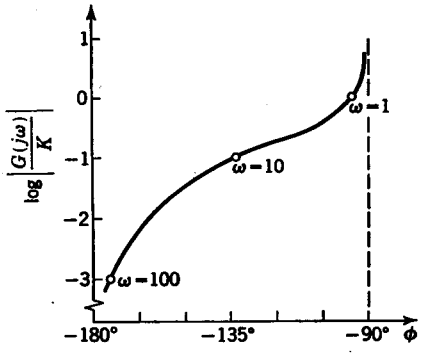


FIG. 9.28. log-modulus plot for $\frac{1}{j\omega(1 + 0.1j\omega)}$

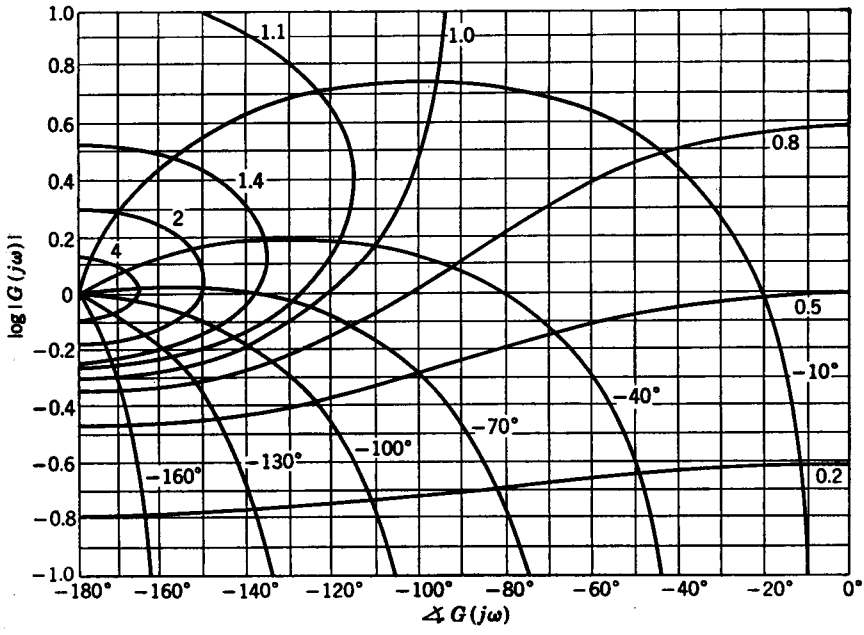


FIG. 9.29. log-modulus representation for lines of constant M and lines of constant α.

information is log-modulus, or Nichols, plots.[1] The log-modulus curve is a plot of $\log |G(j\omega)|$ versus $\measuredangle G(j\omega)$ for various values of angular velocity ω. The log-modulus curve for $G(j\omega)/K = 1/[j\omega(1 + 0.1j\omega)]$ is constructed in Fig. 9.28. From the equation for $G(j\omega)/K$ it is seen that, for a given angular velocity ω, the corresponding phase angle and amplitude ratio can be obtained. These values determine one point on the log-modulus plot. By repeating this process for other values of ω, the log-modulus graph of $G(j\omega)/K$ shown in Fig. 9.28 is obtained. Every point

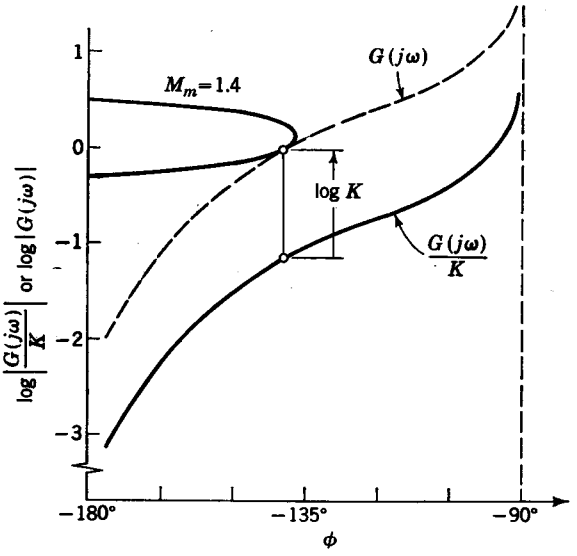


FIG. 9.30. Use of log-modulus plot to yield a desired M_m.

on the log-modulus graph corresponds to a certain value of $G(j\omega)$. By using a procedure similar to that for obtaining M circles and α circles, loci for the closed-loop response may be obtained as shown in Fig. 9.29.

It is now shown how the log-modulus techniques may be used to determine gain K so that a system will have a desired value of M_m. This is the design problem which was just treated by polar-plot techniques. In Fig. 9.30 is shown the log-modulus plot for the function $G(j\omega)/K$ of Fig. 9.28 and also the curve for the desired $M_m = 1.4$. Changing the gain K does not affect the phase angle but merely moves the log-modulus curve vertically up for K greater than 1 and down for K less than 1. In Fig. 9.30, it is to be seen that the original function represented by the solid line must be moved up so that it will be tangent to the desired M_m contour. Because $\log K = 1.2$, it follows that the required gain is $K = 16$.

[1] H. M. James, N. B. Nichols, and R. S. Phillips, "Theory of Servomechanisms," McGraw-Hill Book Company, Inc., New York, 1947.

CHAPTER 10

IMPROVING SYSTEM PERFORMANCE

10.1. Introduction. Additional insight into the correlation between the shape of a polar plot and the dynamic behavior of a system is obtained by the Nyquist stability criterion.[1] For many design problems, it is not only necessary to change the gain K as discussed in the preceding chapter, but it is also necessary to reshape the polar plot. In this chapter, the significance of the Nyquist stability criterion is first presented, and then it is shown how system performance may be improved by reshaping the polar plot.

10.2. Nyquist Stability Criterion. The Nyquist criterion makes extensive use of conformal mapping. The process of conformal mapping is illustrated as follows: Consider the function

$$1 + G(s)H(s) = s^2 - 2s + 2 = (s - 1 - j)(s - 1 + j) \quad (10.1)$$

For each value of s, there is a corresponding value of the function $1 + G(s)H(s)$. It is necessary to specify the range of values of s so that the corresponding values of the function can be computed. For instance, suppose that it is desired to let s follow the path of the circumference of the circle shown in Fig. 10.1*a*. Because the center of the circle is the point $2 + j0$ and the radius is 2, then

$$s = 2 + 2e^{j\beta} \quad (10.2)$$

where β varies from 0 to -90 to $-180°$, etc., as s traverses the circle in a clockwise direction. It follows from Eq. (10.1) that the zeros of $1 + G(s)H(s)$ [that is, values of s for which $1 + G(s)H(s) = 0$] are located inside the circle in Fig. 10.1*a*. Substitution of s from Eq. (10.2) into Eq. (10.1) gives $1 + G(s)H(s)$ as a function of β, that is,

$$1 + G(s)H(s) = 2 + 4e^{j\beta} + 4e^{2j\beta} \quad (10.3)$$

A plot of the corresponding map of $1 + G(s)H(s)$ is shown in Fig. 10.1*b*. It is important to notice that there are two clockwise encirclements of the origin of the $1 + G(s)H(s)$ plot shown in Fig. 10.1*b*. In general, it can be shown that there is one encirclement of the origin for each zero of

[1] H. Nyquist, Regeneration Theory, *Bell System Tech. J.*, vol. 11, pp. 126–147, 1932.

$1 + G(s)H(s)$ which lies within the path of values for s. These encirclements are in the *same* sense as the motion around the path in the s plane, in this case, clockwise. For each pole of $1 + G(s)H(s)$ located within the path of values for s, there is one encirclement of the origin in the

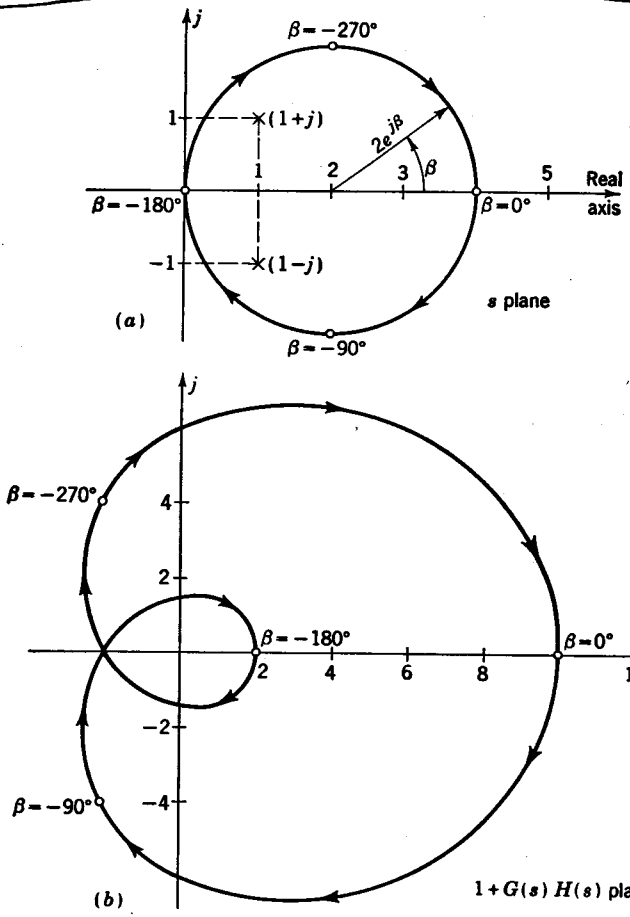


FIG. 10.1. (*a*) Path of values for s; (*b*) map for $1 + G(s)H(s)$.

opposite sense. By a pole is meant a value of s which makes $1 + G(s)H(s)$ infinite. For example, consider the function.

$$1 + G(s)H(s) = \frac{1}{s^2 - 2s + 2} = \frac{1}{(s - 1 - j)(s - 1 + j)} \tag{10.4}$$

The preceding function has two poles, $s = 1 + j$ and $s = 1 - j$.

The following general equation may be formulated,

$$N = Z - P \tag{10.5}$$

$Z = N + P$

where P = no. of poles of $1 + G(s)H(s)$ located inside path of values for s
Z = no. of zeros of $1 + G(s)H(s)$ located inside path of values for s
N = net no. of encirclements of origin of $1 + G(s)H(s)$ plane

When the net number of encirclements N is in the same sense as motion around the path of the s contour, an excess of zeros Z is indicated. The opposite sense signifies an excess of poles P. From Eq. (10.5), it follows that, when N is in the same sense, N is a positive number. Similarly, when N is in the opposite sense, N is a negative number.[1] Generally, the path of values for s is traversed in a clockwise direction. Thus, for a net number of clockwise encirclements N is positive, and there is an excess of zeros. For a net number of counterclockwise encirclements N is negative, and there is an excess of poles.

For control work, the path of values for s is usually taken as shown in Fig. 10.2. This contour is seen to proceed from the origin up the imaginary axis to infinity, then an infinite semicircle ($R \to \infty$) sweeps around to the bottom of the imaginary axis, whence it returns to the origin. This contour, in effect, encloses the entire right half plane.

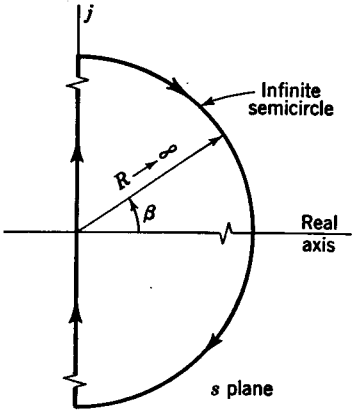


FIG. 10.2. Path of values for s, which encloses the entire right half plane.

As is discussed in Chap. 6, a system is basically unstable if any zeros of the characteristic equation are located in the right half plane. By noting that

$$1 + G(s)H(s) = 1 + \frac{N_{G(s)}N_{H(s)}}{D_{G(s)}D_{H(s)}} = \frac{D_{G(s)}D_{H(s)} + N_{G(s)}N_{H(s)}}{D_{G(s)}D_{H(s)}} \quad (10.6)$$

it is apparent that the zeros of $1 + G(s)H(s)$ are also the zeros of the characteristic equation. From Eq. (10.6), it is also to be noted that the poles of $1 + G(s)H(s)$ are the zeros of $D_{G(s)}D_{H(s)}$. Thus, by letting s assume the values indicated along the contour of Fig. 10.2, it follows that

$$Z = N + P \quad (10.7)$$

where Z = no. of zeros of characteristic equation [i.e., zeros of $1 + G(s)H(s)$] in right half plane
P = no. of zeros of $D_{G(s)}D_{H(s)}$ [that is, poles of $1 + G(s)H(s)$] in right half plane
N = net no. of encirclements of origin of $1 + G(s)H(s)$ map

[1] Often Eq. (10.5) is written in the form $N = P - Z$. When this form is used, N is negative for a net number of encirclements in the same sense and is positive for the opposite sense.

This technique is now illustrated for a third-order system for which

$$G(s)H(s) = \frac{10}{s(s^2 + 5s + 4)} = \frac{10}{s(s + 4)(s + 1)} \tag{10.8}$$

Because $G(s)H(s)$ is usually obtained in factored form, it is more convenient to construct the map for $G(s)H(s)$ rather than $1 + G(s)H(s)$. The effect of adding $+1$ to each point of the $G(s)H(s)$ map to obtain the $1 + G(s)H(s)$ map is accomplished simply by adding $+1$ to the scale of

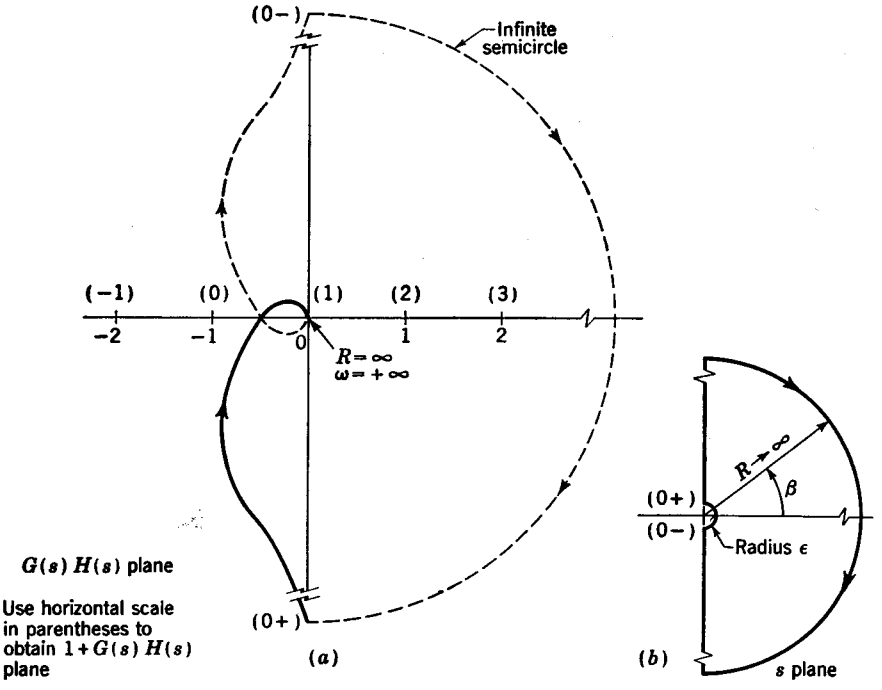


FIG. 10.3. (a) Map for $G(s)H(s)$, in which the origin is not enclosed in the path of values for s; (b) path of values for s, excluding the origin.

the real axis, as is shown by the numbers in parentheses in Fig. 10.3a. The -1 point of the $G(s)H(s)$ map is seen to correspond to the origin of the $1 + G(s)H(s)$ map. Thus, N is equal to the net number of encirclements of the -1 point of the $G(s)H(s)$ plot.

For values of s along the positive imaginary axis, the map for $G(s)H(s)$ is the polar plot $G(j\omega)H(j\omega)$, where ω takes on values from 0 to ∞. The map of $G(s)H(s)$ for this region begins at $-\infty$ on the negative imaginary axis for $\omega = 0+$, and then goes to the origin for $\omega = +\infty$, as is shown in Fig. 10.3a. For values of s along the infinite semicircle of Fig. 10.3b, the map of $G(s)H(s)$ is a point at the origin. This is verified by sub-

stituting $s = Re^{j\beta}$ into Eq. (10.8) and noting that, as R becomes infinite, $G(s)H(s)$ approaches zero. For values of s along the negative imaginary axis of Fig. 10.3*b*, then $G(s)H(s)$ is equal to $G(-j\omega)H(-j\omega)$, which is the complex conjugate of $G(j\omega)H(j\omega)$. Thus, as shown in Fig. 10.3*a*, the map for s along the negative imaginary axis is the mirror image of the map for s along the positive imaginary axis.

As is illustrated by this example, a unique problem exists when a pole of $G(s)H(s)$ (that is, $s = 0$) lies on the path of assumed values for s. In general, when $G(s)H(s)$ has a pole along the imaginary axis, the map becomes indeterminant when s assumes the value at the pole. To avoid this situation, a small semicircle is constructed around the pole, as is shown in Fig. 10.3*b*. As is soon to be demonstrated, it makes no difference in the result if the semicircle is drawn so as to exclude the pole as shown in Fig. 10.3*b* or if it is drawn to include the pole as shown in Fig. 10.4*b*.

For the case shown in Fig. 10.3*b*, in which the radius of the small semicircle is ϵ ($\epsilon \ll 1$) and β varies from -90 to 0 to $+90°$ in going from $s = j(0-)$ to $j(0+)$, substitution of $s = \epsilon e^{j\beta}$ into Eq. (10.8) gives

$$G(s)H(s) = \frac{10}{\epsilon^3 e^{j3\beta} + 5\epsilon^2 e^{j2\beta} + 4\epsilon e^{j\beta}} \approx \frac{10}{4\epsilon e^{j\beta}} \approx \infty e^{-j\beta} \qquad (10.9)$$

In the preceding expression, advantage is taken of the fact, that for small values of ϵ, then $\epsilon e^{j\beta} \gg \epsilon^2 e^{j2\beta} \gg \epsilon^3 e^{j3\beta}$, so that terms with higher powers of ϵ become negligible. For β going from -90 to 0 to $+90°$, Eq. (10.9) describes an infinite semicircle as shown in Fig. 10.3*a*. Application of the Nyquist criterion to determine Z is accomplished as follows: Because none of the poles of $1 + G(s)H(s)$ (that is, $s = 0, -1, -4$) is inside the contour of the values of s shown in Fig. 10.3*b*, then P is zero. The value of N is also zero because there are no encirclements of the -1 point of the $G(s)H(s)$ map and therefore $Z = N + P = 0$. Thus, there are no zeros of the characteristic equation in the right half plane.

As was previously mentioned, it would make no difference if the path for values of s had been chosen to include the pole of $1 + G(s)H(s)$, as is shown in Fig. 10.4*b*. For this case, the equation for the very small semicircle is $\epsilon e^{j\beta}$, where β varies from -90 to -180 to $-270°$ as s goes from $j(0-)$ to $j(0+)$. For this region, Eq. (10.8) becomes

$$G(s)H(s) = \frac{10}{\epsilon^3 e^{j3\beta} + 5\epsilon^2 e^{j2\beta} + 4\epsilon e^{j\beta}} \approx \frac{10}{4\epsilon e^{j\beta}} \approx \infty e^{-j\beta} \qquad (10.10)$$

The values of β to be used in Eq. (10.10) yield the infinite semicircle shown in Fig. 10.4*a*. There is one net counterclockwise encirclement of the -1 point so that N is -1. Because the pole at $s = 0$ is included within the path of values of s as shown in Fig. 10.4*b*, the value of P is 1,

so that $Z = N + P = -1 + 1 = 0$. This is the result that was previously obtained by excluding the origin.

If the polar plot for $G(s)H(s)$ were to go through the -1 point, this would indicate that zeros of the characteristic equation are located on the imaginary axis. For example, from Fig. 10.3*a*, it is to be noticed that the polar plot crosses the negative real axis at a value of -0.5. If the

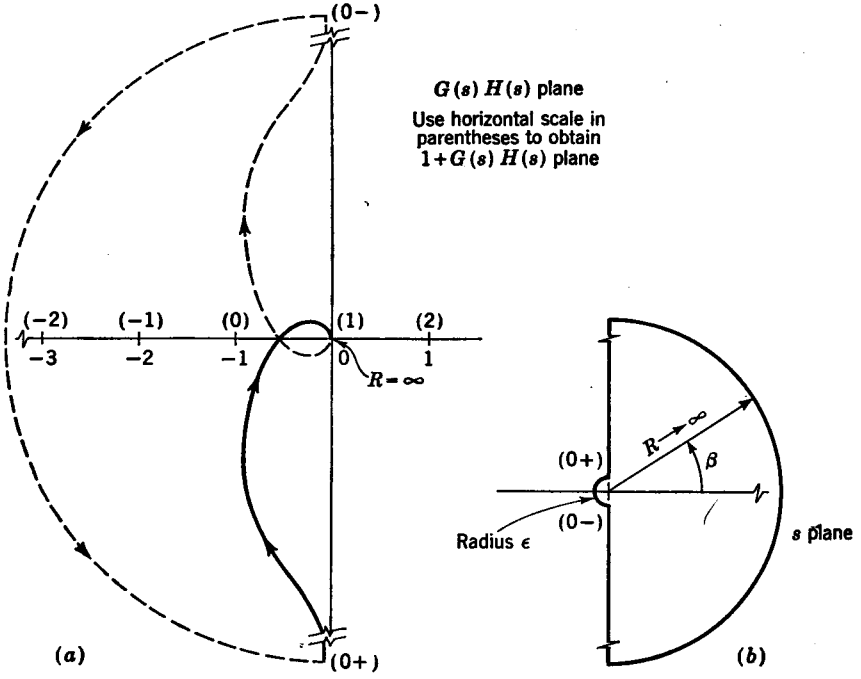


FIG. 10.4. (*a*) Map for $G(s)H(s)$, in which the origin is enclosed in the path of values for s; (*b*) path of values for s, including the origin.

gain which is 10 in Eq. (10.8) were doubled, the polar plot would go through the -1 point. For this case, the characteristic equation is

$$D_{G(s)}D_{H(s)} + N_{G(s)}N_{H(s)} = s(s^2 + 5s + 4) + 20 = (s + 5)(s^2 + 4) \qquad (10.11)$$

and the two zeros $s = \pm j2$ are seen to lie on the imaginary axis.

The open-loop frequency response $G(j\omega)H(j\omega)$ can be determined experimentally only when the open loop is stable. For a stable open loop, all the zeros of D_GD_H lie in the left half plane. Because the zeros of D_GD_H are the poles of $1 + G(s)H(s)$, for this case there are no poles in the right half plane and $P = 0$. When the open loop is stable, the

Nyquist stability criterion becomes

$$Z = N \tag{10.12}$$

For a stable system (i.e., stable closed loop, or $Z = 0$), there can be no net encirclements N of the -1 point.

It is well to note that the closed loop may be stable even though the open loop is unstable, and vice versa. For an unstable open loop, the

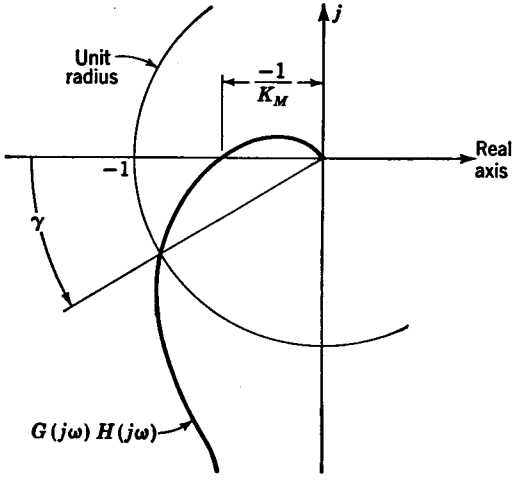


FIG. 10.5. Determination of gain margin and phase margin from the polar plot.

open-loop response may be calculated from a known closed-loop response by solving Eq. (9.52) for $G(j\omega)$, thus,

$$G(j\omega) = \frac{c(j\omega)/r(j\omega)}{1 - c(j\omega)/r(j\omega)} \tag{10.13}$$

In applying frequency-response methods, one always works with the open-loop response.

10.3. Gain Margin and Phase Margin. From the preceding discussion, the -1 point of the $G(s)H(s)$ map was seen to have great significance with regard to the stability of a system. In Fig. 10.5 is shown a typical $G(j\omega)H(j\omega)$ plot in the vicinity of the -1 point. If the gain were multiplied by an amount K_M, called the gain margin, the $G(j\omega)H(j\omega)$ plot would go through the -1 point. Thus, the gain margin is an indication of how much the gain can be increased before the curve goes through the critical point. In Fig. 10.6 is shown the log-modulus plot corresponding to Fig. 10.5. Because $\log 1 = 0$, the $-1 + j0$ point on the log-modulus plot is defined by the ordinate $\log |G(j\omega)H(j\omega)| = 0$ and abscissa

$\phi = -180°$. From Fig. 10.6, the vertical distance that the $G(j\omega)H(j\omega)$ plot may be raised before it goes through the -1 point is log K_M.

The angle γ in Fig. 10.5 is the angle measured from the negative real axis to where the polar plot crosses a circle of unit radius. If the angle γ is zero, the polar plot goes through the -1 point. The angle γ, called the phase margin, is thus seen to be another indication or measure of the closeness of the polar plot to the critical point. The value of the phase margin may be obtained from a log-modulus plot as follows: The horizontal line in Fig. 10.6 of log $|G(j\omega)H(j\omega)|$ = log 1 = 0 corresponds to the unit circle of Fig. 10.5. The angle γ between the point where this horizontal line intersects the $G(j\omega)H(j\omega)$ plot and the value $\phi = -180°$ is the phase margin γ.

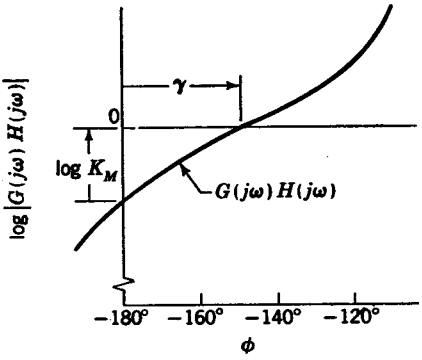


FIG. 10.6. Determination of gain margin and phase margin from the log-modulus plot.

When the gain margin is less than 1, the gain must be decreased to make the curve pass through the -1 point. An unstable system is indicated when the gain margin is less than one.

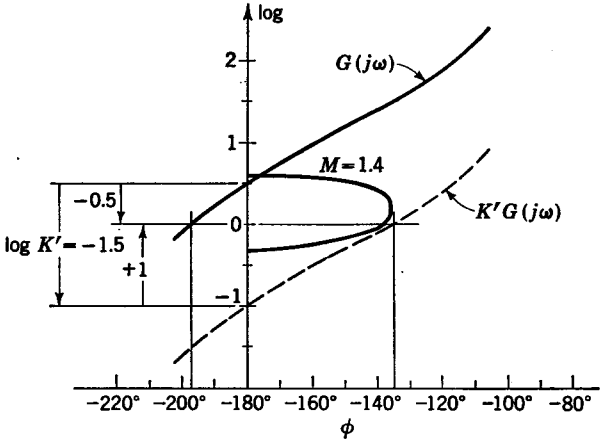


FIG. 10.7. log-modulus plot for $G(j\omega)$.

Illustrative Example 1. The log-modulus plot of $G(j\omega)$ for a unity-feedback system is shown in Fig. 10.7. What is the value of the gain margin and the phase margin for this system? By what factor K' should the gain of the system be changed so that M_m will be 1.4? What is the new value for the gain margin and the phase margin?

SOLUTION. From the $G(j\omega)$ plot for the original system, it follows that $\log K_M = -0.5$, or $K_M = 0.316$, and $\gamma = 180° - 198° = -18°$. It is to be noted that a negative phase margin indicates an unstable system, as does a gain margin of less than 1.

The factor K' is obtained by moving the $G(j\omega)$ locus straight down until the new locus $K'G(j\omega)$ is tangent to the desired $M_m = 1.4$ contour. From Fig. 10.7, it follows that

$$\log K' = -1.5 \quad \text{or} \quad K' = 0.0316$$

The new gain margin is $\log K_M = 1$, or $K_M = 10$, and the new phase margin is $\gamma = 180° - 135° = 45°$.

Illustrative Example 2. The log-magnitude plots of $G(j\omega)$ for a unity-feedback system are shown in Fig. 10.8. What is the value of the gain margin and the phase margin? By what factor should the gain of the system be changed to obtain a gain margin of 2?

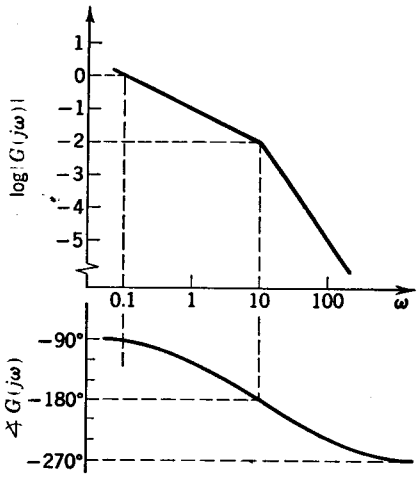


FIG. 10.8. log-magnitude diagrams for $G(j\omega)$.

SOLUTION. On a log-magnitude plot, the effect of increasing the gain is to shift the entire plot of $\log |G(j\omega)|$ vertically upward. Similarly, decreasing the gain shifts it down. The phase angle is unaffected by a change in gain. The gain margin is the factor by which the gain must be increased so that $|G(j\omega)|$ will equal 1 when ϕ is 180°. From Fig. 10.8, at $\phi = 180°$, the value of ω is 10, and $\log |G(j\omega)|$ is -2, or $G(j\omega) = 0.01$. Thus, the gain margin is 100. The phase margin γ is equal to $180° - 93° = 87°$, where $-93°$ is $\angle G(j\omega)$ when $|G(j\omega)|$ is 1, which occurs at $\omega = 0.1$.

Because the original gain margin is 100, increasing the gain by a factor of 50 would yield a gain margin of 2.

10.4. Series Compensation. In Sec. 9.9, it is shown how the gain K is selected in order to obtain a desired value of M_m. A change in the gain K in effect changes the scale factor of the polar plot but does not change the basic shape of the plot. In the design of control systems, it is often necessary to change the shape of the polar plot in order to achieve the desired dynamic performance. A common means of doing this is to insert elements in series with the feedforward portion of the control. This method of compensating the performance of the control system is called series compensation.

In general, the frequency-response characteristics of a component which is used to provide series compensation are such that the output of the component either lags or leads the input. In some cases, it is advantageous to use a component in which the output lags the input for a certain range of frequencies and then the output leads the input for other frequencies. This is known as lag-lead series compensation. A component which is used to provide series compensation is sometimes referred to as a series equalizer.

Phase Lag. The output lags the input for any component which has a transfer function of the form

$$\frac{E_o(s)}{E_{in}(s)} = \frac{Y(s)}{X(s)} = \frac{1 + \tau_2 s}{1 + \tau_1 s} \qquad \tau_1 > \tau_2 \tag{10.14}$$

The frequency response for the preceding transfer function is

$$\frac{E_o(j\omega)}{E_{in}(j\omega)} = \frac{Y(j\omega)}{X(j\omega)} = \frac{1 + j\tau_2\omega}{1 + j\tau_1\omega} \qquad \tau_1 > \tau_2 \tag{10.15}$$

In Fig. 10.9 are shown both an electrical and a mechanical component in which the output lags the input, as is described by Eq. (10.15). The

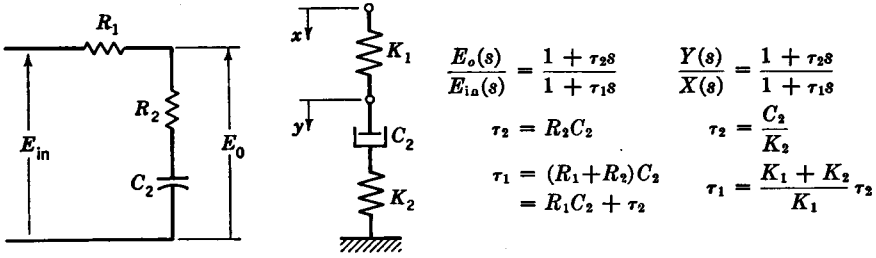


FIG. 10.9. An electrical and a mechanical component used to obtain phase lag.

construction of the log-magnitude plot for Eq. (10.15) for the case in which $\tau_2 = \tau_1/10$ is illustrated in Fig. 10.10. For the term $1/(1 + j\tau_1\omega)$ the break frequency occurs at $\omega = 1/\tau_1$, and for the numerator $1 + j\tau_2\omega$ the break frequency occurs at $\omega = 1/\tau_2 = 10/\tau_1$. The addition of the log-magnitude diagram for the numerator to that for the denominator yields the resulting diagram for $(1 + j\tau_2\omega)/(1 + j\tau_1\omega)$.

The resulting diagrams for three typical cases in which $\tau_2 = \tau_1/2$, $\tau_2 = \tau_1/10$, and $\tau_2 = \tau_1/\infty = 0$ are shown in Fig. 10.11. The amount of phase lag which is introduced depends upon the spread between the time constants τ_1 and τ_2. Also, the greater the phase lag introduced, the less will be the ratio $|(1 + j\tau_2\omega)/(1 + j\tau_1\omega)|$ at higher frequencies.

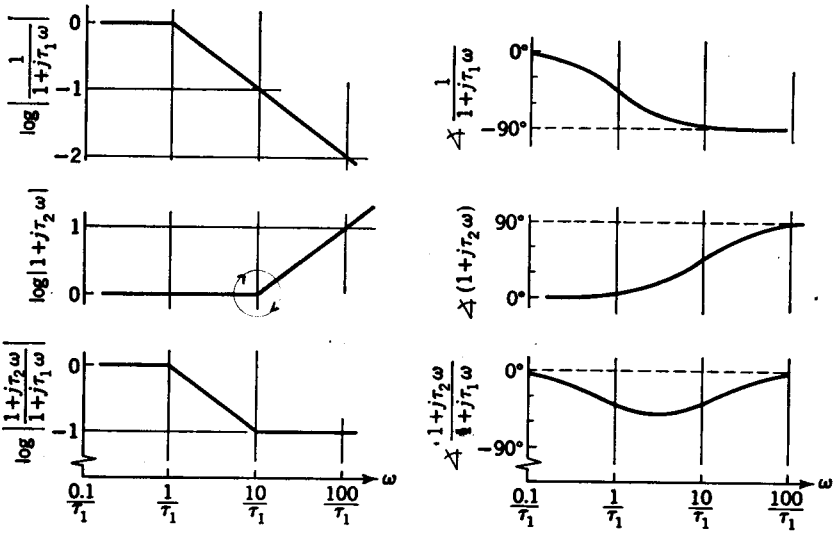


FIG. 10.10. Construction of the log-magnitude diagram for $(1 + j\tau_2\omega)/(1 + j\tau_1\omega)$, in which $\tau_2 = \tau_1/10$.

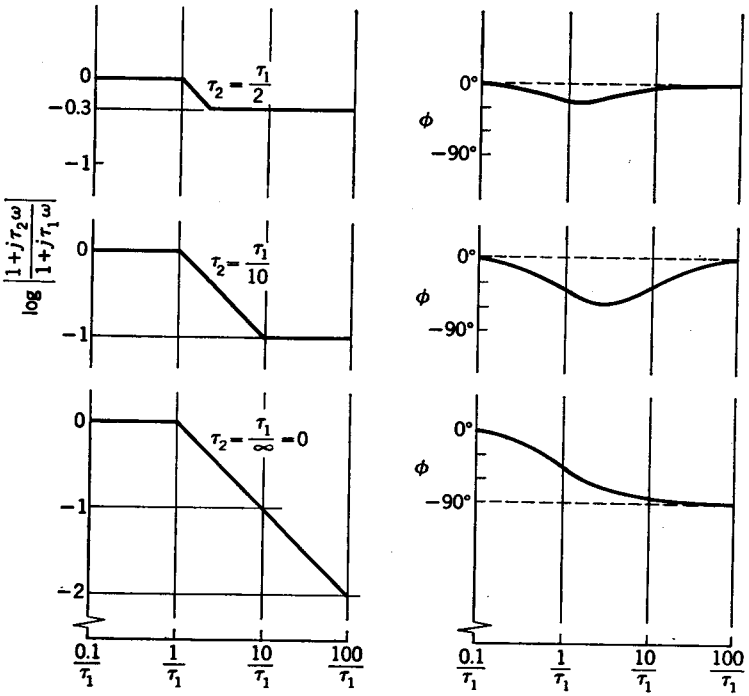


FIG. 10.11. log-magnitude diagrams for $(1 + j\tau_2\omega)/(1 + j\tau_1\omega)$, in which $\tau_2 = \tau_1/2$, $\tau_2 = \tau_1/10$, and $\tau_2 = \tau_1/\infty = 0$.

The use of a phase-lag component placed in series with the feedforward portion of a control system to improve stability is now illustrated. In Fig. 10.12*a*, the dashed curve is the frequency response of an uncompensated control system for which the transfer function is $G(s)$ as shown

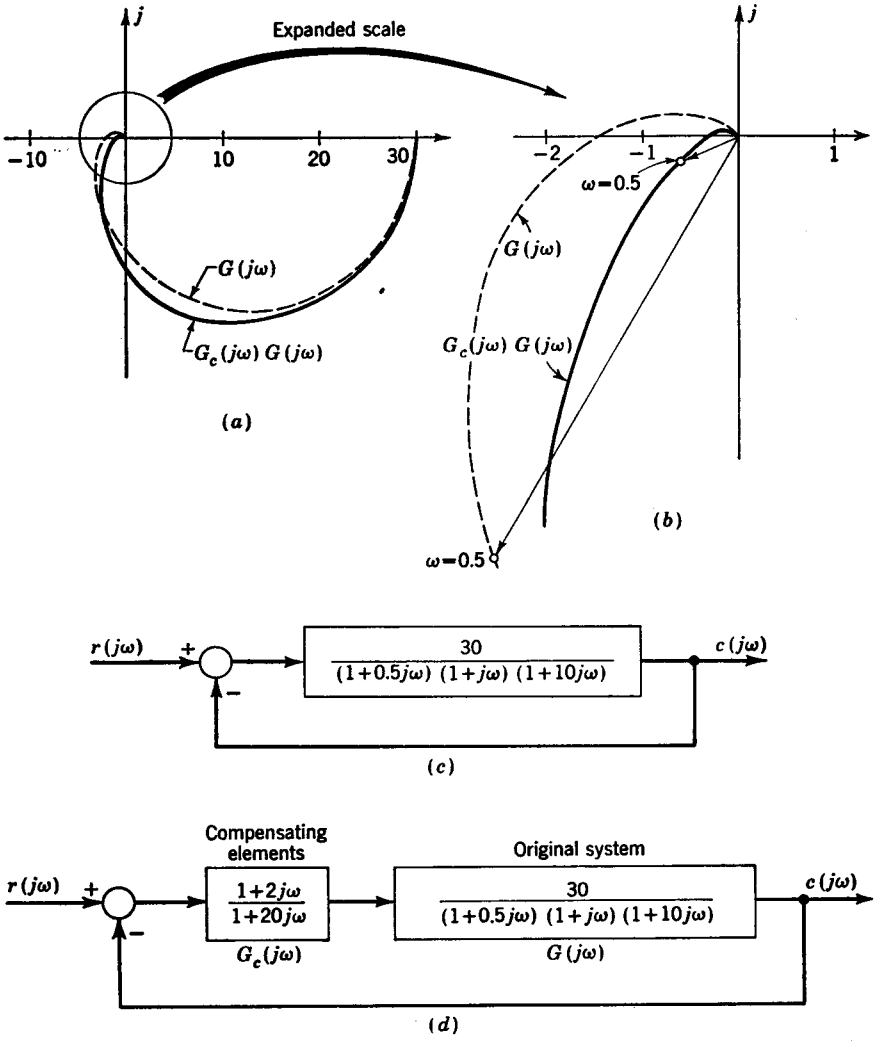


FIG. 10.12. Use of phase lag to reshape a polar plot.

in Fig. 10.12*c*. This control is one which would inherently be unstable. The addition of series compensation $G_c(s)$ is shown in Fig. 10.12*d*. This reshapes the high-frequency portion of the polar plot as is shown by the solid-line curves of Fig. 10.12*a* and *b*. The resultant system has good

dynamic response. From the expanded view of the polar plot in Fig. 10.12*b*, it is to be noted that the effect of lag compensation is to shorten a typical vector such as that for $\omega = 0.5$ and also to rotate it in a clockwise direction. The shortening is due to the attenuation. By attenuation is

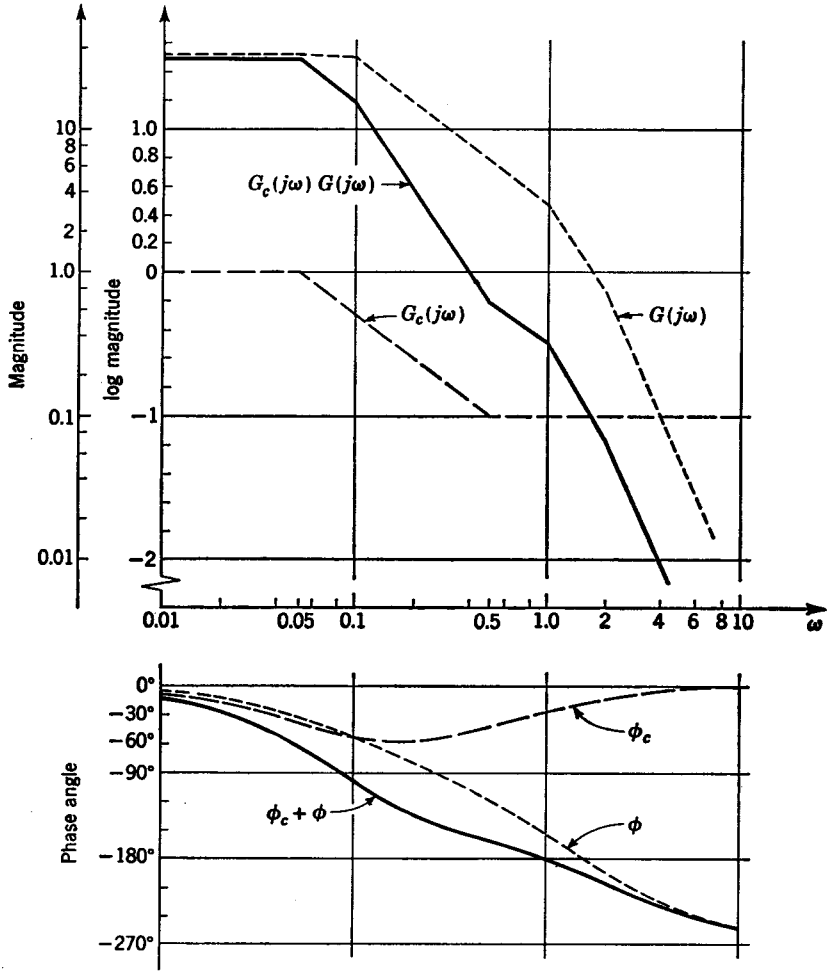
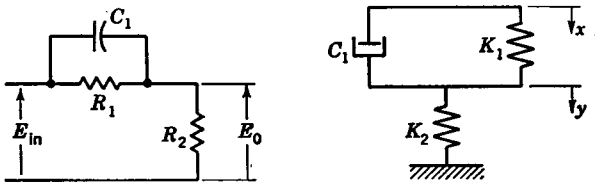


FIG. 10.13. log-magnitude diagrams for system which is compensated by inserting a phase-lag component.

meant multiplication by a factor less than 1. The attenuation caused by use of lag compensation can be seen from Fig. 10.11. The greater the spread in time constants τ_1 and τ_2, the more pronounced is the attenuation which occurs at higher frequencies. Series lag compensation has little effect on the low-frequency portion of the curve. By reshaping the polar

plot, it has been possible to achieve good dynamic performance without changing the value of the gain K. Although it would have been possible to make this system stable by decreasing the gain K only, errors caused by friction, hysteresis, backlash, etc., tend to predominate as the gain is decreased; thus, in general, the higher the value of K, the more accurate will be the control system.

Phase-lag compensation is also used to permit increasing the value of the gain K. For example, if a system initially has a satisfactory value of M_m, the addition of lag compensation will attenuate the high-frequency



$$\frac{E_o(s)}{E_{in}(s)} = \frac{\tau_2}{\tau_1}\frac{1+\tau_1 s}{1+\tau_2 s} \qquad \frac{Y(s)}{X(s)} = \frac{\tau_2}{\tau_1}\frac{1+\tau_1 s}{1+\tau_2 s}$$

$$\tau_1 = R_1C_1 \qquad \tau_1 = \frac{C_1}{K_1}$$

$$\tau_2 = \frac{R_2}{R_1+R_2}R_1C_1 \qquad \tau_2 = \frac{C_1}{K_1+K_2}$$

$$= \frac{R_2}{R_1+R_2}\tau_1 \qquad = \frac{K_1}{K_1+K_2}\tau_1$$

FIG. 10.14. An electrical and a mechanical component used to obtain phase lead.

portion of the plot. Thus, the gain K must now be increased so that this compensated polar plot will be tangent to the original M_m circle.

In Fig. 10.13 are shown the log-magnitude diagrams corresponding to the polar plot of Fig. 10.12. In general, log-magnitude diagrams are better suited for problems in which it is desired to have a certain gain margin or phase margin. Polar plots are usually employed when it is desired to have a certain value of M_m.

Phase Lead. In Fig. 10.14 are shown both an electrical and a mechanical component which have the general phase-lead characteristic given by

$$\frac{E_o(s)}{E_{in}(s)} = \frac{Y(s)}{X(s)} = \frac{\tau_2}{\tau_1}\frac{1+\tau_1 s}{1+\tau_2 s} \qquad \tau_1 > \tau_2 \tag{10.16}$$

or

$$\frac{E_o(j\omega)}{E_{in}(j\omega)} = \frac{Y(j\omega)}{X(j\omega)} = \frac{\tau_2}{\tau_1}\frac{1+j\tau_1\omega}{1+j\tau_2\omega} \qquad \tau_1 > \tau_2 \tag{10.17}$$

In using phase-lead compensation, additional amplification equal to τ_1/τ_2 must be provided to maintain the original system gain. The construction of the log-magnitude diagram for $(1 + j\tau_1\omega)/(1 + j\tau_2\omega)$ for the case where $\tau_2 = \tau_1/10$ is shown in Fig. 10.15. The break frequency

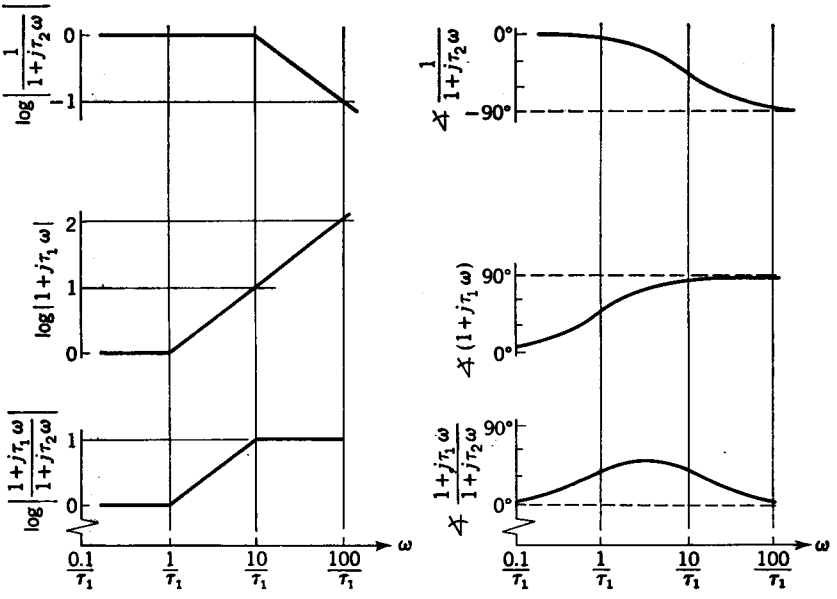


FIG. 10.15. Construction of the log-magnitude diagram for $(1 + j\tau_1\omega)/(1 + j\tau_2\omega)$, in which $\tau_2 = \tau_1/10$.

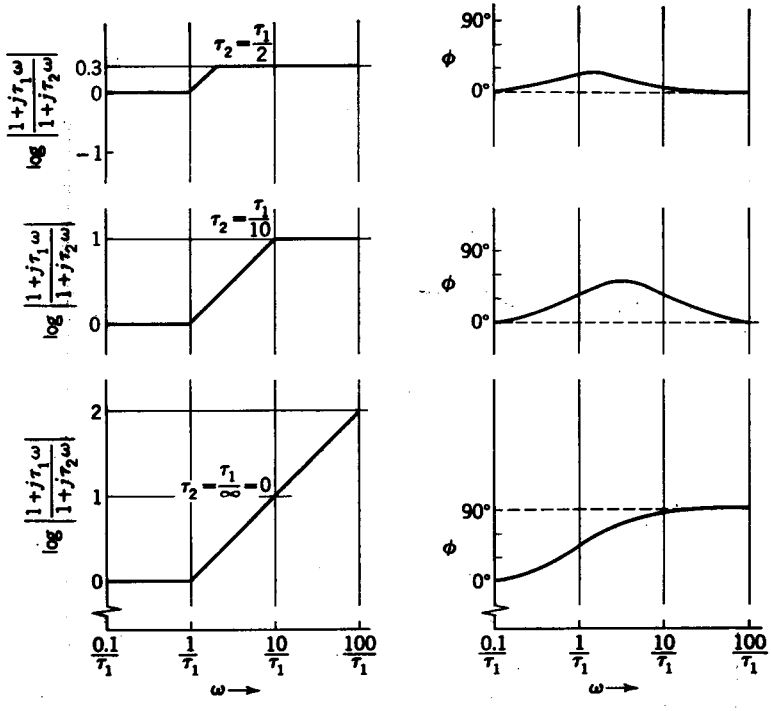


FIG. 10.16. log-magnitude diagrams for $(1 + j\tau_1\omega)/(1 + j\tau_2\omega)$, in which $\tau_2 = \tau_1/2$, $\tau_2 = \tau_1/10$, and $\tau_2 = \tau_1/\infty = 0$.

for the numerator occurs at $\omega = 1/\tau_1$ and for the denominator at $\omega = 1/\tau_2 = 10/\tau_1$. Adding the diagram for the numerator to that for the denominator yields the resultant diagram for phase lead. Log-magnitude diagrams for $(1 + j\tau_1\omega)/(1 + j\tau_2\omega)$ when $\tau_2 = \tau_1/2, \tau_2 = \tau_1/10$, and $\tau_2 = \tau_1/\infty = 0$ are shown in Fig. 10.16.

In Fig. 10.17*a* is shown the application of lead compensation to stabilize the basic system $G(s)$ which was just discussed in relation to the use of lag compensation. From the expanded scale of Fig. 10.17*b*, it is to

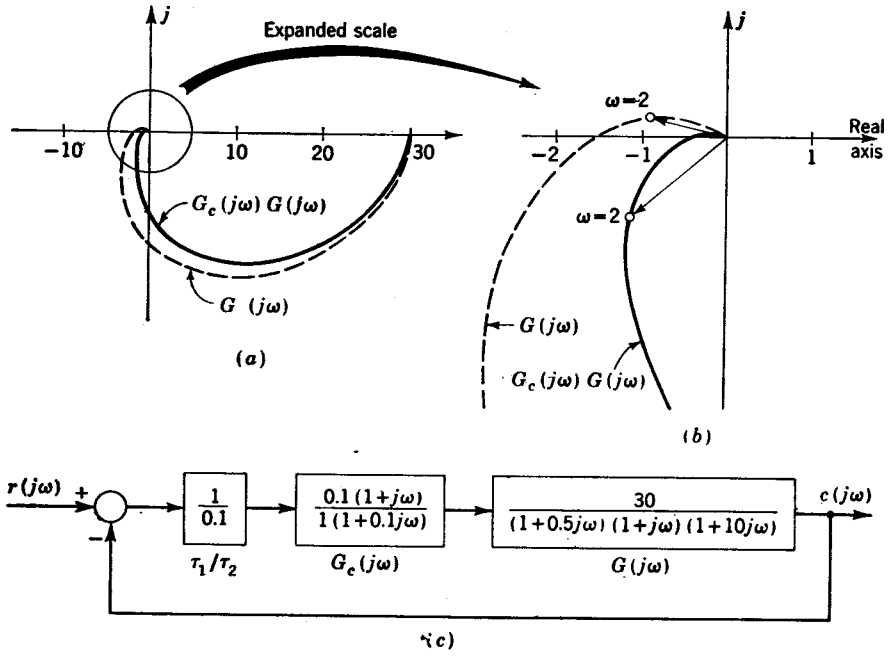


FIG. 10.17. Use of phase lead to reshape a polar plot.

be noted that lead compensation rotates the vector at $\omega = 2$ in a counterclockwise direction away from the -1 point. It is also to be noted that the length of this vector is increased. The complete log-magnitude diagrams corresponding to the polar plots of Fig. 10.17 are shown in Fig. 10.18.

For a system which initially has a satisfactory value of M_m, the addition of lead compensation generally permits increasing the gain K. Because of the counterclockwise rotation of a typical vector as illustrated in Fig. 10.17*b*, the effect of lead compensation is to increase ω_m. As was discussed in Sec. 9.8, this increases the speed of response of the system. On the other hand, lag compensation rotates the polar plot in a clockwise

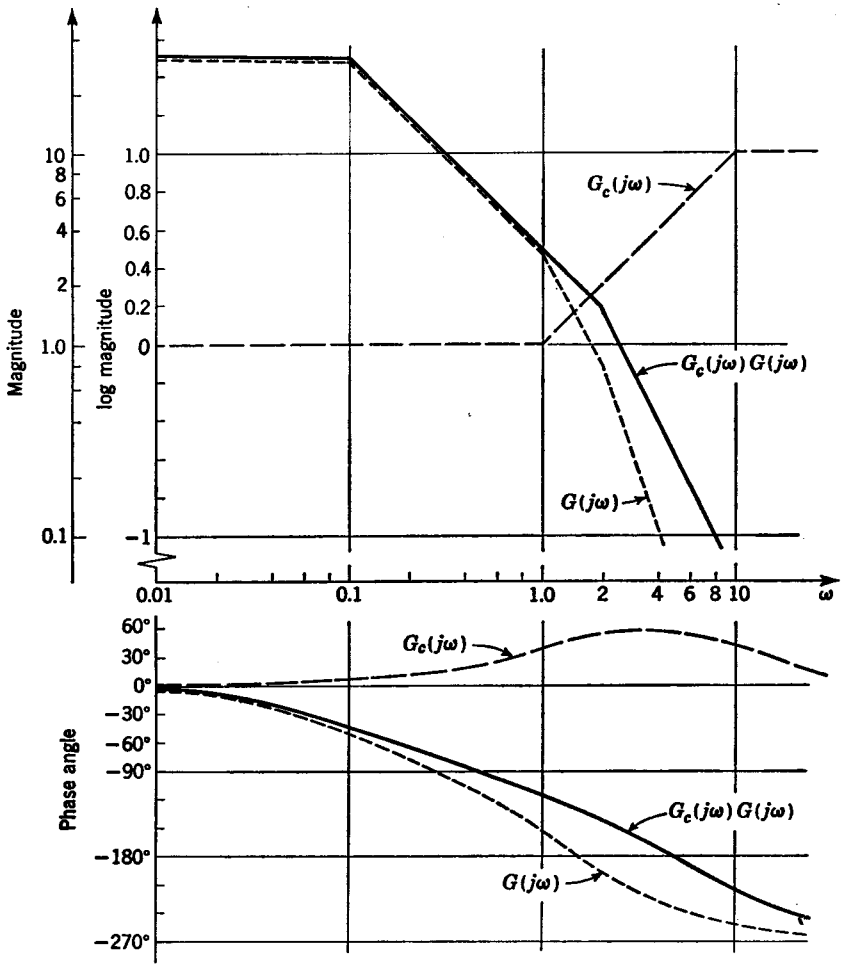


FIG. 10.18. log-magnitude diagrams for system which is compensated by inserting a phase-lead component.

direction, which tends to decrease the value of ω_m. Thus, lag compensation usually decreases the speed of response of a system.

Both series lag and series lead compensation affect primarily the high-frequency characteristics. They have little or no effect upon the low-frequency portion of the frequency response plots.

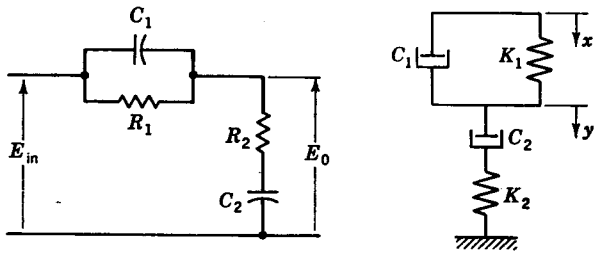
Lag-Lead Compensation. Lag-lead compensation is in effect a series or cascaded combination of a lag and a lead network. The general transfer function for a lag-lead compensator is

$$\frac{E_o(s)}{E_{in}(s)} = \frac{Y(s)}{X(s)} = \frac{1 + \tau_2 s}{1 + \tau_1 s} \frac{1 + c\tau_1 s}{1 + c\tau_2 s} \tag{10.18}$$

where $\tau_1 > \tau_2$ and c is a constant less than 1 and such that $\tau_2 > c\tau_1$ or $\tau_2/\tau_1 > c$. The substitution of $j\omega$ for s gives

$$\frac{E_o(j\omega)}{E_{in}(j\omega)} = \frac{Y(j\omega)}{X(j\omega)} = \frac{1 + j\tau_2\omega}{1 + j\tau_1\omega}\frac{1 + jc\tau_1\omega}{1 + jc\tau_2\omega} \tag{10.19}$$

Rather than using a lag and a lead compensator in series, it is possible to use a single compensator, as shown in Fig. 10.19.



$$\frac{E_o(s)}{E_{in}(s)} = \frac{(1 + \tau_2 s)(1 + c\tau_1 s)}{(1 + \tau_1 s)(1 + c\tau_2 s)} \qquad \frac{Y(s)}{X(s)} = \frac{(1 + \tau_2 s)(1 + c\tau_1 s)}{(1 + \tau_1 s)(1 + c\tau_2 s)}$$

$$\tau_2 = R_2C_2 \qquad \tau_1 = \frac{R_1C_1}{c} \qquad \tau_2 = \frac{C_2}{K_2} \qquad \tau_1 = \frac{C_1}{cK_1}$$

$$c = 1 - \frac{R_1C_2}{\tau_1 - \tau_2} \qquad c = 1 - \frac{C_2/K_1}{\tau_1 - \tau_2}$$

FIG. 10.19. An electrical and a mechanical component used to obtain lag-lead compensation.

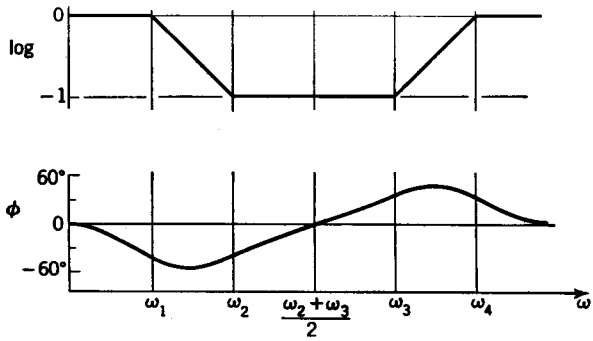


FIG. 10.20. log-magnitude diagrams for lag-lead compensation,

$$\frac{1 + j\tau_2\omega}{1 + j\tau_1\omega}\frac{1 + c\tau_1\omega}{1 + c\tau_2\omega}$$

The log-magnitude diagrams for a typical lag-lead compensator are shown in Fig. 10.20. Because $\tau_1 > \tau_2 > c\tau_1 > c\tau_2$, the first break frequency occurs at $\omega_1 = 1/\tau_1$. This break frequency belongs to a denominator term so that the magnitude plot has a slope of -1 log units/decade

between ω_1 and ω_2. The second break frequency $\omega_2 = 1/\tau_2$ is associated with a numerator term; and thus the magnitude plot again becomes horizontal. The third break frequency $\omega_3 = 1/c\tau_1 = \omega_1/c$ also occurs in the numerator. This results in a slope of $+1$ log unit/decade in the region from ω_3 to ω_4. Finally, the break frequency $\omega_4 = 1/c\tau_2 = \omega_2/c$ which occurs in the denominator causes the magnitude curve to become horizontal again. It is to be noted that for $\omega > (\omega_2 + \omega_3)/2$ attenuation is accompanied by phase lead. This feature makes it possible to obtain a considerably greater increase in the gain K than is usually possible with either a lag or a lead compensator. This fact may be substantiated by noting from Fig. 10.12*b* or Fig. 10.17*b* that a counterclockwise rotation and attenuation of a vector have a greater effect on reshaping the polar plot than does the compensation which is indicated on these polar

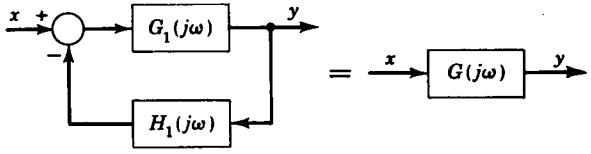


FIG. 10.21. Internal feedback $H_1(s)$ placed about an element $G_1(s)$.

plots. The use of lag-lead compensation has about the same effect of increasing ω_m and the corresponding speed of response as does lead compensation.

As is indicated by Fig. 10.20, lag-lead compensation does not affect the low- or high-frequency regions but rather the mid-frequency region.

10.5. Internal Feedback. Another method commonly used to alter frequency-response characteristics is that of providing a separate internal-feedback path about certain components. In employing log-magnitude diagrams to investigate the effect of internal feedback, the use of a few simple approximations affords much simplification. This approximate analysis in effect puts the designer in the right "ball park." In the latter design stages, it may be desirable to make an exact analysis. The approximations which are used to evaluate the effect of placing a feedback element $H_1(s)$ around a component $G_1(s)$ as shown in Fig. 10.21 are that, when $|G_1(j\omega)H_1(j\omega)| \ll 1$,

$$G(j\omega) = \frac{G_1(j\omega)}{1 + G_1(j\omega)H_1(j\omega)} \approx G_1(j\omega) \tag{10.20}$$

For the case when $|G_1(j\omega)H_1(j\omega)| \gg 1$,

$$G(j\omega) = \frac{G_1(j\omega)}{1 + G_1(j\omega)H(j\omega)} \approx \frac{G_1(j\omega)}{G_1(j\omega)H_1(j\omega)} = \frac{1}{H_1(j\omega)} \tag{10.21}$$

When $|G_1(j\omega)H_1(j\omega)| = 1$, then $|G_1(j\omega)| = 1/|H_1(j\omega)|$; thus the inter-

section of the log-magnitude diagram for $|G_1(j\omega)|$ and that for $1/|H_1(j\omega)|$ determines the point at which $|G_1(j\omega)H_1(j\omega)| = 1$. It follows from Eqs. (10.20) and (10.21) that the frequency response is altered only when $|G_1(j\omega)H_1(j\omega)| \gg 1$. The application of the use of internal feedback is now demonstrated.

In Fig. 10.22*a* is shown a component which has a unity-feedback path $[H_1(j\omega) = 1]$ around it. The log-magnitude diagram for $G_1(j\omega)$ is the

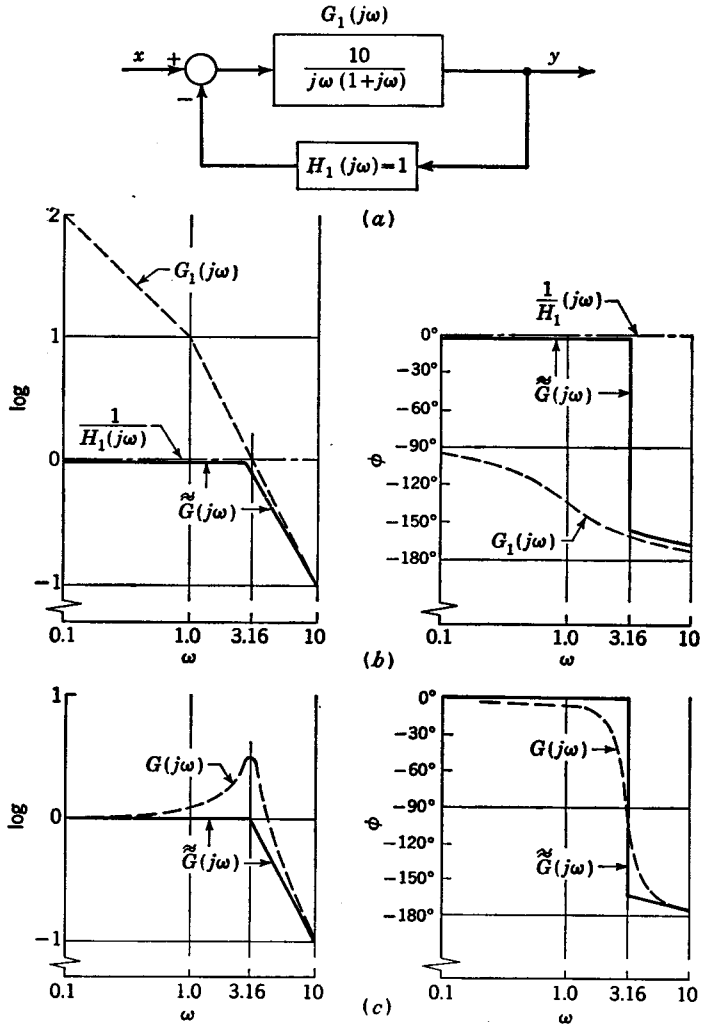


FIG. 10.22. (*a*) Unity feedback placed about a component; (*b*) determination of the approximate frequency response for a component with unity feedback; (*c*) comparison of the approximate and exact frequency response for a component with unity feedback.

dashed line in Fig. 10.22*b*. The plot for $1/H_1(j\omega) = 1$ is indicated by the long- and short-dashed line. As shown in Fig. 10.22*b*, the curve of $\log |G_1(j\omega)|$ intersects that for $\log [1/|H_1(j\omega)|]$ at $\omega = 3.16$. Thus, for $\omega < 3.16$, the response is approximated by $1/H_1(j\omega)$, and for $\omega > 3.16$, the response is approximated by $G_1(j\omega)$. The exact value for $G(j\omega)$ is obtained by solving the equation $G(j\omega) = G_1(j\omega)/[1 + G_1(j\omega)H_1(j\omega)]$. In Fig. 10.22*c* are shown both the exact value of $G(j\omega)$ and also the approximation $\tilde{\tilde{G}}(j\omega)$.

In applying this procedure, it makes no difference whether $H_1(j\omega)$ is a constant or some function of the frequency ω. The intersection of the log-magnitude plot for $\log |G_1(j\omega)|$ and that for $\log [1/|H_1(j\omega)|]$ determines the frequency ω at which $|G_1(j\omega)H_1(j\omega)| = 1$, whence the approximations are made.

Usually at high frequencies $|G_1(j\omega)H_1(j\omega)| \ll 1$, and thus it follows that internal feedback does not affect the high-frequency response. However, the low-frequency response $1/H_1(j\omega)$ is determined entirely by the feedback component.

In summary, this approximation converts an internal-feedback path to an approximate open-loop element for which a standard analysis can be made. With the use of inverse polar plots, as is next described, no approximations are employed.

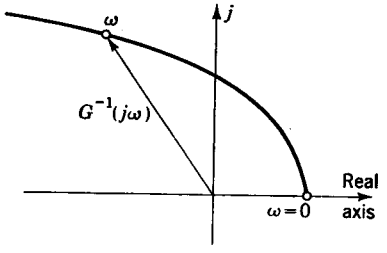


FIG. 10.23. Typical inverse polar plot, $G^{-1}(j\omega)$.

Inverse Polar Plots. A plot of the function $G^{-1}(j\omega) = 1/G(j\omega)$ is called an inverse polar plot. In Fig. 10.23 is shown a typical inverse polar plot of the function $G^{-1}(j\omega)$. At any frequency ω, the vector from the origin to the graph defines the vector $G^{-1}(j\omega)$ for that frequency. The length of the vector is $|G^{-1}(j\omega)| = |1/G(j\omega)|$, and the angle is

$$\measuredangle G^{-1}(j\omega) = \measuredangle \frac{1}{G(j\omega)} = -\measuredangle G(j\omega)$$

A plot of M circles and α lines for inverse polar plots is accomplished by first taking the reciprocal of Eq. (9.52), i.e.,

$$\frac{r(j\omega)}{c(j\omega)} = \frac{1 + G(j\omega)}{G(j\omega)} = G^{-1}(j\omega) + 1 \tag{10.22}$$

A typical vector $G^{-1}(j\omega)$ as shown in Fig. 10.23 may be written in the general form $G^{-1}(j\omega) = x + jy$. Substitution of this general representa-

tion for $G^{-1}(j\omega)$ into Eq. (10.22) gives

$$\frac{r(j\omega)}{c(j\omega)} = x + 1 + jy \tag{10.23}$$

Because $M = |c(j\omega)/r(j\omega)|$, from the magnitude of Eq. (10.23) it follows that

$$\left|\frac{r(j\omega)}{c(j\omega)}\right| = \frac{1}{M} = \sqrt{(x+1)^2 + y^2}$$

Squaring this result gives

$$(x+1)^2 + y^2 = \frac{1}{M^2} \tag{10.24}$$

Thus, on the inverse plane, lines of constant M are circles of radius $1/M$. The center of these concentric M circles is at the point $x = -1$ and

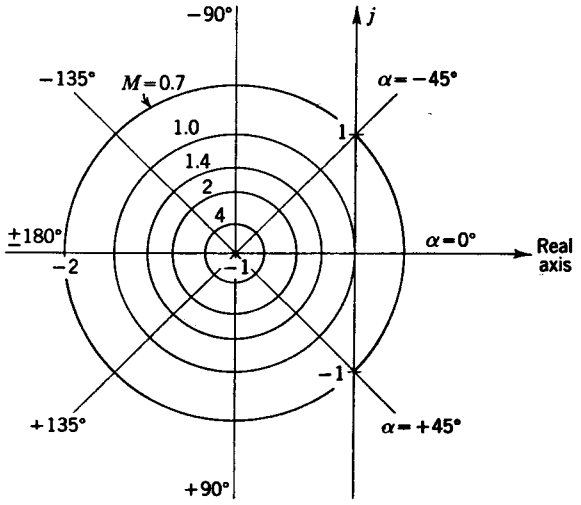


FIG. 10.24. M circles and α lines on the inverse plane.

$y = 0$, that is, the -1 point. A plot of these M circles on the inverse plane is shown in Fig. 10.24. Because the reciprocal of -1 is still -1, this point has the same significance for an inverse polar plot as for a direct polar plot. Polar plots are referred to as direct polar plots when it is necessary to distinguish them from inverse polar plots.

The lines of constant $\alpha = \angle[c(j\omega)/r(j\omega)] = -\angle[r(j\omega)/c(j\omega)]$ are determined from Eq. (10.23) as follows:

$$\alpha = -\angle\frac{r(j\omega)}{c(j\omega)} = -\tan^{-1}\frac{y}{x+1} \tag{10.25}$$

When plotted on the inverse plane, as is shown in Fig. 10.24, constant α contours are radial straight lines which pass through the $-1 + j0$ point.

As is illustrated in Fig. 10.25, the angle ψ of a radial line drawn from the origin tangent to any M circle is

$$\sin \psi = \frac{|1/M|}{|-1|} = \frac{1}{M} \quad (10.26)$$

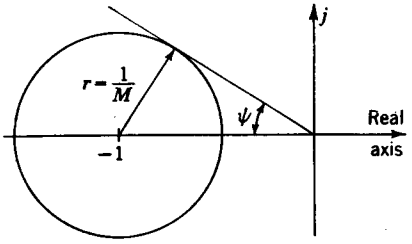


FIG. 10.25. Tangent line to an M circle.

The use of the inverse polar plot for determining the gain K to yield a desired value of M_m is similar to that for the direct polar plot. Consider the same function $G(j\omega) = K/[j\omega(1 + 0.1j\omega)]$ discussed in Chap. 9. The plot of the inverse function

$$\frac{K}{G(j\omega)} = KG^{-1}(j\omega) = j\omega(1 + 0.1j\omega) \quad (10.27)$$

is shown in Fig. 10.26. For a desired $M_m = 1.4$, the angle ψ of the tangent line is

$$\psi = \sin^{-1} \frac{1}{M_m} = \sin^{-1} \frac{1}{1.4} = 45.6° \quad (10.28)$$

Next construct by trial and error the circle which is tangent to both the $KG^{-1}(j\omega)$ plot and the tangent line. In order that this circle be the desired M_m circle, its center must be at the -1 point. From Fig. 10.26, it is to be seen that the center is at $-A = -16.7$. To convert this to the -1 point, the scale factor must be multiplied by $1/A = 1/16.7 = 0.06$.

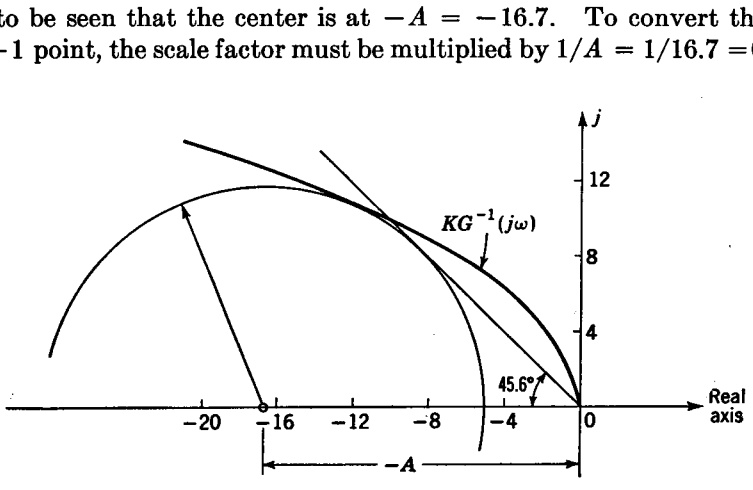


FIG. 10.26. Inverse polar plot $KG^{-1}(j\omega) = j\omega(1 + 0.1j\omega)$.

The resulting function $G^{-1}(j\omega)$, which is tangent to the desired M circle, is

$$G^{-1}(j\omega) = \frac{1}{A}[KG^{-1}(j\omega)] = 0.06j\omega(1 + 0.1j\omega)$$

or

$$G(j\omega) = \frac{16.7}{j\omega(1 + 0.1j\omega)} \tag{10.29}$$

Thus, the value of A yields directly the required gain K.

The general procedure in obtaining K by use of the inverse polar plot is:

1. Plot the inverse function $KG^{-1}(j\omega)$.
2. Construct the tangent line in accordance with Eq. (10.26).
3. By trial and error, determine the circle which is tangent to both the $KG^{-1}(j\omega)$ plot and the tangent line.
4. The desired gain is $K = A$.

The major advantage of using the inverse plane is realized for systems with internal feedback. The reciprocal of Eq. (10.20) is

$$G^{-1}(j\omega) = \frac{1 + G_1(j\omega)H_1(j\omega)}{G_1(j\omega)} = G_1^{-1}(j\omega) + H_1(j\omega) \tag{10.30}$$

The vectors $G_1^{-1}(j\omega)$ and $H_1(j\omega)$ may be added as vector quantities to yield $G^{-1}(j\omega)$, as is illustrated in Fig. 10.27.

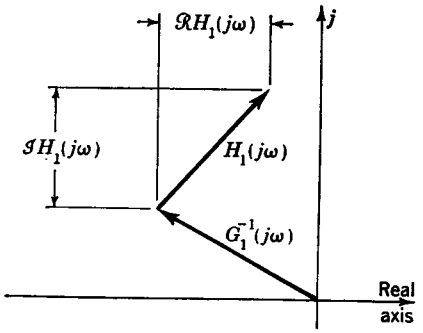


FIG. 10.27. Vector addition of $G_1^{-1}(j\omega)$ and $H_1(j\omega)$ to yield $G^{-1}(j\omega)$.

Illustrative Example 1. For the system shown in Fig. 10.28*a*, let it be desired to determine K_1 so that $M_m = 1.4$.

SOLUTION. For this system

$$G(j\omega) = K_2 \frac{G_1(j\omega)}{1 + G_1(j\omega)H_1(j\omega)} \tag{10.31}$$

or

$$G^{-1}(j\omega) = \frac{1}{K_2}[G_1^{-1}(j\omega) + H_1(j\omega)] = \frac{1}{K_1}\frac{K_1G_1^{-1}(j\omega)}{K_2} + \frac{H_1(j\omega)}{K_2} \tag{10.32}$$

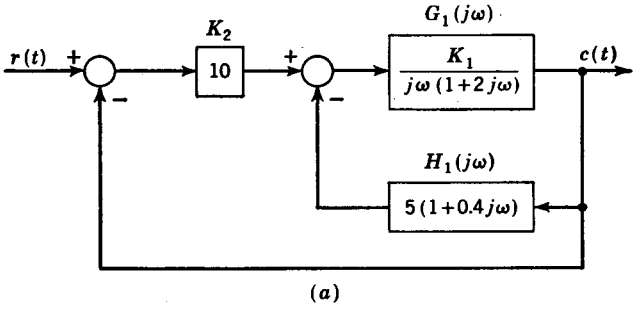
First construct the plot for

$$\frac{H_1(j\omega)}{K_2} = 5\frac{(1 + 0.4j\omega)}{10} = 0.5 + 0.2j\omega \tag{10.33}$$

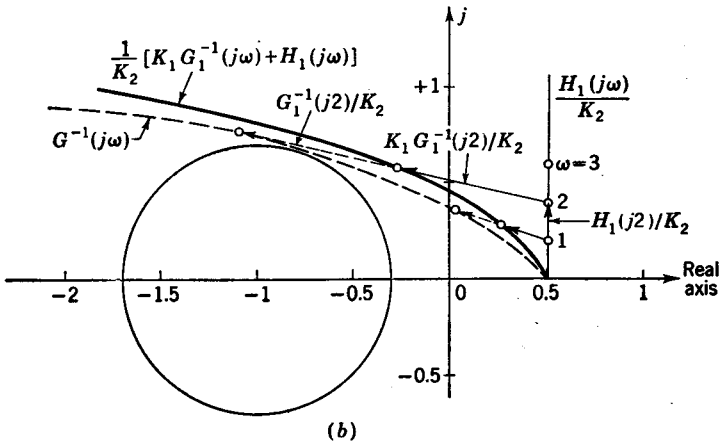
As is shown in Fig. 10.28*b*, the real part of $H_1(j\omega)/K_2$ is always 0.5 and the imaginary part is $0.2j$ for $\omega = 1$, $0.4j$ for $\omega = 2$, etc. The function $K_1G_1^{-1}(j\omega)/K_2$ is

$$\frac{1}{K_2}[K_1G_1^{-1}(j\omega)] = \frac{j\omega(1 + 2j\omega)}{10} = -0.2\omega^2 + 0.1j\omega \tag{10.34}$$

The value of $K_1G_1^{-1}(j\omega)/K_2$ for $\omega = 1$ and 2 is added to the $H_1(j\omega)/K_2$ plot, as is shown by the solid arrows in Fig. 10.28*b*. Multiplication of each $K_1G_1^{-1}(j\omega)/K_2$ vector by $1/K_1$ to obtain $G_1^{-1}(j\omega)/K_2$ is effected by changing the length of each of these vectors by the factor $1/K_1$, as indicated by the dashed extension of each vector. The required value of $1/K_1$ is the



(*a*)



(*b*)

FIG. 10.28. (*a*) System with internal feedback; (*b*) determination of K_1 by inverse polar plots.

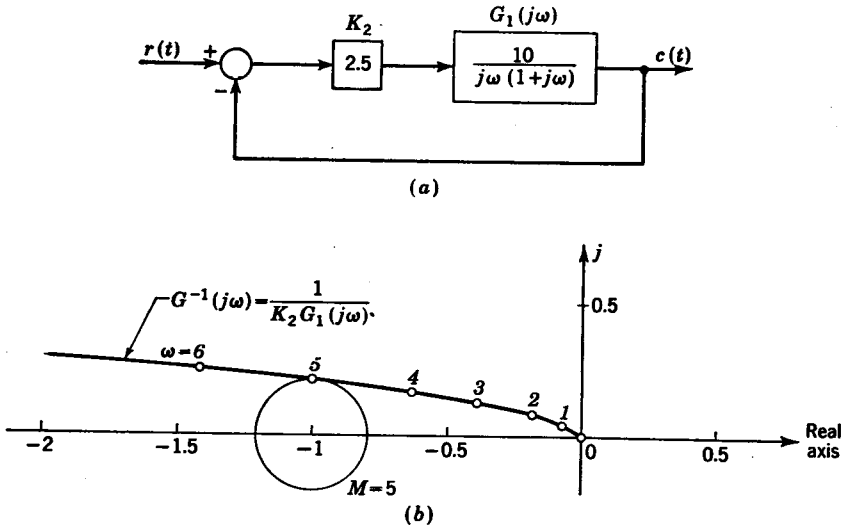


FIG. 10.29. System without internal feedback.

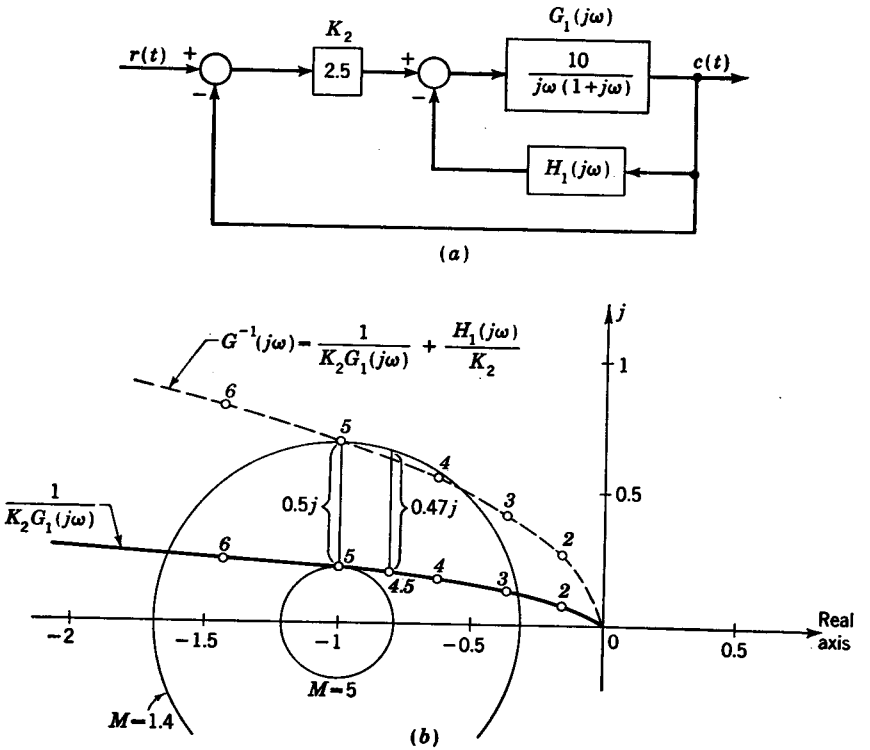


FIG. 10.30. System of Fig. 10.29 with internal feedback.

factor which makes the resulting inverse polar plot $G^{-1}(j\omega)$ tangent to the desired M_m circle. In this case, $1/K_1 = 2$, or $K_1 = 0.5$.

Illustrative Example 2. For the system shown in Fig. 10.29*a*, the value of K_1 is 10 and K_2 is 2.5. From the inverse polar plot $1/K_2G_1(j\omega)$ of Fig. 10.29*b*, the value of M_m is found to be 5. If it be desired to obtain an M_m of 1.4 by the use of an internal-feedback loop as shown in Fig. 10.30*a*, what feedback element $H_1(j\omega)$ should be used?

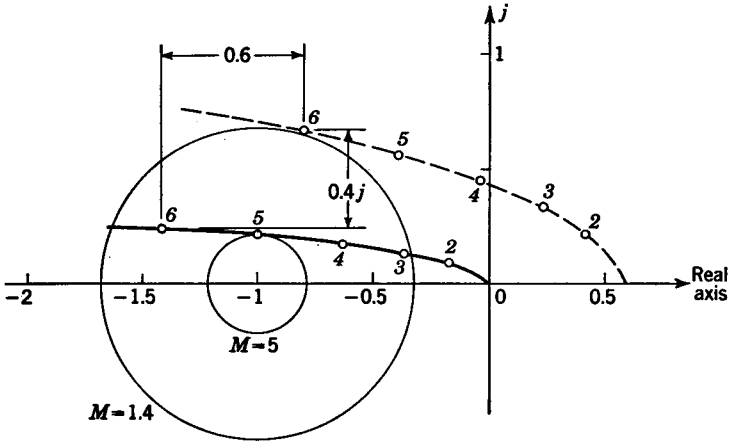


FIG. 10.31. Use of inverse polar plot to increase the speed of response.

SOLUTION. The equation for the inverse polar plot with internal feedback is

$$\frac{1}{G(j\omega)} = \frac{1}{K_2}\frac{1 + G_1(j\omega)H_1(j\omega)}{G_1(j\omega)} = \frac{1}{K_2G_1(j\omega)} + \frac{H_1(j\omega)}{K_2} \qquad (10.35)$$

The quantity $H_1(j\omega)/K_2$ must be such that when it is added to $1/K_2G_1(j\omega)$ the resulting plot will be tangent to the $M = 1.4$ circle. To do this, it is necessary only to raise the original plot vertically. It is to be seen from Fig. 10.27 that the horizontal component of $H_1(j\omega)$ is its real part $\Re\ H_1(j\omega)$ and the vertical component is its imaginary part $\jmath\ H_1(j\omega)$. It is thus necessary only that $H_1(j\omega)$ be purely imaginary, i.e., of the form $H_1(s) = \beta s$ or $H_1(j\omega) = j\beta\omega$. In Fig. 10.30*b*, it is to be noted that, at $\omega = 5$, the addition of $H_1(j5)/K_2 = 0.5j$ to the $1/K_2G_1(j5)$ plot causes the resulting curve to pass through the top of the $M = 1.4$ circle. For this case, the value of β is $j\beta 5/2.5 = 0.5j$, or $\beta = 0.25$. The resulting curve may now be constructed as indicated by the dashed line in Fig. 10.30*b*. Because this curve is not tangent to the $M = 1.4$ circle, another trial value must be taken. From the dashed loci of Fig. 10.30*b*, it now appears that the point of tangency is more likely to occur in the neighbor-

hood of $\omega = 4.5$. The addition of $H_1(j4.5)/K_2 = 0.47j$ to $1/K_2G_1(j4.5)$ causes the new resulting curve to be tangent to the desired M circle. In this case $j\beta 4.5/2.5 = 0.47j$, or $\beta = 0.26$. Thus, the desired result is $H_1(j\omega) = 0.26j\omega$.

When $H_1(j\omega)$ is a constant, $H_1(j\omega)$ is entirely real. As may be seen from Fig. 10.30*b*, the effect of a constant $H_1(j\omega)$ is to shift the inverse plot horizontally to the right. Suppose in the preceding problem that it is desired to increase the speed of response by having ω_m equal to 6 rather than 4.5. After assuming a few trial values for $H_1(j6)/K_2$, it is found that $H_1(j6)/K_2 = 0.6 + 0.4j$ makes the resultant plot tangent to the $M = 1.4$ circle at $\omega = 6$, as is illustrated in Fig. 10.31. Because $H_1(j6)/K_2 = 0.6 + 0.4j = 0.6 + j\beta 6$, then $\beta = 0.4/6 = 0.067$, whence the required $H_1(j\omega)$ is

$$H_1(j\omega) = K_2(0.6 + j\beta\omega) = 1.5 + j0.167\omega \qquad (10.36)$$

The general procedure followed in this illustrative example was to assume a value of $H_1(j\omega)$ which makes a point lie on the desired M circle. From this assumed value, the general equation for the resulting plot was obtained. When the assumed point is not the point of tangency, then another trial point must be selected.

CHAPTER 11

HYDRAULIC SYSTEMS

11.1. Introduction. Hydraulic devices are used extensively in control systems. With high-pressure hydraulic systems, very large forces are obtained. Such forces provide power for rapid accelerations, accurate positioning of heavy loads, etc. Because hydraulic motors are much smaller than equivalent electric motors, considerable size and weight savings can be realized.

In electrical systems, magnetic devices such as motors, solenoids, etc., are used to provide the "muscle" for doing work. The operation of magnetic equipment is characterized by relatively long time lags. In general, hydraulic components are very rapid-acting. Another feature is that hydraulic equipment is more rugged than corresponding electrical components. This can be a major factor in applications such as aircraft, where vibrations and shock may cause fine wires and delicate tubes or transistors to fail. In addition, the noise pickup from such vibrations may adversely effect the normal operation of electrical equipment.

Electrical equipment is better suited for applications in which components must be located far apart, as in remote-control positioning systems. The reason for this is that electrical signals may be transmitted long distances via wires or microwaves.

In general, pneumatic and hydraulic systems are quite similar. An advantage of pneumatic equipment is the accessibility and convenience of using air. However, because of the compressibility of air, pneumatic systems do not have the positive action afforded by hydraulic systems, which employ an incompressible fluid as the working medium.

Electrical components are inherently better suited for certain operations, hydraulic components for others, and pneumatic components for still others. However, for many applications, it is possible to accomplish a desired function almost equally well by electric, hydraulic, or pneumatic equipment. Thus, the designer must weigh the relative importance of size, weight, cost, accuracy, ruggedness, reliability, etc. Frequently, it is necessary to make a detailed design analysis of corresponding electric, hydraulic, or pneumatic components to obtain a good evaluation of these various factors. Sometimes, it may even be necessary to build models for testing before a final decision can be made. It is not uncommon to

utilize all three types, i.e., electric, hydraulic, and pneumatic equipment, in one control system.

To be able to select the components which are best suited to his particular requirements, a designer must be familiar with all types of commonly used control apparatus. Characteristics of pneumatic and electric apparatus are presented in the next two chapters, respectively. This chapter presents basic considerations involved in the design of hydraulic systems. Emphasis is given to explaining the basic laws and equations which govern the operation of hydraulic equipment. Typical examples of commonly used hydraulic devices and the manner in which their performance may be evaluated are also presented. These basic principles and techniques may then be applied to any hydraulic equipment. In one chapter it is not possible to indicate the innumerable practical applications and uses for hydraulic systems.[1–4] Thus, of necessity, the primary effort of this chapter is limited to basic considerations.

Three major classifications of elements for hydraulic circuits are pumps, valves, and receiving units. The basic functions of these elements are distinguished as follows: pumps supply the high-pressure fluid for the system, valves control the direction and amount of flow, and receiving units utilize this flow of fluid to accomplish the desired objective.

11.2. Pumps. The three types of pumps most commonly used for hydraulic power purposes are the gear pump, vane pump, and piston pump.

Gear Pump. A gear pump is shown in Fig. 11.1. Because of the direction of rotation of the gears, the inlet fluid is carried around the outer periphery of the gears to the high-pressure discharge side of the pump. The meshing gear teeth provide a seal to prevent return of the fluid to the low-pressure side of the pump. This type of pump is less expensive than other pumps and is also very rugged. Its best efficiency occurs at lower speeds and operating pressures than those for vane or piston pumps.

There is a leakage between the tips of the gear teeth and the housing and also between the sides or faces of the gear blanks and housing. At high pressures and operating speeds, this leakage increases considerably, which tends to decrease the over-all efficiency of this pump.

[1] Publishers of *Applied Hydraulics*, "Fluid Power Directory 1958/1959," Industrial Publishing Corp., Cleveland, 1957.

[2] J. J. Pippenger and R. F. Koff, "Fluid-power Controls," McGraw-Hill Book Company, Inc., New York, 1959.

[3] J. G. Truxal, "Control Engineers' Handbook," McGraw-Hill Book Company, Inc., New York, 1958.

[4] I. McNeil, "Hydraulic Operation and Control of Machines," The Ronald Press Company, New York, 1958.

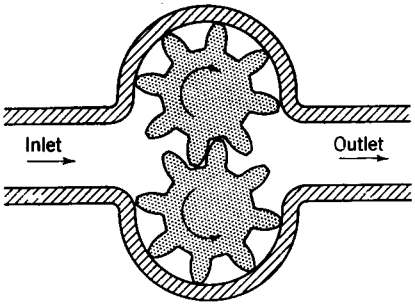


FIG. 11.1. Gear pump.

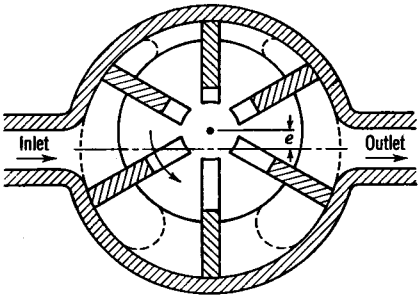
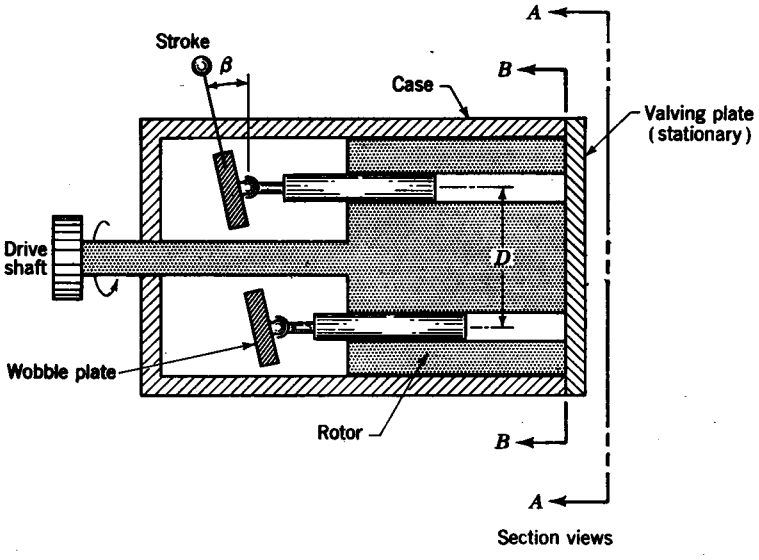


FIG. 11.2. Vane pump.



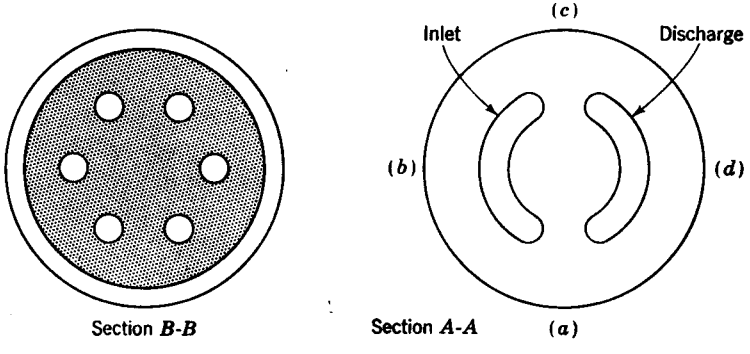


FIG. 11.3. Axial piston pump.

Vane Pump. A vane pump is shown in Fig. 11.2. The centrifugal force on each of the vanes is usually sufficient to maintain contact between the vanes and the housing. For a counterclockwise rotation of the rotor as shown in Fig. 11.2, a large amount of fluid is carried from the inlet side of the pump to the discharge side. Because of the eccentricity e of the center of the rotor with respect to the housing, more fluid is carried to the high-pressure side of the pump than is returned to the low-pressure side. The net flow is seen to depend upon the amount of

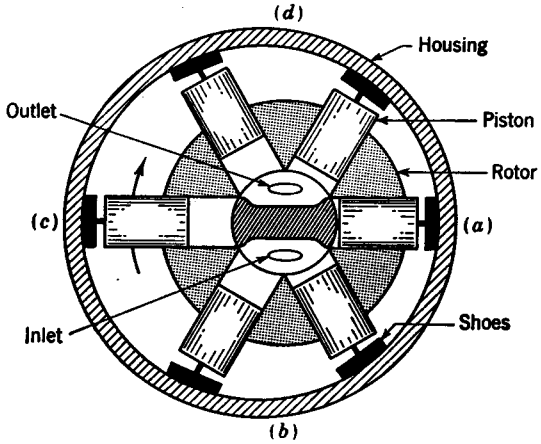


FIG. 11.4. Radial piston pump.

eccentricity e. By varying the eccentricity, a vane pump can be used as a variable-delivery pump.

Axial Piston Pump. An axial piston pump is shown in Fig. 11.3. The pistons are located in the rotor (i.e., rotating cylinder block), which is driven by the drive shaft. Because the pistons are parallel to the drive shaft, this type of pump is called an axial piston pump. The wobble plate does not rotate, and furthermore its angle of inclination β is set by the position of the stroke-adjusting lever. The axial displacement of each piston ($X = D \tan \beta$) is varied by changing the angle of inclination β of the wobble plate. As a piston in the cylinder block rotates in a clockwise direction from (a) to (b) to (c) (pump viewed from right end), fluid is being admitted as the stroke is increasing. Then in going from (c) to (d) to (a) the fluid is forced out to the high-pressure, or discharge, side of the pump as the stroke is decreasing.

Radial Piston Pump. The pistons of a radial piston pump move in a radial direction, as is illustrated in Fig. 11.4. The inner shaft (i.e., pintle), which contains the inlet and discharge passageways, is fixed, as is the housing. As a piston in the rotor rotates clockwise from (a) to

(*b*) to (*c*), fluid is being admitted as the stroke increases. Then in going from (*c*) to (*d*) to (*a*) fluid is forced out at high pressure as the stroke is decreasing. The stroke and thus the amount of fluid delivered are varied by changing the eccentricity between the center of the housing and the center of the rotor.

The selection of a pump for a particular application depends upon many factors, such as quantity of fluid to be delivered, discharge pressure, reliability desired, cost, operating efficiency, size, etc. As is the case with most specific equipment, a manufacturer's catalogue should be consulted to obtain data regarding such details.

In general, the power output for any pump may be computed from the equation

$$\text{hp} = \frac{Q(P_o - P_{in})}{(12)(550)} \tag{11.1}$$

where Q = net rate of flow delivered, in.^{3}/sec
$P_o - P_{in}$ = change in pressure, psi
The over-all pump efficiency is equal to the hydraulic power output of the pump as given by Eq. (11.1) divided by the power supplied to drive the pump.

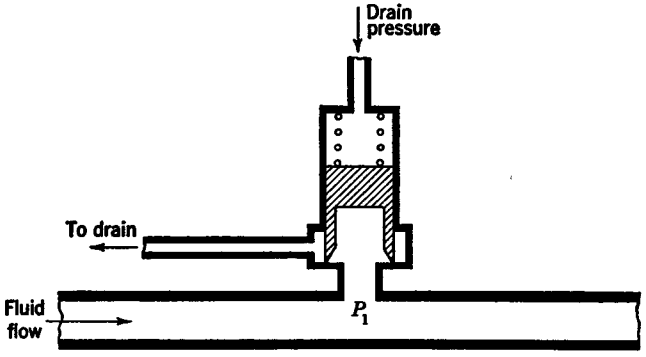


FIG. 11.5. Relief valve.

11.3. Valves. From the time the fluid leaves the pump until it reaches the receiving units, the flow is controlled and directed by valves.

Relief Valve. A relief valve is shown in Fig. 11.5. When the pressure P_1 in the main flow passage is high enough to overcome the spring force tending to close the valve, the valve opens. This connects the main flow passage to the return to sump (i.e., pump reservoir), which is at drain pressure. This relief valve remains open until the pressure P_1 decreases to the value that was required to open the valve. In effect, then, a relief valve limits the maximum obtainable pressure P_1. Because of

its function, this type of valve is often referred to as a pressure-regulating valve.

If A is the cross-sectional area of the plunger in the relief valve, then the pressure force P_1A tends to raise the valve. The force F_c is the force exerted on the plunger by the spring when the valve is closed. When the plunger is raised a distance X from its closed position, an additional force KX is exerted by the spring. Thus

$$P_1A = F_c + KX \tag{11.2}$$

or

$$P_1 = \frac{F_c}{A} + \frac{KX}{A} \tag{11.3}$$

To minimize the variation in pressure P_1 due to the opening X, then K/A should be as small as possible. In addition, by making the diameter D of the plunger large, the plunger does not have to open so wide to bypass the flow. This minimizes the opening X. For most applications, the pressure P_1 may be regarded as the cracking pressure F_c/A, that is, the pressure at which the valve just opens or cracks.

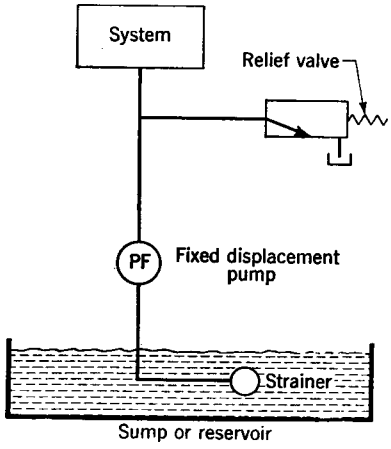


FIG. 11.6. Relief valve used to provide a constant-pressure power supply.

In Fig. 11.6 is shown the fluid supply system for a typical hydraulic circuit. The oil in the sump is strained before going to the pump, which is driven by a motor. A fixed-displacement pump is indicated by the circle with PF in it. The representation for the spring-loaded relief valve indicates that the pressure in the main line is opposed by the spring shown on the opposite side of the rectangle. The arrow in the box indicates the direction of flow through the valve. The arrow is drawn in the position that indicates the flow when the spring force is greater than the opposing pressure. In this case, the flow is blocked off. When the opposing pressure exceeds the spring force, it is connected to the sump. The drain, or return to sump, is indicated by drawing a little sump under the valve.

A fixed-displacement pump which is driven at a constant speed delivers a relatively constant amount of flow regardless of the discharge pressure. The excess flow not used by the system is bypassed by the relief valve. This bypass flow represents wasted power because it was supplied at high pressure. The resulting wasted energy increases the temperature

of the fluid in the system. If the rise in temperature is sufficient, it may be necessary to provide a cooler.

Accumulator and Unloading Valve. A constant-pressure supply may also be obtained by the use of an accumulator and an unloading valve, as is shown in Fig. 11.7*a*. The accumulator holds the supply of fluid at constant pressure for use as the system needs it. An accumulator has an elastic diaphragm or bladder to separate the top portion, which.

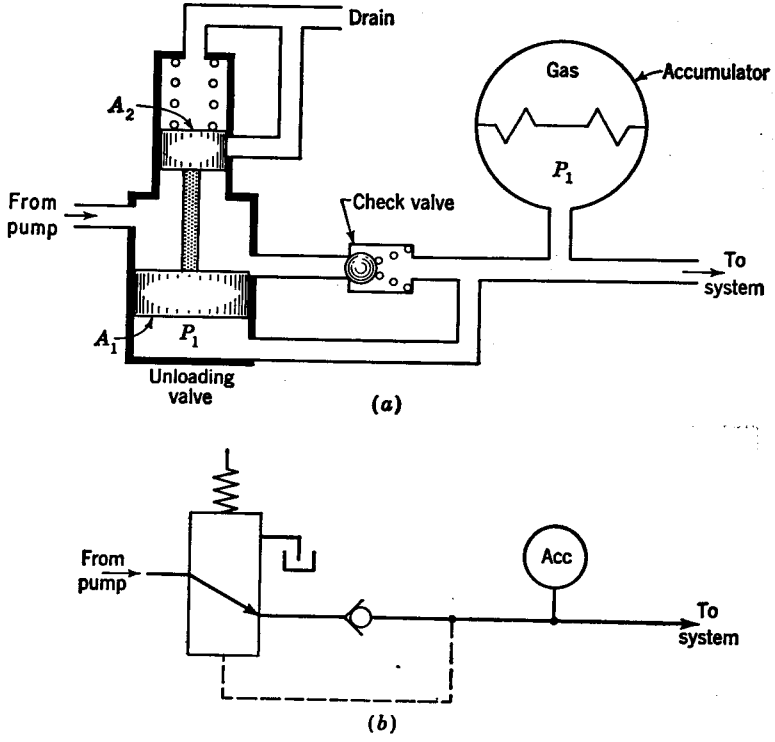


FIG. 11.7. Accumulator and unloading valve used to provide a constant-pressure power supply.

contains a gas under pressure, from the supply of fluid on the bottom side. The purpose of the gas is to maintain the fluid pressure. As the fluid level drops, the gas expands and the pressure of the fluid decreases slightly. The same effect may also be obtained by using a spring to load the top side of the diaphragm.

The unloading valve in Fig. 11.7*a* is drawn in the position in which the pump flow is refilling the accumulator. The check valve offers no resistance to flow to the accumulator, but it closes off to prevent reverse, or leakage, flow. As the accumulator fills, the pressure P_1 increases and the plunger rises. Just before the drain port is uncovered, the pressure P_1

is acting on both sides of the bottom land. The net upward force on the valve is P_1A_2. After the drain is uncovered, both sides of the top land are connected to drain, as is the top side of the bottom land, so that the upward force is now P_1A_1. Because A_1 is greater than A_2, the upward force suddenly increases to unload the pump by opening the drain port wide. Thus the pump is unloaded when

$$P_{1,\max} = \frac{F_c}{A_2} \tag{11.4}$$

where F_c is the force of the spring when the drain port is just uncovered and $P_{1,\max}$ is the maximum value of P_1. As the fluid in the accumulator is used, the pressure P_1 decreases and the plunger moves down to close off the drain. Just before the drain is closed off, the forces acting on the plunger are

$$P_{1,\min} = \frac{F_c}{A_1} \tag{11.5}$$

where $P_{1,\min}$ is the minimum value of P_1. After the drain port is closed off, the pressure force is suddenly decreased from P_1A_1 to P_1A_2. The spring then pushes the plunger down to open wide the passageway to the accumulator in order to refill it. For a large accumulator, more fluid can be used before the supply of fluid must be replenished. The purpose of employing a differential-area plunger is to prevent the pump from continually being unloaded and then loaded, i.e., to provide a spread between $P_{1,\max}$ and $P_{1,\min}$.

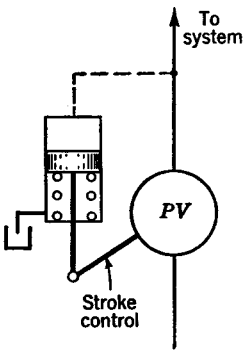


FIG. 11.8. Variable-delivery pump to provide a constant-pressure power supply.

In the symbolic representation of the unloading valve in Fig. 11.7*b*, the pressure of the fluid in the accumulator opposes the spring force. The arrow indicates the direction of flow through the valve when the spring force is greater than the pressure force, i.e., when the accumulator is being refilled. When the pressure force exceeds the spring force, the pump is connected to drain.

The use of a variable-delivery pump to provide a constant pressure supply is illustrated in Fig. 11.8. When the pressure to the system exceeds the nominal value, the piston in the cylinder is forced down. This actuates the pump-delivery control linkage to decrease the flow. The maximum variation in supply pressure is equal to the variation in pressure required to move the pump-delivery control linkage (e.g., wobble-plate control) from maximum to no flow. There are numerous types of hydraulic power supplies; most of these, by far, are of the constant-pressure type.

Much time and effort are saved by using the line-diagram method for drawing hydraulic circuits. In Fig. 11.9 is shown a partial listing of symbolic representations for various circuit elements. These are standard symbols which have been adopted for hydraulic circuits by the Joint Hydraulic Industrial Conference (JIC).

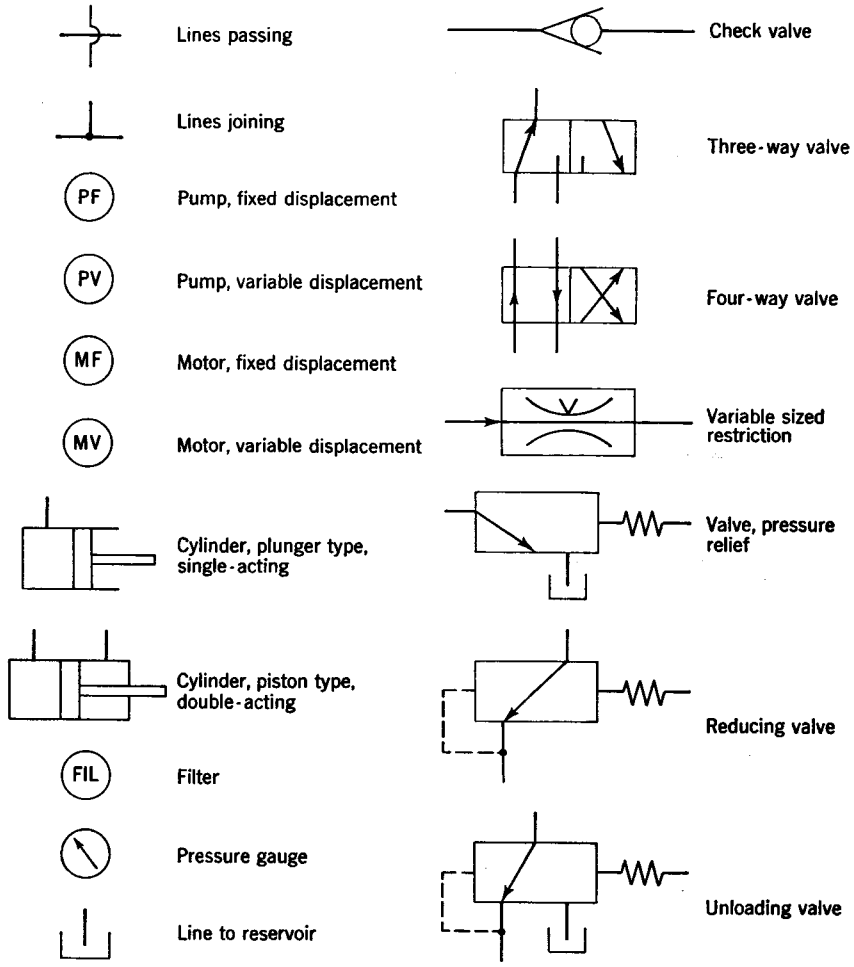


FIG. 11.9. JIC standard symbols for hydraulic circuits.

Differential-pressure-regulating Valve. A differential-pressure-regulating valve maintains a constant pressure difference $P_1 - P_2$ between two points in a system. In Fig. 11.10*a*, the pressure P_2 is determined by the downstream characteristics of the system. This system must be supplied by a constant-flow power supply such as is obtained by the use of a fixed-

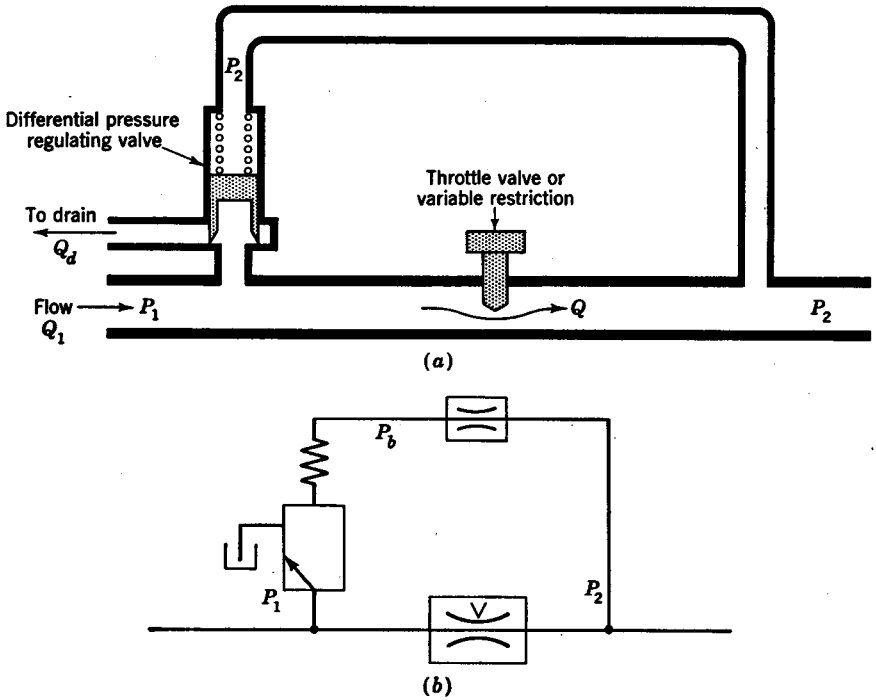


FIG. 11.10. Differential-pressure-regulating valve.

displacement pump. When $P_1 - P_2$ exceeds the nominal value set by the spring, the plunger rises to bypass more flow to drain. The flow through the throttle valve is decreased because of this additional bypass flow. This decrease in flow through the throttle valve in turn decreases the pressure drop $P_1 - P_2$ across it. The force equation for the plunger is

$$(P_1 - P_2)A = F_c + KX$$

or

$$P_1 - P_2 = \frac{F_c}{A} + \frac{KX}{A} \tag{11.6}$$

As for the case of the relief valve, the quantity K/A should be small and the diameter D of the plunger should be large to minimize the variation $P_1 - P_2$ from the cracking value F_c/A. The pressure drop $P_1 - P_2$ may be regarded as remaining essentially constant. Thus, the flow across the variable-sized restriction, or throttle valve, is controlled entirely by the area of the throttle valve. In effect, a flow control device results.

If one writes the equation of motion for the plunger and considers

the mass M of the plunger, it follows that

$$Mp^2X = (P_1 - P_2)A - KX - F_c$$

or

$$X = \frac{1/M[(P_1 - P_2)A - F_c]}{p^2 + K/M} \tag{11.7}$$

The term in brackets is the difference between the pressure force acting on the valve and the spring force at cracking. From Eq. (11.7), it follows that the plunger will tend to oscillate or chatter at its natural frequency $\omega_n = \sqrt{K/M}$. This undesirable effect is avoided by inserting a viscous damper in the line between the spring side of the plunger and the discharge line at pressure P_2, as illustrated in the symbolic representation of Fig. 11.10*b*. Viscous damping is provided by an orifice whose length is sufficiently greater than its diameter so as to cause laminar flow. The flow through such a restriction is given by the equation

$$Q = \frac{\pi D^4 \Delta P}{128 \mu L} = C \Delta P \tag{11.8}$$

where Q = rate of flow, in.^{3}/sec
D = diameter of restriction, in.
ΔP = pressure drop across restriction, psi
L = length of restriction, in.
μ = absolute viscosity, reyns (lb-sec/in.^{2})

If the pressure behind the plunger between the spring and restriction is designated as P_b, the pressure drop across the restriction is $\Delta P = P_b - P_2$. Because the fluid is incompressible, the flow through this restriction is equal to the change of volume $A(dX/dt)$, on the spring side of the plunger. Thus

$$C(P_b - P_2) = ApX \tag{11.9}$$

The force equation for the plunger is now

$$Mp^2X + KX = (P_1 - P_b)A - F_c \tag{11.10}$$

Eliminating P_b from Eqs. (11.9) and (11.10) and solving for X gives

$$X = \frac{1/M[(P_1 - P_2)A - F_c]}{p^2 + (A^2/CM)p + K/M} \tag{11.11}$$

whence $\omega_n = \sqrt{K/M}$ and $\zeta = A^2/(2C\sqrt{KM})$. The viscous damper thus provides a convenient means for eliminating the chatter. For steady-state operation, there is no flow through the damper so that $P_b = P_2$. Thus, the basic function of maintaining a constant pressure drop $P_1 - P_2$ is not altered by the damper.

Reducing Valve. A reducing valve is used to lower the pressure to a predetermined value. That is, it regulates the outlet pressure from the

valve. A reducing valve is shown in Fig. 11.11*a*. When the outlet pressure is equal to the cracking value set by the spring, the plunger just shuts off the outlet from the inlet passageway. A decrease in the outlet pressure lowers the valve to admit high-pressure fluid, which in turn increases the outlet pressure. This then returns the valve to its line-on-line position. An increase in the outlet pressure closes off the outlet passageway. Thus, the outlet pressure is maintained at its cracking value.

Although a relief valve could be used to regulate the outlet pressure, such a valve would also lower the inlet pressure. A pressure-reducing

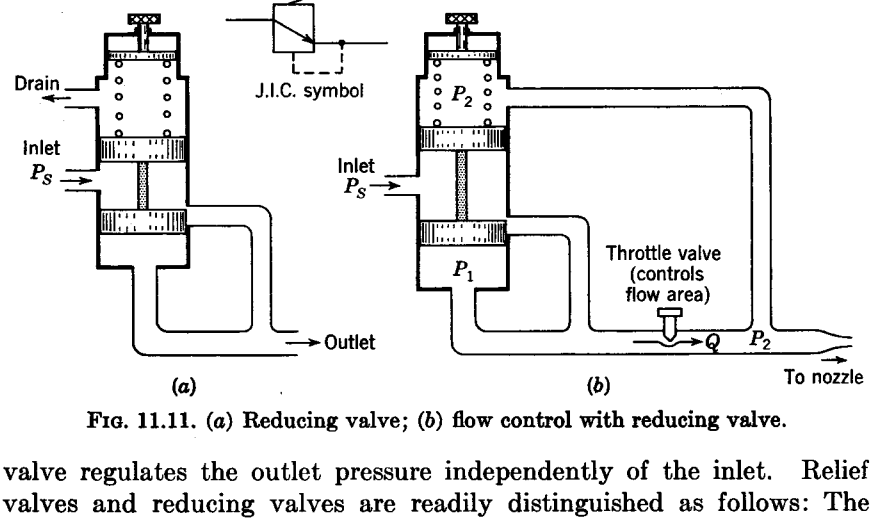


FIG. 11.11. (*a*) Reducing valve; (*b*) flow control with reducing valve.

valve regulates the outlet pressure independently of the inlet. Relief valves and reducing valves are readily distinguished as follows: The spring of a relief valve tends to close it, while that of a reducing valve tends to open it. The pressure force acting on a relief valve tends to open it, while that of a reducing valve tends to close it. In the symbolic representation for the reducing valve, the outlet pressure opposes the spring force. When the outlet pressure gets high enough, the inlet is disconnected from the outlet.

The use of a reducing valve is illustrated by the flow control system shown in Fig. 11.11*b*. The discharge pressure P_2 is controlled independently by the nozzle characteristics which are downstream from the flow control portion of the system shown in Fig. 11.11*b*. To have the flow Q a function of the position of the throttle valve only, it is desired to maintain a constant pressure drop across the throttle valve. From Fig. 11.11*b*, it is seen that the pressure drop P_1-P_2 across the throttle valve acts across the reducing valve. The reducing valve tends to

maintain this pressure drop constant because if P_1 decreases, then the valve opens the outlet-pressure port wider to increase P_1. Similarly, if P_1-P_2 exceeds the nominal value determined by the spring force, the reducing valve moves up to close off the outlet port, which in turn returns P_1-P_2 to its nominal value. For this system, the value of the inlet pressure is independent of the operation of the system, whereas with the differential-pressure-regulating valve the inlet pressure was $P_s = P_1$. To eliminate chatter of the plunger of the reducing valve, it

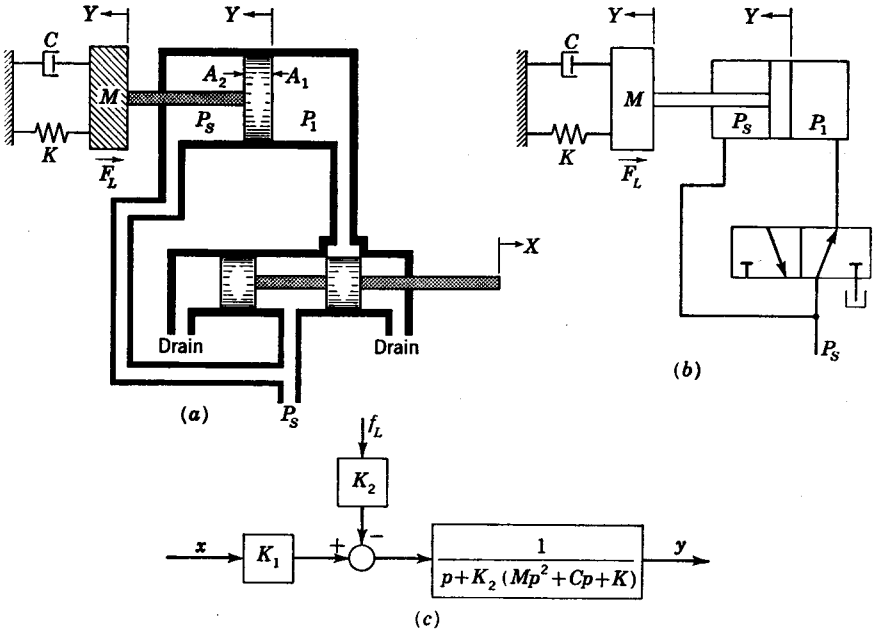


FIG. 11.12. Three-way-valve–cylinder circuit.

is necessary to insert a viscous damper in the line connecting the spring side of the plunger to the discharge line at pressure P_2, as was done for the differential-pressure-regulating valve.

11.4. Three-way Pilot Valve. The purpose of a pilot valve is to control the direction and amount of flow to a receiving unit. A spool-type pilot valve is one in which the shape of the valve resembles that of a spool. A circuit which utilizes a spool-type valve is illustrated in Fig. 11.12*a*. This valve has three external ports: a high-pressure, (or supply) port, a cylinder port, and a drain port. Thus, this valve is further classified as a three-way pilot valve. Pilot valves are sometimes referred to as control valves, servo valves, or proportional valves. When the valve is moved to the right, the high-pressure line is connected to the cylinder port and the drain is blocked off. Although high pressure is

now connected to both sides of the piston, there is a larger area exposed to this pressure on the right side of the piston than on the left. Thus, the greater force on the right side causes the piston and load to move to the left. When the valve is moved to the left of its line-on-line position, the cylinder port is connected to drain. The high pressure which always acts on the left side of the piston now forces the piston and load to move to the right.

The symbolic representation for this circuit is shown in Fig. 11.12*b*. The right half of the valve representation indicates the line connections when the valve is moved to the right. The direction of the arrow indicates the direction of flow from the high-pressure port to the cylinder. The drain connection is blocked off. It is not necessary to reproduce the location of the ports for the left position. With the valve in its left position, the arrow indicates that the direction of flow is from the cylinder port to drain. The high-pressure port is blocked off in this position.

Because turbulent flow exists at a sharp-edged orifice such as a valve, the flow equation when the valve is moved to the right is

$$Q = C_d W X \sqrt{\frac{2g}{\rho}(P_s - P_1)} = K_d X \sqrt{P_s - P_1} \qquad X > 0 \quad (11.12)$$

where $K_d = C_d W \sqrt{2g/\rho}$
Q = flow through valve
C_d = discharge coefficient
WX = port area in which X = length of opening and W = circumference
ρ = fluid density
g = gravitational conversion constant

When the valve is to the left of its line-on-line position, the flow rate is given by the equation

$$Q = K_d X \sqrt{P_1} \qquad X < 0 \quad (11.13)$$

From Eq. (11.12), it is to be noted that flow into the cylinder is represented as a positive quantity. Because X is negative in Eq. (11.13), the return flow is also negative.

From Fig. 11.12*a*, it is to be seen that the flow rate through the valve is equal to the area A_1 of the right side of the piston times the piston velocity pY. Thus

$$Q = A_1 pY = A_1 \dot{Y} \quad (11.14)$$

The pressure force acting on the piston is

$$F = P_1 A_1 - P_s A_2 \quad (11.15)$$

For positive values of X, the over-all equation relating F, X, and $\dot{Y}$ is obtained by substituting P_1 from Eq. (11.15) into Eq. (11.12) and then substituting Q from this resulting expression into Eq. (11.14). Thus

$$\dot{Y} = \frac{Q}{A_1} = \frac{K_d X}{A_1}\sqrt{P_s - \frac{F + P_s A_2}{A_1}}$$
$$= \frac{K_d X}{A_1^{3/2}}\sqrt{P_s(A_1 - A_2) - F} \qquad X > 0 \tag{11.16}$$

For negative values of X, Eq. (11.13) must be used rather than Eq. (11.12). Thus, substituting P_1 from Eq. (11.15) into Eq. (11.13) and then substituting Q from this resulting expression into Eq. (11.14) gives

$$\dot{Y} = \frac{Q}{A_1} = \frac{K_d X}{A_1}\sqrt{\frac{F + P_s A_2}{A_1}}$$
$$= \frac{K_d X}{A_1^{3/2}}\sqrt{F + P_s A_2} \qquad X < 0 \tag{11.17}$$

In both Eqs. (11.16) and (11.17), it is to be noticed that $\dot{Y}$ is a function of X and F. Thus, linearization of Eqs. (11.16) and (11.17) gives

$$\dot{y} = py = \left.\frac{\partial \dot{Y}}{\partial X}\right|_i x + \left.\frac{\partial \dot{Y}}{\partial F}\right|_i f$$

or

$$y = \frac{K_1 x - K_2 f}{p} \tag{11.18}$$

where for $X > 0$

$$\left.\frac{\partial \dot{Y}}{\partial X}\right|_i = \left.\frac{K_d}{A_1^{3/2}}\sqrt{P_s(A_1 - A_2) - F}\right|_i = K_1 \tag{11.19}$$

$$\left.\frac{\partial \dot{Y}}{\partial F}\right|_i = \left.\frac{-K_d X}{2A_1^{3/2}\sqrt{P_s(A_1 - A_2) - F}}\right|_i = -K_2 \tag{11.20}$$

and for $X < 0$

$$\left.\frac{\partial \dot{Y}}{\partial X}\right|_i = \left.\frac{K_d}{A_1^{3/2}}\sqrt{F + P_s A_2}\right|_i = K_1 \tag{11.21}$$

$$\left.\frac{\partial \dot{Y}}{\partial F}\right|_i = \left.\frac{K_d X}{2A_1^{3/2}\sqrt{F + P_s A_2}}\right|_i = -K_2 \tag{11.22}$$

In the general case, the pressure force F acting on the piston is resisted by the inertia Mp^2Y of the load, viscous friction CpY, a spring force KY, an external load F_L, thus,

$$F = (Mp^2 + Cp + K)Y + F_L$$

Linearization gives

$$f = (Mp^2 + Cp + K)y + f_L \tag{11.23}$$

The substitution of f from Eq. (11.23) into Eq. (11.18) yields the following over-all equation of operation for the valve-cylinder circuit:

$$[p + K_2(Mp^2 + Cp + K)]y = K_1x - K_2f_L \tag{11.24}$$

The block-diagram representation for Eq. (11.24) is shown in Fig. 11.12*c*.

For smooth operation, it is desirable that K_1 be the same for positive as well as negative values of X, and in addition K_2 should be the same for positive or negative values of X. By comparison of Eqs. (11.19) and (11.21) and by comparison of Eqs. (11.20) and (11.22) it follows that both the preceding conditions are satisfied when the square-root terms are the same, i.e.,

$$P_sA_1 - P_sA_2 - F = F + P_sA_2$$

or

$$P_sA_1 = 2(P_sA_2 + F) \tag{11.25}$$

When the nominal load F is quite small, the preceding result is satisfied when $A_2 = A_1/2$. For this case, it follows from Eq. (11.15) that $P_1 = P_s/2$, and thus Eqs. (11.19) and (11.21) reduce to

$$K_1 = \frac{K_d}{A_1}\sqrt{\frac{P_s}{2}} = \frac{C_1}{A_1} \tag{11.26}$$

The normal operation of most valves is about the reference point $X_i = 0$. In this case $K_2 = 0$ in accordance with Eqs. (11.20) and (11.22). When $K_2 = 0$ and Eq. (11.26) is applicable for K_1, Eq. (11.18) reduces to the result that was given by Eq. (2.56), i.e.,

$$y = \frac{C_1}{A_1p}x$$

In Fig. 11.13*a* is shown a typical family of curves of F versus $\dot{Y}$ with constant valve positions X. These curves may be obtained analytically from Eqs. (11.16) and (11.17). If an actual valve is available, it is possible to obtain the curves experimentally.

One method of obtaining the curves experimentally is to disconnect the cylinder and place a flowmeter in the P_1 pressure line. Then vary the pressure P_1, and plot the flow through the valve for various fixed (constant) positions X of the valve. A typical family of flow curves is shown in Fig. 11.13*b* (note that $\dot{Y} = Q/A_1$). For a given $\dot{Y}$ and X, the corresponding value of P_1 may be determined, whence F may be computed in accordance with Eq. (11.15). This process is repeated to obtain corresponding values of $\dot{Y}$, X, and F for plotting Fig. 11.13*a*. The family of curves shown in Fig. 11.13*a* is a plot of $\dot{Y}$ as a function of X and F. Linearization of $\dot{Y} = f(X,F)$ again yields Eq. (11.18). The value of $K_1 = \partial\dot{Y}/\partial X|_i$ is obtained directly from a vertical interpolation of the curves about the reference operating point. Similarly

$K_2 = \partial \dot{Y}/\partial F|_i$ is the slope of the line of constant X at the reference point. Thus, these curves provide a very convenient means for evaluating K_1 and K_2 for a particular valve.

Overlapped or Underlapped Valves. Because of manufacturing tolerances, it is not possible to achieve an idealized line-on-line valve. An actual valve will either be overlapped as shown in Fig. 11.14*a* or underlapped as illustrated in Fig. 11.15*a*. For an overlapped valve, a dead

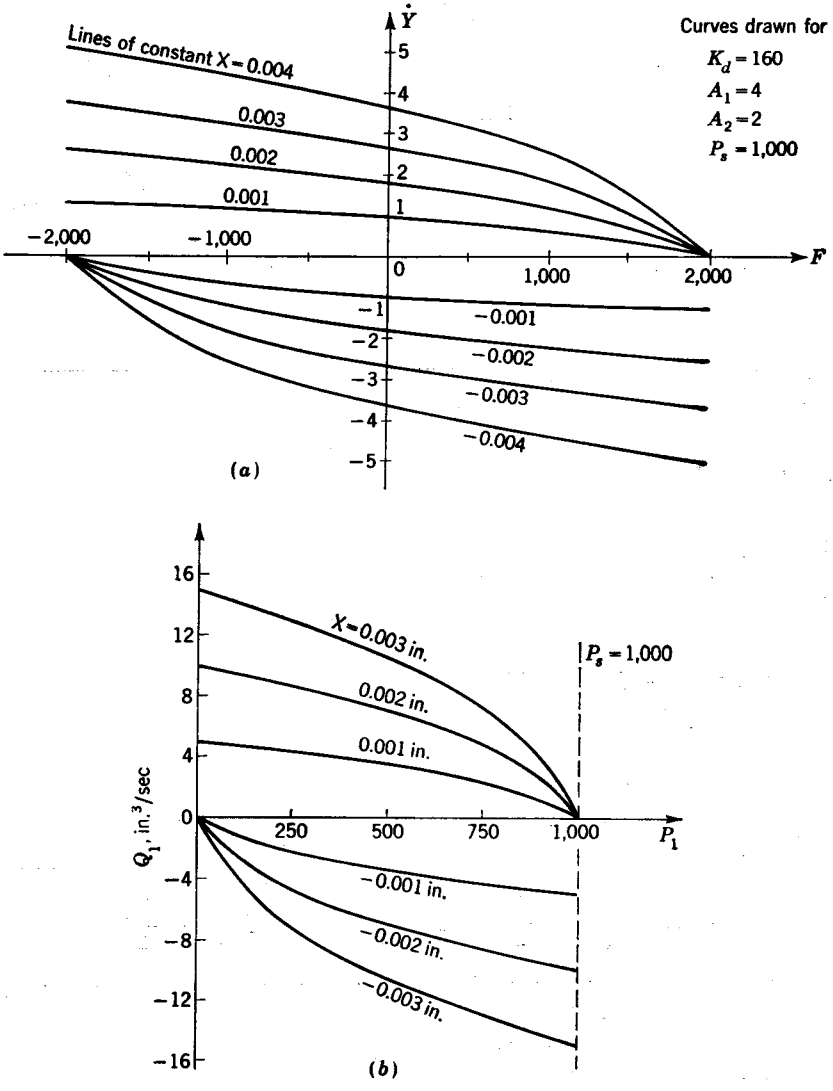


FIG. 11.13. Flow curves for three-way valve.

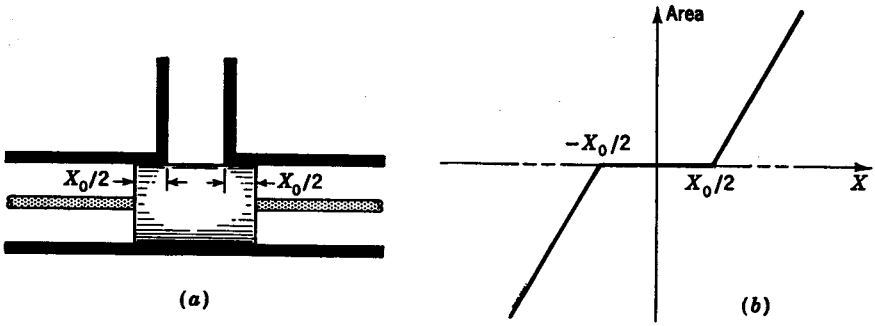


FIG. 11.14. Overlapped valve.

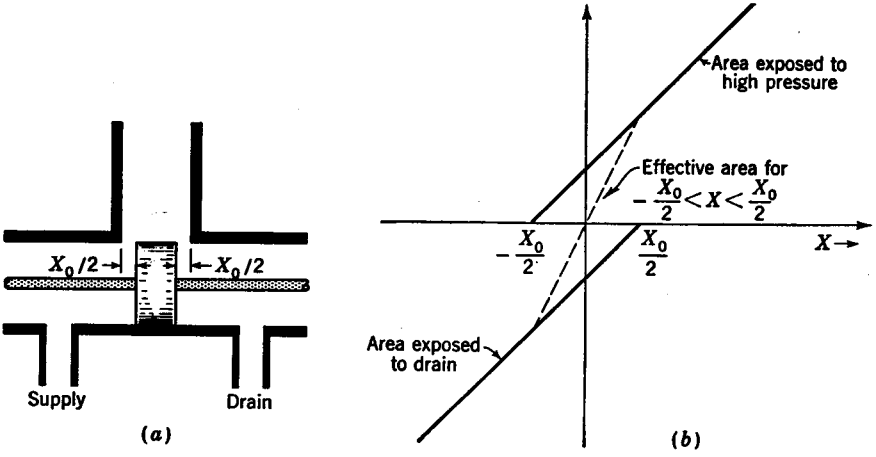


FIG. 11.15. Underlapped valve.

zone occurs for $-X_0/2 < X < X_0/2$. The plot of uncovered port area A versus X is shown in Fig. 11.14*b*. The flow equations for an overlapped valve are

$$\begin{aligned} Q &= K_d W \left(X - \frac{X_0}{2}\right) \sqrt{P_s - P_1} && X \geq \frac{X_0}{2} \\ Q &= 0 && \frac{-X_0}{2} \leq X \leq \frac{X_0}{2} \\ Q &= K_d W \left(X + \frac{X_0}{2}\right) \sqrt{P_1} && X \leq -\frac{X_0}{2} \end{aligned} \tag{11.27}$$

For the underlapped valve of Fig. 11.15*a*, the valve must be moved to the left a distance $-X_0/2$ in order to close off the high-pressure, or inlet, passageway. In Fig. 11.15*b*, the inlet port area increases linearly for $X > -X_0/2$. To close off the return flow, the valve must be moved to the right a distance $X_0/2$. As illustrated by the dashed line of Fig.

11.15*b*, the effective port area for the underlapped region is the sum of that exposed to inlet flow and that exposed to return flow so that the resulting slope doubles in this region. For $-X_0/2 \leq X \leq X_0/2$ the flow equation is

$$Q = K_d\left(X + \frac{X_0}{2}\right)\sqrt{P_s - P_1} + K_d\left(X - \frac{X_0}{2}\right)\sqrt{P_1} \quad (11.28)$$

When Q is positive, there is a net flow into the cylinder, and when Q is negative, there is a net flow from the cylinder. In Fig. 11.16 are shown the flow curves in the underlapped region for the case in which

$$\frac{X_0}{2} = 0.001 \text{ in.}$$

If an underlapped or overlapped valve is used, the partial derivatives K_1 and K_2 must be evaluated from the appropriate equations. Nothing

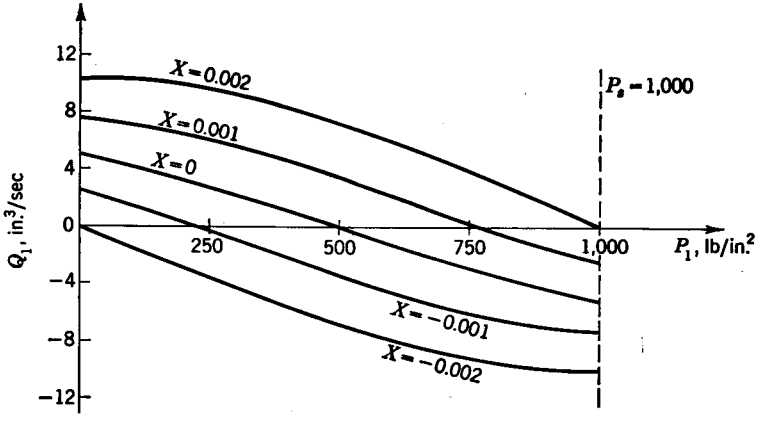


FIG. 11.16. Flow curves for underlapped valve.

else is affected in the preceding analysis of the valve-piston circuit. When experimental curves are available for the valve, K_1 and K_2 may be evaluated graphically from these curves.

For control applications in which X is proportional to the error signal, it is usually undesirable to have a dead zone. The reason is that, for small errors such that X is within the dead zone, changes have no effect. The error must become sufficiently large before it is detected. Control valves are usually dimensioned to ensure a slight underlapped condition. This increases the valve sensitivity when X is within the underlapped region.

Spring-loaded Piston. In Fig. 11.12*a*, the high-pressure line which bypasses the valve and acts on the left side of the piston serves to provide a constant force P_sA_2. This force moves the piston to the right when

the cylinder line is connected to drain. This high-pressure bypass line can be eliminated as shown in Fig. 11.17 by inserting a large spring behind the piston. The preceding analysis applies equally well to this circuit except that the spring rate K of the load is increased by that of the spring inserted behind the piston.

11.5. Four-way Pilot Valve. A four-way pilot valve is illustrated in Fig. 11.18*a*. The four ports are the high-pressure (or supply) port, a

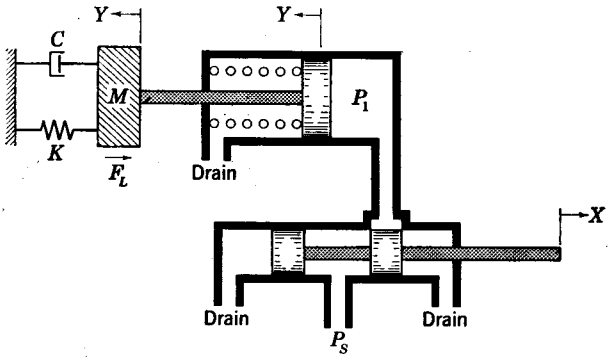


FIG. 11.17. Spring-loaded cylinder.

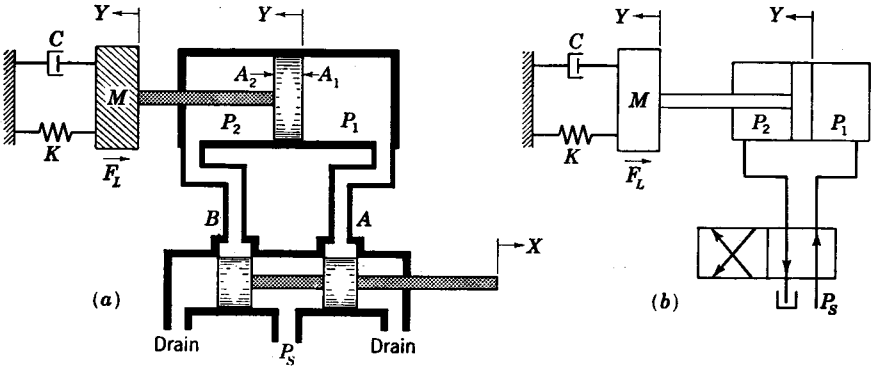


FIG. 11.18. Four-way-valve–cylinder circuit.

port each for both ends of the cylinder, and a drain port. When the valve is moved to the right, port A is connected to high pressure and port B to drain. This causes the piston to move to the left. When the valve is moved to the left, the reverse action occurs. The receiving unit shown in Fig. 11.18*a* is called a double-acting cylinder. Because of the four-way valve, it is possible to have the full supply pressure acting on either side of the piston with the other side connected to drain. The cylinders used with a three-way valve are single-acting. With a single-acting cylinder essentially a constant force is applied to one side. This force

is generally about one-half the maximum force that can be applied to the other side. Thus, with the single-acting piston which is necessitated by the use of a three-way valve, only about one-half the force can be developed as for the case of a double-acting cylinder. Consequently, four-way valves, which permit the use of double-acting cylinders, are more commonly employed than three-way valves.

To analyze the operation of this circuit, it should first be noted that the velocity $pY = \dot{Y}$ of the piston is

$$\dot{Y} = \frac{Q_1}{A_1} = \frac{K_d X \sqrt{P_s - P_1}}{A_1} \qquad X > 0 \tag{11.29}$$

$$\dot{Y} = \frac{Q_1}{A_1} = \frac{K_d X \sqrt{P_1}}{A_1} \qquad X < 0 \tag{11.30}$$

where Q_1 is the rate of flow through port A. In addition, it follows that

$$\dot{Y} = \frac{Q_2}{A_2} = \frac{K_d X \sqrt{P_2}}{A_2} \qquad X > 0 \tag{11.31}$$

$$\dot{Y} = \frac{Q_2}{A_2} = \frac{K_d X \sqrt{P_s - P_2}}{A_2} \qquad X < 0 \tag{11.32}$$

where Q_2 is the flow rate at port B.

The resultant pressure force on the piston is

$$F = P_1 A_1 - P_2 A_2 \tag{11.33}$$

For $X > 0$, the desired functional relationship between $\dot{Y}$, X, and F is obtained by eliminating P_1 and P_2 from Eqs. (11.29), (11.31), and (11.33). Thus

$$\dot{Y} = K_d X \sqrt{\frac{P_s A_1 - F}{A_1^3 + A_2^3}} \qquad X > 0 \tag{11.34}$$

Similarly, the corresponding relationship for $X < 0$ is obtained by eliminating P_1 and P_2 from Eqs. (11.30), (11.32), and (11.33). Thus

$$\dot{Y} = K_d X \sqrt{\frac{P_s A_2 + F}{A_1^3 + A_2^3}} \qquad X < 0 \tag{11.35}$$

In both Eqs. (11.34) and (11.35), $\dot{Y}$ is a function of X and F; thus linearization gives

$$y = \frac{K_1 x - K_2 f}{p} \tag{11.18}$$

The preceding result is the same as that given by Eq. (11.18) except that the constants K_1 and K_2 are as follows:

For $X > 0$

$$\left.\frac{\partial \dot{Y}}{\partial X}\right|_i = K_d \left.\sqrt{\frac{P_s A_1 - F}{A_1^3 + A_2^3}}\right|_i = K_1$$

$$\left.\frac{\partial \dot{Y}}{\partial F}\right|_i = -\frac{K_d X}{2} \left.\sqrt{\frac{A_1^3 + A_2^3}{P_s A_1 - F}}\right|_i = -K_2 \tag{11.36}$$

For $X < 0$

$$\left.\frac{\partial \dot{Y}}{\partial X}\right|_i = K_d \left.\sqrt{\frac{P_s A_2 + F}{A_1^3 + A_2^3}}\right|_i = K_1$$

$$\left.\frac{\partial \dot{Y}}{\partial F}\right|_i = \frac{K_d X}{2} \left.\sqrt{\frac{A_1^3 + A_2^3}{P_s A_2 + F}}\right|_i = -K_2 \tag{11.37}$$

From the preceding expressions, it follows that the term K_1 will be the same for positive or negative values of X and so will the term K_2, if

$$P_s A_1 - F = P_s A_2 + F$$

or

$$P_s(A_1 - A_2) = 2F \tag{11.38}$$

For the case in which the nominal load F is quite small and $A_1 = A_2$, the resulting expression for K_1 reduces to that given by Eq. (11.26). If, in addition, $K_2 = 0$, as is the case for $X_i = 0$, the equation of operation for this valve and cylinder reduces to the result that was given by Eq. (2.56).

A plot of the family of curves determined by Eqs. (11.34) and (11.35) is shown in Fig. 11.19*a*. This plot greatly facilitates the determination of $K_1 = \partial \dot{Y}/\partial X$ and $K_2 = \partial \dot{Y}/\partial F$. The curves for an underlapped valve are as shown in Fig. 11.19*b*. To determine these curves for a valve experimentally, first disconnect the cylinder, and then install a flowmeter in the line for port A and one in the line for port B. For port A, vary the pressure P_1, and obtain the flow curves of $\dot{Y} = Q_1/A_1$ versus P_1 for constant values of X. Similarly for port B obtain the flow curves of $\dot{Y} = Q_2/A_2$ versus P_2 for constant valve positions. Then by assuming a $\dot{Y}$ and X the corresponding value of P_1 is obtained from the curves for port A and the value of P_2 from the port B curves. The force can now be evaluated from Eq. (11.33). Thus, repeating this process to obtain corresponding values of $\dot{Y}$, F, and X gives the desired curves shown in Fig. 11.19*b*.

11.6. Valve Forces. When fluid is flowing through a spool valve as shown in Fig. 11.20, an axial force is exerted on the valve spool due to the change in momentum of the fluid.[1,2] As is shown in the following

[1] J. F. Blackburn, G. Reethof, and J. L. Shearer, "Fluid Power Control," Technology Press, M.I.T., Cambridge, Mass., and John Wiley & Sons, Inc., New York, 1960.

[2] J. E. Gibson and F. B. Tuteur, "Control System Components," McGraw-Hill Book Company, Inc., New York, 1958.

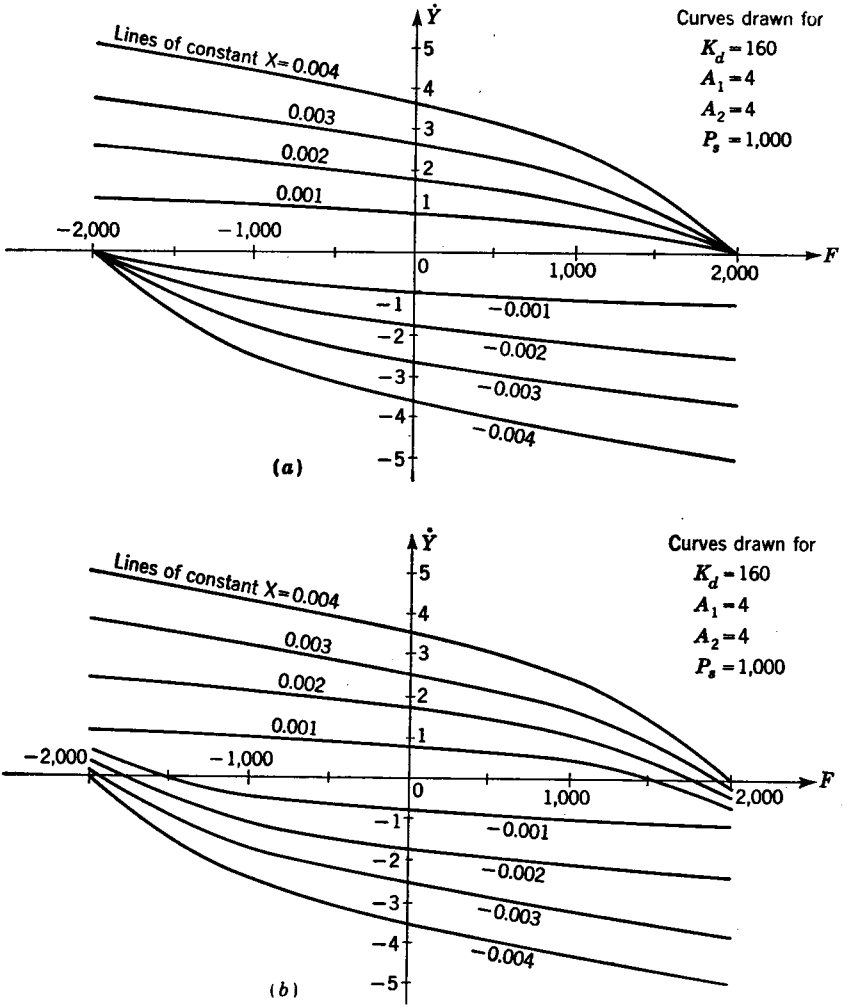


FIG. 11.19. Operating curves for a four-way valve.

analysis, this hydraulic force is always such as to tend to close the valve. This reactive force acts as a spring in that it increases linearly with the opening X of the valve. Because of the large inlet port of width de for the supply pressure, the inlet velocity is negligible, as is the change in momentum. For the right, or exit, port, the effective width ab is small so that the fluid leaves with a very high velocity. The change in momentum is

$$\frac{d}{dt}(Mv) = \rho \frac{QV}{g} \tag{11.39}$$

where Mv = momentum
ρ = density
Q = flow rate
V = exit velocity at vena contracta, ab

The axial component of force exerted on the valve spool is

$$F_1 = -\frac{Q\rho V}{g}\cos\theta \tag{11.40}$$

where θ is the exit angle of the fluid. Lee and Blackburn[1] have found that for the type valve shown in Fig. 11.20, in which the valve is square, the value of θ is 69°. In their analysis, it was assumed that the flow is

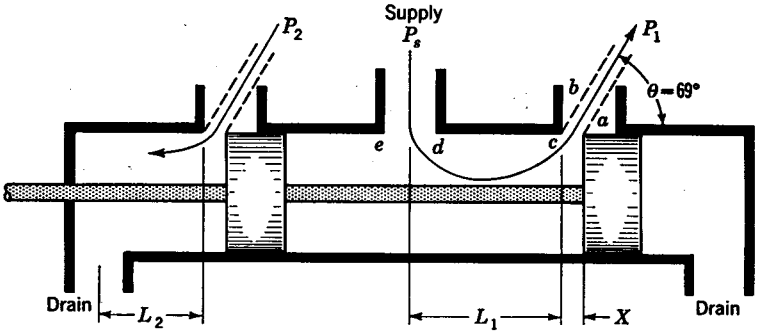


FIG. 11.20. Flow through a spool valve.

irrotational, nonviscous, and incompressible and that there is no radial clearance. The velocity V is given by Bernoulli's equation

$$V = \sqrt{\frac{2g}{\rho}(P_s - P_1)} \tag{11.41}$$

By noting that $Q = C_d A V = C_d W X V$, then

$$F_1 = -2C_d W X(P_s - P_1)\cos\theta \tag{11.42}$$

The force F_1 acts as a spring which is extended to its free length when X is zero. When the valve is to the right of center and X is positive, the force acts to the left to tend to close the valve. When the valve is to the left of center and X is negative, the force acts in the opposite direction, again to close the valve. The axial force may be written in the form

$$F_1 = -K_{h_1} X \tag{11.43}$$

where $K_{h_1} = 2C_d W(P_s - P_1)\cos 69°$ is the equivalent hydraulic spring rate. Similarly, it may be shown that the hydraulic force acting on the

[1] S. Y. Lee and J. F. Blackburn, Contributions to Hydraulic Control, I, Steady-state Axial Forces on Control Valve Pistons, *Trans. ASME*, vol. 74, p. 1005, 1952.

left land is

$$F_2 = -2C_dWXP_2 \cos\theta = -K_{h_2}X \tag{11.44}$$

where $K_{h_2} = 2C_dWP_2 \cos 69°$. Thus, the total axial force is

$$F = -(K_{h_1} + K_{h_2})X = -K_hX \tag{11.45}$$

where $K_h = 2C_dW[P_s - (P_1 - P_2)] \cos 69°$.

Lee and Blackburn[1] have also investigated the case of a valve with radial clearance and with land edge of a finite radius. These factors tend to increase the flow area and decrease the exit angle θ so that the axial force tends to be somewhat greater than the analytical predictions indicate.

In addition to the axial valve force which depends on the opening X, Lee and Blackburn[2] have shown that a force is developed due to a velocity dx/dt of the valve. This force may be visualized as follows: Consider the right valve land in Fig. 11.20, in which the supply flow is moving from left to right. With the valve at rest, this fluid has some average velocity. If the valve is in motion to the left so as to close it, the amount of flow is being decreased by the closing of the port, thus decreasing the average velocity of the fluid in the chamber. The force exerted on the valve because of this decrease of momentum tends to resist the motion of the valve. When the valve is in motion to the right, the change in momentum of the fluid also tends to resist the motion of the valve.

If the direction of flow across the valve land is in the opposite direction, as is the case for the left valve land, then the change in momentum is in the opposite direction. The resulting force acting on the valve now aids rather than resists the motion. Lee and Blackburn have shown that the reactive force due to motion of the valve is

$$F' = -\frac{\rho}{g} L \frac{dQ}{dt} \tag{11.46}$$

where L is the axial distance between the inlet and discharge ports as shown in Fig. 11.20. This distance L is positive when the flow exits from a port (for example, L_1), and the distance L is negative when flow enters the chamber through a port (for example, L_2).† This force may

[1] *Ibid.*

[2] S. Y. Lee and J. F. Blackburn, Contributions to Hydraulic Control, II, Transient Flow Forces and Valve Instability, *Trans. ASME*, vol. 74, pp. 1013–1016, 1952.

† When the force F' is written in the form $F' = (\rho/g)L(dQ/dt)$, then L must be regarded as negative when the flow exits and positive when the flow enters. The advantage of the form given by Eq. (11.46) is that positive L corresponds to positive damping and negative L results in negative damping.

be expressed as a function of the velocity of the valve as follows,

$$\frac{dQ}{dt} = \frac{d}{dt}\left(C_d W X \sqrt{\frac{2g}{\rho} P}\right) = C_d W \sqrt{\frac{2g}{\rho} P}\, px \tag{11.47}$$

where P is the pressure drop across the restriction. In this analysis, it is assumed that the pressure drop across the valve may be considered to remain constant. If P is not constant, the preceding differentiation yields an additional term proportional to the rate of change of the pressure drop. Substitution of the preceding value for dQ/dt in Eq. (11.46) gives

$$F' = -C_d W L \sqrt{\frac{2\rho}{g} P}\, px \tag{11.48}$$

Because this force is proportional to the velocity, then in effect a positive-damping term results for positive values of L and negative damping for

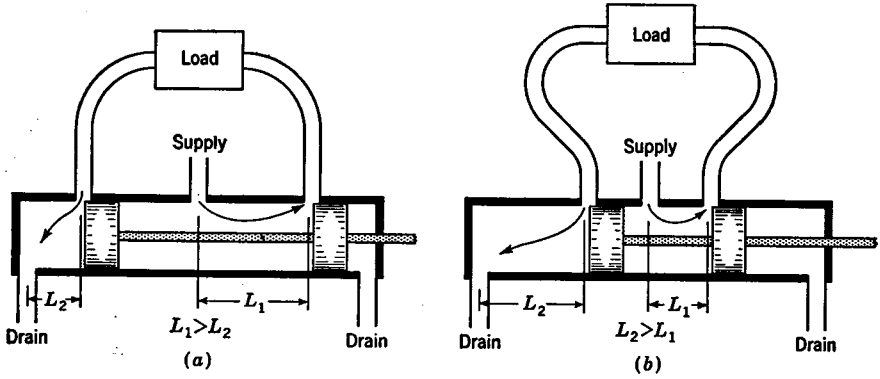


FIG. 11.21. Four-way valves.

negative L. The coefficient of viscous damping is $C_d W L \sqrt{(2\rho/g)P}$. Negative damping tends to decrease the valve stability and may cause the valve to sing (high-frequency chatter). The total hydraulic forces acting on the four-way valve of Fig. 11.20 are

$$F = -[(K_{h_1} + K_{h_2}) + C_d W \sqrt{(2\rho/g)P}\,(L_1 + L_2)p]x \tag{11.49}$$

For the valve shown in Fig. 11.21*a*, $L_1 > L_2$ so that this valve has more positive damping than the valve in Fig. 11.21*b*, in which $L_2 > L_1$. Thus, the valve in Fig. 11.21*a* tends to be more stable than that in Fig. 11.21*b*. The spring rate and damping caused by hydraulic forces should be taken into account when a change in the external forces acting on a valve causes a corresponding change in the position, as is the case in Figs. 3.12, 4.6, and 4.9. When the valve is, in effect, positively or manually positioned,

as is the case in Fig. 3.6, hydraulic reactions on the valve have a negligible effect.

11.7. Flapper Valves. Flapper valves have been developed in an effort to eliminate some of the disadvantages of spool valves. For instance, spool valves are quite expensive to manufacture because axial tolerances between valve lands must be held very closely. Also, the radial clearance must be kept as small as possible, and the corners of the valve lands should be very sharp. If a particle of dirt gets stuck in the small radial clearance, the valve becomes inoperative. The combination of friction, hydraulic forces, inertia forces, changes in fluid viscosity due

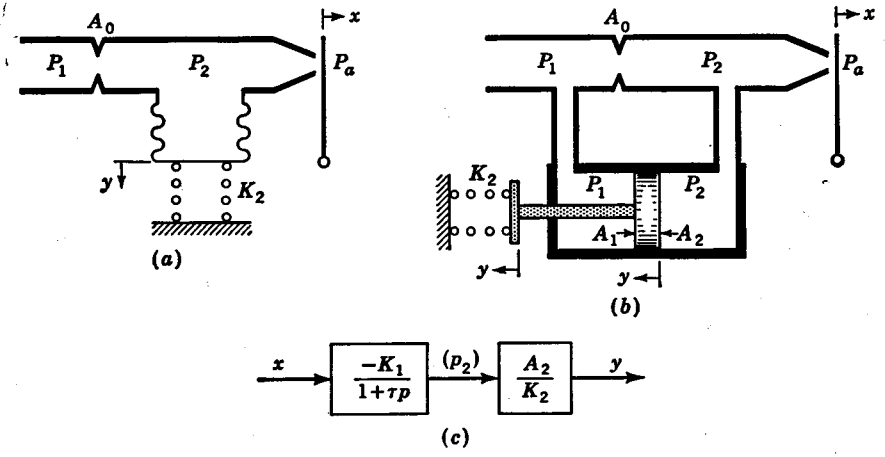


FIG. 11.22. Flapper valve.

to changes in temperature, etc., can cause a spool-type valve to chatter, or sing, under various operating conditions.

A flapper valve as shown in Fig. 11.22*a* is one in which small changes in the position X of the flapper cause large variations in the controlled pressure P_2 in the chamber. When the flapper is closed off so that there is no flow, the pressure in the chamber is equal to the supply pressure P_1. If the flapper is opened wide, the chamber pressure approaches the ambient pressure P_a. In the following analysis, it is assumed that the fluid flowing through this system is a hydraulic, or an incompressible, fluid. By attaching a spring-loaded bellows which is free to expand as the chamber pressure changes, this becomes a position control device as well as a pressure controller.

In Fig. 11.22*b* is shown a flapper valve, which serves the same purpose as the three-way valve previously discussed. The following analysis applies equally well to the systems of Fig. 11.22*a* and *b*.

Usually the area A_0 of the inlet orifice and the inlet pressure P_1 are

held constant so that the pressure P_2 is controlled by the position of the flapper only. Thus, with A_0 and P_1 constant, the volume rate of flow Q_{in} into the chamber is a function of the chamber pressure only. That is,

$$Q_{in} = F(P_2) = C_{d_0} A_0 \sqrt{\frac{2g}{\rho}(P_1 - P_2)} \tag{11.50}$$

where C_{d_0} is the discharge coefficient for orifice A_0. Linearization of Eq. (11.50) gives

$$q_{in} = \left.\frac{\partial Q_{in}}{\partial P_2}\right|_i p_2 = -C_1 p_2 \tag{11.51}$$

where

$$\left.\frac{\partial Q_{in}}{\partial P_2}\right|_i = \left.\frac{-C_{d_0} A_0 \sqrt{2g/\rho}}{2\sqrt{P_1 - P_2}}\right|_i = -C_1$$

Because the ambient or discharge pressure is usually constant, the volume rate of flow Q_o leaving the chamber is seen to be a function of the position of the flapper X and also the chamber pressure P_2.

$$Q_o = F(X, P_2) = C_{d_f} W X \sqrt{\frac{2g}{\rho} P_2} \tag{11.52}$$

where C_{d_f} is the discharge coefficient for the flapper orifice and W is the circumference of the flapper hole. Linearization gives

$$q_o = \left.\frac{\partial Q_o}{\partial X}\right|_i x + \left.\frac{\partial Q_o}{\partial P_2}\right|_i p_2 = C_2 x + C_3 p_2 \tag{11.53}$$

where

$$\left.\frac{\partial Q_o}{\partial X}\right|_i = C_{d_f} W \sqrt{\frac{2g}{\rho} P_2}\Bigg|_i = C_2$$

and

$$\left.\frac{\partial Q_o}{\partial P_2}\right|_i = \left.\frac{C_{d_f} W X \sqrt{2g/\rho}}{2\sqrt{P_2}}\right|_i = C_3$$

When equilibrium exists at the reference condition, $Q_{in}|_i = Q_o|_i$ and the variation $q_{in} - q_o$ is the net rate of flow into or out of the chamber. This net rate of flow is equal to the rate of change of volume of the chamber. For Fig. 11.22*a*, the rate of change of volume is the effective area of the bellows A_2 times the velocity py, and for Fig. 11.22*b* this is the area A_2 of the piston times the velocity py. Thus

$$q_{in} - q_o = A_2 py \tag{11.54}$$

The substitution of q_{in} from Eq. (11.51) and q_o from Eq. (11.53) into Eq. (11.54) gives

$$-C_2 x - (C_1 + C_3) p_2 = A_2 py \tag{11.55}$$

Flapper valves are frequently employed in applications where the load is primarily a spring. For Fig. 11.22*a*, it follows that

$$P_2 A_2 = K_2 Y \tag{11.56}$$

Similarly for Fig. 11.22*b*, the summation of forces acting on the piston is

$$P_2A_2 - P_1A_1 = K_2Y \tag{11.57}$$

Because P_1A_1 is a constant, linearization of Eq. (11.56) or Eq. (11.57) gives the same result, i.e.,

$$y = \frac{A_2}{K_2} p_2 \tag{11.58}$$

The substitution of p_2 from Eq. (11.58) into Eq. (11.55) yields the following over-all equation of operation,

$$y = \frac{A_2(-K_1)x}{K_2(1 + \tau p)} \tag{11.59}$$

where $K_1 = \dfrac{C_2}{C_1 + C_3}$

and $\tau = \dfrac{A_2^2}{K_2(C_1 + C_3)}$

The block-diagram representation for Eq. (11.59) is shown in Fig. 11.22*c*.

The $-K_1$ indicates that, as X increases, Y decreases. This fact is evident from Fig. 11.22*a* or *b* because as X increases, the pressure P_2 decreases and thus Y moves in the negative direction. For most flapper valves C_1 and C_3 are very large so that the time constant τ may generally be regarded as negligible.

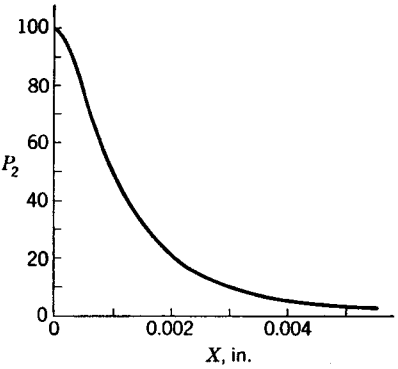


FIG. 11.23. Equilibrium curve of P_2 versus X for a flapper valve.

Equilibrium Operation. Equilibrium exists for a flapper valve when the flow in is equal to the flow out, or

$$Q_{in} - Q_o = C_{d_0}A_0\sqrt{\frac{2g}{\rho}(P_1 - P_2)} - C_{d_f}WX\sqrt{\frac{2g}{\rho}P_2} = 0 \tag{11.60}$$

Solving for P_2 gives

$$P_2 = \frac{P_1}{1 + (C_{d_f}WX/C_{d_0}A_0)^2} \tag{11.61}$$

In Fig. 11.23 is shown a typical plot of P_2 versus X for a flapper valve. For steady-state or equilibrium operation, it follows from Fig. 11.22*c* that $p_2 = -K_1x$. Thus, the constant $-K_1$ is equal to the slope of the curve of P_2 versus X, which as shown in Fig. 11.23 remains quite constant over a considerable portion of the operating range. The fact that the slope of this curve is equal to $-K_1$ may also be ascertained as follows: For the implicit function F as given by Eq. (11.60), it follows from calculus that

$$\frac{p_2}{x} \approx \frac{\partial P_2}{\partial X} = -\frac{\partial F/\partial X}{\partial F/\partial P_2} = \frac{-C_2}{C_1 + C_3} = -K_1 \tag{11.62}$$

where $\partial F/\partial X = -C_2$ and $\partial F/\partial P_2 = -(C_1 + C_3)$.

Although flapper valves find numerous applications in hydraulic systems, they are more extensively used in pneumatic controls. A major reason for this is that spool valves do not lend themselves to pneumatic applications because of the excessive leakage of air that is a result of the very low viscosity of air. Various types of pneumatic flapper valves are treated in detail in Chap. 12.

11.8. Receiving Units. The most commonly employed receiving units for hydraulic circuits are cylinders and motors. Cylinders may be single- or double-acting. Only motors are discussed in the following, because cylinders were included in preceding sections.

Motors. When a pump is supplied with a high-pressure fluid, the pressure force acting on the pistons, vanes, or gears causes the drive

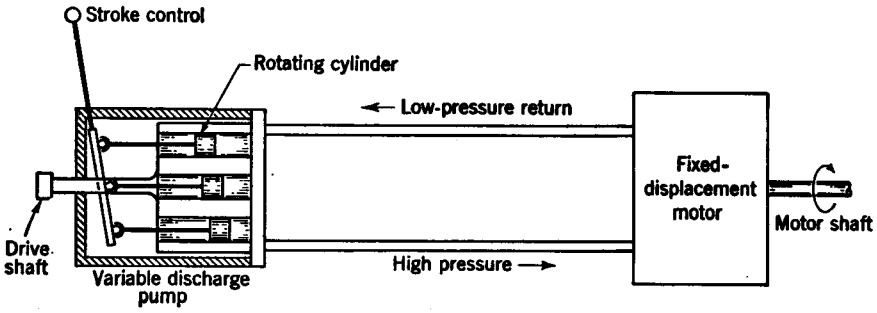


FIG. 11.24. Hydraulic transmission.

shaft to rotate. This is the reverse process of supplying power to the drive shaft of a pump to obtain a flow of high-pressure fluid. Thus, hydraulic motors are basically pumps which are supplied with a flow of high-pressure fluid which transmits power to the drive shaft. When pumps are used as motors, it is usually necessary to modify the design to reduce undesirable unbalanced forces.

In Fig. 11.24 is shown a hydraulic transmission in which the receiving unit is a fixed-displacement motor. Hydraulic transmissions are commonly used as the power element for hydraulic servomechanisms to provide a very fast and accurate control of speed.[1,2] As is shown in Fig. 11.24, a hydraulic transmission consists of a variable-displacement pump which supplies high-pressure oil to a fixed-displacement motor. The direction of rotation is reversed by moving the stroke adjustment to the other side of its neutral position.

The operational representation for this hydraulic transmission is

[1] Blackburn, Reethof, and Shearer, *op. cit.*
[2] Gibson and Tuteur, *op. cit.*

obtained as follows: The ideal volume of flow Q_i coming from the pump is

$$Q_i = na\frac{\dot{\theta}_p}{2\pi}X = K_pX \tag{11.63}$$

where Q_i = ideal pump flow, in.^{3}/sec
n = number of pistons in pump
a = area of each piston, in.^{2}
$\dot{\theta}_p/2\pi$ = pump speed, rps
X = length of stroke, in.

Because of the pressure drop across the pistons in the pump, a portion of this ideal flow leaks back past the pistons. For a given pump, the leakage flow Q_L is proportional to the pressure P developed by the pump. That is,

$$Q_L = K_LP \tag{11.64}$$

For a fixed-displacement motor, the volume rate of fluid flow Q_n delivered to the motor is

$$Q_n = D_m\dot{\theta}_m \tag{11.65}$$

where D_m is the volumetric displacement of the motor per radian and $\dot{\theta}_m$ is the motor speed (radians per second). The mechanical power developed by a motor, $T_m\dot{\theta}_m$, is equal to the hydraulic power delivered to the motor, PQ_n.

$$PQ_n = T_m\dot{\theta}_m \tag{11.66}$$

By substituting Q_n from Eq. (11.65) into Eq. (11.66), it follows that the pump discharge pressure P is proportional to the motor torque T_m, that is,

$$P = \frac{1}{D_m}T_m \tag{11.67}$$

Substitution of the preceding result into Eq. (11.64) shows that the leakage flow Q_L is also proportional to the motor torque T_m,

$$Q_L = \frac{K_L}{D_m}T_m \tag{11.68}$$

In addition to leakage through the pump, the net volume rate of flow Q_n delivered to the motor is less than the ideal volume flow Q_i displaced by the pump because of compressibility effects.

$$Q_n = Q_i - Q_L - Q_c \tag{11.69}$$

Compressibility of a hydraulic fluid may usually be considered negligible except when the fluid is at a very high pressure and when there is a large quantity of fluid in the system. The change in volume ΔV_c of a volume

of fluid V is given by the equation

$$\Delta V_c = \frac{V}{B} \Delta P \tag{11.70}$$

where B = bulk modulus of the fluid. The equivalent compressibility flow is the rate of change of volume ΔV_c with respect to time.

$$Q_c = \lim_{\Delta t \to 0} \frac{\Delta V_c}{\Delta t} = \frac{V}{B} \frac{dP}{dt} = \frac{V}{BD_m} \frac{dT_m}{dt} = \frac{K_c}{D_m} pT_m \tag{11.71}$$

where $K_c = V/B$. Substitution of Eqs. (11.63), (11.65), (11.68), and (11.71) into Eq. (11.69) gives

$$D_m \dot{\theta}_m = K_p X - \frac{K_L}{D_m} T_m - \frac{K_c}{D_m} pT_m \tag{11.72}$$

If the load on the motor shaft consists of inertia, viscous damping, and an arbitrary load torque,

$$T_m = (Jp + C_v)\dot{\theta}_m + T_L \tag{11.73}$$

Substitution of T_m from Eq. (11.73) into Eq. (11.72) and solving for $\dot{\theta}_m$ yields

$$\dot{\theta}_m = \frac{D_m K_p X - (K_L + K_c p) T_L}{K_c J p^2 + (K_c C_v + K_L J) p + (D_m^2 + K_L C_v)} \tag{11.74}$$

After making the coefficient of the p^2 term unity by factoring $K_c J$ out of the denominator, the natural frequency ω_n for the second-order characteristic equation is found to be

$$\omega_n = \sqrt{\frac{D_m^2 + K_L C_v}{K_c J}} \tag{11.75}$$

and similarly it follows that

$$2\zeta\omega_n = \left(\frac{C_v}{J} + \frac{K_L}{K_c}\right) \tag{11.76}$$

whence the damping ratio may be evaluated.

The preceding hydraulic transmission is primarily a speed control device. The position of the stroke control lever determines the flow to the fixed-displacement motor and thus controls the speed of rotation. Slight variations in speed result because of leakage and compressibility effects.

Numerous circuits are available for utilizing fluid motors. For example, if a fixed-displacement motor is supplied from a constant-pressure power supply, a constant torque is exerted on the motor. Thus a constant-torque drive results. The addition of a pressure-regulating device to vary the pressure permits torque control. In a similar manner, a constant-horsepower drive may be obtained by supplying a variable-displacement motor with a fixed-displacement pump.

CHAPTER 12

PNEUMATIC SYSTEMS

12.1. Introduction. Pneumatic systems are distinguished from hydraulic systems in that the fluid medium for pneumatic systems is a compressible fluid (usually air) and that for hydraulic systems is an incompressible liquid. An advantage of using air as the working medium is its availability. After completion of its work cycle, the air may be exhausted to the atmosphere so that there is no need for return lines as there is with hydraulic fluid. Most hydraulic fluids are flammable so that leaks in such systems present fire hazards. No such danger results when air is employed as the working medium. Because the viscosity of hydraulic fluids changes considerably with temperature, variations in temperature of the working fluid have a marked effect upon the performance of such systems. The change of viscosity of the working medium used in pneumatic systems is usually negligible.

A fundamental advantage of hydraulic systems is that the incompressibility of the fluid results in positive action, or motion, and faster response. With pneumatic systems some of the flow is used to change the density of the fluid (i.e., to compress the fluid) so that pneumatic systems are characterized by longer time delays and less positive action. These disadvantages are diminished by the use of higher operating pressures.

The analysis of pneumatic systems is similar to that for hydraulic systems. The main difference is that the mass rate of flow of the fluid must be considered in pneumatic systems, whereas the volume rate of flow suffices for hydraulic systems.[1] It is not uncommon to have the same type of component used for both high-pressure pneumatic systems and hydraulic systems. To account for the increase in volume at the low-pressure outlets of a pneumatic component, the outlet ports are larger than the inlet ports.

12.2. Pneumatic Power Supplies. Although there are many different forms and kinds of air compressors used to supply pneumatic power,

[1] J. F. Blackburn, G. Reethof, and J. L. Shearer, "Fluid Power Control," Technology Press, M.I.T., Cambridge, Mass., and John Wiley & Sons, Inc., New York, 1960.

usually such compressors may be classified as centrifugal, axial-flow, or positive-displacement compressors.

In Fig. 12.1 is shown a centrifugal-type air compressor. The air enters at the center, or eye, of the impeller. By centrifugal action the impeller throws the fluid into the volute, whence it goes to the diffuser. The fluid leaving the impeller has considerable kinetic energy, which is

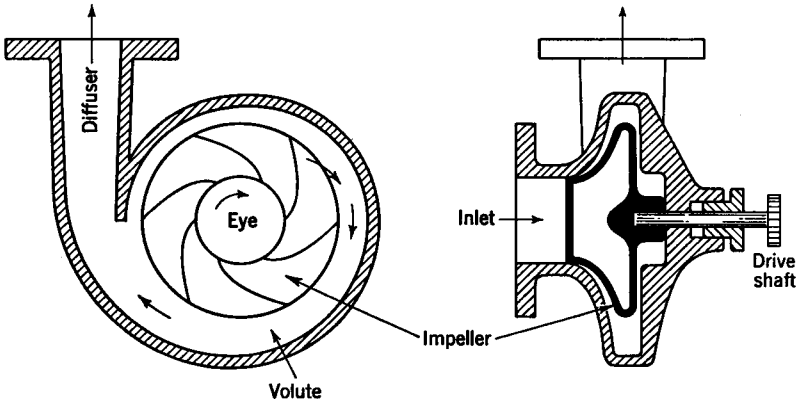


FIG. 12.1. Centrifugal compressor.

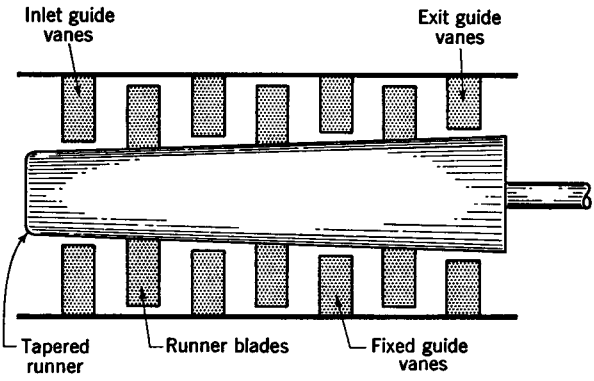


FIG. 12.2. Axial-flow compressor.

gradually changed to static pressure as the fluid travels through the volute and the diffuser. Such compressors deliver relatively small amounts of flow. The discharge pressure from a centrifugal compressor seldom exceeds 50 psia.

An axial-flow type of compressor is shown in Fig. 12.2. The annular area at the inlet to an axial-flow compressor is much larger than the relatively small area at the eye of the impeller of a centrifugal compressor. Thus, an axial compressor can deliver much more flow than a centrifugal

compressor. The blades that are attached to the rotor of an axial compressor impart kinetic energy to the fluid, and the fixed blades in the housing act as diffusers to change this kinetic energy to static pressure. Because the fluid is continually being compressed as it flows through the compressor, the blades gradually become smaller to account for the decrease in specific volume of the fluid. The discharge pressure for such compressors is usually less than 100 psia.

Positive-displacement compressors are used to supply pneumatic power for high-pressure systems. In Fig. 12.3*a* is shown the PV diagram for an ideal positive-displacement compressor which has no clearance.

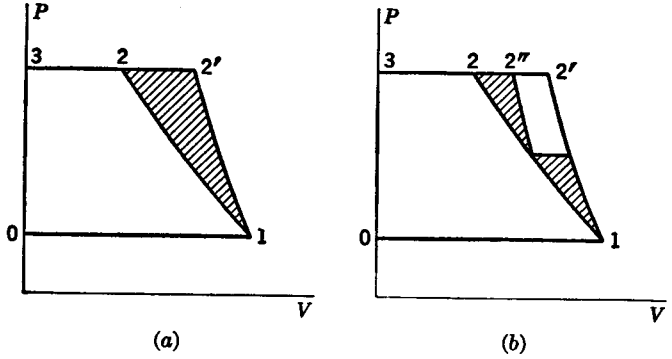


FIG. 12.3. PV diagram for ideal positive-displacement compressor.

Air is brought into the cylinder as indicated by the line 0-1. If the compression is accomplished isothermally (i.e., the temperature of the air remains constant), then the line 1-2 is the path of compression. The equation for this line is

$$PV = \text{constant} \tag{12.1}$$

To have the compression occur isothermally, the heat due to compression must be continuously extracted. Actually, the cylinder, or housing, around the piston insulates the fluid so that the actual compression process is nearly adiabatic (i.e., no heat is transferred). The equation of adiabatic compression which follows the path 1-2′ is

$$PV^{1.4} = \text{constant} \tag{12.2}$$

When the discharge pressure $P_2 = P_2'$ is reached, the discharge valve opens and air is expelled from the cylinder until the end of the stroke as indicated by point 3. Most compressor discharge valves are pressure-operated, so that when the air is compressed to the desired discharge pressure, the exhaust valve is forced open.

The total work of compression is equal to the area enclosed by the closed path of the compression cycle. The shaded area 1-2′-2-1 in Fig. 12.3*a* is the additional work needed for adiabatic compression as

compared with isothermal compression. By using interstage cooling as represented in Fig. 12.3*b*, the ideal isothermal compression is more closely approximated. This requires the use of a multistage compressor in which the air is first compressed to some intermediate pressure in the low pressure stage. Then the air is directed through an intercooler before it enters the high-pressure stage to complete the compression. For pressures of 500 psia or higher, three or more stages of compression may be used. To obtain clean, dry air, an intake filter and dehydrator are used. The filter removes foreign particles, dust, and dirt, while the dehydrator removes excess moisture.

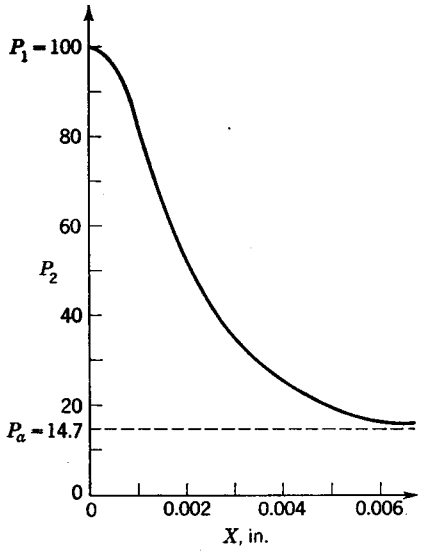


FIG. 12.4. Graph of P_2 versus X for pneumatic flapper valve.

12.3. Flapper Amplifier. When the flapper valves of Fig. 11.22*a* and *b* are supplied with high-pressure air, pneumatic flapper amplifiers result. The basic operation of a pneumatic amplifier is the same as that for the corresponding hydraulic amplifier. That is, the pressure P_2 in the chamber is controlled by the position X of the flapper. When X is zero so that the flapper is closed, the pressure P_2 is equal to the supply pressure P_1. When the flapper is opened wide so that X is large, P_2 approaches the ambient pressure P_a. A typical graph of P_2 versus X is shown in Fig. 12.4. Small changes in the input motion X cause large changes in the controlled pressure P_2. A position controller is obtained by providing a spring-loaded bellows or piston.

The procedure used to obtain the operational form of the differential equation for this pneumatic amplifier is similar to that used for the hydraulic amplifier, with the exception that, because of compressibility effects, the mass rate of flow must be considered rather than the volume rate. With a constant supply pressure P_1 and fixed area of inlet orifice, A_0, the mass rate of flow into the chamber, M_{in}, is a function of the chamber pressure P_2 only. Thus,

$$M_{in} = F(P_2)$$

and

$$m_{in} = \left.\frac{\partial M_{in}}{\partial P_2}\right|_i p_2 = -C_1 p_2 \tag{12.3}$$

In Sec. 12.7, equations for the flow of a compressible fluid through an orifice are developed, whence the corresponding partial derivatives may be evaluated. It is also shown how to determine the equilibrium or reference operating conditions such as are illustrated by Fig. 12.4.

The mass rate of flow out from the chamber, M_o, is a function of X and P_2. Thus

$$M_o = F(X,P_2)$$

and

$$m_o = \frac{\partial M_o}{\partial X}\bigg|_i x + \frac{\partial M_o}{\partial P_2}\bigg|_i p_2 = C_2x + C_3p_2 \tag{12.4}$$

The change in mass w of air in the chamber is the integral of $m_{in} - m_o$, that is,

$$w = \frac{-C_1p_2 - C_2x - C_3p_2}{p} \tag{12.5}$$

From the equation of state, the pressure P_2 in the chamber is

$$P_2 = \frac{WRT_2}{V_2} \tag{12.6}$$

where V_2 is the volume of the chamber and T_2 is the stagnation temperature of the air in the chamber. For the usual case of adiabatic flow, the stagnation temperature T_2 is equal to the stagnation temperature T_1 of the supply, which is constant. Linearization of Eq. (12.6) yields the following expression for the variation p_2 of pressure in the chamber:

$$p_2 = \frac{\partial P_2}{\partial W}\bigg|_i w + \frac{\partial P_2}{\partial V_2}\bigg|_i v_2 = C_4w - C_5v_2 \tag{12.7}$$

The change of pressure in the chamber depends upon the change of mass w and the change of volume v_2. By use of Eq. (12.6) to evaluate the partial derivatives, it follows that

$$\frac{\partial P_2}{\partial W}\bigg|_i = \frac{RT_2}{V_2}\bigg|_i = C_4 \tag{12.8}$$

$$\frac{\partial P}{\partial V}\bigg|_i = \frac{-WRT_2}{V_2^2}\bigg|_i = \frac{-P_2}{V_2}\bigg|_i = -C_5 \tag{12.9}$$

The change in volume v_2 of the chamber is equal to the area A_2 times the change in length y. That is,

$$v_2 = A_2y \tag{12.10}$$

The change in the pressure force p_2A_2 is equal to the change in spring force K_2y. Thus

$$p_2A_2 = K_2y \tag{12.11}$$

The over-all relationship between the input x and output y is obtained by substituting w from Eq. (12.5), v_2 from Eq. (12.10), and p_2 from

Eq. (12.11) into Eq. (12.7). Thus

$$y = \frac{A_2}{K_2} \frac{-K_1}{1 + \tau p} x \tag{12.12}$$

where $K_1 = C_2/(C_1 + C_3)$ and $\tau = (1 + C_5A_2^2/K_2)/[(C_1 + C_3)C_4]$. Equation (12.12) has the same general form as Eq. (11.59), and thus the block-diagram representation for this pneumatic amplifier is the same as that for the corresponding hydraulic amplifier drawn in Fig. 11.22*c*. The constant $-K_1$ is the slope of Fig. 12.4 at the reference operating condition. Although τ is slightly larger for a pneumatic amplifier than for a corresponding hydraulic amplifier, it is still extremely small and is generally regarded as negligible.

12.4. Two-stage Amplifier. For the control of large industrial processes where it is necessary to have large quantities of a controlled pneumatic pressure, it is customary to use a two-stage amplifier as shown in Fig. 12.5*a*. The first stage of amplification consists of a flapper-type amplifier in which the pressure P_2 is controlled by the flapper position X. The controlled pressure P_2 determines the position Y of the metering valve for the second amplifier. The second stage of amplification is capable of handling large quantities of flow. This second unit is called an air relay.

The actuating signal e coming from the comparator is the input signal for the two-stage controller shown in Fig. 12.5*a*. An increase in the actuating signal moves the top of the error link to the right, and thus the flapper also moves to the right. This in turn decreases the pressure P_2, which causes the metering valve to move up to admit more flow to the main part of the system. If the actuating signal decreases, then the metering valve moves down to block off the supply flow and open wider the bleed passageway. Bleeding off more flow to the atmosphere in turn causes P_o to decrease. By proper contouring of the metering valve, a linear relationship may be obtained between the actuating signal e and the output pressure P_o.

The first stage (i.e., flapper amplifier) for a two-stage amplifier is usually such that the time constant τ is negligibly small, in which case Eq. (12.12) becomes

$$y = \frac{A_2}{K_2}(-K_1)x \tag{12.13}$$

Because of the compressibility of air, a certain time lag is associated with the change of pressure in any portion of the system. However, pneumatic systems are generally proportioned so that the time constants associated with certain portions are negligibly small, and thus the larger time constants have the predominant effect on the system behavior.

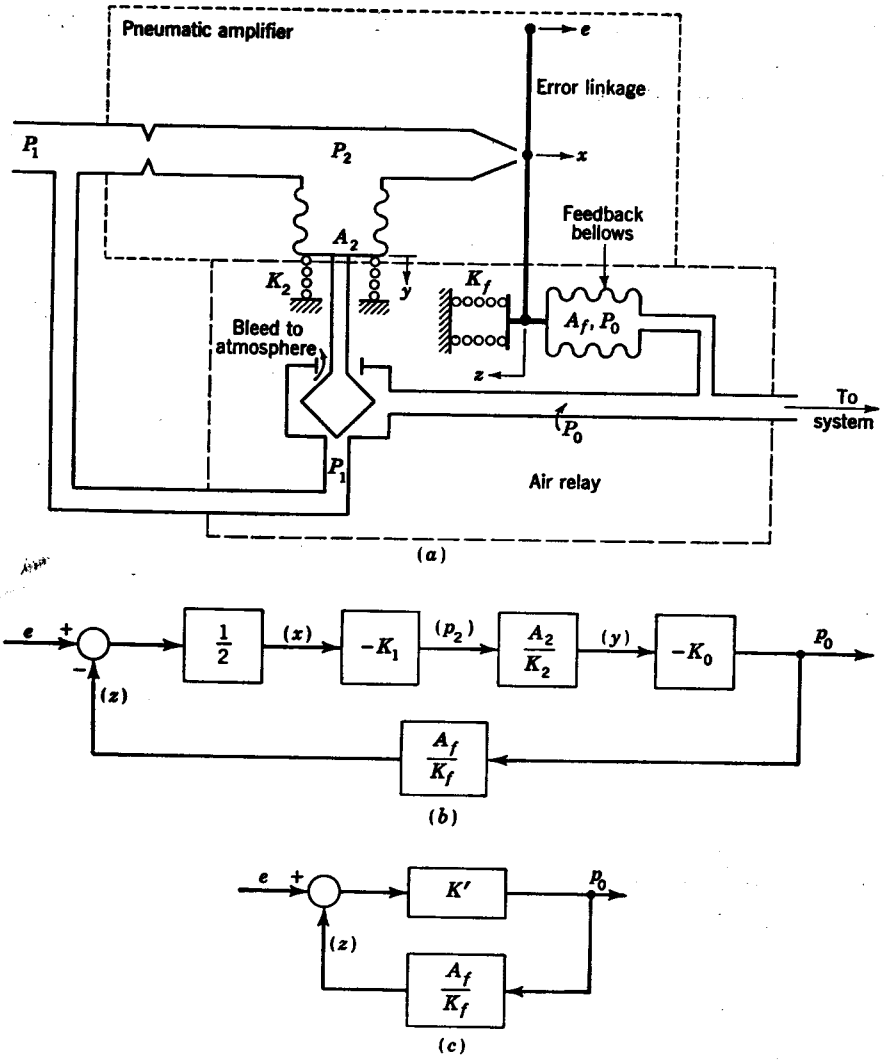


FIG. 12.5. (*a*) Two-stage amplifier; (*b*) block diagram for two-stage amplifier; (*c*) block diagram for two-stage amplifier in which $K' = K_0K_1A_2/2K_2$.

The position Y of the air relay determines the pressure P_o, or $P_o = F(Y)$. Because the air relay handles a large quantity of flow, the pressure P_o changes almost instantaneously for changes in the position Y of the metering valve, and hence there is a negligible time delay associated with this device. Thus, the equation of operation is

$$p_o = -K_o y \tag{12.14}$$

where $-K_o = \partial P_o/\partial Y\Big|_i$ is the slope of the curve of P_o versus y for the air relay.

For the error linkage

$$x = \frac{e - z}{2} \tag{12.15}$$

The summation of forces acting on the feedback bellows gives

$$K_f Z = A_f P_o$$

or

$$z = \frac{A_f}{K_f} p_o \tag{12.16}$$

From Eqs. (12.13) through (12.16), the block-diagram representation for this controller is drawn as shown in Fig. 12.5*b*. By letting $K' = K_0 K_1 A_2/2K_2$ the block diagram of Fig. 12.5*c* results. The relationship between the output p_o and the input e for this controller is

$$p_o = \frac{K'}{1 + K'(A_f/K_f)} e \tag{12.17}$$

This is recognized as a proportional controller because

$$G_1(p) = \frac{K'}{[1 + K'(A_f/K_f)]}$$

is a constant. For the usual case in which $K'A_f/K_f \gg 1$, the preceding expression reduces to

$$p_o \approx \frac{K'}{K'(A_f/K_f)} e = \frac{K_f}{A_f} e \tag{12.18}$$

The relationship given by Eq. (12.18) can be deduced directly from a closer examination of Fig. 12.5*a*. The position x changes only a very small amount to produce a large change in P_2, which in turn controls P_o. Because of the very small motion at x, this point may be considered as a fixed pivot point, or

$$z \approx e \tag{12.19}$$

The substitution of $z = (A_f/K_f)p_o$ into Eq. (12.19) gives the result of Eq. (12.18).

Proportional plus Derivative Action. The proportional controller of Fig. 12.5*a* is changed to a proportional plus derivative controller by inserting a restriction between the outlet line P_o and the feedback bellows, as is shown in Fig. 12.6*a*. The area of this restriction is small so that the time constant associated with changing the feedback pressure P_f in the bellows is appreciable. The equation relating the pressure P_f in this bellows to the pressure P_o on the other side of the restriction is obtained

as follows: For a given sized restriction at the input to this bellows, the mass rate of flow, M, going to the bellows is a function of the pressure P_o and the pressure P_f in the bellows.

$$M = F(P_o,P_f)$$

Linearization gives

$$m = \frac{\partial M}{\partial P_o}\bigg|_i p_o + \frac{\partial M}{\partial P_f}\bigg|_i p_f = C_1 p_o - C_2 p_f \tag{12.20}$$

where $\partial M/\partial P_o\big|_i = C_1$ and $\partial M/\partial P_f\big|_i = -C_2$. The fact that C_1 is equal to C_2 is ascertained as follows: Suppose that p_o and p_f are changed by the same amount so that there is no resulting pressure drop and no flow

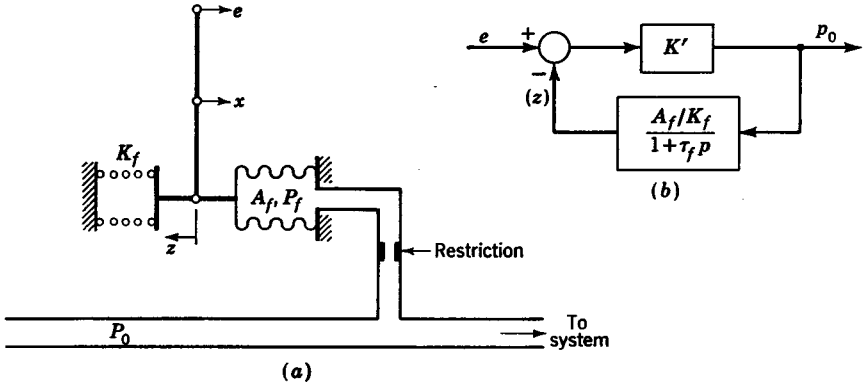


FIG. 12.6. (*a*) Two-stage amplifier with proportional plus derivative action; (*b*) block diagram for two-stage amplifier with proportional plus derivative action.

($m = 0$) across the restriction. From Eq. (12.20), it follows that C_1 must equal C_2 to substantiate the known result that m is zero. The pressure P_f is obtained from the equation of state, i.e.,

$$P_f = \frac{WRT_f}{V_f} \tag{12.21}$$

where V_f is the volume of the bellows and T_f is the corresponding temperature. The change of pressure p_f is

$$p_f = \frac{\partial P_f}{\partial W}\bigg|_i w + \frac{\partial P_f}{\partial V_f}\bigg|_i v_f = C_3 w - C_4 v_f \tag{12.22}$$

The change in volume v_f is equal to the effective area of the bellows times the displacement z.

$$v_f = A_f z \tag{12.23}$$

A force balance on the feedback bellows gives $A_f P_f = K_f Z$ or for small

variations

$$A_f p_f = K_f z \tag{12.24}$$

The over-all relationship between z and p_o is obtained as follows: Substitute w, which is obtained by integration of Eq. (12.20), substitute v_f from Eq. (12.23) into Eq. (12.22), and then eliminate p_f by use of Eq. (12.24). Thus

$$z = \frac{(A_f/K_f)p_o}{1 + \tau_f p} \tag{12.25}$$

where $\tau_f = (1 + C_4 A_f^2/K)/C_2 C_3$

This relationship between z and p_o is the only difference between the proportional controller and the proportional plus derivative controller.

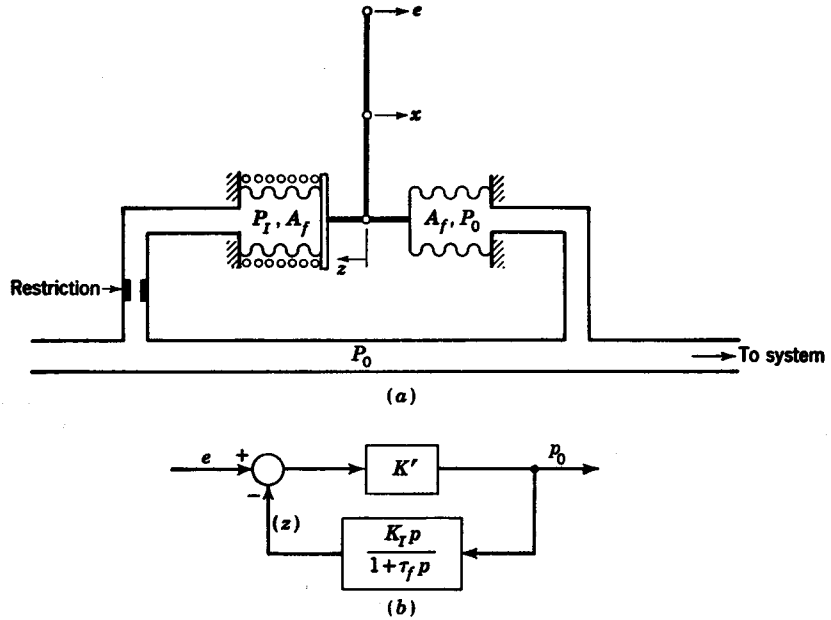


FIG. 12.7. (*a*) Two-stage amplifier with proportional plus integral action; (*b*) block diagram for two-stage amplifier with proportional plus integral action.

Thus, the insertion of the block diagram for Eq. (12.25) into the feedback path between p_o and z of Fig. 12.5*c* gives the over-all block diagram for the proportional plus derivative controller shown in Fig. 12.6*b*. The over-all equation of operation for this controller is

$$p_o = \frac{K'e}{1 + (K'A_f/K_f)/(1 + \tau_f p)} = \frac{K'(1 + \tau_f p)e}{(1 + K'A_f/K_f) + \tau_f p} \tag{12.26}$$

Because K' is very large, $1 + K'A_f/K_f \approx K'A_f/K_f \gg 1$, in which case

Eq. (12.26) becomes

$$p_o = \frac{(K_f/A_f)(1 + \tau_f p)e}{1 + [\tau_f/(K'A_f/K_f)]p} \tag{12.27}$$

Because $K'A_f/K_f \gg 1$, the time constant in the denominator may be regarded as negligible. Thus

$$p_o = \frac{K_f}{A_f} e + \frac{K_f \tau_f}{A_f} pe \tag{12.28}$$

The preceding expression shows the proportional plus derivative action of this controller.

Proportional plus Integral Action. The proportional controller of Fig. 12.5*a* is changed to a proportional plus integral controller by the addition of a second bellows as shown in Fig. 12.7*a*. The effective area A_f of this bellows is the same as that of the feedback bellows. Because the bellows on the left provides the integrating action, it is called the integrating bellows. The pressure inside of this bellows is P_I. The feedback bellows on the right has no restriction, so that its pressure is always P_o. The equation of operation for this proportional plus integral controller is obtained by applying the method of analysis just described for the proportional plus derivative controller. Thus

$$m = \frac{\partial M}{\partial P_o}\bigg|_i p_o + \frac{\partial M}{\partial P_I}\bigg| p_I = C_1 p_o - C_2 p_I$$

where $C_1 = C_2$

$$p_I = \frac{\partial P_I}{\partial W}\bigg| w + \frac{\partial P_I}{\partial V_I}\bigg| v_I = C_3 w - C_4 v_I \tag{12.29}$$

where $P_I = \dfrac{WRT_I}{V_I}$

$$v_I = -A_f z$$

and $(p_o - p_I)A_f = K_f z$

From the preceding expressions, the over-all relationship between z and p_o is found to be

$$z = \frac{K_I p}{1 + \tau_f p} p_o \tag{12.30}$$

where $K_I = \dfrac{A_f}{C_2 C_3 K_f}$ and $\tau_f = \dfrac{1 + C_4 A_f^2/K_f}{C_2 C_3}$

Figure 12.7*b* shows the over-all block diagram for this controller, in which Eq. (12.30) is seen to describe the operation of the internal feedback elements. The over-all equation of operation is

$$p_o = \frac{K'e}{1 + K'K_I p/(1 + \tau_f p)} = \frac{K'(1 + \tau_f p)e}{1 + (\tau_f + K'K_I)p} \tag{12.31}$$

Because the time constant for the denominator, $\tau_f + K'K_I \approx K'K_I \gg 1$, then $1 + (\tau_f + K'K_I)p$ is closely approximated by $K'K_I p$. Thus Eq.

(12.31) becomes

$$p_o = \frac{(1 + \tau_f p)e}{K_I p} = \frac{1}{K_I p} e + \frac{\tau_f p}{K_I p} e \tag{12.32}$$

The first term on the right-hand side of Eq. (12.32) provides the integrating effect, and the second term contributes the proportional action.

By adding derivative action to this proportional plus integral controller, a proportional plus integral plus derivative type of controller is obtained, as is shown in Fig. 12.8.

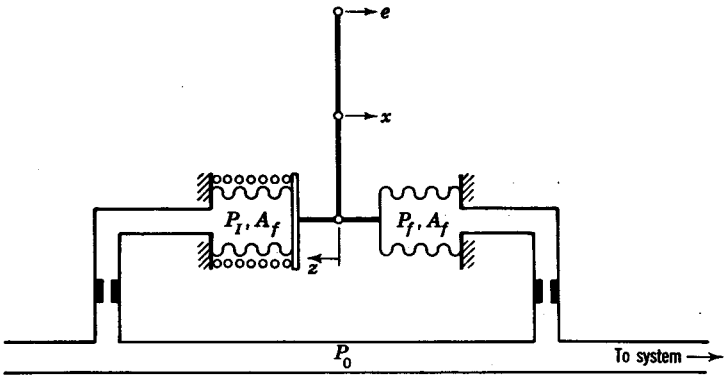


FIG. 12.8. Proportional plus integral plus derivative controller.

12.5. Pneumatic Controllers (Force Type). A force-type pneumatic controller operates only on pressure signals, and so it is necessary to convert the reference input and controlled variable to corresponding pressures. Simple adjustments make it easy to modify the operating characteristics of a force-type controller. Such industrial controllers are sometimes called "stack controllers."[1,2]

In Fig. 12.9*a* is shown a force-type controller which has proportional action. The subscripts refer to the signal to which the pressure corresponds, i.e.,

P_r = pressure corresponding to reference input
P_c = pressure corresponding to controlled variable
P_o = pressure corresponding to output of controller
P_f = pressure corresponding to an internal-feedback signal

The operation of the controller shown in Fig. 12.9*a* may be summarized as follows: An increase in the pressure P_r which is proportional to the

[1] D. P. Eckman, "Automatic Process Control," John Wiley & Sons, Inc., New York, 1958.

[2] J. E. Gibson and F. B. Tuteur, "Control System Components," McGraw-Hill Book Company, Inc., New York, 1958.

reference input causes the valve stem to move down to close off the bleed restriction for P_o. This in turn causes the output pressure P_o to increase.

To obtain the equation of operation for this controller, first sum up the forces acting on the pilot stem.

$$P_f A_2 + P_r(A_1 - A_2) - P_c(A_1 - A_2) - P_o A_2 = 0 \qquad (12.33)$$

or

$$(P_r - P_c)(A_1 - A_2) + P_f A_2 = P_o A_2 \qquad (12.34)$$

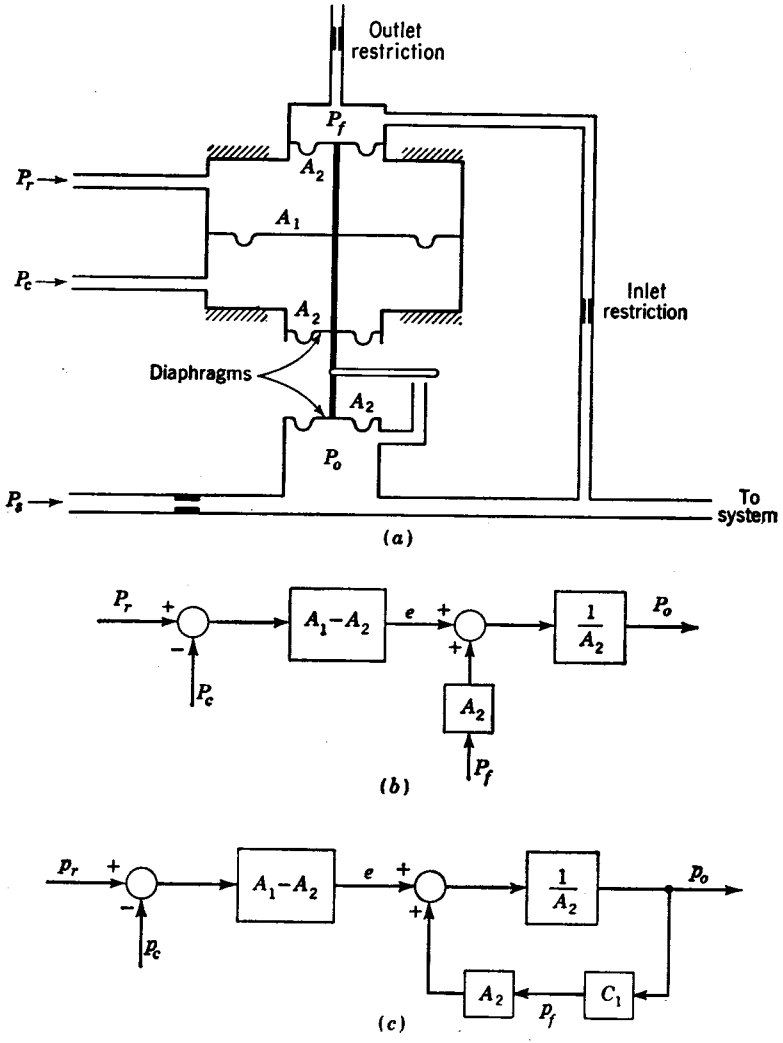


FIG. 12.9. (*a*) Force-type controller with proportional action; (*b*) block diagram of force balance for force-type controller; (*c*) over-all block diagram for force-type controller.

The block-diagram representation for Eq. (12.34) is given in Fig. 12.9*b*. This shows that the comparator which produces the actuating signal e is automatically incorporated into this controller. The two-stage amplifier was separately excited by the actuating signal e. To complete the feedback between the output signal P_o and the feedback signal P_f, it is necessary to determine the relationship between p_o and p_f. On the

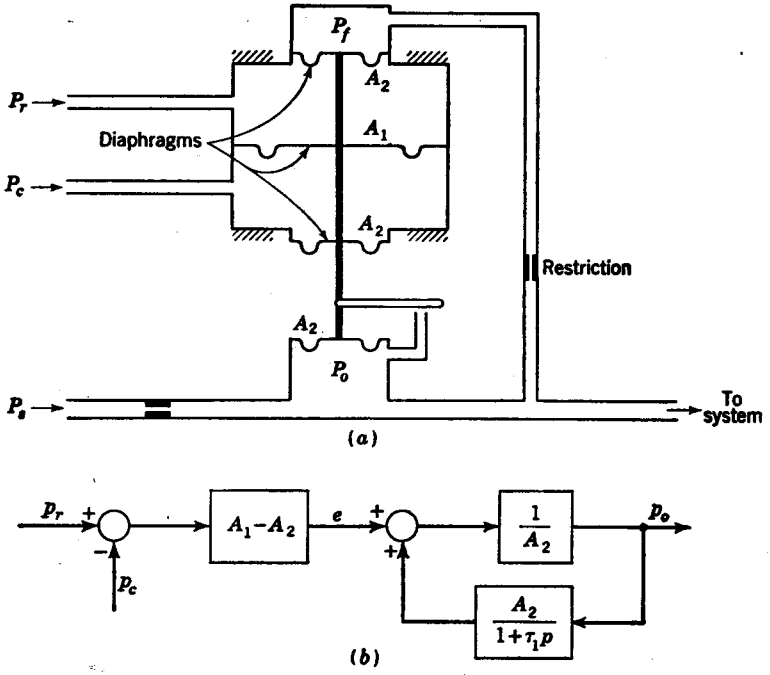


FIG. 12.10. (*a*) Force-type controller with proportional plus integral action; (*b*) over-all block diagram for force-type controller with proportional plus integral action.

assumption that there is a negligible time lag between p_f and p_o, it follows that

$$p_f = C_1 p_o \tag{12.35}$$

where

$$C_1 = \left.\frac{\partial P_f}{\partial P_o}\right|_i$$

Because p_f is less than p_o, the value of C_1 is less than 1. The completed block diagram is shown in Fig. 12.9*c*. Because of the one linearized relationship, Eq. (12.35), the entire block diagram must be linearized. The over-all relationship between p_o and e is

$$p_o = \frac{(1/A_2)e}{1 - C_1} = \frac{1}{A_2(1 - C_1)}\, e = K'e \tag{12.36}$$

This is recognized as the operating equation for a proportional con-

troller. The value of K' increases as C_1 approaches unity. This is accomplished by making the inlet restriction to the P_f chamber large compared with the outlet restriction.

This controller is converted to a proportional plus integral controller by closing off the outlet restriction as shown in Fig. 12.10*a*. For this case, it may be demonstrated that

$$p_f = \frac{p_o}{1 + \tau_1 p} \tag{12.37}$$

The block-diagram representation for this controller is given in Fig. 12.10*b*, from which it follows that

$$p_o = \frac{(1/A_2)e}{1 - 1/(1 + \tau_1 p)} = \frac{1}{A_2}\frac{1 + \tau_1 p}{\tau_1 p}e$$
$$= \left(\frac{1}{A_2 \tau_1 p} + \frac{p}{A_2 p}\right)e \tag{12.38}$$

The preceding expression describes the operation of a proportional plus integral controller.

Caldwell[1] has compiled a table of numerous force-type controllers which may be used to obtain proportional plus integral action, proportional plus derivative action, or proportional plus integral plus derivative action. Williamson[2] has designed a compact "plug-in" type of unit for obtaining compound action.

12.6. Pneumatic Receiving Units. A commonly employed receiving unit for position control is the pneumatic actuator shown in Fig. 12.11*a*. The pressure force acting on the diaphragm is $A_d P_o$, where A_d is the area of the diaphragm. For the general case in which this actuator is used to position a load of mass M, coefficient of viscous damping C, and spring constant K and upon which an external load F_L acts it follows that

$$A_d P_o = (Mp^2 + Cp + K)Y + F_L$$

Because the preceding expression is linear, for small variations

$$A_d p_o = (Mp^2 + Cp + K)y + f_L$$

or

$$y = \frac{A_d p_o - f_L}{Mp^2 + Cp + K} \tag{12.39}$$

A typical control system utilizing an actuator is shown in Fig. 12.11*b*. The reference input signal is x, and the controlled variable is y. The error signal e could be the input signal to a two-stage pneumatic amplifier. The over-all block diagram for such a controller is given in Fig. 12.11*c*.

[1] W. I. Caldwell, Generating Control Functions Pneumatically, *Control Eng.*, vol. 1, no. 1, p. 58, 1954.

[2] H. Williamson, Theory and Design of Compound Action Pneumatic Controllers, *Trans. Soc. Instrument Technol.*, vol. 6, no. 4, p. 153, 1954.

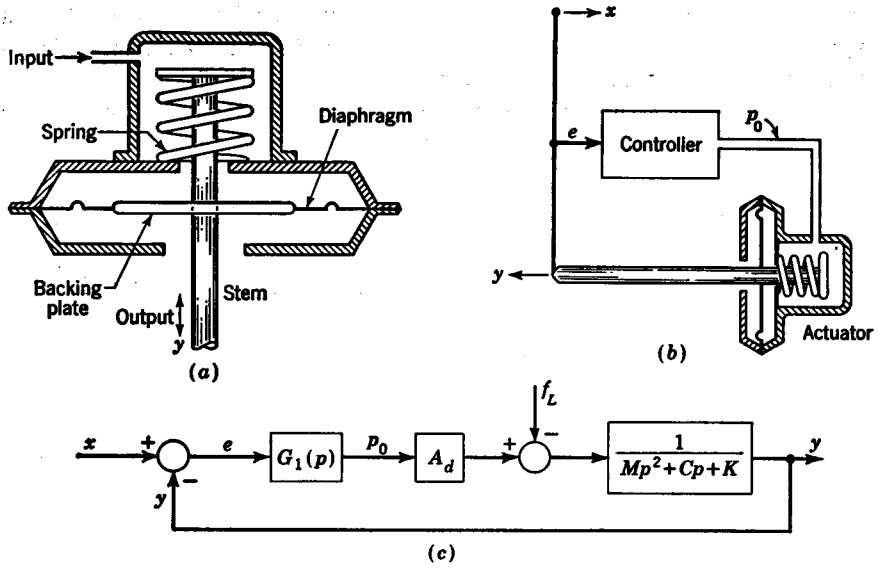


FIG. 12.11. (*a*) Pneumatic actuator; (*b*) pneumatic position control system; (*c*) block diagram for pneumatic position control system.

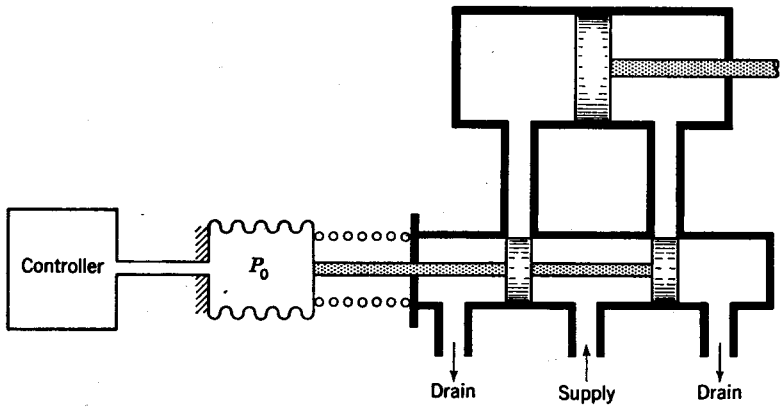


FIG. 12.12. Pneumatic four-way valve positioned by a pneumatic controller.

Single- or double-acting cylinders similar to the hydraulic cylinders discussed in Chap. 11 are also frequently employed as pneumatic receiving units. Actually, a pneumatic actuator is basically a single-acting cylinder in which the piston has been replaced by a diaphragm. In pneumatic systems, as was the case with hydraulic systems, double-acting cylinders are usually directed by a four-way valve. However, the four-way valve may in turn be positioned by the output pressure P_o

of a pneumatic controller, as is illustrated in Fig. 12.12. The analysis of a pneumatic four-way valve and cylinder combination is similar to that for the corresponding hydraulic unit with the exception that mass rate of flow must be considered rather than volume rate of flow. When the variation of pressure is small, compressibility effects become negligible. In this case, the analysis of the pneumatic system is the same as that for the corresponding hydraulic system.

Another type of receiving unit is the pneumatic motor. Although theoretically it should be possible to utilize a gear-type or piston-type motor, only the vane-type motor is now commercially available. A characteristic feature of a vane-type motor is that the torque developed

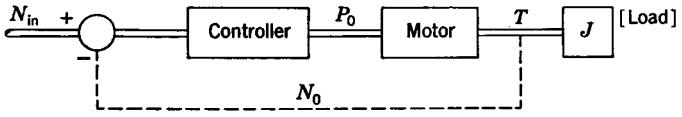


FIG. 12.13. Pneumatic speed control system.

by the motor is proportional to the supply pressure P_o and is independent of the speed of rotation, i.e.,

$$T = KP_o \tag{12.40}$$

This result is not astonishing, because a force balance on the blades shows clearly that the torque depends only on the pressure that is acting against them. The use of such a motor is illustrated in the speed control system shown in Fig. 12.13.

12.7. Equilibrium Flow through Series Restrictions. Many pneumatic components use two orifices in series to obtain a controlled pressure in the chamber between the orifices. If the component is available, the chamber pressure may be experimentally determined for various operating conditions. However, in the initial design stages, before any parts have been manufactured, it is desirable to be able to predict the value of the chamber pressure. This may be accomplished by use of the nondimensional family of curves shown in Fig. 12.14*a*. As is illustrated by the insert above Fig. 12.14*a*, the pressures P_1, P_{1_t}, P_2, P_{2_t}, and P_3 represent the inlet pressure, throat pressure at first orifice, chamber pressure, throat pressure at second orifice, and discharge pressure, respectively. The symbol A_{1_t} is the area of the first orifice times the coefficient of discharge, and A_{2_t} is the area of the second orifice times the coefficient of discharge.

Usually, the over-all pressure ratio P_1/P_3 is known, and also the ratio A_{2_t}/A_{1_t} is known, so that the ratio P_2/P_1 can be found from Fig. 12.14*a*. The value of the chamber pressure P_2 is then computed as the product of the ratio P_2/P_1 and the value of the inlet pressure P_1. In using

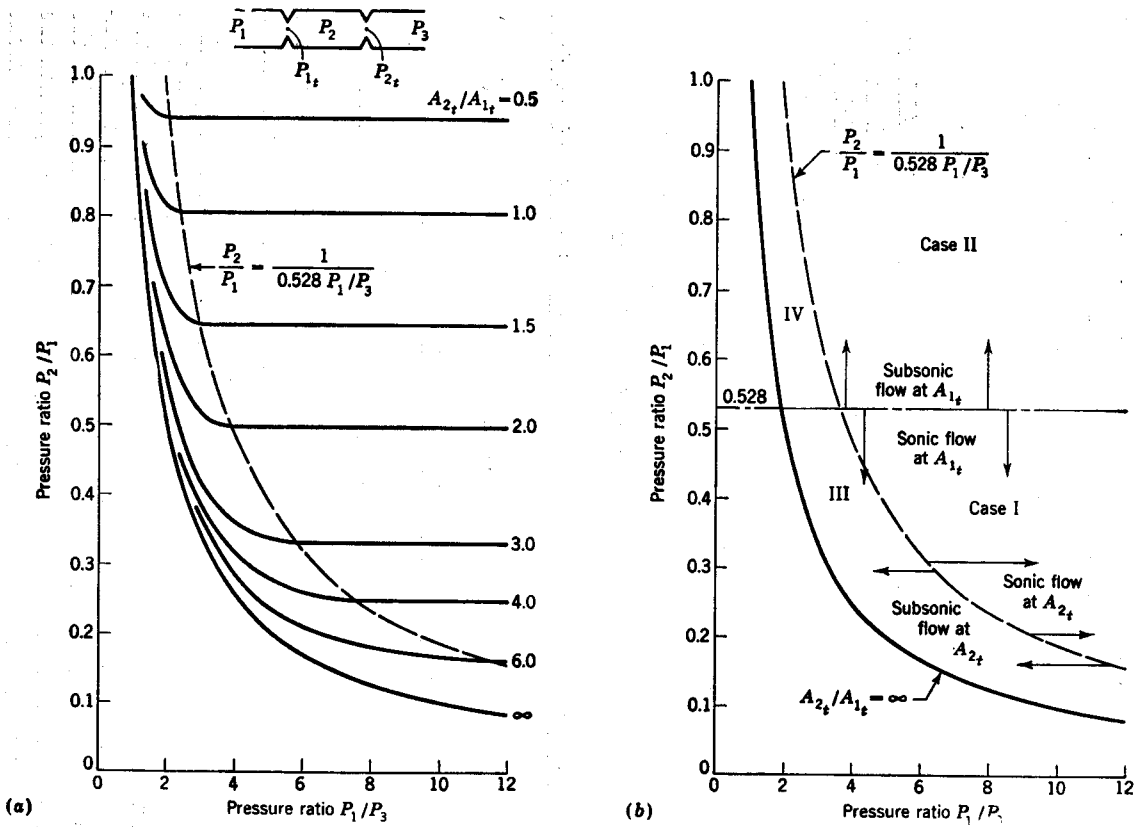


FIG. 12.14. Nondimensional curve for determining equilibrium conditions for flow through two orifices in series.

Fig. 12.14*a*, it is necessary to use absolute pressures. Because these are nondimensional curves, any consistent set of units may be used.

The derivation of this nondimensional family of curves is accomplished as follows: By assuming that the fluid is a perfect gas and that the kinetic energy at the inlet is negligible compared with other terms in the energy equation, the mass rate of flow through the first orifice is

$$M_1 = \frac{A_{1_t} P_1}{\sqrt{T_1}} \left\{ 2g \frac{k}{k-1} \frac{1}{R} \left[\left(\frac{P_{1_t}}{P_1} \right)^{2/k} - \left(\frac{P_{1_t}}{P_1} \right)^{(k+1)/k} \right] \right\}^{1/2} \tag{12.41}$$

where T_1 = stagnation temperature at inlet
g = gravitational conversion factor
k = ratio of specific heat at constant pressure to that at constant volume
R = gas constant

By replacing the subscript 1_t in Eq. (12.41) by 2_t and the subscript 1 by 2, the equation for the mass rate of flow through the second orifice is obtained. For equilibrium to exist, the mass rate of flow in equals that out, so that

$$\frac{A_{1_t} P_1}{\sqrt{T_1}} B_1^{1/2} = \frac{A_{2_t} P_2}{\sqrt{T_2}} B_2^{1/2} \tag{12.42}$$

where

$$B_1 = \left(\frac{P_{1_t}}{P_1} \right)^{2/k} - \left(\frac{P_{1_t}}{P_1} \right)^{(k+1)/k}$$

$$B_2 = \left(\frac{P_{2_t}}{P_2} \right)^{2/k} - \left(\frac{P_{2_t}}{P_2} \right)^{(k+1)/k}$$

Because there is little time for heat transfer to take place, the flow may be considered to be adiabatic so that $T_1 = T_2$. Thus Eq. (12.42) becomes

$$A_{1_t} P_1 B_1^{1/2} = A_{2_t} P_2 B_2^{1/2} \tag{12.43}$$

In the following analysis, it is assumed that the fluid is air, for which $k = 1.4$, and that the critical pressure ratio is $P_2/P_1 = P_3/P_2 = 0.528$. By using the appropriate value of k and the critical ratio, this analysis is applicable for any gas.

When sonic flow exists at the first orifice, $B_1 = 0.259$ and similarly for sonic flow at the second orifice $B_2 = 0.259$. Thus, for sonic flow at both orifices Eq. (12.43) reduces to

$$\frac{P_2}{P_1} = \frac{1}{A_{2_t}/A_{1_t}} \tag{12.44}$$

When sonic flow exists at the first orifice,

$$\frac{P_2}{P_1} \leq 0.528$$

Above the line $P_2/P_1 = 0.528$ shown in Fig. 12.14*b*, subsonic flow exists at A_{1_t}, and below this line sonic flow exists at A_{1_t}.

The equation for the line of separation between subsonic and sonic flow at A_{2_t} is obtained by noting that the critical ratio is $P_3/P_2 = 0.528$; thus

$$\frac{P_2}{P_1} = \frac{1}{(P_3/P_2)(P_1/P_3)} = \frac{1}{0.528(P_1/P_3)} \tag{12.45}$$

The curve defined by the preceding expression is shown in Fig. 12.14*b*. To the right of this curve sonic flow exists at A_{2_t}, and to the left of this curve subsonic flow exists at A_{2_t}. The regions in which each of the four possible combinations of sonic or subsonic flow may exist at the first and second orifices are shown in Fig. 12.14*b*, that is,

Case I sonic flow at both orifices
Case II subsonic flow at the first orifice and sonic at the second
Case III sonic flow at the first orifice and subsonic at the second
Case IV subsonic flow at both orifices

For case I, from Eq. (12.44) it follows that the lines of constant values of A_{2_t}/A_{1_t} are horizontal straight lines as shown in Fig. 12.14*a*.

For case II, $P_2 = P_{1_t}$, and $B_2 = 0.259$, so that Eq. (12.43) reduces to

$$\frac{P_2}{P_1}\frac{1}{B_1^{1/2}} = \frac{1}{0.259(A_{2_t}/A_{1_t})} \tag{12.46}$$

For a given area ratio A_{2_t}/A_{1_t}, there is but one value of P_2/P_1 which makes the left-hand side of Eq. (12.46) equal to the right-hand side. Thus, for case II, lines of constant A_{2_t}/A_{1_t} are also horizontal.

By applying these techniques to case III and case IV, the complete family of curves shown in Fig. 12.14*a* is obtained. This method of analysis may be extended to determine equilibrium flow conditions for three or more orifices in series.[1]

Equation (12.41) is in an awkward form for computing partial derivatives. However, for the case of sonic flow, this expression reduces to

$$M_1 = \frac{0.528}{\sqrt{T_1}} A_{1_t} P_1 \tag{12.47}$$

where the preceding expression has units of inches, pounds, degrees Rankine, and seconds. For the usual design case in which the stagnation temperature of the inlet air is 60°F or 520°R, then $M_1 = 0.0074 A_{1_t} P_1$.

[1] The author wishes to express his appreciation to Mr. Stanley Best, Chief of Analysis, Hamilton Standard Division, United Aircraft Corporation, Windsor Locks, Conn., for his fine comments and suggestions concerning the development of the preceding method for determining equilibrium flow conditions.

For the case of subsonic flow $P_{1_t} = P_2$, Eq. (12.41) may be approximated by noting that a plot of the function B_1 versus $(P_1 - P_2)P_2/P_1^2$ is very nearly a straight line. The slope of this line is such that

$$B_1 \approx \frac{0.261(P_1 - P_2)P_2}{P_1^2} \tag{12.48}$$

Substitution of the preceding approximation into Eq. (12.41) gives

$$\begin{aligned} M_1 &\approx \frac{2.06}{\sqrt{T_1}} A_{1_t} \sqrt{0.261(P_1 - P_2)P_2} \\ &= \frac{1.05}{\sqrt{T_1}} A_{1_t} \sqrt{(P_1 - P_2)P_2} \end{aligned} \tag{12.49}$$

where $\sqrt{\frac{2g}{R}\frac{k}{k-1}} = 2.06$ when units of inches, pounds, degrees Rankine, and seconds are employed. For a stagnation temperature of 520°R, Eq. (12.49) is

$$M_1 = 0.0146 A_{1_t} \sqrt{(P_1 - P_2)P_2} \tag{12.50}$$

CHAPTER 13

ELECTRICAL SYSTEMS

13.1. Introduction. The availability of electrical power and the ease of transmitting signals via wires or microwaves are desirable aspects of electrical equipment. The characteristics of an electrical component may usually be altered by a simple adjustment such as changing the size of a resistor or a capacitor, as was the case for the general-purpose analog computer discussed in Chap. 8. Thus, electrical equipment tends to be versatile and convenient to use.

This chapter discusses the characteristics of electrical components which are frequently used in automatic control systems. It is shown how to obtain the operational form of the differential equation for commonly used electrical devices such as motors, generators, vacuum tubes, transistor amplifiers, etc.[1–3]

13.2. D-C Motors. A major reason for the use of d-c machines in electromechanical control systems is the ease with which speed can be controlled. The polarity of the applied voltage determines the direction of rotation. Also d-c machines are capable of providing large power amplifications.

The field and armature windings of d-c motors may be shunt-connected, series-connected, compounded, or separately excited. The motors used in control systems are generally separately excited. There are two types of separate excitation, field control with fixed armature current, and armature control with fixed field.

Field Control. A separately excited motor in which the armature current I_a is maintained constant is shown in Fig. 13.1*a*. The constant current I_a may be supplied by a d-c generator or from an a-c line. The latter method requires the use of transformers and rectifiers to obtain the proper rectification. The voltage E_f applied to the field is obtained from the output of an amplifier in low-power application or from a d-c

[1] J. E. Gibson and F. B. Tuteur, "Control System Components," McGraw-Hill Book Company, Inc., New York, 1958.

[2] N. R. Ahrendt and C. J. Savant, "Servomechanism Practice," 2d ed., McGraw-Hill Book Company, Inc., New York, 1960.

[3] J. G. Truxal, "Control Engineers' Handbook," McGraw-Hill Book Company, Inc., New York, 1958.

generator when greater power is needed. In the field circuit, the resistance of the windings is R_f, and the inductance is designated by L_f. The torque T developed by a motor is proportional to the product of the armature current I_a and the magnetic flux ϕ of the field.

$$T = K_1 \phi I_a \tag{13.1}$$

where K_1 is a constant for any motor and depends upon the total number of armature conductors, the number of field poles, etc.

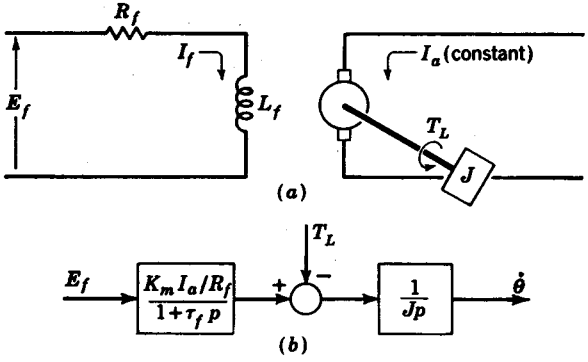


FIG. 13.1. Field-controlled d-c motor.

A typical curve of flux ϕ versus field current I_f is shown in Fig. 13.2. When the field current I_f becomes great enough to cause the iron to saturate, the flux ϕ no longer increases linearly with the current. Motors used in control systems usually operate over the linear portion of this curve, in which case

$$\phi = K_2 I_f \tag{13.2}$$

where K_2 is the slope of the linear portion of the curve as shown in Fig. 13.2. The substitution of the preceding result into Eq. (13.1) yields

$$T = K_1 K_2 I_a I_f = K_m I_a I_f \tag{13.3}$$

where $K_m = K_1 K_2$.

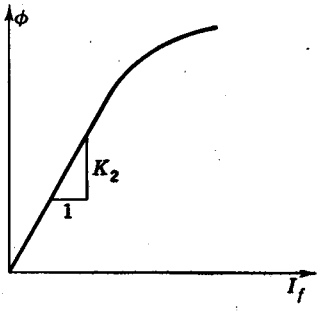


FIG. 13.2. Plot of flux vs. field current.

If the moment of inertia of the armature is J, the coefficient of viscous friction C_v, and the load torque T_L, from a summation of torques acting on the armature it follows that

$$T = (C_v p + J p^2)\theta + T_L \tag{13.4}$$

where θ is the angular position of the armature, or motor shaft.

The equation for the field current I_f is obtained from the equivalent field circuit of Fig. 13.1*a*.

$$I_f = \frac{E_f}{R_f + L_f p} = \frac{E_f}{R_f(1 + \tau_f p)} \tag{13.5}$$

where $\tau_f = L_f/R_f$ is the time constant of the field circuit.

Substituting T from Eq. (13.4) and I_f from Eq. (13.5) into Eq. (13.3) and solving for θ gives

$$\theta = \frac{1}{p(C_v + Jp)} \left[\frac{(K_m I_a/R_f)E_f}{1 + \tau_f p} - T_L \right] \tag{13.6}$$

Multiplication of the preceding expression by p gives the angular velocity $\dot{\theta} = p\theta$,

$$\dot{\theta} = \frac{1}{C_v + Jp} \left[\frac{(K_m I_a/R_f)E_f}{1 + \tau_f p} - T_L \right] \tag{13.7}$$

Generally, the damping C_v is negligible, so that the block-diagram representation for the speed of this field-controlled d-c motor is as shown in Fig. 13.1*b*.

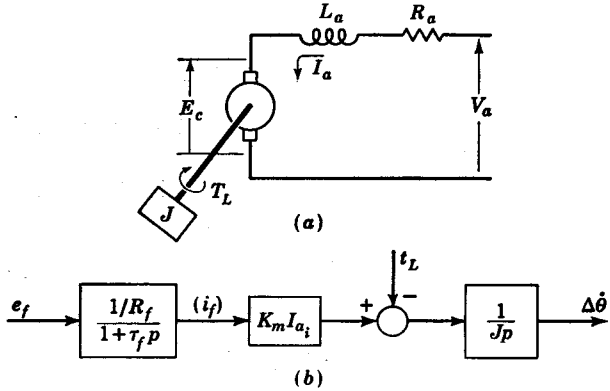


FIG. 13.3. Approximately constant current source.

Because of the difficulty and expense of obtaining a constant-current source, this type of motor is often operated with an approximately constant-current source. This is accomplished by supplying the armature with a constant voltage V_a and inserting a very large resistance in series with the armature. The resulting schematic representation is shown in Fig. 13.3*a*. The resistance R_a is the sum of the inserted series resistance and that of the armature. The term L_a represents the inductance of the armature. The voltage E_c is the counter emf induced by the rotation of the armature windings in the magnetic field. The counter emf is proportional to the product of the armature speed $\dot{\theta}$ and the field

strength ϕ. Thus

$$E_c = K_3\phi\dot{\theta} \tag{13.8}$$

where K_3 is a constant for any particular motor. The substitution of ϕ from Eq. (13.2) into Eq. (13.8) yields

$$E_c = K_2K_3I_f\dot{\theta} = K_cI_f\dot{\theta} \tag{13.9}$$

where $K_c = K_2K_3$. Solving the circuit of Fig. 13.3*a* for the armature current I_a yields

$$I_a = \frac{V_a - E_c}{R_a + L_ap} \tag{13.10}$$

Equation (13.9) is nonlinear in that I_f and $\dot{\theta}$ multiply. To obtain a linear representation for this system, it is necessary to linearize the equations which describe its operation. The linearized form of Eq. (13.9) is

$$e_c = \left.\frac{\partial E_c}{\partial\dot{\theta}}\right|_i \Delta\dot{\theta} + \left.\frac{\partial E_c}{\partial I_f}\right|_i i_f = K_cI_{f_i}\,\Delta\dot{\theta} + K_c\dot{\theta}_i i_f \tag{13.11}$$

where from Eq. (13.9) it follows that

$$\left.\frac{\partial E_c}{\partial\dot{\theta}}\right|_i = K_cI_{f_i} \qquad \text{and} \qquad \left.\frac{\partial E_c}{\partial I_f}\right|_i = K_c\dot{\theta}_i$$

Because Eq. (13.10) is linear, all that is necessary is to replace the capital-letter representation of the independent variables by their lowercase counterparts.

$$i_a = \frac{v_a - e_c}{R_a + L_ap} \tag{13.12}$$

From the assumption that the armature supply voltage V_a is constant, v_a is zero. The preceding expression thus becomes

$$i_a = \frac{-e_c}{R_a + L_ap} \tag{13.13}$$

The torque T developed by a motor as given by Eq. (13.3) is linearized as follows,

$$t = \left.\frac{\partial T}{\partial I_f}\right|_i i_f + \left.\frac{\partial T}{\partial I_a}\right|_i i_a = K_mI_{a_i}i_f + K_mI_{f_i}i_a \tag{13.14}$$

where from Eq. (13.3) it follows that

$$\left.\frac{\partial T}{\partial I_f}\right|_i = K_mI_{a_i} \qquad \text{and} \qquad \left.\frac{\partial T}{\partial I_a}\right|_i = K_mI_{f_i}$$

The torque balance for the armature as given by Eq. (13.4) is already linear. Replacing the independent variable terms designated by capital

letters in Eq. (13.4) by their lower-case counterparts yields the following torque balance for small departures:

$$t = (C_v + Jp)\,\Delta\dot{\theta} + t_L \tag{13.15}$$

By substituting e_c from Eq. (13.11) into Eq. (13.13), and then substituting i_a from this resulting equation into Eq. (13.14), one obtains an equation for the torque variation t as a function of i_f and $\Delta\dot{\theta}$. Equating

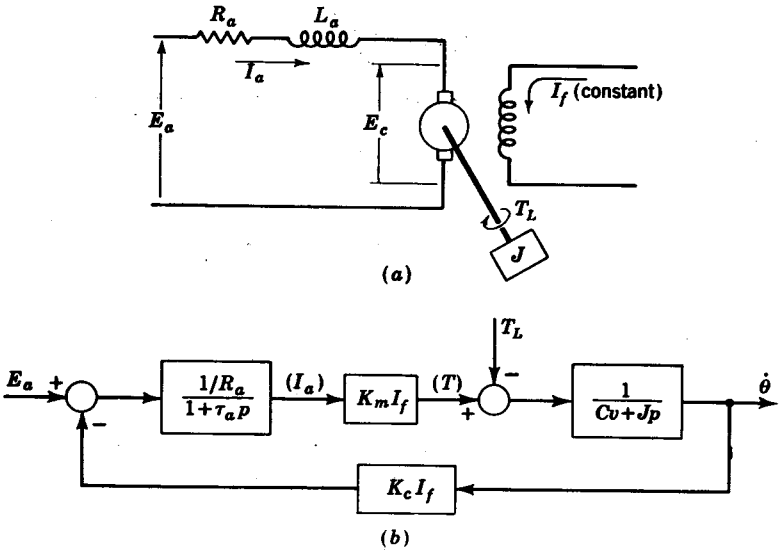


FIG. 13.4. Armature-controlled d-c motor.

this value of t to Eq. (13.15) and solving for $\Delta\dot{\theta}$ yields

$$\Delta\dot{\theta} = \frac{K_m[(I_{a_i}R_a - K_cI_{f_i}\dot{\theta}_i) + I_{a_i}L_ap]i_f - (R_a + L_ap)t_L}{(R_a + L_ap)(C_v + Jp) + K_mK_cI_{f_i}^2} \tag{13.16}$$

For the usual case in which the equilibrium operating condition is about the point $\dot{\theta}_i = I_{f_i} = 0$ and the damping C_v is negligible, the preceding expression reduces to

$$\Delta\dot{\theta} = \frac{1}{Jp}\,(K_mI_{a_i}i_f - t_L) \tag{13.17}$$

From Eq. (13.17) and the linearized form of Eq. (13.5), the over-all block-diagram representation for this field-controlled motor which is supplied by an approximately constant armature current may be constructed as shown in Fig. 13.3*b*.

Armature Control. A d-c motor with armature control is one in which the speed is controlled by the armature voltage E_a. An armature-

controlled motor is shown in Fig. 13.4*a*, in which the field current I_f is kept constant. It is much easier to maintain a constant field current I_f than a constant armature current I_a because there is no counter emf generated in the fixed field windings. The armature voltage E_a is usually supplied by a generator, which in turn may be supplied by an amplifier.

The circuit equation for the armature portion of Fig. 13.4*a* is

$$E_a - K_c I_f \dot{\theta} = R_a I_a + L_a p I_a = R_a(1 + \tau_a p) I_a \tag{13.18}$$

where $\tau_a = L_a/R_a$ and the term $K_c I_f \dot{\theta}$ is the counter emf developed in the armature. The torque developed by the motor is given by the equation

$$T = K_m I_f I_a \tag{13.19}$$

and the torque balance for the output shaft is

$$T = (C_v p + J p^2)\theta + T_L \tag{13.20}$$

The block-diagram representation for this armature-controlled motor is obtained by combining the block-diagram representations for Eqs.

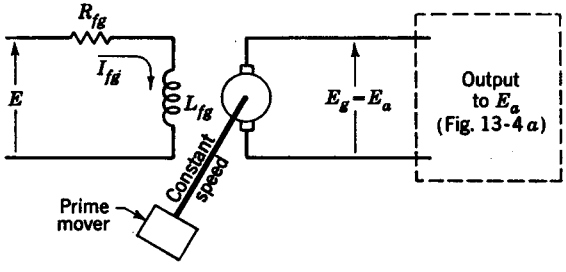


FIG. 13.5. Complete generator and armature-controlled motor combination.

(13.18), (13.19), and (13.20) as shown in Fig. 13.4*b*. The counter emf is responsible for the minor feedback.

A complete generator and armature-controlled motor combination is shown in Fig. 13.5. This is in effect a Ward-Leonard system, or motor-generator set. The voltage E supplied to the generator may be quite small, as in the case of that coming from an amplifier. The resistance of the field of the generator is R_{fg}, and the inductance is L_{fg}. The armature of the generator is driven at a constant speed by a prime mover. The output voltage of the generator, E_g, goes directly to the armature of the motor so that $E_g = E_a$.

The circuit equation for the generator field is

$$E = (R_{fg} + L_{fg} p) I_{fg} \tag{13.21}$$

The voltage induced in the armature is the generated voltage $E_g = E_a$, which is

$$E_a = K_{cg}\dot{\theta}_g I_{fg} = K'_c I_{fg} \tag{13.22}$$

where $\dot{\theta}_g$ is the angular velocity of the prime mover, which is constant, so that $K_{cg}\dot{\theta}_g = K'_c$. The substitution of I_{fg} from Eq. (13.21) into Eq. (13.22) yields

$$E_a = \frac{K'_c E}{R_{fg} + L_{fg}p} = \frac{(K'_c/R_{fg})E}{1 + \tau_{fg}p} \tag{13.23}$$

where $\tau_{fg} = L_{fg}/R_{fg}$ is the time constant of the generator field.

The over-all block diagram relating the input voltage E and the velocity $\dot{\theta}$ of this armature-controlled motor-generator system is obtained

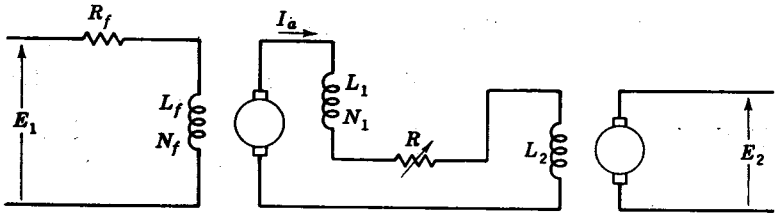


FIG. 13.6. Rototrol generator.

by connecting the block-diagram representation for the output E_a from Eq. (13.23) to that for E_a in Fig. 13.4*b*.

D-C Tachometer. A d-c tachometer is a generator in which the magnetic flux is supplied by a permanent magnet. Because the flux is maintained constant, the equation of operation for a tachometer is obtained from Eq. (13.8) as follows,

$$E_c = E_g = K_3\phi\dot{\theta} = K_g\dot{\theta} \tag{13.24}$$

where $K_g = K_3\phi$ is a constant and E_g is the generated voltage.

Thus a tachometer is seen to supply a voltage E_g which is proportional to the speed at which it is driven.

Rototrol Generator. When very large power amplification is required, it is customary to place two generators in series. The Rototrol generator is a combination of two generators in which part of the field for the first, or pilot, generator is in series with the armature. In Fig. 13.6 this series portion of the field is designated by L_1, and the portion with separate excitation is L_f. The series field L_1 causes a self-energizing effect in that as the armature current I_a increases so also does the flux caused by the series winding L_1. An analysis of the system shown in Fig. 13.6 yields the following equation relating the output voltage E_2 to

the input E_1,

$$E_2 = \frac{K_v K_p / R_f}{(1 + \tau_f p)(R - K_p N_1 / N_f + Lp)} E_1 \tag{13.25}$$

where K_v is the constant relating the voltage output E_2 of the main generator to the armature current I_a (that is, $E_2 = K_v I_a$) and K_p is the corresponding constant for the pilot generator. The term $L = L_1 + L_2$ is the total inductance of the armature circuit, and R is the total resistance. The number of turns on the series field winding is N_1, and that on the separately excited winding is N_f. The time constant of the separately excited pilot-generator field is $\tau_f = L_f / R_f$.

The critical value of R is

$$R = R_c = \frac{K_p N_1}{N_f} \tag{13.26}$$

When the total resistance R is adjusted to the critical value, Eq. (13.25) becomes

$$\frac{E_2}{E_1} = \frac{K_v K_p}{R_f L} \frac{1}{p(1 + \tau_f p)} \tag{13.27}$$

The primary advantage of the Rototrol is the integrating effect that is obtained when it is critically tuned; otherwise there would be little if any advantage in using a Rototrol in preference to two simple generators in series.

Amplidyne. An amplidyne is essentially a two-stage generator, but the two stages are combined in a single machine. The main advantage of an amplidyne is that it is smaller and more compact than a Rototrol of the same rating. Because of coupling effects and various interactions which take place in the one machine, the performance of the amplidyne is poorer than that obtained with two separate machines. Also, the use of two generators permits greater flexibility in the adjustment of operating characteristics.

Remote-control Positional Servomechanism. A remote-control positional servomechanism is shown in Fig. 13.7*a*. The wiper arm of the input potentiometer is positioned by the desired input position θ_r, so that the voltage E_r is proportional to θ_r (that is, $E_r = K_r \theta_r$). Similarly, the controlled shaft position θ_c determines the position of the wiper arm for the other potentiometer so that $E_c = K_c \theta_c$. The error signal $E_e = E_r - E_c$ is amplified by the amplifier, and the resultant voltage is applied to the field of a field-controlled motor so that $E_f = K_1 E_e$. The operational representation of the differential equation for the motor is given by Eq. (13.6). The over-all block diagram for this system is shown in Fig. 13.7*b*.

The motor must be located at the output shaft, while the input potenti-

ometer is usually situated in any convenient location. A major advantage in using such electrical equipment for position control systems is the ease of connecting the input and output by means of wires.

The preceding position controller may be converted to a speed control system by connecting the output shaft to a tachometer rather than to a potentiometer. In this case, the voltage signal E_c coming from the tachometer is proportional to the speed $\dot{\theta}_c$ (that is, $E_c = K_c'\dot{\theta}_c$). Similarly, each wiper position of the input potentiometer corresponds to a desired

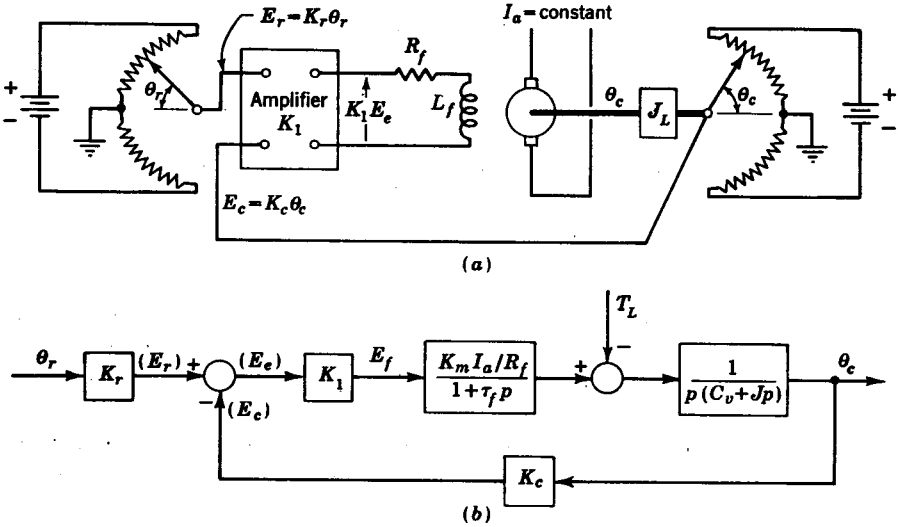


FIG. 13.7. Remote-control positional servomechanism.

speed setting $\dot{\theta}_r$ rather than position θ_r, so that the reference voltage is $E_r = K_r'\dot{\theta}_r$.

13.3. A-C Two-phase Motor. An a-c two-phase motor is used for simple low-power applications. One of the phases is supplied with a fixed a-c voltage which acts as the reference voltage. The other phase is connected to the controlled voltage. A schematic representation of a two-phase motor is shown in Fig. 13.8*a*. Because the reference voltage E_R is constant, the speed depends upon the control voltage E. The direction of rotation is reversed by changing the polarity of the control voltage.

As is shown in Fig. 13.8*a*, the reference and control windings are displaced by 90° in the stator of the motor. Thus, although the voltage applied to each winding has the same frequency, there is a 90° phase shift of one with respect to the other.

Typical performance curves relating the developed motor torque T and

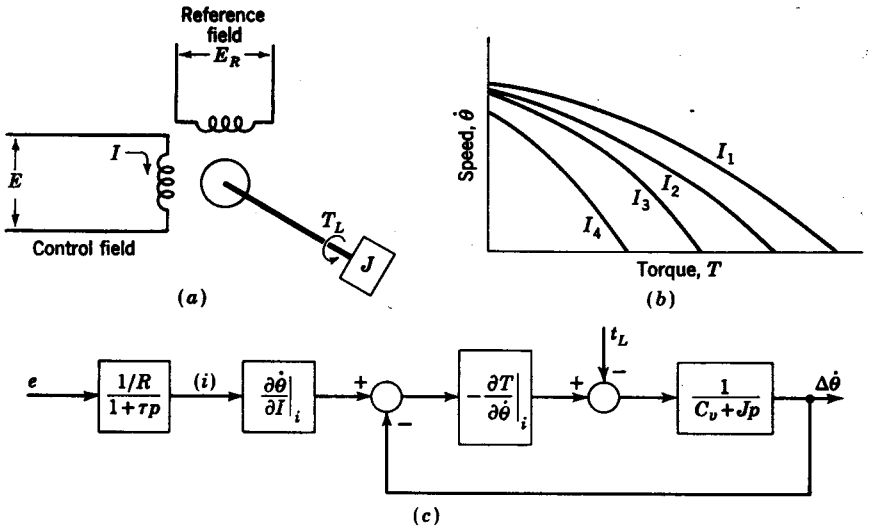


FIG. 13.8. (*a*) Two-phase a-c motor; (*b*) typical performance curves for a two-phase a-c motor; (*c*) block diagram representation for a two-phase a-c motor.

the angular velocity $\dot{\theta}$ for the constant values of control current I are shown in Fig. 13.8*b*. The equation describing the operation of a two-phase motor about some equilibrium point of operation is derived as follows: From Fig. 13.8*b* it is to be noticed that the speed is a function of T and I.

$$\dot{\theta} = F(T,I) \tag{13.28}$$

$$\Delta\dot{\theta} = \left.\frac{\partial\dot{\theta}}{\partial T}\right|_i t + \left.\frac{\partial\dot{\theta}}{\partial I}\right|_i i \tag{13.29}$$

The torque balance for the armature is

$$T = (C_v + Jp)\dot{\theta} + T_L \tag{13.30}$$

The form of the preceding equation for small departures is

$$t = (C_v + Jp)\,\Delta\dot{\theta} + t_L \tag{13.31}$$

The circuit equation for the control winding is

$$E = (R + Lp)I = R(1 + \tau p)I \tag{13.32}$$

For small departures this expression becomes

$$e = R(1 + \tau p)i \tag{13.33}$$

The block-diagram representation for this two phase a-c motor is obtained from Eqs. (13.29), (13.31), and (13.33) as shown in Fig. 13.8*c*.

13.4. Synchro Error-detecting Device. Synchro systems are usually employed as error-sensing devices. A synchro system as shown in Fig. 13.9*a* consists of a synchro generator and synchro motor. The rotor of both the generator and motor are connected to the same a-c supply. When the rotors are in the same relative position such that θ_1 is equal to θ_2, identical flux patterns are set up in each stator. Because the stator winding for each unit is connected as shown in Fig. 13.9*b*, the voltage

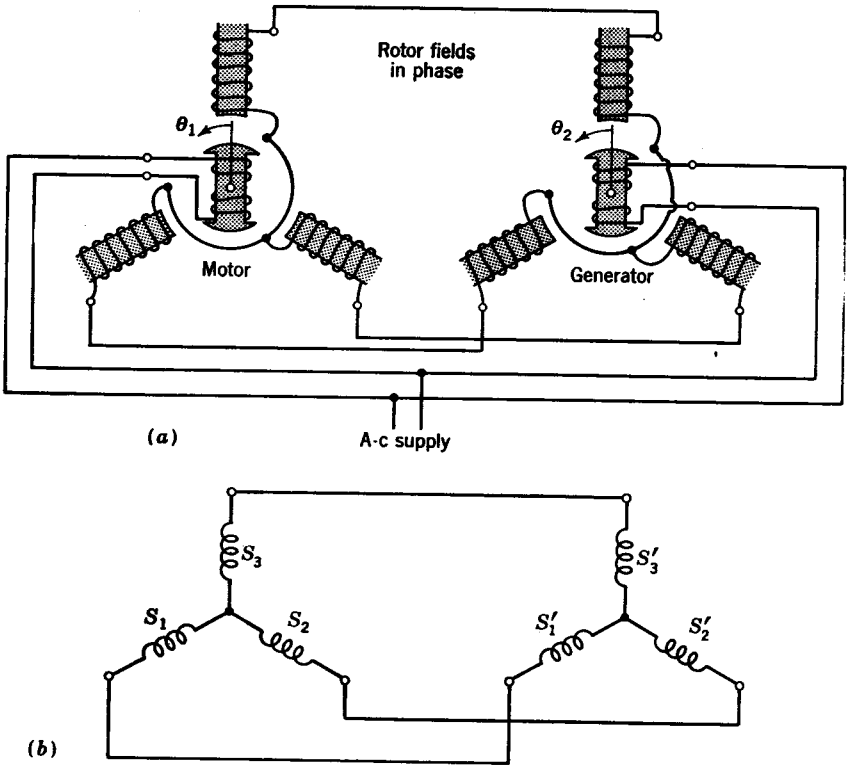


FIG. 13.9. (*a*) Synchro generator and motor combination; (*b*) wiring of synchro units in correspondence.

induced in each corresponding phase will be the same for the generator and the motor. In this case, the voltage induced in the first stator winding S_1' of the generator is the same as that induced in the first winding S_1 of the motor, etc. Because all corresponding coils have the same induced voltage, no current flows through the stator coils for this balanced position.

When the rotors are rotated with respect to one another so that θ_1 is not equal to θ_2, different voltages are induced across corresponding phases

of the stator windings. This causes currents to flow through the stator windings. The reaction of the rotor flux and the current produces a torque on each rotor which tends to align them. In many applications, the rotor of the generator is held in a fixed, or reference, position; thus the relative position of the rotor in the motor will depend upon the torque applied to the motor shaft. The relation between torque and relative angular position is reasonably linear up to about 75° and is given by the equation

$$T = k(\theta_2 - \theta_1) \tag{13.34}$$

A device that produces a voltage which is a function of the relative rotor positions may be obtained by a slight modification of the synchro

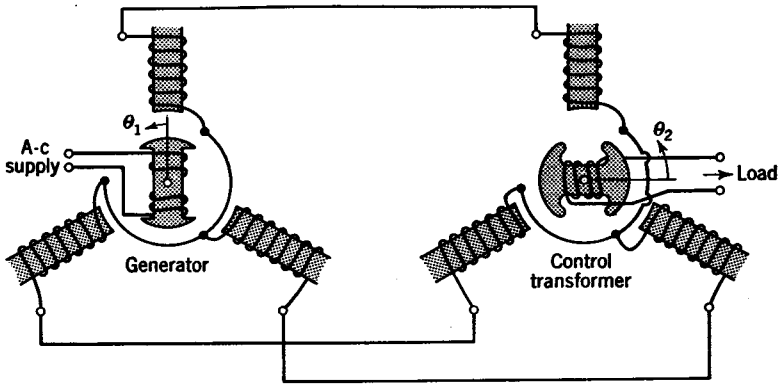


FIG. 13.10. Synchro control transformer.

generator-motor system. Such a system is called a synchro control transformer and is shown in Fig. 13.10. The generator is the same as before, but the motor has been replaced by a synchro control transformer. Only the rotor of the generator is connected to the a-c supply. The stator of the generator and that of the control transformer are identical, as shown in Fig. 13.10. Thus, the field produced in the stator of the control transformer is the same as that of the generator. The voltage E_o induced in the rotor of the control transformer is a maximum, E_m, when the two rotors are aligned. The voltage decreases sinusoidally as the rotors are displaced. When the rotors are 90° apart, the control transformer rotor is at right angles to the field and the output voltage E_o is zero. To take into account the fact that the output is zero when the rotors are displaced by 90°, it is convenient to measure θ_2 from a reference position which is at right angles to the reference position for θ_1, as shown in Fig. 13.10. The operation of the synchro control transformer is then described by the equation

$$E_o = E_m \sin (\theta_2 - \theta_1) \tag{13.35}$$

In Fig. 13.11 is shown a typical remote-control positional servomechanism in which the error-sensing device is a synchro control transformer. The output voltage coming from the control transformer is a measure of the error

$$E_e = E_m \sin (\theta_r - \theta_c) \tag{13.36}$$

For normal operation in which the error $\theta_r - \theta_c$ is small, the preceding expression becomes

$$E_e \approx E_m(\theta_r - \theta_c) \tag{13.37}$$

13.5. Vacuum-tube Amplifier. Vacuum tubes are frequently used to amplify voltage signals in electrical control systems. A triode, as shown

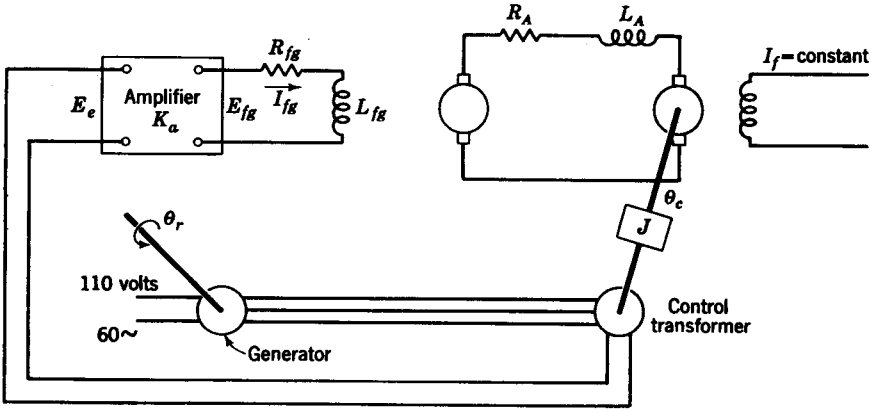


FIG. 13.11. Remote-control positional servomechanism.

in Fig. 13.12*a*; is the simplest type of vacuum-tube amplifier. The three basic elements of a triode are the cathode, grid, and plate. The cathode serves as the source of electrons. A heater element is incorporated in the cathode for the purpose of increasing the number of free electrons on the surface of the cathode. Because electrons are negatively charged particles and the plate is charged positively with respect to the cathode, the number of electrons flowing from the cathode to the plate will depend upon the value of the voltage drop from the plate to the cathode, E_{pk}. (In accordance with standard practice, the direction of current flow is opposite to that of the flow of electrons.) When the grid is charged negatively with respect to the cathode (that is, E_{gk} is negative), electrons on the cathode are repelled, which decreases the current flow. Because the grid is located close to the cathode, a small negative charge E_{gk} will cause a rather substantial decrease in current flow. For example, in Fig. 13.12*b* with the plate voltage E_{pk} held constant at 80 volts, changing

the grid voltage E_{gk} from 0 to -2 volts results in a change in the plate current I_p from 7.8 to 3.4 ma.

For a vacuum-tube amplifier, the plate current I_p is a function of the voltage drop E_{pk} from the plate to the cathode and also the grid-bias

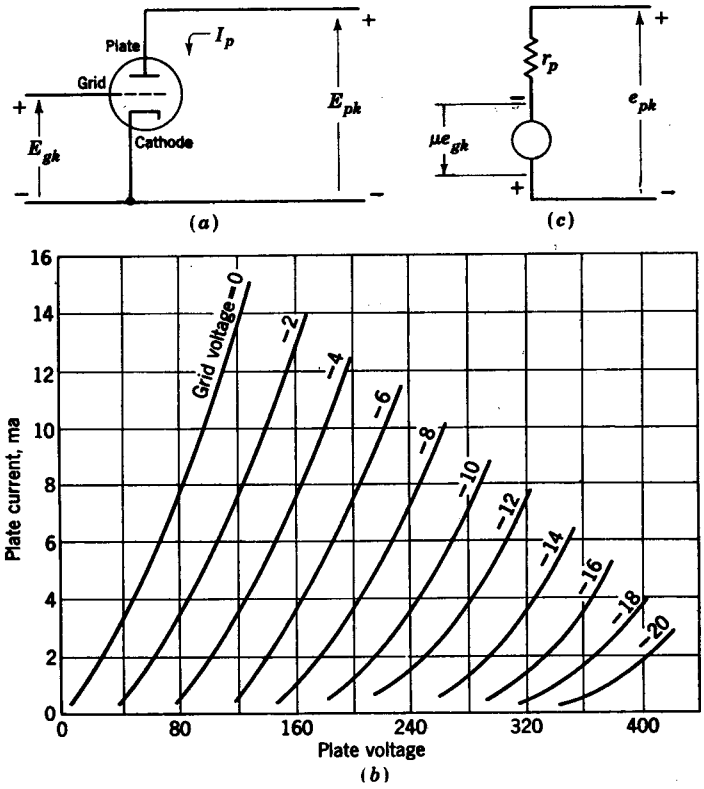


FIG. 13.12. (*a*) Triode; (*b*) typical tube characteristics for a triode; (*c*) equivalent circuit for a triode.

voltage E_{gk}, that is,

$$I_p = F(E_{pk}, E_{gk}) \tag{13.38}$$

Linearization of the preceding expression yields

$$i_p = \left.\frac{\partial I_p}{\partial E_{pk}}\right|_i e_{pk} + \left.\frac{\partial I_p}{\partial E_{gk}}\right|_i e_{gk} \tag{13.39}$$

Solving for e_{pk} gives

$$e_{pk} = \left.\frac{\partial E_{pk}}{\partial I_p}\right|_i i_p - \left.\frac{\partial E_{pk}}{\partial I_p}\right|_i \left.\frac{\partial I_p}{\partial E_{gk}}\right|_i e_{gk} \tag{13.40}$$

The product of the partial derivatives in the last term may be simplified

by first writing Eq. (13.38) in the implicit form

$$G(I_p,E_{pk},E_{gk}) = 0 \tag{13.41}$$

From calculus, it is known that the product of all the partial derivatives for an implicit function is equal to -1, that is,

$$\frac{\partial I_p}{\partial E_{pk}} \frac{\partial E_{pk}}{\partial E_{gk}} \frac{\partial E_{gk}}{\partial I_p} = -1 \tag{13.42}$$

where

$$\frac{\partial E_{pk}}{\partial E_{gk}} = -\frac{\partial E_{pk}}{\partial I_p} \frac{\partial I_p}{\partial E_{gk}} \tag{13.43}$$

Substitution of the preceding result into Eq. (13.40) gives

$$e_{pk} = \frac{\partial E_{pk}}{\partial I_p}\bigg|_i i_p + \frac{\partial E_{pk}}{\partial E_{gk}}\bigg|_i e_{gk}$$

or

$$e_{pk} = r_p i_p - \mu e_{gk} \tag{13.44}$$

where $\partial E_{pk}/\partial I_p = r_p$ is the change in voltage across the plate per change in current, which is the effective resistance of the plate. Thus, r_p is called the dynamic plate resistance. The term $\partial E_{pk}/\partial E_{gk} = -\mu$ is the change in voltage drop across the plate per change in grid to cathode voltage, with I_p remaining constant. Because E_{pk} increases as E_{gk} decreases, the negative sign is used so that μ will always be a positive number. The term μ is called the amplification factor. The equation for the operation of the circuit shown in Fig. 13.12*c* is the same as Eq. (13.44). Thus, Fig. 13.12*c* is the equivalent circuit for a vacuum-tube amplifier.

In Fig. 13.13*a* is shown a vacuum tube in which the load is a pure resistance. The d-c voltage E_{cc} determines the equilibrium value of the grid bias E_{gk}. The total grid bias is given by the equation

$$E_{gk} = E_s - E_{cc} \tag{13.45}$$

where E_s is the input-voltage signal which is to be amplified. The plate bias is provided by the d-c voltage E_{bb}. The loop equation for load and plate circuit is

$$E_{bb} = I_p R_L + E_{pk}$$

or

$$I_p = -\frac{1}{R_L} E_{pk} + \frac{E_{bb}}{R_L} \tag{13.46}$$

Equation (13.46) is the equation of the straight line which is superimposed on the curve for the tube characteristics as shown in Fig. 13.13*b*. From Eq. (13.46), it follows that with I_p as the ordinate and E_{pk} as the abscissa, the slope of this line is $-1/R_L$, the I_p intercept is E_{bb}/R_L (that is, when $E_{pk} = 0$, then $I_{pk} = E_{bb}/R_L$), and the E_{pk} intercept is E_{bb}. This straight line is called the load line. The reference point of operation of

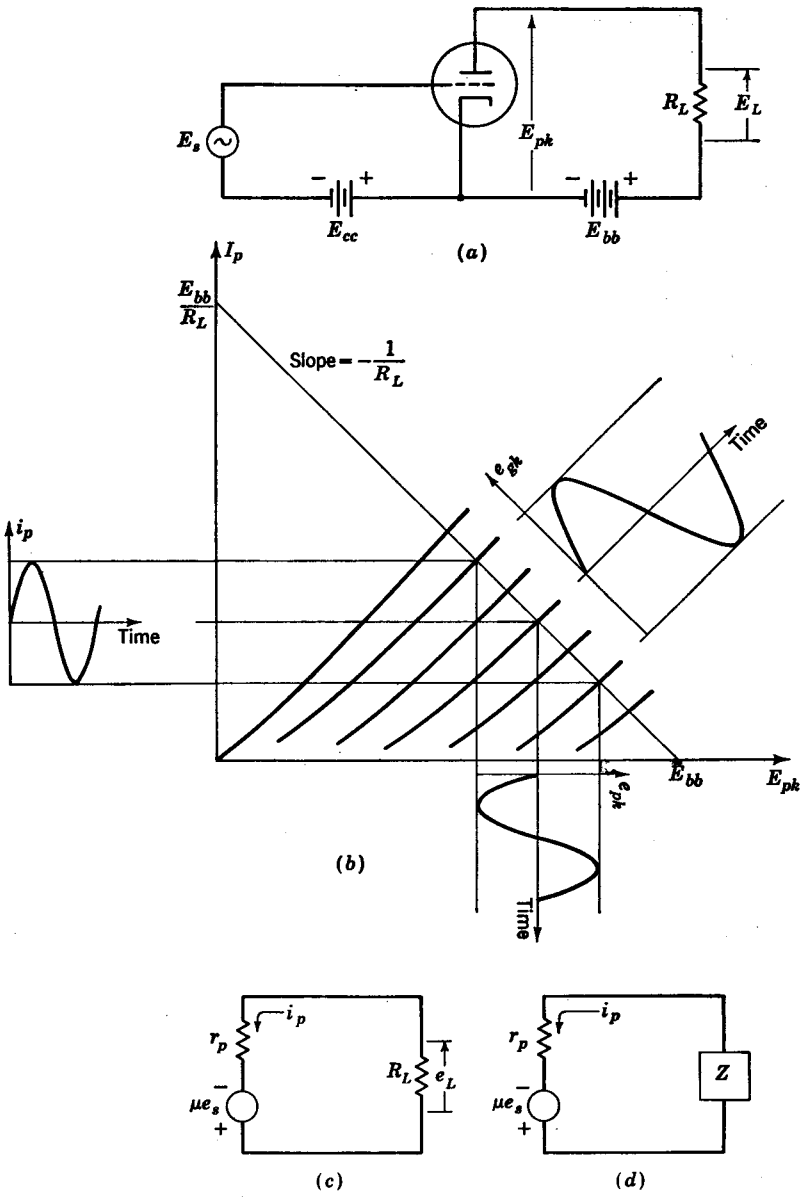


FIG. 13.13. (*a*) Vacuum tube with resistive load; (*b*) load line superimposed upon operating curves; (*c*) equivalent amplifier circuit with a resistive load; (*d*) equivalent amplifier circuit with a general impedance load.

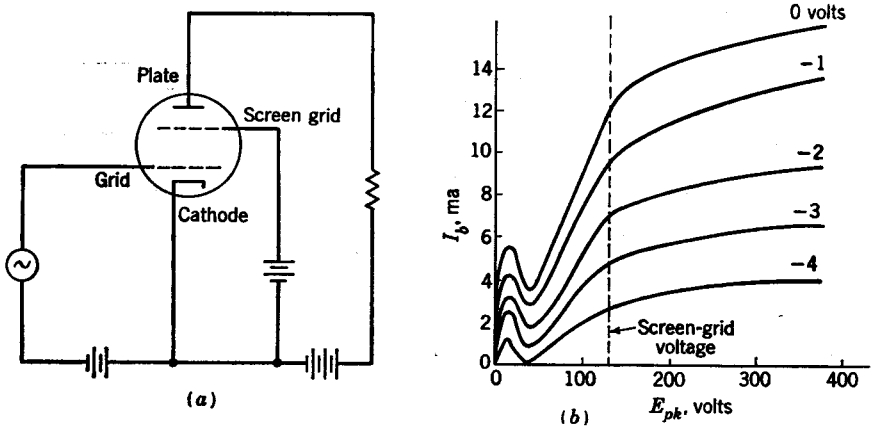


FIG. 13.14. (*a*) Tetrode; (*b*) typical tube characteristics for a tetrode.

the tube is determined by the intersection of this load line and the tube-characteristic line for $E_s = 0$, which is $E_{gk} = -E_{cc}$. From Fig. 13.13*b*, it is to be seen that, for a small sinusoidal input e_{gk}, the variation in the plate current i_p and the variation e_{pk} are also sinusoidal. The linearization of Eqs. (13.45) and (13.46) yields

$$\begin{aligned} e_{gk} &= e_s \\ i_p &= -\frac{e_{pk}}{R_L} \end{aligned} \tag{13.47}$$

The substitution of the preceding results for e_{gk} and e_{pk} into Eq. (13.44) yields

$$-R_L i_p = r_p i_p - \mu e_s$$

or

$$\mu e_s = (r_p + R_L) i_p \tag{13.48}$$

Equation (13.48) is the basis for the equivalent amplifier circuit shown in Fig. 13.13*c*. When the load is some general impedance Z rather than a pure resistance, the equivalent circuit is that shown in Fig. 13.13*d*. The over-all voltage amplification for a tube is the ratio of output voltage e_L to the input signal e_s. Thus, the over-all voltage amplification, or gain, is

$$\text{Gain} = \frac{e_L}{e_s} = \frac{-R_L i_p}{e_s} \tag{13.49}$$

The substitution of i_p from Eq. (13.48) into the preceding expression yields

$$\text{Gain} = -R_L \frac{\mu e_s}{(r_p + R_L)/e_s} = \frac{-\mu R_L}{r_p + R_L} = -K \tag{13.50}$$

The minus sign in the preceding expression indicates that the voltage e_L decreases as e_s increases (i.e., there is a 180° phase shift).

Tetrode and Pentode. A tetrode, which is a four-element tube, and a pentode, which has five elements, both have much higher gains than a triode, which has only three elements. A tetrode has a screen grid in addition to the usual cathode, grid, and plate of a triode. A tetrode is shown in Fig. 13.14*a*, and typical tube characteristics are shown in

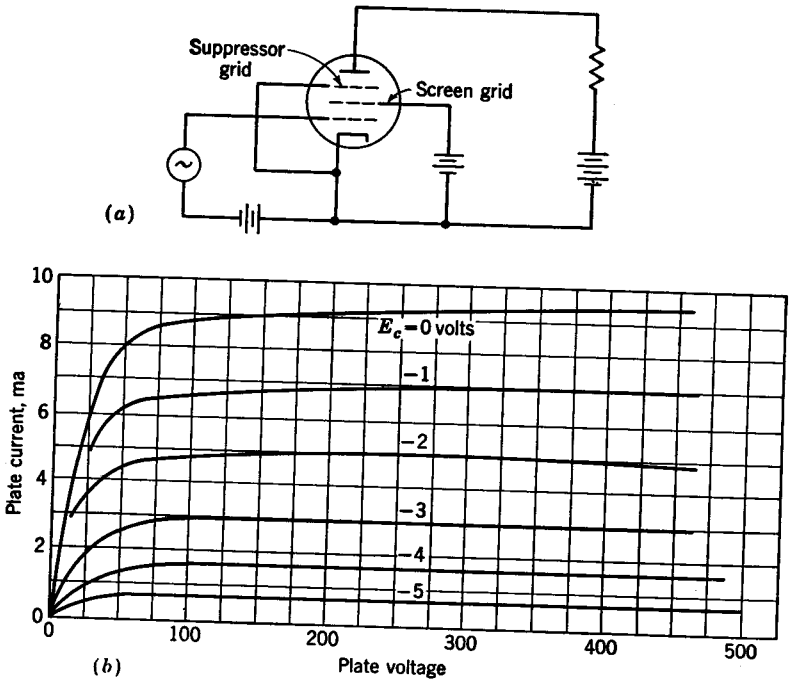


FIG. 13.15. (*a*) Pentode; (*b*) typical tube characteristics for a pentode.

Fig. 13.14*b*. When the plate-to-cathode voltage E_{pk} is less than the screen-grid voltage, the tube characteristics become quite erratic. Thus, tetrodes are usually operated in the range where the voltage E_{pk} is greater than the screen-grid voltage.

A pentode has a fifth element, the suppressor grid, in addition to the four elements of the tetrode. A pentode tube is shown in Fig. 13.15*a*, and a family of curves of typical tube characteristics is shown in Fig. 13.15*b*. Because of the suppressor grid the pentode does not have the irregular tube characteristics that were found in the tetrode when the voltage E_{pk} was less than the screen-grid voltage. The suppressor grid is located between the plate and the screen grid. Because the suppressor

grid is at cathode potential, it is negative with respect to the plate. Thus, it repels the flow of electrons from the plate to the screen grid even when the plate voltage is less than that of the screen grid.

The plate resistance r_p for a pentode is much greater than that for a triode, so that for a pentode r_p is generally much greater than R_L. Thus, for a pentode $r_p + R_L$ may be approximated by r_p, in which case from Eq. (13.50) it is seen that the amplifier gain K for a pentode is

$$K \approx \frac{\mu}{r_p} R_L \tag{13.51}$$

where $\mu/r_p = (\partial E_{pk}/\partial E_{gk})/(\partial E_{pk}/\partial I_{pk}) = -\partial I_{pk}/\partial E_{gk} = g_m$ is the change in plate current I_p per change in grid-to-cathode voltage E_{gk}, with E_{pk}

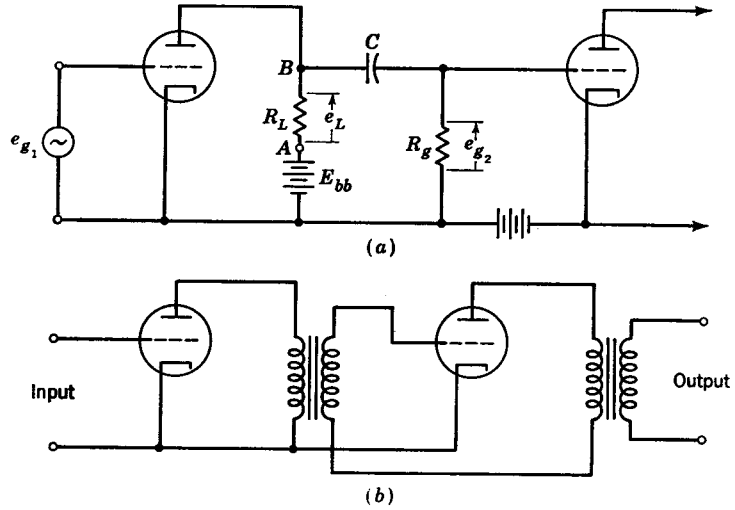


FIG. 13.16. (*a*) Capacitance-coupled amplifiers; (*b*) transformer-coupled amplifier.

held constant. This coefficient is called the mutual conductance, or the transconductance, of the tube. The pentode is capable of much greater amplification than is a triode.

Coupling. To obtain a greater amplification, it is customary to place amplifiers in series so that the voltage output from the first tube is the input to the second, etc. The over-all gain is the product of the gain of each amplifier. For a-c amplifiers, a capacitor C as shown in Fig. 13.16*a* is frequently used as the coupling. Only the a-c component e_L at the output of the first amplifier is transmitted through the capacitor to the second amplifier. For example, the voltage at point A is E_{bb}, and that at point B is $E_{bb} - e_L$. Because it is desired to have only e_L appear as the input signal to the second tube rather than e_L plus the d-c bias

E_{bb}, it is necessary to block the d-c component by means of the capacitor. For practical purposes the value of e_{g2} may be assumed equal to e_L if the resistor R_g is at least ten times greater than R_L and the reactance of the capacitor is one-tenth or less than the value of R_g.

Another method for isolating the a-c component e_L of the output tube is to use a transformer. Two amplifiers which are coupled through a transformer are shown in Fig. 13.16*b*. An advantage of transformer coupling is that the over-all gain per stage is increased by the turns ratio. However, flux leakage soon becomes excessive for a large turns ratio. The major disadvantage to transformer coupling is that at higher frequencies leakage reactance and capacitance between turns limit the operation. Resistance-capacitance-coupled amplifiers have a greater frequency range than transformer-coupled amplifiers.

When the input signal is a direct current or a very low frequency, the voltage e_L cannot be isolated by the preceding techniques. To amplify

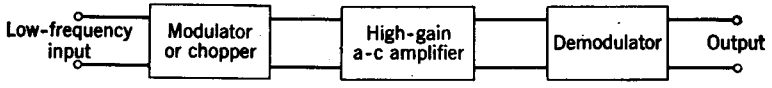


FIG. 13.17. Chopper-stabilized d-c amplifier.

d-c or very-low-frequency signals, it is customary to use chopper stabilization to achieve high performance. In Fig. 13.17 is shown a chopper-stabilized d-c amplifier. The modulator, or chopper, converts the input signal to a relatively high-frequency a-c signal, which may be amplified by a standard a-c amplifier. The demodulator than restores this amplified signal to its original d-c, or low-frequency, form.

Push-Pull Amplifier. A push-pull amplifier is in effect two tubes in parallel. When considerable a-c power is required, as is often the case with electrical control systems, a push-pull type of amplifier is capable of doubling the power output. A more important factor is that the nonlinearities of one tube tend to be compensated for by the second, with the result that distortion is minimized. This enables the push-pull amplifier to be operated over a wider range than would be possible with a single tube, thus permitting an even greater power gain. Often a push-pull amplifier is used as the last stage of a high-gain amplifier system in order to reduce distortion. In Fig. 13.18*a* is shown a push-pull amplifier. If there is no input voltage e_{in}, then the grid-to-cathode bias for each tube is $-E_{cc}$. Thus, each tube has the same initial equilibrium point of operation. An input voltage causes a signal voltage e_{s_1} of the polarity indicated to appear on the grid of tube 1 so that the resultant grid-to-cathode voltage is $E_{gk_1} = e_{s_1} - E_{cc}$. This causes a current i_{p_1} to flow in the direction indicated. The signal voltage $e_{s_2} = -e_{s_1}$

applied to tube 2 is of the opposite polarity. Thus the current i_{p_2} flows through the second tube in the opposite direction to that in the first. However, the directions of the currents flowing through the output transformer are additive. In Fig. 13.18*b* is shown the equivalent circuit for a push-pull amplifier, in which $i_p = i_{p_1} = i_{p_2}$. Because the current i_{p_2} flowing through the power supply E_{bb} is opposite to i_{p_1}, there is no net a-c

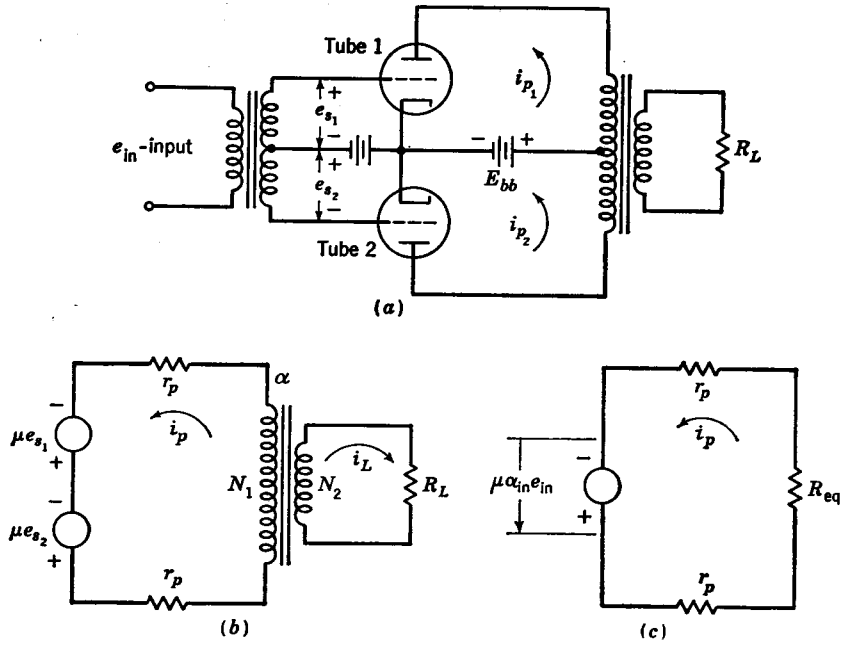


FIG. 13.18. (*a*) Push-pull amplifier; (*b*) equivalent circuit for a push-pull amplifier; (*c*) simplified equivalent circuit for push-pull amplifier.

current flowing through E_{bb}. Thus, the E_{bb} line is omitted in the equivalent-circuit representation.

For an ideal transformer, the sum of the magnetomotive forces (mmfs) around the core is zero. Thus for the output transformer it follows that

$$i_p N_1 - i_L N_2 = 0$$

or

$$i_L = \frac{N_1}{N_2} i_p = \frac{i_p}{\alpha} \tag{13.52}$$

where α is the ratio of the total secondary turns N_2 to the total primary turns N_1 (that is, $\alpha = N_2/N_1$). The power delivered to the load resistor is

$$P = i_L^2 R_L = i_p^2 R_{eq} \tag{13.53}$$

Thus, the equivalent resistance R_{eq}

$$R_{eq} = \left(\frac{i_L}{i_p}\right)^2 R_L = \frac{R_L}{\alpha^2} \tag{13.54}$$

Similarly, if it is assumed that the input transformer is also ideal,

$$e_{s_1} = \frac{\alpha_{in}}{2} e_{in} \qquad \text{and} \qquad e_{s_2} = \frac{\alpha_{in}}{2} e_{in} \tag{13.55}$$

Thus, the voltage signal in Fig. 13.18*b* may be written as

$$\mu(e_{s_1} + e_{s_2}) = \mu\alpha_{in}e_{in} \tag{13.56}$$

The resulting equivalent a-c circuit is shown in Fig. 13.18*c*.

13.6. Transistor Amplifier. In many respects transistors are superior to vacuum-tube amplifiers, as is evidenced by their increasing use in electronic controls. An obvious advantage is their small size. In addition, they require much less power than vacuum tubes. For example, low-level transistor amplifier stages operate with collector voltages of only a few volts, whereas vacuum-tube plate voltages are usually several hundred volts. Another source of power savings is that transistors have no heater element such as is required for the cathode of vacuum tubes. A strong feature is that the life of a well-designed transistor is about 50,000 hr, as compared with 2,000 hr for a vacuum tube.

A disadvantage of transistors is that their characteristics are sensitive to temperature variations. Also, there is a rather large variation in characteristics of successive transistor units as they come from the production line. However, these disadvantages are continually being minimized as manufacturing techniques are improved.

The three major parts of a junction transistor are the emitter, collector, and base. A junction transistor is, in effect, a sandwich of three sections of semiconductor crystal. The section in the center is the base, while the outer sections are the emitter and collector, respectively.[1–3]

As is shown in Fig. 13.19*a*, a transistor is represented by a circle with a dash in it. The base is indicated by the line which is perpendicular to the dash. The emitter is the line with the arrowhead pointing toward the dash. The collector is represented by the line which is symmetrical to the emitter, but the collector does not have an arrowhead.

In Fig. 13.19*a* is shown a transistor circuit in which the base is grounded. The current I_e is the current which flows from the emitter

[1] Gibson and Tuteur, *op. cit.*

[2] L. P. Hunter, "Handbook of Semiconductor Electronics," McGraw-Hill Book Company, Inc., New York, 1956.

[3] R. F. Shea, "Transistor Circuit Engineering," John Wiley & Sons, Inc., New York, 1957.

through the base, and the current I_c flows from the collector through the base. The emitter voltage with respect to ground is V_1, and the collector voltage is V_2. As is indicated in Fig. 13.19*a*, the emitter current I_e is a function of the emitter-to-ground voltage V_1 and the current I_c flowing through the collector, thus,

$$I_e = F(V_1, I_c) \tag{13.57}$$

For small variations, the preceding expression becomes

$$i_e = \frac{\partial I_e}{\partial V_1}\bigg|_i v_1 + \frac{\partial I_e}{\partial I_c}\bigg|_i i_c \tag{13.58}$$

The equivalent-circuit representation for the preceding expression in which v_1 is the applied voltage is obtained by first solving for v_1.

$$v_1 = \frac{\partial V_1}{\partial I_e}\bigg|_i i_e - \frac{\partial V_1}{\partial I_e}\bigg|_i \frac{\partial I_e}{\partial I_c}\bigg|_i i_c \tag{13.59}$$

By writing Eq. (13.57) in the implicit form $G(I_e, V_1, I_c) = 0$ and utilizing

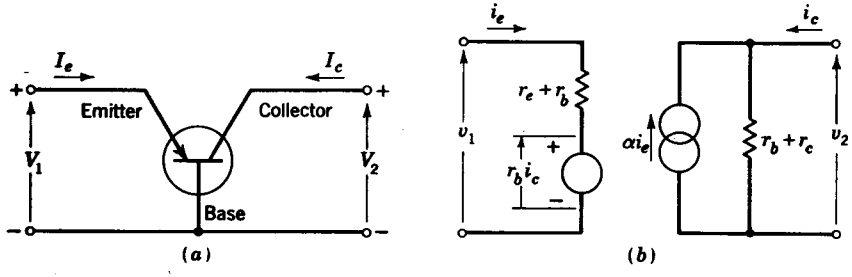


FIG. 13.19. (*a*) Grounded-base connection; (*b*) equivalent circuit for grounded-base connection.

the fact that the product of the partial derivatives for an implicit function is equal to -1, it follows that

$$\frac{\partial I_e}{\partial V_1}\frac{\partial V_1}{\partial I_c}\frac{\partial I_c}{\partial I_e} = -1$$

or

$$\frac{\partial V_1}{\partial I_c} = -\frac{\partial V_1}{\partial I_e}\frac{\partial I_e}{\partial I_c} \tag{13.60}$$

Substitution of the preceding result into Eq. (13.59) gives

$$v_1 = \frac{\partial V_1}{\partial I_e}\bigg|_i i_e + \frac{\partial V_1}{\partial I_c}\bigg|_i i_c = (r_b + r_e)i_e + r_b i_c \tag{13.61}$$

where

$$r_b + r_e = \frac{\partial V_1}{\partial I_e}\bigg|_i \quad \text{and} \quad r_b = \frac{\partial V_1}{\partial I_c}\bigg|_i \tag{13.62}$$

The equivalent-circuit representation for Eq. (13.61) is shown on the left side of Fig. 13.19*b*. The resistances r_b and r_e do not exist physically

in the transistor but are defined by Eq. (13.62). The dynamic resistance $r_b + r_e$ is, in effect, the ratio of the change in voltage from the emitter to the base per change in current. The collector current does not flow through the emitter, so that the change in voltage per change in collector current is designated as r_b.

From Fig. 13.19*a*, it follows that the current I_c is a function of the voltage drop V_2, from the collector to the base, and the emitter current I_e, that is,

$$I_c = F(V_2, I_e) \tag{13.63}$$

Linearizing and solving for v_2 gives

$$v_2 = \frac{\partial V_2}{\partial I_c}\bigg|_i i_c - \frac{\partial V_2}{\partial I_c}\bigg|_i \frac{\partial I_c}{\partial I_e}\bigg|_i i_e = (r_b + r_c)i_c + (r_b + r_c)\alpha i_e \tag{13.64}$$

where $$r_b + r_c = \frac{\partial V_2}{\partial I_c}\bigg|_i \quad \text{and} \quad \alpha = -\frac{\partial I_c}{\partial I_e}\bigg|_i \tag{13.65}$$

The reason for the minus sign in the preceding equation for α is that, with the collector voltage V_2 held constant, I_c decreases as I_e increases.

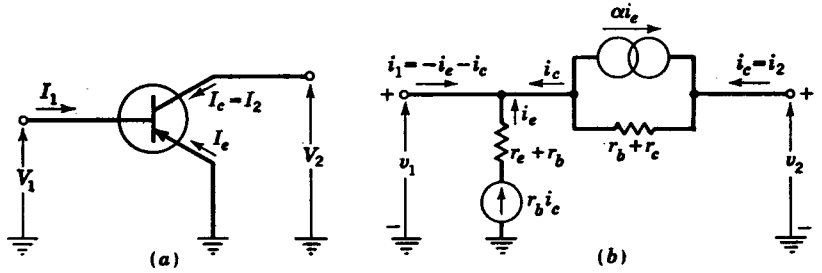


FIG. 13.20. (*a*) Grounded-emitter connection; (*b*) equivalent circuit for grounded-emitter connection.

The equivalent circuit for the collector as described by Eq. (13.64) is shown in the right portion of Fig. 13.19*b*. The two interlocking circles represent a constant-current source of αi_e. The term α is a very important transistor parameter. For junction transistors, the value of α is usually slightly less than unity. Typical values of α range from 0.95 to 0.98. For a grounded-base transistor, it follows from Eq. (13.65) that the current amplification ratio I_c/I_e can never exceed unity because α is always less than 1.

To achieve a high amplification ratio, the grounded-emitter transistor connection shown in Fig. 13.20*a* is commonly used. The input current is I_1, and the output current I_2 is also the collector current I_c. The potential of the collector, or output voltage, is V_2 volts. As is indicated in Fig. 13.20*a*, the collector current I_c is a function of the voltage drop

$V_2 - V_1$ from the collector to the base, and also of the current I_e which flows through the emitter.

$$I_c = F(V_2 - V_1, I_e) \tag{13.66}$$

Linearization yields

$$i_c = \left.\frac{\partial I_c}{\partial (V_2 - V_1)}\right|_i (v_2 - v_1) + \left.\frac{\partial I_c}{\partial I_e}\right|_i i_e \tag{13.67}$$

Solving for the voltage drop $v_2 - v_1$ gives

$$v_2 - v_1 = \left.\frac{\partial (V_2 - V_1)}{\partial I_c}\right|_i i_c - \left.\frac{\partial (V_2 - V_1)}{\partial I_c}\right|_i \left.\frac{\partial I_c}{\partial I_e}\right|_i i_e \tag{13.68}$$

$$v_2 - v_1 = (r_b + r_c)(i_c + \alpha i_e) \tag{13.69}$$

where

$$r_b + r_c = \left.\frac{\partial (V_2 - V_1)}{\partial I_c}\right|_i \quad \text{and} \quad \alpha = -\left.\frac{\partial I_c}{\partial I_e}\right|_i \tag{13.70}$$

As is indicated in Fig. 13.20*a*, the emitter current I_e is a function of the voltage drop from the emitter to the base, $-V_1$, and also of the collector current I_c. Thus

$$I_e = F(V_1, I_c) \tag{13.71}$$

In the preceding general functional relationship for I_e, it is not necessary to introduce the minus sign in front of V_1. The correct sign in the final linearized expression is automatically accounted for in the evaluation of the partial derivatives. The resultant linearized expression for v_1 is

$$v_1 = \left.\frac{\partial V_1}{\partial I_c}\right|_i i_c + \left.\frac{\partial V_1}{\partial I_e}\right|_i i_e = -r_b i_c - (r_b + r_e) i_e \tag{13.72}$$

where

$$-r_b = \left.\frac{\partial V_1}{\partial I_c}\right|_i \quad \text{and} \quad -(r_b + r_e) = \left.\frac{\partial V_1}{\partial I_e}\right|_i \tag{13.73}$$

The reason for the minus signs in Eq. (13.73) is that v_1 decreases as either i_c or i_e increases. Equations (13.69) and (13.72) form the basis for the equivalent grounded-emitter circuit shown in Fig. 13.20*b*.

The much simplified equivalent circuit of Fig. 13.21*a* is obtained in the following manner: From Fig. 13.20*a*, it is to be noted that

$$I_1 = -(I_e + I_c) \tag{13.74}$$

or

$$i_1 = -(i_e + i_c) \tag{13.75}$$

It should also be noted that $I_c = I_2$, and thus

$$i_c = i_2 \tag{13.76}$$

h Parameters for the Grounded-emitter Connection. By rearranging Eqs. (13.69) and (13.72) so that v_1 and i_2 are the dependent variables and i_1 and v_2 are the independent variables, that is,

$$v_1 = h_{ie} i_1 + h_{re} v_2 \tag{13.77}$$

$$i_2 = h_{fe} i_1 + h_{oe} v_2 \tag{13.78}$$

The simplified equivalent circuit of Fig. 13.21*a* follows directly from Eqs. (13.77) and (13.78). The values of the constant-h parameters are obtained as follows: Substituting i_e from Eq. (13.75) into Eq. (13.72) and noting that $i_o = i_2$ yields

$$v_1 = -r_b i_2 + (r_b + r_e)(i_1 + i_2) = (r_b + r_e)i_1 + r_e i_2 \qquad (13.79)$$

Similarly, substituting v_1 from Eq. (13.79) into Eq. (13.69) and noting that $i_e = -(i_1 + i_2)$ yields

$$v_2 = [r_b + r_e - \alpha(r_b + r_c)]i_1 + [r_b + r_c + r_e - \alpha(r_b + r_c)]i_2 \qquad (13.80)$$

The elimination of i_2 from Eqs. (13.79) and (13.80) yields Eq. (13.77), in which it is seen that

$$h_{ie} = r_e + r_b - \frac{r_e[r_e + r_b - \alpha(r_b + r_c)]}{r_b + r_c + r_e - \alpha(r_b + r_c)} \approx \frac{r_e}{1 - \alpha} + r_b \qquad (13.81)$$

and

$$h_{re} = \frac{r_e}{r_b + r_c + r_e - \alpha(r_b + r_c)} \approx \frac{r_e}{r_c(1 - \alpha)} \qquad (13.82)$$

In making these approximations, use is made of the fact that r_c is much larger than r_b or r_e. The h-parameter subscripts i, r, f, and o refer to

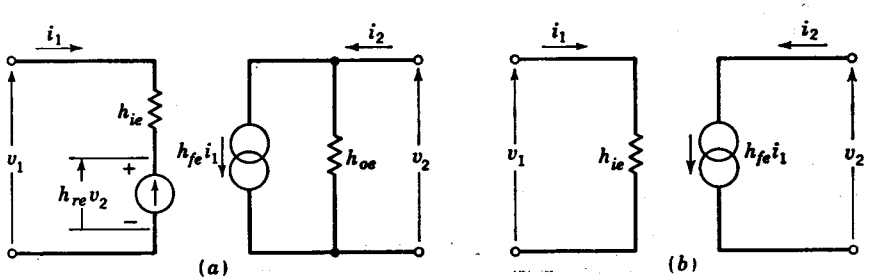


FIG. 13.21. (*a*) Grounded emitter in terms of h parameters; (*b*) approximate circuit for grounded emitter in terms of h parameters.

input, reverse, forward, and output, respectively, while the second subscript, e, indicates that these parameters are defined for the grounded-emitter transistor. A second subscript e could also be added to the voltage and current terms of Eqs. (13.77) and (13.78).

Because Eq. (13.80) is in the same general form as Eq. (13.78), all that is necessary is to solve Eq. (13.80) for i_2 and compare coefficients of like terms to obtain directly

$$h_{fe} = -\frac{r_b + r_e - \alpha(r_b + r_c)}{r_b + r_c + r_e - \alpha(r_b + r_c)} \approx \frac{\alpha}{1 - \alpha} \qquad (13.83)$$

$$h_{oe} = \frac{1}{r_b + r_c + r_e - \alpha(r_b + r_c)} \approx \frac{1}{r_c(1 - \alpha)} \qquad (13.84)$$

A further simplification in the equivalent circuit shown in Fig. 13.21*a* is obtained by utilizing the fact that h_{ie} is generally much greater than $h_{re}v_2$. Also, the load resistance R_L is usually much less than h_{oe} so that h_{oe} is effectively shunted and may thus be neglected. The resulting equivalent circuit for a grounded-emitter transistor amplifier is shown in Fig. 13.21*b*. The input circuit and output circuit are seen to be decoupled. For the output circuit, it is seen that

$$i_2 = h_{fe}i_1$$

or

$$\frac{i_2}{i_1} = h_{fe} \tag{13.85}$$

Thus, the current gain is simply h_{fe}.

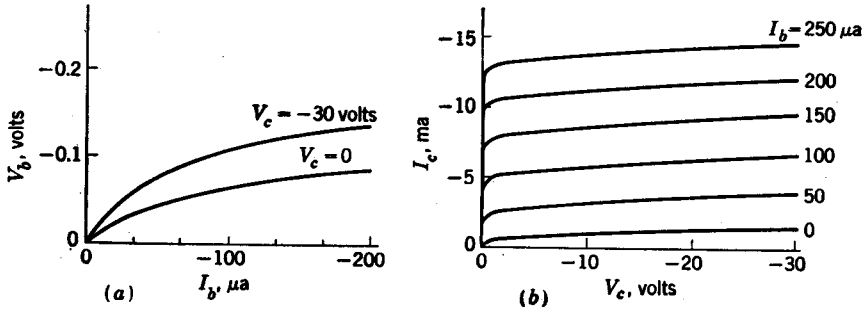


FIG. 13.22. Characteristic curves for grounded-emitter connection.

The voltage gain is obtained by substituting $i_2 = -v_2/R_L$ into Eq. (13.78) and solving Eqs. (13.77) and (13.78) simultaneously. Thus

$$\frac{v_2}{v_1} = \frac{-h_{fe}(R_L/h_{ie})}{1 + R_L h_{oe}(1 - h_{fe}h_{re}/h_{oe}h_{ie})} \tag{13.86}$$

For most amplifiers $R_L h_{oe} \ll 1$, and $0 < (h_{fe}h_{re}/h_{oe}h_{ie}) < 1$, so that the voltage amplification is approximately

$$\frac{v_2}{v_1} \approx -h_{fe}\frac{R_L}{h_{ie}} \tag{13.87}$$

The resulting power amplification is the product of the current and voltage amplifications.

An important advantage of the *h*-parameter designation is that the values of *h* may be obtained directly from characteristic curves which are usually supplied by the manufacturer. From Fig. 13.23*a*, it is to be seen that for a grounded-emitter connection the base is the input and the collector the output, so that the subscripts 1 and *b* may be used interchangeably, as well as the subscripts 2 and *c*. A typical set of curves for

a grounded-emitter connection is shown in Fig. 13.22. The geometric interpretation of the h parameters from Fig. 13.22a and b gives, respectively

$$\begin{aligned} V_1 &= F(I_1, V_2) \\ I_2 &= F(I_1, V_2) \end{aligned} \tag{13.88}$$

Linearization of the preceding equations yields

$$v_1 = \left.\frac{\partial V_1}{\partial I_1}\right|_i i_1 + \left.\frac{\partial V_1}{\partial V_2}\right|_i v_2 \tag{13.89}$$

$$i_2 = \left.\frac{\partial I_2}{\partial I_1}\right|_i i_1 + \left.\frac{\partial I_2}{\partial V_2}\right|_i v_2 \tag{13.90}$$

Equating coefficients of like terms of Eqs. (13.89) and (13.77) and also coefficients of Eqs. (13.90) and (13.78) yields

$$\begin{aligned} h_{ie} &= \left.\frac{\partial V_1}{\partial I_1}\right|_i = \left.\frac{\partial V_b}{\partial I_b}\right|_i & h_{re} &= \left.\frac{\partial V_1}{\partial V_2}\right|_i = \left.\frac{\partial V_b}{\partial V_c}\right|_i \\ h_{fe} &= \left.\frac{\partial I_2}{\partial I_1}\right|_i = \left.\frac{\partial I_c}{\partial I_b}\right|_i & h_{oe} &= \left.\frac{\partial I_2}{\partial V_2}\right|_i = \left.\frac{\partial I_c}{\partial V_c}\right|_i \end{aligned} \tag{13.91}$$

The problem of coupling transistor amplifiers is usually much simpler than that for vacuum-tube amplifiers. For example, it is possible to

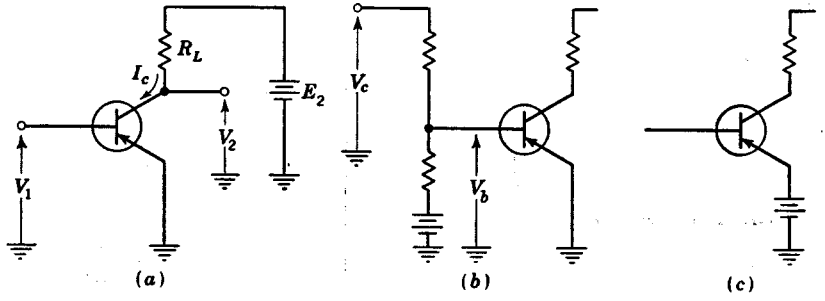


FIG. 13.23. (a) Grounded-emitter amplifier with a resistive load; (b) voltage divider; (c) biased emitter.

operate the base of a second stage directly from the collector voltage V_c of the preceding stage, etc. If the quiescent collector voltage V_{co} of the first stage is different from the desired value V_{bo} for the second stage, V_{co} may be run through a biased voltage divider as shown in Fig. 13.23b to achieve the proper V_{bo}. Another technique is to bias the emitter as illustrated in Fig. 13.23c. A biased emitter in effect decreases V_b and V_c for the particular transistor by the amount of the biasing voltage.

The direction of current flow and the signs for the various voltages used in the preceding discussion is appropriate for the p-n-p type of transistors. In a p-n-p transistor, the current is carried through the

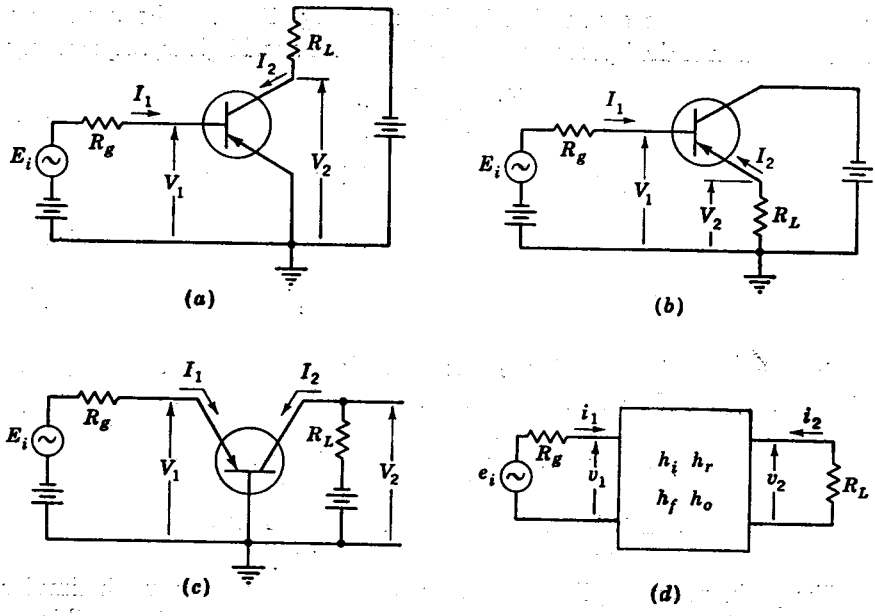


FIG. 13.24. (*a*) Grounded-emitter connection; (*b*) grounded-collector connection; (*c*) grounded-base connection; (*d*) equivalent four-terminal network.

TABLE 13.1

Circuit	G_v	G_i	Approximation
Grounded emitter	$-h_{fe}\dfrac{R_L}{h_{ie}}$	h_{fe}	$R_L h_{oe} \ll 1$ $\dfrac{R_g}{h_{ie}} \gg 1$ $0 < \dfrac{h_{fe}h_{re}}{h_{ie}h_{oe}} < 1$
Grounded collector	1	$-(1 + h_{fe})$	$R_L h_{oe} \ll 1$ $\dfrac{R_g}{h_{ie}} \gg 1$ $\dfrac{(1 + h_{fe})R_L}{h_{ie}} \gg 1$ $\dfrac{1 + h_{fe}}{R_g h_{oe}} \gg 1$
Grounded base	$h_{fe}\dfrac{R_L}{h_{ie}}$	$-\dfrac{h_{fe}}{1 + h_{fe}}$	$R_L h_{oe} \ll 1$ $\dfrac{R_g}{h_{ie}} \gg 1$ $0 < \dfrac{h_{fe}h_{re}}{h_{ie}h_{oe}} < 1$

emitter and the collector by positive carriers called holes, while the current in the base is transmitted by means of negative carriers, or electrons. For the other transistor type, the *n-p-n* transistor, the current is carried in the emitter and the collector by negative carriers, while the base has positive carriers. To convert from a *p-n-p* to a *n-p-n* transistor, all that is necessary is to reverse the polarity of the battery voltages.

h Parameters for the Grounded-collector and the Grounded-base Connections. Figure 13.24*a*, *b*, and *c* show the grounded-emitter, grounded-collector, and grounded-base transistor circuits, respectively. Each of these circuits may be represented by the equivalent four-terminal network shown in Fig. 13.24*d*. Table 13.1 gives the equations for the voltage gain $G_v = v_2/v_1$ and the current gain $G_i = i_2/i_1$ for each of these circuits.

By comparison of Fig. 13.24*a* and *b*, it follows that the voltage drop from the base to the emitter is

$$V_{1e} = V_{1c} - V_{2c}$$

and thus

$$v_{1e} = v_{1c} - v_{2c} \tag{13.92}$$

The second subscript refers to the type of circuit, that is, *e* for grounded emitter and *c* for grounded collector. Similarly, corresponding values of the voltage drop from the collector to the emitter, the current flowing through the base, and the current flowing through the collector are, respectively,

$$\begin{aligned} v_{2e} &= -v_{2c} \\ i_{1e} &= i_{1c} \\ i_{2e} &= -i_{1c} - i_{2c} \end{aligned} \tag{13.93}$$

Substitution of these results into Eqs. (13.77) and (13.78) gives

$$\begin{aligned} v_{1c} &= h_{ie}i_{1c} + (1 - h_{re})v_{2c} \\ i_{2c} &= -(1 + h_{fe})i_{1c} + h_{oe}v_{2c} \end{aligned} \tag{13.94}$$

Thus, the *h* parameters for a grounded-collector circuit are obtained in terms of the *h* parameters for a grounded-emitter circuit, that is,

$$\begin{aligned} h_{ic} &= h_{ie} \\ h_{rc} &= 1 - h_{re} \\ h_{fc} &= -(1 + h_{fe}) \\ h_{oc} &= h_{oe} \end{aligned} \tag{13.95}$$

Similarly, the characteristics of a grounded-base circuit may be expressed as a function of the *h* parameters for a grounded-emitter circuit. An advantage of using *h* parameters is that the characteristics of the grounded-emitter circuit (which are usually supplied by the manufacturer) may be used to evaluate the performance of the grounded-collector or grounded-base connections.

CHAPTER 14

INERTIAL GUIDANCE

Gyroscopes are basically the position-measuring or -indicating devices for inertial-guidance systems. The ability of a controlled vehicle to maintain its desired path depends upon the accuracy with which the gyroscopes indicate its position.

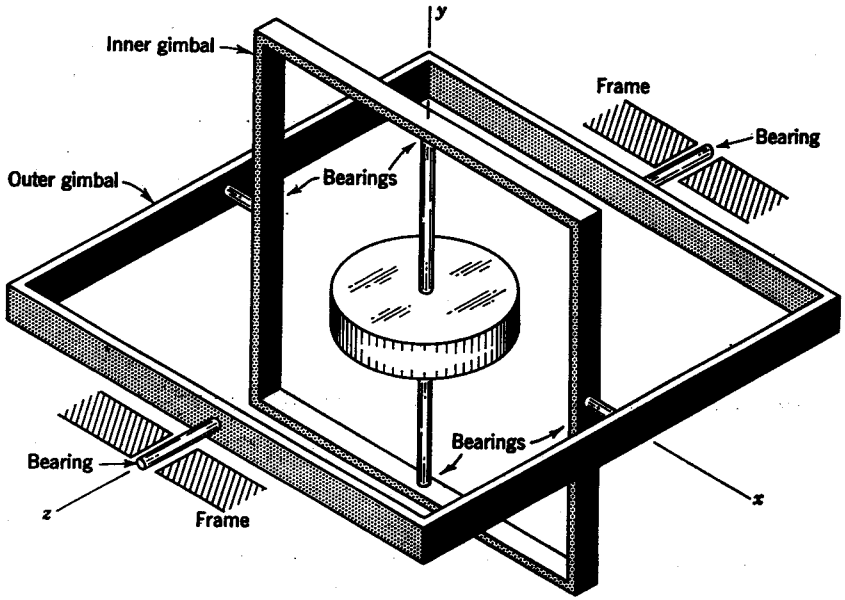


FIG. 14.1. Free gyroscope.

As is shown in Fig. 14.1, a gyroscope consists of a spinning disk (i.e., gyrowheel) which is supported by linkages called gimbals. The axis of rotation of the gyrowheel is referred to as the rotor spin axis. This is the y axis in Fig. 14.1.

14.1. Free Gyros. A basic characteristic of gyroscopes is that the rotor spin axis always points in the same direction unless it is acted upon by an external torque. The term "free gyro" is used to describe a gyrowheel which is suspended in such a manner that external torques cannot

be transmitted to it. The gimbals for the free gyro shown in Fig. 14.1 are supported by bearings so that, regardless of the motion of the frame, the direction of the spin axis tends to remain fixed in space.

This tendency for the rotor spin axis to maintain a fixed direction in inertial space is illustrated in Fig. 14.2. Initially, the spin axis of this gyroscope is perpendicular to the earth, as is shown in Fig. 14.2*a*. If the vehicle in which the gyroscope is mounted is not moving relative to the earth (i.e., remains at point A, which is fixed), in 3 hr the earth turns one-eighth its daily revolution, or 45°. Thus, in 3 hr, the angle of inclination of the spin axis is 45° with respect to the earth, as shown in Fig. 14.2*b*. The gyroscope is still located at point A, and its rotor spin axis has not

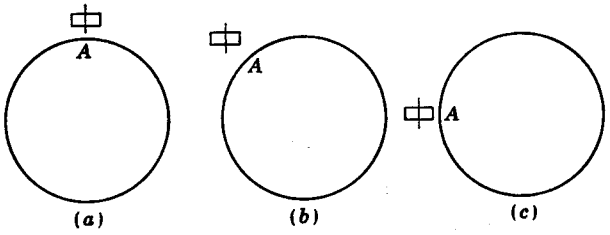


FIG. 14.2. Relative motion of the earth with respect to the rotor spin axis.

changed direction in inertial space. The angle of inclination is due to the earth's rotation in space. After 6 hr, the spin axis is tangent to the earth, as shown in Fig. 14.2*c*.

In a practical gyro, the spin axis tends to drift slowly from its initial direction because of torque acting on the gyrowheel due to friction in the gimbal bearings. This phenomenon is referred to as drift. It should be noted that the angle of inclination of the spin axis with respect to the earth depends not only on the rotation of the earth and drift but also on the movement relative to the earth of the vehicle in which the gyroscope is mounted.

14.2. Vertical and Directional Gyros. A vertical gyro is one in which the spin axis is aligned with the gravitational field of the earth (i.e., perpendicular to the earth's surface). To compensate for the effects of drift, earth's rotation, and motion of the gyroscope relative to the earth, it is necessary to reset the reference position of the spin axis. This is generally accomplished by applying an external torque to realign the gyrowheel to the desired reference position. An external torque can be applied to the gyrowheel shown in Fig. 14.1 by extending the gimbal axes and connecting torque motors to these shaft extensions as shown in Figs. 14.4 and 14.5, rather than letting them rotate freely in bearings. The rotor of the torque motor is wound around the shaft extension and is free to rotate inside its stator, which is fixed to the

adjoining gimbal or frame. An electrical signal applied to the torque motor causes a torque to be exerted on the shaft extension.

A gyro used to establish a vertical reference can automatically be controlled by, or slaved to, a more accurate primary reference such as the average position of a pendulum. This is accomplished by comparing the direction of the gyro spin axis with the direction of the pendulum and using the resulting error signal to make the gyro position correspond to that of the pendulum. By having a small torque applied to the gyro so that the spin axis cannot follow rapid fluctuations of the pendulum,

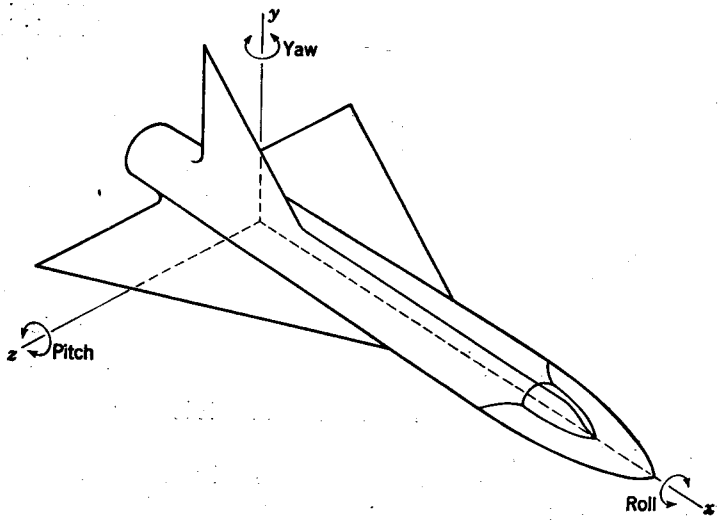


FIG. 14.3. Pitch, roll, and yaw motions of a vehicle.

the gyro spin axis tends to remain aligned to the average pendulum position (i.e., the vertical).

A directional gyroscope is one in which the spin axis lies in a horizontal plane and holds a given direction, which is usually north-south. A directional gyroscope may be slaved to a magnetic compass, which provides the primary north-south reference.

By maintaining the spin axis in some reference direction, then it is possible to measure the angular displacement of the frame (which is the angular displacement of the vehicle to which the frame is mounted). Gyroscopes which are used to measure such displacements of a vehicle from some reference direction, as a vertical or a north-south orientation, are frequently called displacement gyroscopes.

A vertical gyroscope can measure pitch and roll, while a directional gyroscope can measure yaw (azimuth). From Fig. 14.3 it is to be noted that pitch is angular motion about the lateral, or z, axis, roll is angular

motion about the longitudinal, or x, axis, and yaw is angular motion about the normal, or y, axis. The frame and the pitch and roll scales of the vertical gyroscope shown in Fig. 14.4 are fastened to the vehicle. A rotation of the vehicle about the x, y, or z axis causes a corresponding rotation of the frame and the attached pitch and roll scales. For example, a rotation of the vehicle about the z axis causes a corresponding rotation of the pitch scale, but the outer gimbal and attached pointer cannot rotate about the z axis, because the spin axis is maintained in a

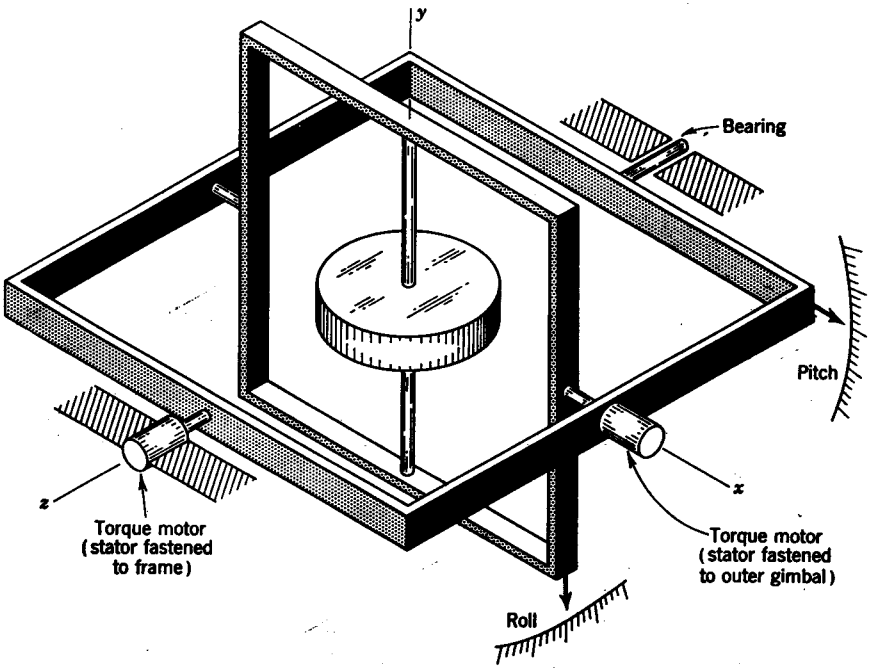


FIG. 14.4. Vertical gyroscope.

vertical position. Thus the relative motion between the pitch scale, which rotates with the vehicle, and the pointer on the outer gimbal is a measure of the angle of inclination, or pitch, of the vehicle. The pointer and scale may be replaced by an electrical pick-off in order to obtain a voltage signal proportional to the pitch angle. This change in pitch also causes the roll scale to rotate about the z axis, while the inner gimbal remains fixed. However, there is no change in the scale reading, because this motion is perpendicular to the pointer on the inner gimbal. For a rotation about the x axis the roll scale rotates relative to the pointer on the inner gimbal, which maintains its vertical inclination. This relative motion is a measure of the roll. The pitch reading is unaffected,

because both the outer gimbal and the pitch scale rotate the same amount, and thus there is no relative motion. Similarly, a motion of the vehicle about the y axis (i.e., yaw) has no effect upon either the roll or the pitch readings.

For the directional gyroscope shown in Fig. 14.5, the rotor spin axis maintains a fixed horizontal direction (usually a north-south direction).

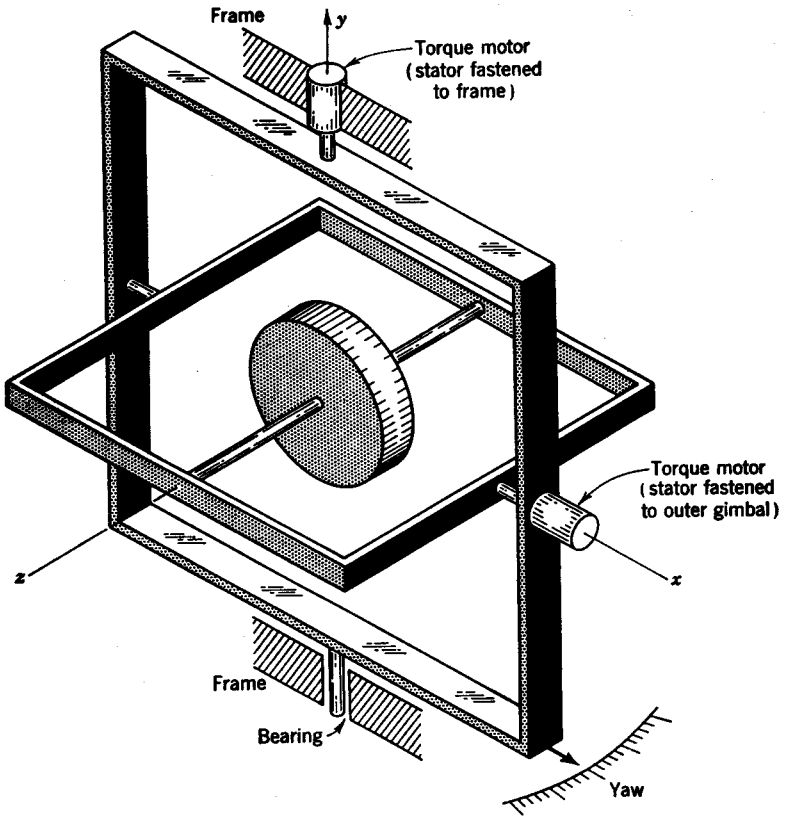


FIG. 14.5. Directional gyroscope.

A rotation of the vehicle about the y axis is seen to cause a corresponding rotation of the attached yaw scale. However, the outer gimbal and attached pointer cannot rotate, because the spin axis is maintained in its reference direction. This relative motion is thus a measure of the direction of heading, or yaw. Other motions of the vehicle have no effect upon the measured angle of yaw.

Vertical and directional gyros are used in automatic pilots, roll-stabilizing equipment, artificial horizons, inertial-navigation equipment, etc. To obtain a better understanding of the operation of other types of

gyroscopes which play an essential role in inertial navigation, some attention is first given to the dynamics of gyroscopes.

14.3. Dynamics of Gyroscopes. A gyrowheel is essentially a spinning wheel, as is shown in Fig. 14.6. Frequently, the gyrowheel is the rotor of a synchronous motor which rotates the rotor at a constant angular velocity. The disk, or rotor, rotates about the x axis (i.e., its spin axis) with an angular velocity ω_s. In accordance with standard procedure, the vector representing the angular velocity ω_s is in the direction in which a right-hand thread would move if the thread were rotating in the same direction as the gyrowheel. Thus, the vector representing ω_s is in the positive direction of the x axis. If a positive torque $\mathbf{T}_z$ is applied about the z axis as shown in Fig. 14.6, from Newton's law this torque

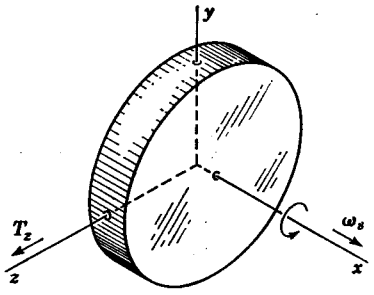


FIG. 14.6. Free-body diagram of a gyrowheel.

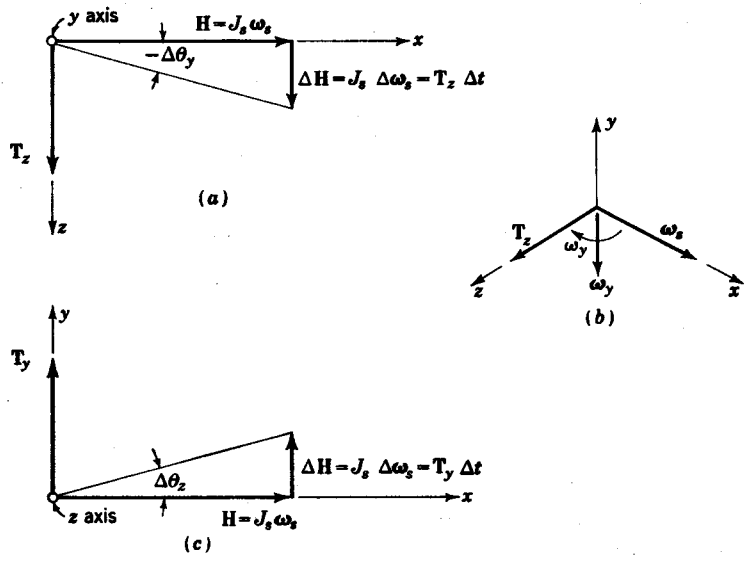


FIG. 14.7. (*a*) Gyroscopic precession caused by T_z; (*b*) precession velocity; (*c*) gyroscopic precession caused by T_y.

is equal to the time rate of angular momentum. That is,

$$\mathbf{T}_z = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{H}}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Delta(J_s \omega_s)}{\Delta t} = \lim_{\Delta t \to 0} \frac{J_s \Delta \omega_s}{\Delta t} \tag{14.1}$$

where J_s is the polar moment of inertia of the gyrowheel about its spin axis and $\mathbf{H} = J_s\boldsymbol{\omega}_s$ is the angular momentum of the disk. Because J_s and Δt are scalar quantities, $\Delta\boldsymbol{\omega}_s$ is in the same direction as $\mathbf{T}_x$ and $\Delta\mathbf{H}$, as shown in Fig. 14.7*a*. In this figure, the y axis is perpendicular to the plane of the paper and coming out from the paper. From the geometry of Fig. 14.7*a*, it follows that

$$\tan(-\Delta\theta_y) = \frac{T_x\,\Delta t}{J_s\omega_s} \tag{14.2}$$

Because $\tan(-\Delta\theta_y) \approx -\Delta\theta_y$,

$$\omega_y = \lim_{\Delta t\to 0}\frac{\Delta\theta_y}{\Delta t} = -\frac{T_x}{J_s\omega_s} \tag{14.3}$$

The angular velocity imparted to a gyrowheel because of an external torque is referred to as the velocity of precession. The vector representing the velocity of precession is perpendicular to the plane determined by the spin vector and the torque vector. In particular, the direction of rotation of the precession velocity is the same as that required to rotate the spin vector $\boldsymbol{\omega}_s$ into the torque vector $\mathbf{T}_x$, as is shown in Fig. 14.7*b*.

For a positive torque T_y applied about the y axis, the resultant vector diagram is as shown in Fig. 14.7*c*. From the geometry of this figure, it follows that

$$\tan(\Delta\theta_z) = \frac{T_y\,\Delta t}{J_s\omega_s} \tag{14.4}$$

and therefore

$$\omega_z = \frac{d\theta_z}{dt} = \frac{T_y}{J_s\omega_s} \tag{14.5}$$

where ω_z is the velocity of precession about the z axis. Again, it is to be noted that the gyro tends to precess so as to line up the spin axis with the direction of the torque vector.

In a practical gyroscope, torque is not applied directly to the gyrowheel but rather is applied to the gimbals, usually by means of either the y-axis torque motor or the z-axis torque motor. Thus, Eqs. (14.3) and (14.5) must be modified to include the effects of the gyro rest inertia and viscous friction of the gimbal bearings. That is, in addition to causing the precession, a torque T_y applied by the y-axis torque motor is utilized in accelerating the rest inertia $J_y(d^2\theta_y/dt^2)$ of the gyro about the y axis and in overcoming the viscous friction of the y-axis gimbal bearings, $C_y(d\theta_y/dt)$. Thus

$$T_y = J_y\frac{d^2\theta_y}{dt^2} + C_y\frac{d\theta_y}{dt} + J_s\omega_s\frac{d\theta_z}{dt} \tag{14.6}$$

Similarly, it follows that the torque T_z applied by the z-axis torque motor is

$$T_z = J_z\frac{d^2\theta_z}{dt^2} + C_z\frac{d\theta_z}{dt} - J_s\omega_s\frac{d\theta_y}{dt} \tag{14.7}$$

where J_y and J_z are the rest inertias of the gyro and gimbals about the y and z axes, respectively. Similarly, C_y and C_z are the coefficients of viscous damping for rotation about the y axis and the z axis, respectively.

14.4. Restrained Gyro. A restrained gyro is one which has constraints such as springs or dampers attached to the gimbals in such a manner that a motion of the gimbal tends to precess the gyro.

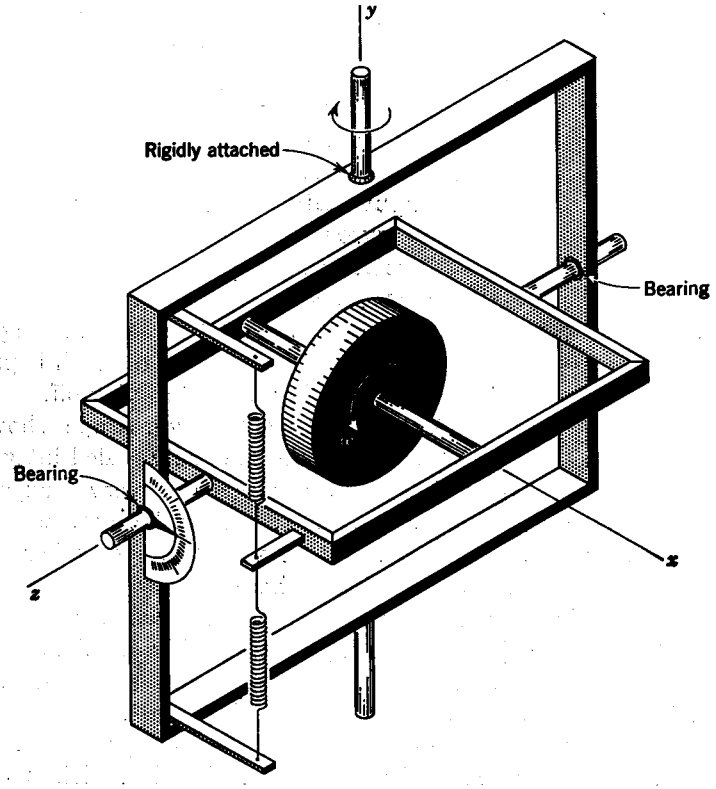


FIG. 14.8. Rate gyro.

For the gyro shown in Fig. 14.8, it is to be noticed that with the frame held fixed the gyrowheel is free to rotate about the z axis but not about the y axis. Such a gyroscope, which can pivot about only one axis, has one degree of freedom. Gyroscopes such as those shown in Figs. 14.1, 14.4, and 14.5 are free to pivot about the two mutually perpendicular y and z axes and have two degrees of freedom.

Rate Gyro. A rate gyro is one in which the motion of a gimbal is restrained by means of springs, as shown in Fig. 14.8. When the frame rotates about the y axis, the y-axis torque is transmitted through the

gimbal arrangement to the gyrowheel. This causes a precession about the z axis, which is seen to be resisted by the restraining springs. Because the torque T_z exerted by the restraining springs opposes the precession θ_z, it follows that

$$T_z = -K_z\theta_z \tag{14.8}$$

The substitution of the preceding value of T_z into Eq. (14.7) yields the following equation of operation for a rate gyro,

$$K_z\theta_z + C_zp\theta_z + J_zp^2\theta_z = J_s\omega_sp\theta_y$$

or

$$\theta_z = \frac{J_s\omega_s}{J_zp^2 + C_zp + K_z}\dot{\theta}_y \tag{14.9}$$

By designing the rate gyro so that the first two terms in the denominator may be considered negligible, the angular position θ_z is a measure of the angular velocity $\dot{\theta}_y$ of the input shaft. In addition, by having the angular momentum $J_s\omega_s$ of the gyrowheel very large, a rate gyro is capable of measuring small angular velocities quite accurately. The value of θ_z can be measured by attaching a scale to the outer gimbal and a pointer to the shaft extension of the inner gimbal, as shown in Fig. 14.8. Usually, it is desired to have an electrical output signal which is proportional to θ_z. This is accomplished by using a synchro pick-off.

Integrating Rate Gyro. The integrating rate gyroscope shown in Fig. 14.9*a* is referred to as an HIG gyro (hermetically sealed integrating gyro).[1] This single-degree-of-freedom gyroscope is very rugged and extremely accurate. The inner gimbal, or can, is hermetically sealed and filled with an inert gas which acts as a neutral atmosphere. The outer case is filled with a viscous fluid whose density is such that the inner gimbal neither floats nor sinks but rather remains suspended. Thus, the load on the jeweled bearings is negligibly small in order to minimize the effects of drift. A schematic diagram for the operation of an integrating rate gyro is shown in Fig. 14.9*b*. This illustrates that a rotation of the outer case about the y axis causes a precession of the gyrowheel about the z axis. By placing an electrical pickup on the output shaft, a signal is obtained which is a measure of the angular position θ_y. Although at first glance the integrating rate gyro looks quite different from the rate gyro of Fig. 14.8, actually both gyros are quite similar in many respects. A motion about the y axis of an integrating rate gyro causes the gyro to precess about the z axis, as was the case for the rate gyro. The major distinction is that an integrating rate gyro does not have a restraining spring, but rather the viscous friction of the fluid between the inner gimbal, which is a cylinder, and the outer

[1] C. S. Draper, W. Wrigley, and L. R. Groke, The Floating Integrating Gyro and Its Application to Geometrical Stabilization Problems on Moving Bases, *Aeronaut. Eng. Rev.*, vol. 15, no. 6, June, 1956.

gimbal, which is also a cylinder, produces a torque T_z which opposes precession. That is,

$$T_z = -C\frac{d\theta_z}{dt} \tag{14.10}$$

where C is the coefficient of viscous friction between the inner gimbal and the outer case. Substituting T_z from Eq. (14.10) into Eq. (14.7) and noting that the viscous friction C_z of the bearings is negligible compared with C,

$$J_z p^2\theta_z + Cp\theta_z = J_z\omega_z p\theta_y$$

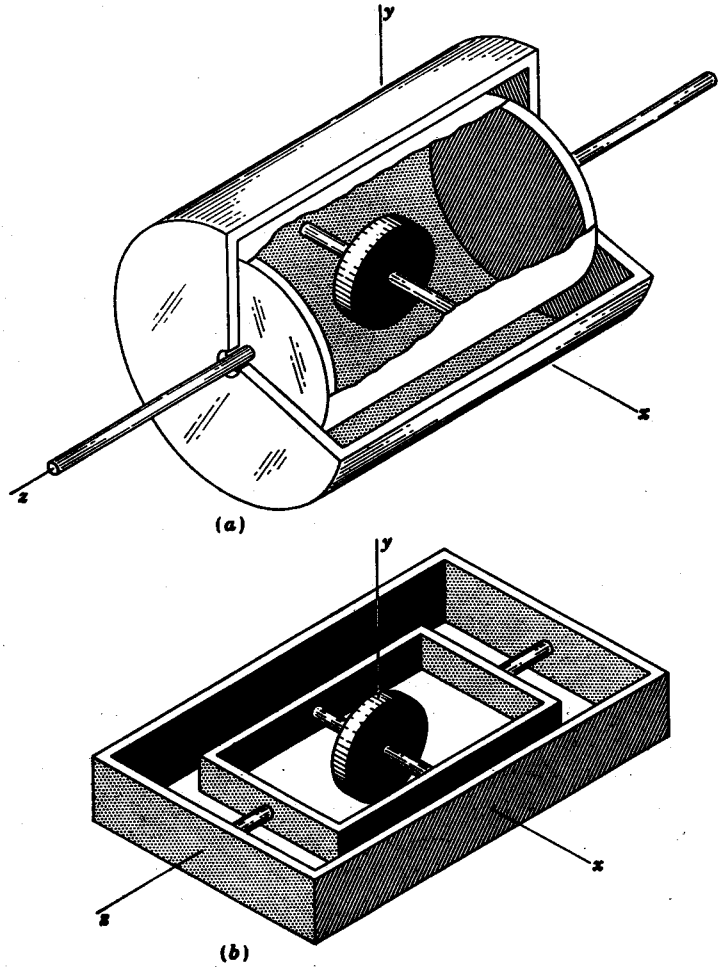


FIG. 14.9. HIG gyro.

Solving for θ_z yields

$$\theta_z = \frac{J_z\omega_z}{J_z p + C}\,\theta_y = \frac{J_z\omega_z/C}{\tau p + 1}\,\theta_y \approx \frac{J_z\omega_z}{C}\,\theta_y \tag{14.11}$$

where $\tau = J_z/C$. By making the value of C large and the value for J_z small so that the time constant is negligible, the output angle θ_z is proportional to the input θ_y rather than the rate of change of θ_y, as is the case for the rate gyro. Thus, this device is called an integrating rate gyro or sometimes an "integrating gyro."

14.5. Stable Platform. The primary member of a stable platform is the stable element. The stable element is orientated so that one axis always points in a north-south direction, another in an east-west direction, and a third in a vertical direction, as is shown in Fig. 14.10. Three accelerometers are mounted on this stable element at right angles to each other to measure acceleration in the north-south, east-west, and vertical directions. Successive integration of the output of each accelerometer yields velocity and then displacement. The three components of displacement determine the position of the vehicle. Because the accelerometers cannot distinguish acceleration of the vehicle from the acceleration of gravity, it is essential that the north-south and east-west accelerometers be mounted in a horizontal plane. Otherwise, each accelerometer would measure a component of the acceleration due to gravity, which would be interpreted by the computer as an acceleration in the direction of motion of the particular accelerometer whence the signal came. Similarly, the vertical accelerometer would yield an erroneous signal if it did not maintain its vertical direction. Thus, it is essential that the stable element be maintained in its proper orientation to obtain an accurate measurement of position.

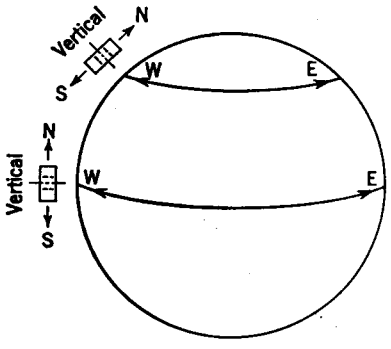


FIG. 14.10. Orientation of the stable element with respect to the earth.

As is shown in Fig. 14.11*a*, a stable platform utilizes a set of gimbals mounted in bearings so that the motion of the vehicle is not transmitted to the stable element. However, because of bearing friction and other extraneous disturbances, it is possible for the stable element to rotate from its reference orientation. Such motion is detected by the use of three gyroscopes, as is shown in Fig. 14.11*b*. The plane of the inner gimbal and gyrowheel for each gyroscope is perpendicular to the axis about which rotation is to be detected. For example, a rotation about

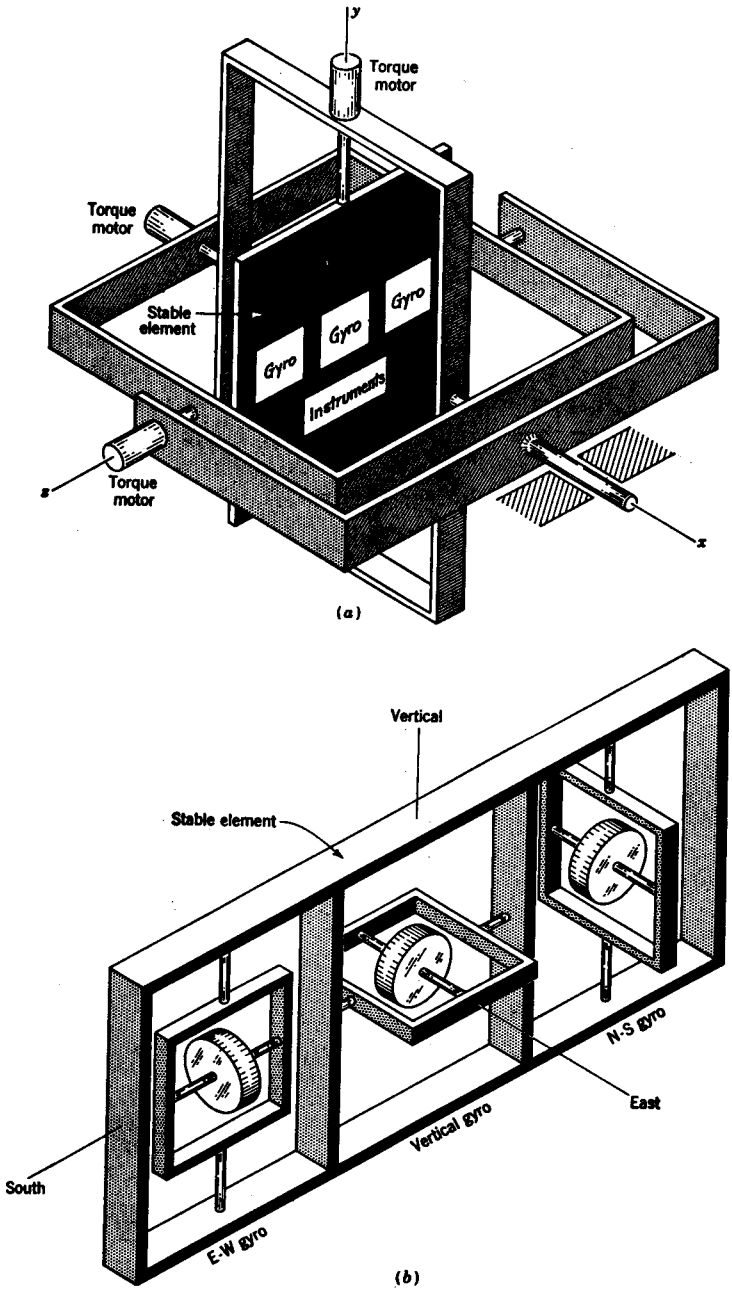


FIG. 14.11. (*a*) Stable platform; (*b*) stable element.

the vertical axis causes a corresponding precession of the vertical gyro. In turn, the electrical pick-off sends a signal to the azimuth gimbal servomotor, which applies a torque to return the stable element to its reference position.

To maintain a fixed attitude as shown in Fig. 14.10, it is necessary to compensate the orientation of the stable element for the effect of the rotation of the earth and motion of the vehicle relative to the earth. A

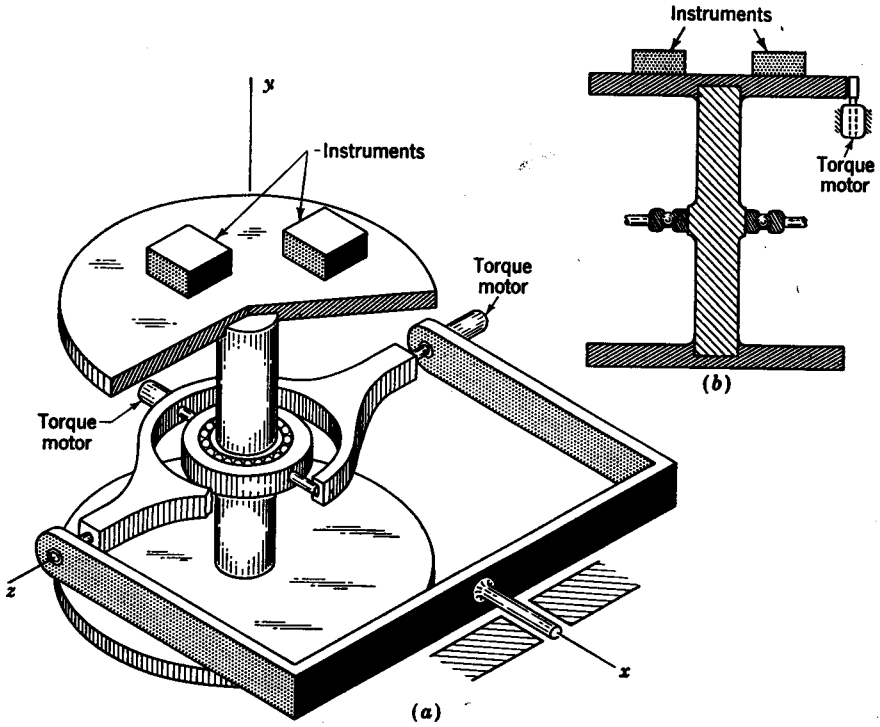


FIG. 14.12. (*a*) Stable platform with internal gimbaling; (*b*) stable element of a stable platform with internal gimbaling.

computer automatically corrects the reference orientation for rotation of the earth by sending an electrical signal to each torque motor which is incorporated in each gyro. The torque motor of a gyro applies a torque to precess the gyro which thus changes the reference position, or orientation, of the gyroscope. The electrical pick-off detects this change and sends an electrical signal to the appropriate gimbal servomotor so that the stable element follows the reference position determined by each gyroscope. In addition to compensating for rotation of the earth, it is also necessary for the computer to take into account the motion of the vehicle with respect to the earth. This motion is obtained by integrating

the three mutually perpendicular accelerations of the vehicle, as was previously discussed.

In the design of a stable platform,[1] often a considerable savings in weight and space may be realized by using the internal-gimbaling scheme shown in Fig. 14.12*a* rather than the external gimbaling illustrated in Fig. 14.11*a*. The outer gimbal is seen to be the same for either internal or external gimbaling. However, the middle gimbal now becomes a split beam, and the inner gimbal is simply a sleeve. The stable element consists of two tables connected by a solid post, as is shown separately in Fig. 14.12*b*.

14.6. The Control Loop. From Figs. 14.11 and 14.12, it is to be noted that a basic similarity exists in the control of the position of the stable element due to roll, pitch, or azimuth. A rotation of the stable element is sensed by the appropriate gyro, which sends an electrical signal to the corresponding torque motor to return the stable element to its reference position. The effect of roll, pitch, or azimuth may be considered independently.

In the following, it is shown how to obtain the block diagram describing the control of the azimuth position of the stable element.[2] Similar techniques may be used to determine the response due to roll or pitch. By designating the reference direction of the azimuth gyro as ϕ_r and the angular position of the stable element as ϕ_c, then $\phi_c - \phi_r$ is the rotation of the stable element from the reference or desired direction. This error causes a precession of the gyro, which in turn sends a signal to the azimuth torque motor to correct the orientation of the stable element, i.e.,

$$T_m = -K_m(\phi_c - \phi_r) = K_m(\phi_r - \phi_c) \tag{14.12}$$

where K_m is the gain for the corrective action of the torque motor.

Because of the precision bearings which support the stable element, the damping may be considered as negligible. The motor torque T_m applied to the table is resisted by the inertia of the table and the load torque.

$$T_m = Jp^2\phi_c + T_L \tag{14.13}$$

where T_L is the sum of the extraneous torques such as that due to gimbal-bearing friction. The block-diagram representation for the over-all control loop is obtained by combining Eqs. (14.12) and (14.13) and substituting s for p, as shown in Fig. 14.13*a*. This is the basic control

[1] R. H. Cannon and D. P. Chandler, Stable Platforms for High Performance Aircraft, *Aeronaut. Eng. Rev.*, vol. 16, no. 12, pp. 42–47, December, 1957.

[2] R. H. Cannon, Jr., Root-locus Analysis of Structural Coupling in Control Systems, *Trans. ASME, J. Appl. Mechanics*, vol. 81, ser. E, no. 2, 1959.

loop whose characteristic equation is

$$Js^2 + K_m = 0$$

or

$$s^2 + \frac{K_m}{J} = s^2 + K = 0 \tag{14.14}$$

where $K = K_m/J$ and $J = J_1 + J_2$ is the combined inertia of the upper table J_1 plus that of the lower table J_2. The $G(j\omega)H(j\omega)$ polar plot for this system is seen to lie entirely along the negative real axis, as is shown by the solid line in Fig. 14.14*a*. The corresponding root-locus plot is shown in Fig. 14.14*b*.

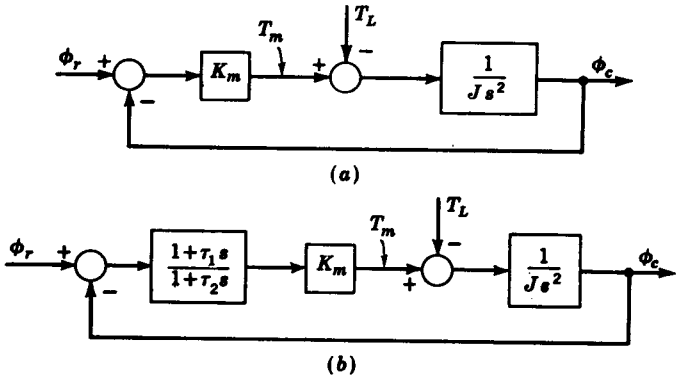


FIG. 14.13. Block diagram for azimuth control of a stable element with a rigid post. (*a*) Uncompensated system; (*b*) system with series phase-lead compensation.

To stabilize the operation of this system, it is necessary to use phase-lead series compensation to reshape the $G(j\omega)H(j\omega)$ plot, as is shown by the dashed-line loci in Fig. 14.14*a*. The addition of series lead compensation to the servo controller results in the block diagram shown in Fig. 14.13*b*, for which the characteristic equation is

$$Js^2(\tau_2 s + 1) + K_m(\tau_1 s + 1) = 0$$

or

$$s^2\left(s + \frac{1}{\tau_2}\right) + K\left(s + \frac{1}{\tau_1}\right) = 0 \tag{14.15}$$

where $K = \tau_1 K_m/\tau_2 J$. The resultant root-locus plot is shown in Fig. 14.14*c*.

In the preceding analysis, the azimuth post connecting the upper and lower tables was considered to be rigid. An investigation of the effects of the limberness of the post is accomplished as follows: The summation of torques acting on the upper table is

$$\Sigma T_e = T_m - T_L + K_s(\phi_{c_2} - \phi_{c_1}) = J_1 s^2 \phi_{c_1}$$

or

$$(J_1 s^2 + K_s)\phi_{c_1} - K_s \phi_{c_2} = T_m - T_L \tag{14.16}$$

where K_s is the torsional spring constant for the azimuth post. The angular position of the upper table is designated by ϕ_{c_1} and that of the lower table by ϕ_{c_2}. Similarly, from the summation of torques acting on the lower table, it is found that

$$\Sigma T_s = -K_s(\phi_{c_2} - \phi_{c_1}) = J_2 s^2 \phi_{c_2}$$

or

$$(J_2 s^2 + K_s)\phi_{c_2} - K_s \phi_{c_1} = 0 \tag{14.17}$$

Elimination of ϕ_{c_2} between Eqs. (14.16) and (14.17) gives

$$\phi_{c_1} = \frac{J_2(s^2 + K_s/J_2)}{J_1 J_2 s^2[s^2 + K_s(J_1 + J_2)/J_1 J_2]}(T_m - T_L)$$

$$= \frac{(s + j\omega_{n_2})(s - j\omega_{n_2})}{J_1 s^2 (s + j\omega_{n_1})(s - j\omega_{n_1})}(T_m - T_L) \tag{14.18}$$

where $\omega_{n_1} = \sqrt{\dfrac{K_s(J_1 + J_2)}{J_1 J_2}}$ and $\omega_{n_2} = \sqrt{\dfrac{K_s}{J_2}}$

The block-diagram representation for Eqs. (14.12) and (14.18) is

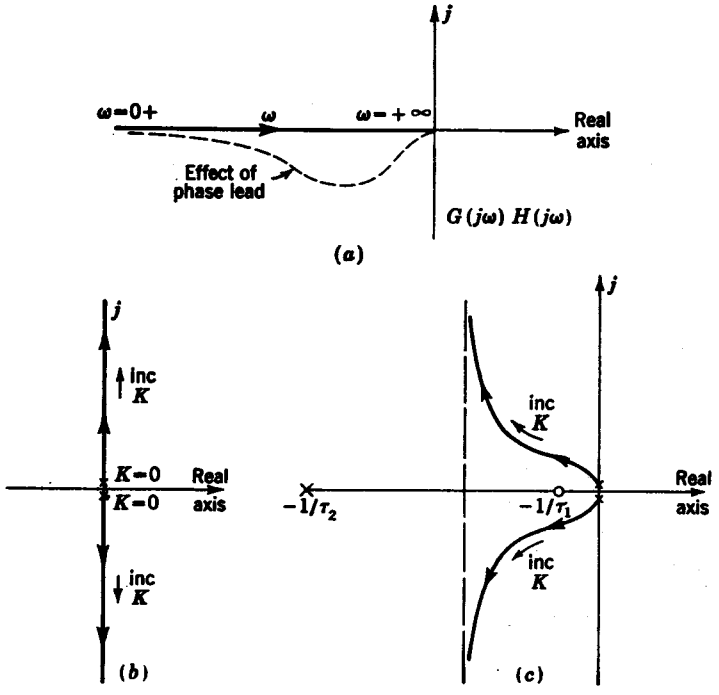


FIG. 14.14. Azimuth control of a stable element with a rigid post. (*a*) Vector loci plot; (*b*) root-locus plot for uncompensated system; (*c*) root-locus plot for system with series phase lead.

shown in Fig. 14.15*a*. The characteristic equation for this system is

$$s^2(s + j\omega_{n_1})(s - j\omega_{n_1}) + K(s + j\omega_{n_2})(s - j\omega_{n_2}) = 0 \qquad (14.19)$$

where $K = K_m/J_1$. The corresponding root-locus plot is shown in Fig. 14.16*a*. It is again necessary to use phase-lead series compensation to stabilize this system, as is indicated in Fig. 14.15*b*.

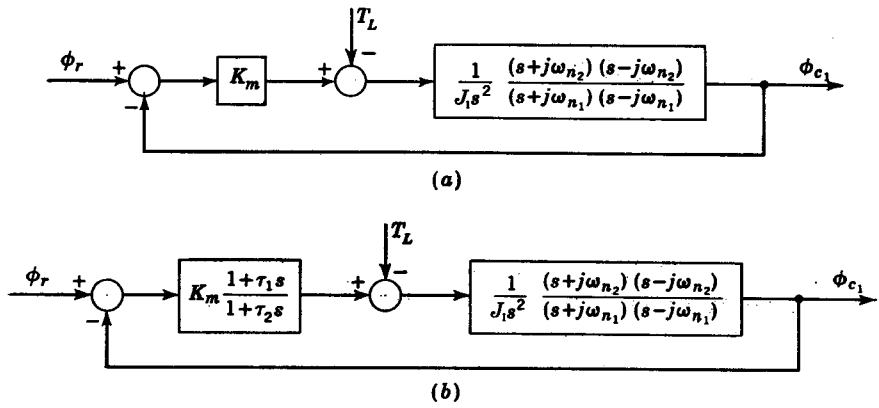


FIG. 14.15. Block-diagram representation for azimuth control of a stable element with a limber post (torque applied to upper table). (*a*) Uncompensated system; (*b*) compensated system.

The corresponding characteristic equation which includes the effect of the series compensation is

$$s^2(s + j\omega_{n_1})(s - j\omega_{n_1})\left(s + \frac{1}{\tau_2}\right) + K(s + j\omega_{n_2})(s - j\omega_{n_2})\left(s + \frac{1}{\tau_1}\right) = 0 \qquad (14.20)$$

where $K = \dfrac{\tau_1}{\tau_2}\dfrac{K_m}{J_1}$. The root-locus plot for Eq. (14.20) is shown in Fig. 14.16*b*.

For a rigid post, it makes no difference whether the torque motor is attached to the upper or lower table. However, for a limber post, the torque motor must be attached to the upper table, on which the gyros are located, for otherwise it becomes impossible to stabilize the system by ordinary means. This is easily proved as follows:

With the torque motor connected to the bottom table, the summation of torques acting on the upper table gives

$$(J_1 s^2 + K_s)\phi_{c_1} - K_s\phi_{c_2} = 0 \qquad (14.21)$$

and similarly the summation of torques acting on the lower table is

$$(J_2 s^2 + K_s)\phi_{c_2} - K_s\phi_{c_1} = T_m - T_L \qquad (14.22)$$

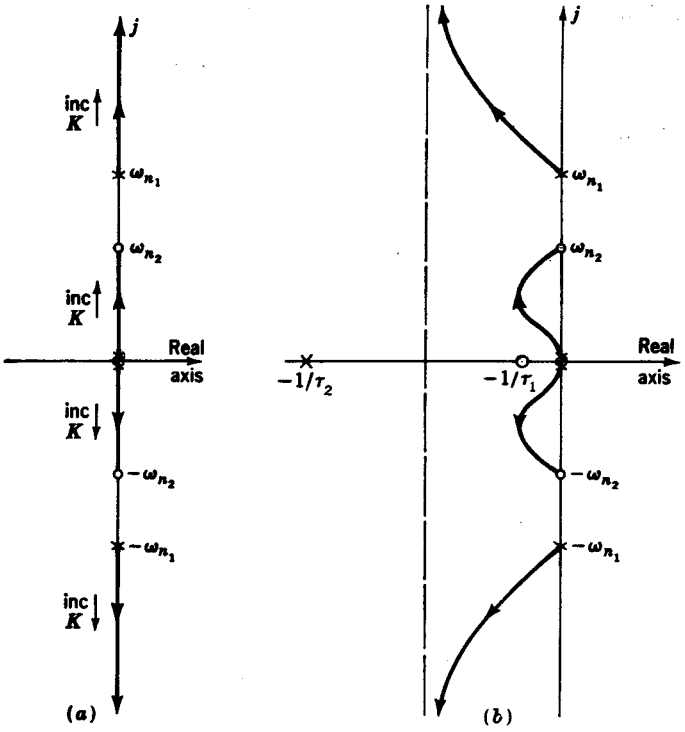


FIG. 14.16. Root-locus plot for control of stable element with a limber post (torque applied to upper table). (*a*) Uncompensated system; (*b*) system with series phase-lead compensation.

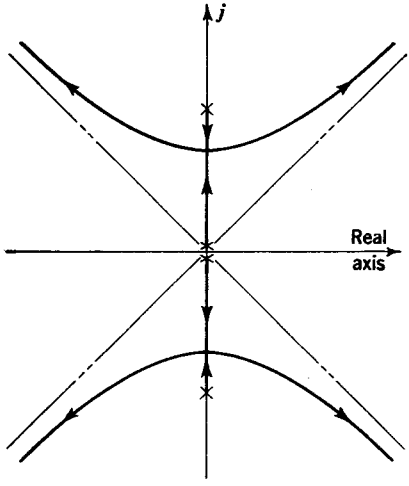


FIG. 14.17. Root-locus plot for uncompensated control of stable element with a limber post (torque applied to lower table).

Elimination of ϕ_{c_2} gives

$$\phi_{c_1} = \frac{K_s/J_1J_2}{s^2(s^2 + \omega_n^2)} (T_m - T_L) \tag{14.23}$$

where $$\omega_n = \sqrt{\frac{K_s(J_1 + J_2)}{J_1J_2}}$$

The root-locus plot for this system in which torque is applied to the lower table is shown in Fig. 14.17. This is a system which cannot be stabilized by simple compensation. Thus, it is interesting to note that,

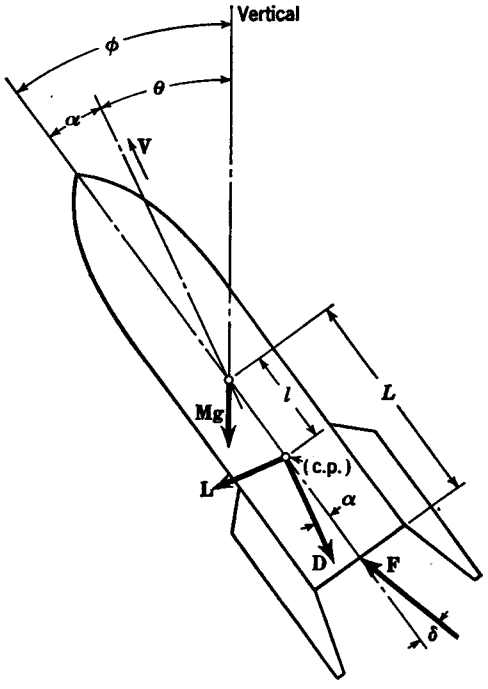


FIG. 14.18. Force acting on a missile.

for a rigid post, it makes no difference whether the correcting torque is applied to the lower or upper platform, whereas for a limber system the effect is somewhat astonishing.

14.7. Missile in Flight. When a missile is in flight, it is subjected to a number of forces such as aerodynamic forces, gravitational forces, and the thrust which propels it, as is shown in Fig. 14.18. Aerodynamic forces are conveniently handled by resolving them into a component parallel to and a component perpendicular to the direction of motion, or heading, of the missile. The aerodynamic component parallel to the direction of motion is designated as the drag D, and the perpendicular

component is called the lift L. The intersection of the line of action of D and L is called the center of pressure (cp).

Although the vehicle is pointed at an angle ϕ with respect to the vertical, the direction of motion is inclined at an angle θ from the vertical. Because $e^{j(\theta+90°)}$ is a unit vector in the direction of motion, as is shown in Fig. 14.19*a*, the velocity may be written

$$\mathbf{V} = Ve^{j(\theta+90°)} \tag{14.24}$$

Differentiation of the velocity with respect to time yields the acceleration

$$\mathbf{A} = \frac{d\mathbf{V}}{dt} = V\frac{d}{dt}e^{j(\theta+90°)} + e^{j(\theta+90°)}\frac{dV}{dt} \tag{14.25}$$

Noting that

$$\frac{d}{dt}e^{j(\theta+90°)} = j\frac{d\theta}{dt}e^{j(\theta+90°)} = \frac{d\theta}{dt}e^{j90°}e^{j(\theta+90°)}$$

then

$$\mathbf{A} = \frac{dV}{dt}e^{j(\theta+90°)} + V\frac{d\theta}{dt}e^{j(\theta+180°)} \tag{14.26}$$

The acceleration has a component dV/dt parallel to the direction of motion and a component $V(d\theta/dt)$ which is perpendicular, as shown in Fig. 14.19*b*.

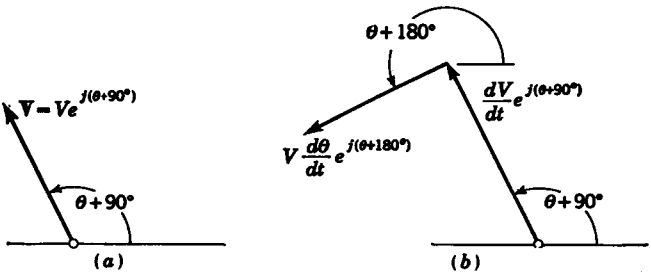


FIG. 14.19. Vector representation of (*a*) missile velocity and (*b*) missile acceleration.

For small angles of attack (i.e., less than 10°) the lift force is approximately proportional to the angle of attack, that is,

$$L = K_L\alpha$$

where K_L is the lift coefficient (pounds lift per unit α).

Unlike lift, the drag force D tends to remain constant for small angles of attack. The summation of forces acting perpendicular to the direction of motion of the missile yields

$$MV\frac{d\theta}{dt} = K_L\alpha + F\sin(\alpha+\delta) + Mg\sin\theta \tag{14.27}$$

which for small values of $\alpha + \delta$ and θ becomes

$$MV\frac{d\theta}{dt} - Mg\theta = (K_L + F)\alpha + F\delta \tag{14.28}$$

where δ is the angle between the thrust force F and the centerline of the missile, as shown in Fig. 14.18.

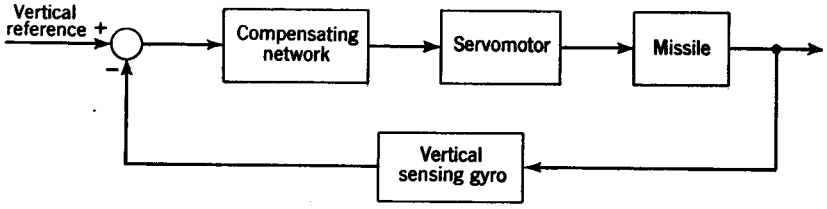


FIG. 14.20. Block-diagram representation for control of missile in vertical flight.

Similarly, the summation of torques acting about the center of gravity yields

$$\sum T_e = J\frac{d^2\phi}{dt^2} = -K_L\alpha\cos\alpha - D\sin\alpha - FL\sin\delta - C_v\frac{d\phi}{dt} \tag{14.29}$$

which for small values of α and δ gives

$$J\frac{d^2\phi}{dt^2} + C_v\frac{d\phi}{dt} = -K_L\alpha - D\alpha - FL\delta \tag{14.30}$$

From Fig. 14.18, it follows that

$$\phi = \theta + \alpha \tag{14.31}$$

Thus, from Eqs. (14.28), (14.30), and (14.31) the over-all relationship between ϕ and δ is

$$\phi = \frac{-F[LMVp + (K_L + F - Mg)L - (K_L + D)]\delta}{JMVp^3 + [J(K_L + F - Mg) + C_vMV]p^2 + [C_v(K_L + F - Mg) + (K_L + D)MV]p - (K_L + D)Mg} \tag{14.32}$$

The characteristic equation for the preceding result is seen to be of the third order.

A typical block diagram for controlling the flight of a missile is shown in Fig. 14.20. The servomotor controls the direction of the angle of thrust, and the vertical sensing gyro measures the angle of inclination ϕ of the missile with respect to the vertical.

CHAPTER 15

NONLINEAR SYSTEMS

15.1. Introduction. In the earlier chapters, it was shown how certain nonlinear relationships could be linearized by approximating the nonlinear function by the tangent to the curve at the point of interest. In this chapter, it is shown how other types of nonlinearities may be

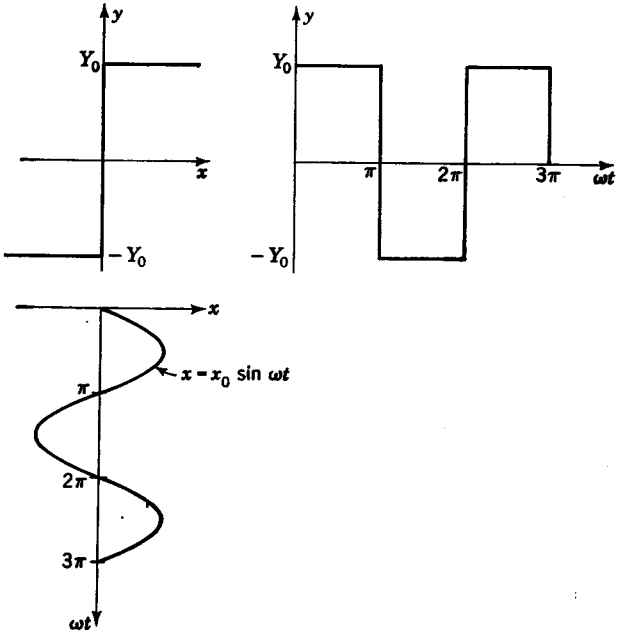


FIG. 15.1. Characteristics of on-off element.

treated. For example, phenomena such as dead band, backlash, saturation, and on-off type of action may be investigated by the use of the describing-function technique. Sampled-data systems may be handled by an interesting extension of the ordinary techniques for linear systems. Phase-plane methods are applicable for systems which are characterized by first- or second-order nonlinear differential equations.

15.2. Describing Functions. Certain types of nonlinear elements are characterized by the feature that, if the input is a sinusoid, then the output will have a periodic waveform. The period of the output wave is the same as that of the input.[1–3] In Fig. 15.1 are shown the operating characteristics of an on-off type of control element. When the input x is greater than zero, the value of the output is Y_0. When x is less than zero,

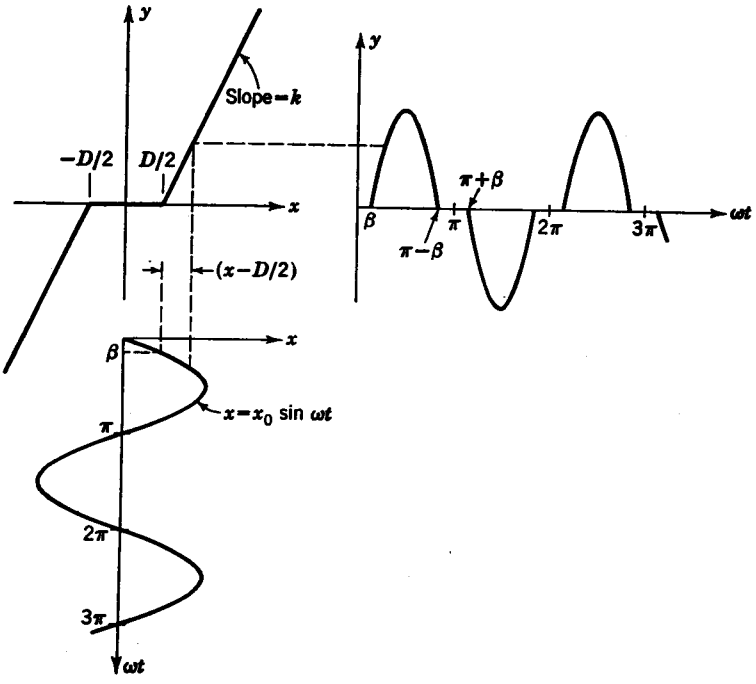


FIG. 15.2. Characteristics of linear element with a dead zone.

the output is $-Y_0$. The system would be in equilibrium at its reference operating condition when $y = 0$, but this is impossible with an on-off type of element. A plot of the sinusoidal input $x = x_0 \sin \omega t$ is drawn vertically down from the on-off contactor characteristics. The corresponding output is shown horizontally to the right.

Another nonlinear characteristic which is frequently encountered is that of a dead zone, or dead band, which is shown in Fig. 15.2. When

[1] R. J. Kochenburger, A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, *Trans. AIEE*, vol. 69, pt. 1, pp. 270–284, 1950.

[2] J. J. D'azzo and C. H. Houpis, "Control System Analysis and Synthesis," McGraw-Hill Book Company, Inc., New York, 1960.

[3] J. G. Truxal, "Automatic Feedback Control System Synthesis," McGraw-Hill Book Company, Inc., New York, 1955.

the value of D is zero, this becomes a linear system. When $-D/2 \leq x \leq D/2$, the value of y is zero. For $x > D/2$ the output is

$$y = k\left(x - \frac{D}{2}\right) = k\left(x_0 \sin \omega t - \frac{D}{2}\right) \tag{15.1}$$

With the aid of Fig. 15.2, it follows that

$$\frac{D}{2} = x_0 \sin \beta$$

or

$$\beta = \sin^{-1} \frac{D}{2x_0} \tag{15.2}$$

For a sinusoidal input to such nonlinear elements, the output y is periodic. Generally, the periodic output wave y which is obtained for a

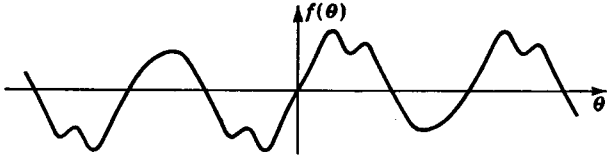


FIG. 15.3. Odd function.

nonlinear component is an odd function. As shown in Fig. 15.3, an odd function is one for which $f(\theta) = -f(-\theta)$. Any odd function may be approximated by the following series of sine terms:

$$f(\theta) = A_0 + B_1 \sin \theta + B_2 \sin 2\theta + B_3 \sin 3\theta + \cdots \tag{15.3}$$

Integration of the preceding expression between the limits $\theta = 0$ and $\theta = 2\pi$ yields

$$\int_0^{2\pi} f(\theta)\, d\theta = A_0 \int_0^{2\pi} d\theta + \sum_{n=1}^{\infty} B_n \int_0^{2\pi} \sin n\theta\, d\theta \tag{15.4}$$

Because

$$B_n \int_0^{2\pi} \sin n\theta\, d\theta = 0 \tag{15.5}$$

then

$$A_0 = \frac{1}{2\pi} \int_0^{2\pi} f(\theta)\, d\theta \tag{15.6}$$

The term A_0 is thus seen to be the average value of the periodic function for one period. That is, the constant A_0 is the area of the function $f(\theta)$ divided by the length of the period.

The multiplication of Eq. (15.3) by $\sin \theta$ and the integration of each term over the period yields

$$\int_0^{2\pi} f(\theta) \sin \theta\, d\theta = A_0 \int_0^{2\pi} \sin \theta\, d\theta + B_1 \int_0^{2\pi} \sin^2 \theta\, d\theta + \sum_{n=2}^{\infty} B_n \int_0^{2\pi} \sin \theta \sin n\theta\, d\theta \tag{15.7}$$

After evaluating each integral on the right-hand side of Eq. (15.7), it is found that each term becomes zero except

$$B_1 \int_0^{2\pi} \sin^2 \theta \, d\theta = B_1 \pi$$

Thus

$$B_1 = \frac{1}{\pi} \int_0^{2\pi} f(\theta) \sin \theta \, d\theta \tag{15.8}$$

Upon multiplying Eq. (15.3) by $\sin n\theta$ and integrating over the period, it may be shown, in general, that

$$B_n = \frac{1}{\pi} \int_0^{2\pi} f(\theta) \sin n\theta \, d\theta \tag{15.9}$$

Series Representation for an On-Off Component. Let it be desired to determine the series expansion for the on-off element shown in Fig. 15.1. The general form of the series is

$$y = A_0 + \sum_{n=1}^{\infty} B_n \sin n\theta \tag{15.10}$$

Because the average value of the function is zero, $A_0 = 0$. The constant coefficient B_n is

$$B_n = \frac{1}{\pi} \int_0^{2\pi} f(\theta) \sin n\theta \, d\theta = \frac{2}{\pi} \int_0^{\pi} f(\theta) \sin n\theta \, d\theta \tag{15.11}$$

Substitution of $f(\theta) = Y_0$ gives

$$B_n = \frac{2Y_0}{\pi} \int_0^{\pi} \sin n\theta \, d\theta = \frac{-2Y_0}{n\pi} [\cos n\theta]_0^{\pi} = \frac{4Y_0}{n\pi} \tag{15.12}$$

Thus

$$B_1 = \frac{4}{\pi} Y_0 \qquad B_3 = \frac{4}{3\pi} Y_0 \qquad B_5 = \frac{4}{5\pi} Y_0 \qquad \cdots \tag{15.13}$$

and

$$0 = A_0 = B_2 = B_4 = B_6 = \cdots \tag{15.14}$$

The coefficients A_0, B_2, B_4, B_6, . . . are always zero when the value of the function from $\pi \leq \theta \leq 2\pi$ is the negative of the function for $0 \leq \theta \leq \pi$, as is the case for the functions shown in Figs. 15.1 and 15.2. This is usually the case for this type of nonlinearity which occurs in control components.

The resultant series expression for y is

$$y = \frac{4}{\pi} Y_0 \left(\sin \omega t + \frac{1}{3} \sin 3\omega t + \frac{1}{5} \sin 5\omega t + \cdots \right) \tag{15.15}$$

In Fig. 15.4 are shown the original square-wave function and the first two terms of the series expansion for this function. The sum of all the terms would yield the original periodic function. However, the fundamental component $B_1 \sin \omega t$ is seen to yield the most significant con-

tribution. Good results are obtained in the analysis of such nonlinear control elements by approximating the output by its fundamental component.

$$y \approx B_1 \sin \omega t \tag{15.16}$$

The two major reasons why the fundamental component yields good results in approximating the characteristics of a nonlinear element are:

1. The fundamental component yields the most significant contribution to the output.

2. The higher-frequency terms such as $B_3 \sin 3\omega t$, $B_5 \sin 5\omega t$, . . . are progressively attenuated more by the other linear components in the

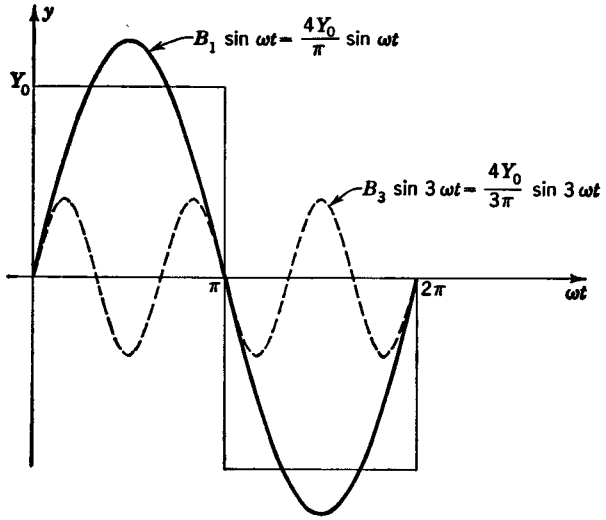


FIG. 15.4. Graphical interpretation of series expansion for a square wave.

system, and therefore they have less of an effect upon the operating characteristics of the system.

Series Representation of Dead Band. Because of the symmetry of the characteristics shown in Fig. 15.2, all the coefficients A_0, B_2, B_4, . . . will be zero. The approximation for this function is given by Eq. (15.16), where

$$B_1 = \frac{1}{\pi} \int_0^{2\pi} f(\theta) \sin \theta \, d\theta = \frac{2}{\pi} \int_0^{\pi} f(\theta) \sin \theta \, d\theta \tag{15.17}$$

From Fig. 15.2, it is to be seen that

$$f(\theta) = 0 \qquad \begin{cases} 0 \leq \theta \leq \beta \\ (\pi - \beta) \leq \theta \leq \pi \end{cases} \tag{15.18}$$

and

$$f(\theta) = k\left(x - \frac{D}{2}\right) = k(x_0 \sin \theta - x_0 \sin \beta) \qquad \beta \leq \theta \leq (\pi - \beta) \tag{15.19}$$

Thus
$$B_1 = \frac{2kx_0}{\pi}\int_{\beta}^{\pi-\beta}(\sin\theta - \sin\beta)\sin\theta\,d\theta$$
$$= \frac{2kx_0}{\pi}\left[\left(\frac{\theta}{2} - \frac{\sin 2\theta}{4}\right) + \cos\theta\sin\beta\right]_{\beta}^{\pi-\beta}$$
$$= \frac{2kx_0}{\pi}\left(\frac{\pi - 2\beta}{2} - \frac{\sin 2\beta}{2}\right) \tag{15.20}$$

The ratio $y/x = (B_1 \sin \omega t)/(x_0 \sin \omega t)$ describes the operation of this nonlinear component. From Eq. (15.20), it follows that

$$N = \frac{y}{x} = \frac{B_1}{x_0} = \frac{k}{\pi}(\pi - 2\beta - \sin 2\beta) \tag{15.21}$$

where $N = y/x$ is called the describing function because it describes the relationship between the fundamental output y and the input x.

Because $\beta = \sin^{-1}(D/2x_0)$, a plot of N/k versus $2x_0/D$ may be constructed as shown in Fig. 15.5. When $x_0 \leq D/2$, that is,

$$\frac{2x_0}{D} \leq 1$$

the output is zero. As the deadband zone D becomes small in comparison with x_0, then N approaches k, which would be the characteristic of a linear element in which $y = kx$.

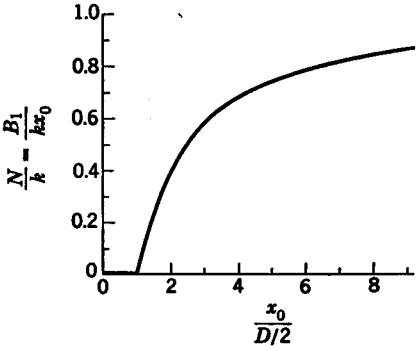


FIG. 15.5. Plot of describing function for linear element with a dead zone.

15.3. Stability Analysis. In Fig. 15.6*a* is shown a control system which has a nonlinear component represented by N. The component N may be considered as a variable gain. The value of N, the variable gain, depends upon the amplitude of the input signal x_0. The frequency response for linear elements such as $G(j\omega)$ depends on frequency only.

In Fig. 15.6*b* is shown a typical polar plot for $G(j\omega)$. If the gain is doubled, the polar plot would go through the -1 point. Thus, the system is unstable if $N \geq 2$.

For $k = 4$, it follows that $N \geq 2$ when $N/k \geq \frac{2}{4} = 0.5$. Thus, from Fig. 15.5 instability results when $x_0/(D/2) \geq 2.5$, or $x_0 \geq 1.25D$. If the value of k is 1, instability results when $N/k \geq 2$. Instability cannot occur for this case because the maximum attainable value of N/k is 1. In effect, a describing function N represents the nonlinear element whose characteristics are a function of the *amplitude* of the input signal but are independent of the frequency. The characteristics of a linear element

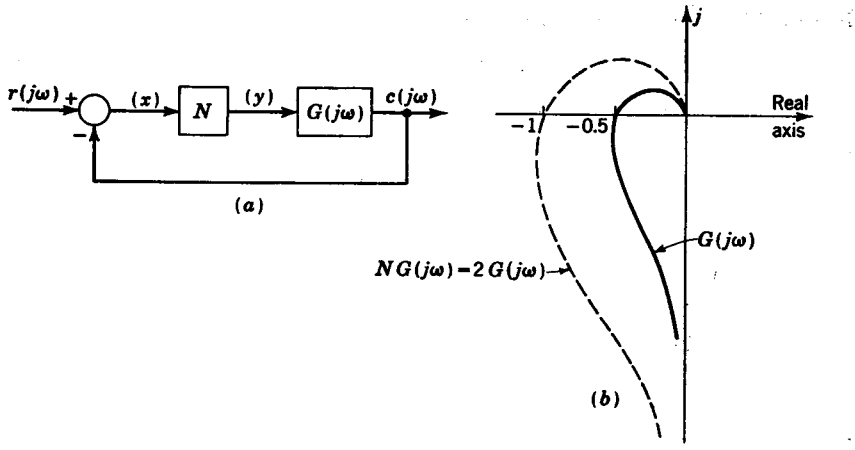


FIG. 15.6. Stability analysis of system with a nonlinear element.

$G(j\omega)$ are a function of frequency only and are independent of the amplitude of the excitation.

Another method for investigating stability is to note that the over-all system response is

$$\frac{c(j\omega)}{r(j\omega)} = \frac{NG(j\omega)}{1 + NG(j\omega)} \tag{15.22}$$

Instability results if

$$NG(j\omega) = -1$$

or

$$G(j\omega) = -\frac{1}{N} \tag{15.23}$$

By plotting $G(j\omega)$ and $-1/N$ as shown in Fig. 15.7, an unstable system is indicated if the two curves intersect. The function $-1/N$ is plotted by obtaining N/k for various values of $x_0/2D$ from Fig. 15.5 and then com-

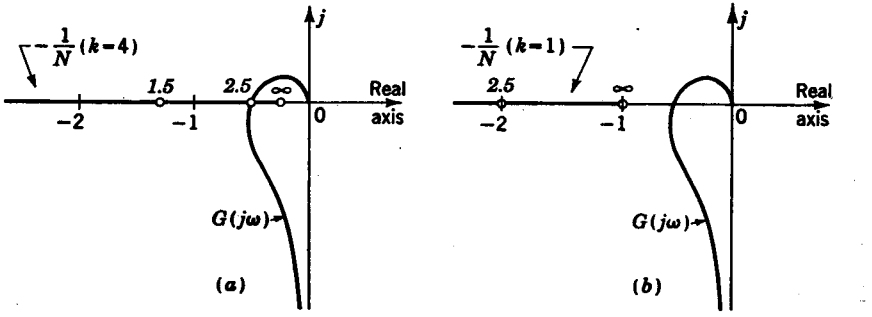


FIG. 15.7. Plots of $G(j\omega)$ and $-1/N$ for (a) $k = 4$ and (b) $k = 1$.

puting $-1/N$ from the equation

$$-\frac{1}{N} = -\frac{1}{k(N/k)} \tag{15.24}$$

For example, when $x_0/2D = \infty$, then $N/k = 1$ so that $-1/N = -1/k$. Thus, for $k = 4$, then $-1/N = -\frac{1}{4}$, and for $k = 1$, then $-1/N = -1$. The values of $x_0/2D$ are italicized in Fig. 15.7. An unstable system results for $k = 4$, as shown in Fig. 15.7*a*. The critical value is $x_0/2D \geq$

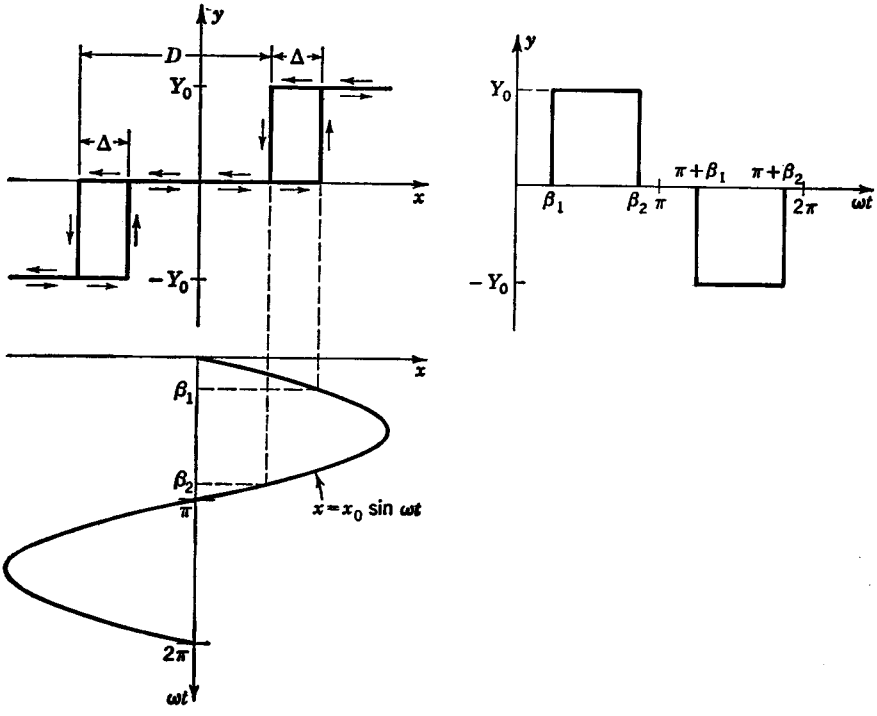


FIG. 15.8. Characteristics of on-off element with hysteresis and dead zone.

2.5, or $x_0 \geq 1.25D$. For the case in which $k = 1$, the system is always stable.

15.4. Describing Functions with a Phase Shift. The preceding discussion was limited to describing functions in which there was no phase shift between the input $x = x_0 \sin \omega t$ and the fundamental component $y = B_1 \sin \omega t$. In Fig. 15.8 is shown an input-output relationship for an on-off element which has a hysteresis loop of width Δ in addition to the dead band D. As the amplitude x of the input is increased from zero, the contactor does not close until x exceeds the value $(D + \Delta)/2$. The contactor then remains closed until the value of the input becomes less than $(D - \Delta)/2$. Thus, in the region between $(D - \Delta)/2 \leq x \leq$

$(D + \Delta)/2$ the value of the output depends upon the past history, or the manner in which the input was varying. A similar phenomenon occurs for negative values of x. The angle β_1 at which the contactor closes and the angle β_2 at which it opens again are obtained from Fig. 15.8 as follows:

$$\sin \beta_1 = \frac{D + \Delta}{2x_0}$$
$$\sin \beta_2 = \frac{D - \Delta}{2x_0} \tag{15.25}$$

In Fig. 15.9 is shown an enlarged view of the contactor output and the first harmonic component. The actual square-wave output is sym-

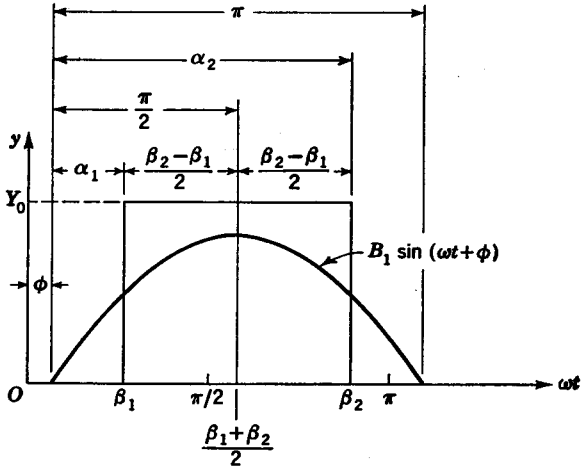


FIG. 15.9. Fundamental component of contactor output.

metrical about the angular position $(\beta_1 + \beta_2)/2$. The phase shift ϕ is given by the equation

$$\phi = \frac{\pi}{2} - \frac{\beta_1 + \beta_2}{2} \tag{15.26}$$

The amplitude of the fundamental harmonic component is

$$B_1 = \frac{2Y_0}{\pi} \int_{\alpha_1}^{\alpha_2} \sin \theta \, d\theta = -\frac{2Y_0}{\pi} \cos \theta \Big]_{\pi/2-(\beta_2-\beta_1)/2}^{\pi/2+(\beta_2-\beta_1)/2}$$
$$= \frac{4Y_0}{\pi} \sin \frac{\beta_2 - \beta_1}{2} \tag{15.27}$$

The describing function which is the ratio of the output (as approximated by the fundamental harmonic) to the input is

$$N = \frac{y}{x} = \frac{B_1 \sin (\omega t + \phi)}{x_0 \sin \omega t}$$

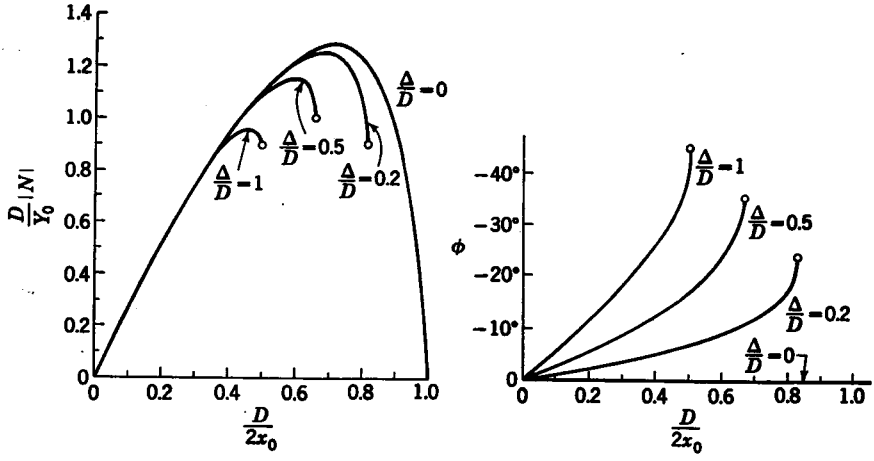


FIG. 15.10. Amplitude and phase angle of describing function.

whence the magnitude of the describing function is

$$|N| = \frac{B_1}{x_0} = \frac{4Y_0}{\pi x_0} \sin \frac{\beta_2 - \beta_1}{2} \tag{15.28}$$

Convenient graphs as shown in Fig. 15.10 for determining the amplitude $|N|$ of the describing function and the phase angle ϕ are obtained by rewriting Eqs. (15.28) and (15.26) in the following form,

$$\frac{D}{Y_0}|N| = \frac{8}{\pi}\frac{D}{2x_0} \sin \frac{\beta_2 - \beta_1}{2} \tag{15.29}$$

and

$$\phi = \frac{\pi}{2} - \frac{1}{2}\left[\sin^{-1}\left(\frac{D}{2x_0} + \frac{\Delta}{2x_0}\right) + \sin^{-1}\left(\frac{D}{2x_0} - \frac{\Delta}{2x_0}\right)\right] \tag{15.30}$$

In Fig. 15.11 is shown the curve of $G(j\omega)$ for the system shown in Fig. 15.6.

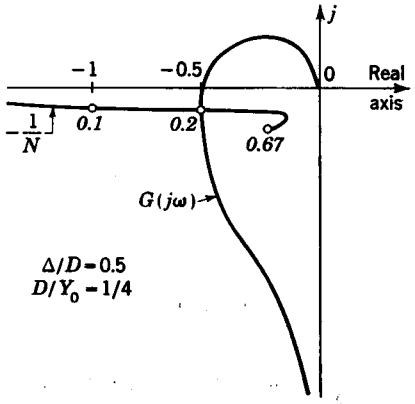


FIG. 15.11. Plot of $G(j\omega)$ and $-1/N$.

Because $N = |N|e^{j\phi}$ and $-N = |N|e^{j(\phi-180°)}$, then

$$-\frac{1}{N} = \left|\frac{1}{N}\right| e^{-j(\phi-180°)}$$

The curve of $-1/N$ of Fig. 15.11 is drawn for the case in which $\Delta/D = 0.5$ and $D/Y_0 = \frac{1}{4}$. Values of $D/2x_0$ are italicized. Instability results when $D/2x_0 \geq 0.2$, or $2x_0/D \leq 5$, or $x_0 \leq 2.5D$.

15.5. Sampled-data Systems. A schematic representation of a sampler and holder is shown in Fig. 15.12*a*. The sampler switch closes every T

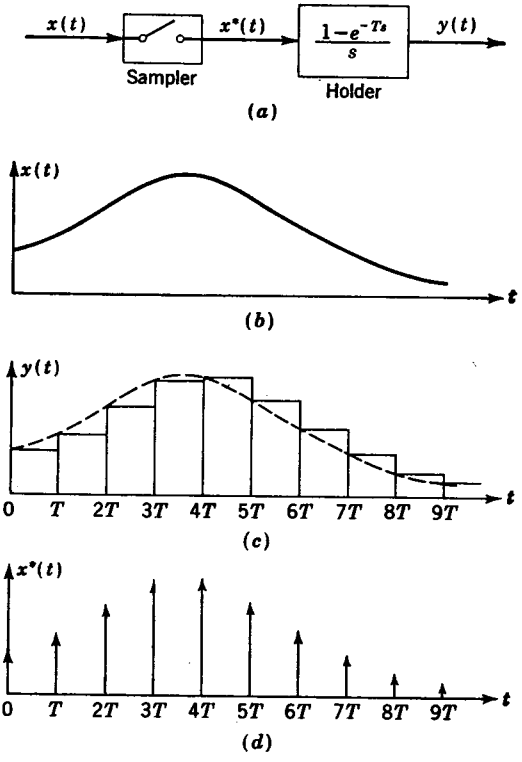


FIG. 15.12. (*a*) Sampled-data system; (*b*) input to switch $x(t)$; (*c*) output at holder $y(t)$; (*d*) output at switch $x^*(t)$.

sec to admit the input signal $x(t)$. The holder retains this value of $x(t)$ until the next sample is taken. A typical plot of $x(t)$ is shown in Fig. 15.12*b*. If $x(t)$ is sampled every T sec, the output $y(t)$ of the holder is a train of pulse functions as shown in Fig. 15.12*c*.

Recently, numerous systems have been devised which utilize digital computers as control elements. Such systems are generally sampled-data systems, because the information fed into a digital computer is the

value (sample) of the corresponding information at some instant of time. The computed output remains unchanged until new information (another sample) is fed into the digital computer. Another example is guidance systems that utilize radar scanning in which signals are obtained at discrete increments of time.

The Laplace transform $Y(s)$ for the train of pulses $y(t)$ is obtained by application of the real-translation theorem. Thus

$$\begin{aligned} Y(s) &= \frac{x(0)}{s}(1 - e^{-Ts}) + \frac{x(T)e^{-Ts}}{s}(1 - e^{-Ts}) \\ &\quad + \frac{x(2T)e^{-2Ts}}{s}(1 - e^{-Ts}) + \cdots \\ &= \frac{1 - e^{-Ts}}{s}[x(0) + x(T)e^{-Ts} + x(2T)e^{-2Ts} + \cdots] \\ &= \frac{1 - e^{-Ts}}{s}\sum_{n=0}^{\infty} x(nT)e^{-nTs} = \frac{1 - e^{-Ts}}{s}X^*(s) \end{aligned} \tag{15.31}$$

where $X^*(s) = \sum_{n=0}^{\infty} x(nT)e^{-nTs}$, $x(0)$ is $x(t)$ at $t = 0$, $x(T)$ is $x(t)$ at $t = T$, and $x(nT)$ is the value of the input $x(t)$ at time $t = nT$. The symbol * should be read "star." The inverse transform of $X^*(s)$ is

$$\begin{aligned} x^*(t) &= \mathcal{L}^{-1}[X^*(s)] = \mathcal{L}^{-1}[x(0) + x(T)e^{-Ts} + x(2T)e^{-2Ts} + \cdots] \\ &= x(0)u_1(t) + x(T)u_1(t - T) + x(2T)u_1(t - 2T) + \cdots \end{aligned} \tag{15.32}$$

The term $x^*(t)$ is a train of impulses, as shown in Fig. 15.12*d*, in which the area of each impulse is the value of $x(nT)$ at the sampling instant. From the preceding analysis, it follows that the signal $x^*(t)$ coming from the sampler switch may be regarded as a train of impulses, and the transform for the holder is in effect $(1 - e^{-Ts})/s$.[1-6]

In Fig. 15.13*a* is shown a sampling switch followed by a linear element whose transfer function is $G(s)$. The transfer function for the holder

[1] Kochenburger, *op. cit.*

[2] J. C. Gille, M. J. Pelegrin, and P. Decaulne, "Feedback Control Systems," McGraw-Hill Book Company, Inc., New York, 1959.

[3] J. R. Ragazzini and L. A. Zadeh, The Analysis of Sampled Data Systems, *Trans. AIEE*, 71, 1951.

[4] J. R. Ragazzini and G. F. Franklin, "Sampled-data Control Systems," McGraw-Hill Book Company, Inc., New York, 1959.

[5] W. K. Linvill and R. W. Sittler, Design of Sampled Data Systems by Extension of Conventional Techniques, *Mass. Inst. Technol. Rept.* R-222, *Digital Computer Lab.*, 1953.

[6] E. I. Jury, Analysis and Synthesis of Sampled Data Control Systems, *Trans. AIEE*, Communications and Electronics, September, 1954.

$(1 - e^{-Ts})/s$ is included in the over-all function $G(s)$. The value of $C(s)$ is

$$C(s) = X^*(s)G(s) \tag{15.33}$$

The response $c(t)$ is that due to the sum of the impulses which occur before time t, thus,

$$\begin{aligned}
c(t) &= x(0)g(t) && 0 \leq t < T \\
c(t) &= x(0)g(t) + x(T)g(t - T) && T \leq t < 2T \\
c(t) &= x(0)g(t) + x(T)g(t - T) + x(2T)g(t - 2T) && 2T \leq t < 3T \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned} \tag{15.34}$$

where $g(t) = \mathfrak{L}^{-1}[G(s)]$ represents the response to an impulse occurring at $t = 0$, $g(t - T)$ is the response to an impulse occurring at $t = \tau$,

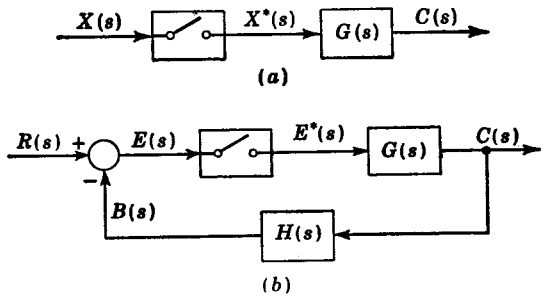


FIG. 15.13. (*a*) Sampling switch followed by a linear element; (*b*) sampled-data system.

etc. The value of $c(nT)$ at the sampling instants is

$$\begin{aligned}
c(0) &= x(0)g(0) \\
c(T) &= x(0)g(T) + x(T)g(0) \\
c(2T) &= x(0)g(2T) + x(T)g(T) + x(2T)g(0) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
c(nT) &= x(0)g(nT) + x(T)g[(n-1)T] + x(2T)g[(n-2)T] + \cdots
\end{aligned} \tag{15.35}$$

It follows from Eq. (15.31) that $C^*(s)$ is

$$C^*(s) = c(0) + c(T)e^{-Ts} + c(2T)e^{-2Ts} + \cdots \tag{15.36}$$

With the aid of Eqs. (15.35) $C^*(s)$ may be written in the form

$$\begin{aligned}
C^*(s) = {} & x(0)[g(0) + g(T)e^{-Ts} + g(2T)e^{-2Ts} + \cdots] \\
& + x(T)e^{-Ts}[g(0) + g(T)e^{-Ts} + g(2T)e^{-2Ts} + \cdots] \\
& + x(2T)e^{-2Ts}[g(0) + g(T)e^{-Ts} + g(2T)e^{-2Ts} + \cdots]
\end{aligned} \tag{15.37}$$

or

$$\begin{aligned}
C^*(s) = {} & [x(0) + x(T)e^{-Ts} + x(2T)e^{-2T} + \cdots][g(0) + g(T)e^{-Ts} \\
& + g(2T)e^{-2Ts} + \cdots]
\end{aligned} \tag{15.38}$$

or

$$C^*(s) = X^*(s)G^*(s) \tag{15.39}$$

In Fig. 15.13*b* is shown a sampled-data system in which the actuating signal is the sampled quantity. For the comparator it follows that

$$E(s) = R(s) - B(s)$$

and thus

$$E^*(s) = R^*(s) - B^*(s) \tag{15.40}$$

The equation of operation for the sampled-data system of Fig. 15.13*b* is obtained as follows:

$$B(s) = E^*(s)G(s)H(s) \tag{15.41}$$

The preceding expression has the same form as Eq. (15.33) and thus may also be expressed in the form given by Eq. (15.39), i.e.,

$$B^*(s) = E^*(s)[G(s)H(s)]^* = E^*(s)GH^*(s) \tag{15.42}$$

where $GH(s)$ is another representation for the product $G(s)H(s)$. It is to be noted that $GH^*(s)$ is quite different from $G^*(s)H^*(s)$. Equation (15.40) describes the operation of the comparator, thus,

$$E^*(s) = R^*(s) - B^*(s) = R^*(s) - E^*(s)GH^*(s)$$

or

$$E^*(s) = \frac{R^*(s)}{1 + GH^*(s)} \tag{15.43}$$

The output $C(s)$ is

$$C(s) = E^*(s)G(s) \tag{15.44}$$

Substitution of $E^*(s)$ from Eq. (15.43) into Eq. (15.44) gives the following over-all relationship for $C(s)$:

$$C(s) = \frac{G(s)R^*(s)}{1 + GH^*(s)} \tag{15.45}$$

The application of Laplace transforms to sampled-data systems yields starred (*) terms which signify infinite series. The simple substitution

$$z = e^{Ts} \tag{15.46}$$

greatly simplifies the analysis of such systems. The equation for $X^*(s)$ is

$$X^*(s) = x(0) + x(T)e^{-Ts} + x(2T)e^{-2Ts} + \cdots \tag{15.47}$$

The substitution of $z = e^{Ts}$ gives

$$X(z) = x(0) + \frac{x(T)}{z} + \frac{x(2T)}{z^2} + \cdots = \sum_{n=0}^{\infty} x(nT)z^{-n} \tag{15.48}$$

where $X(z)$ designates the z transform of $X^*(s)$.

The z transform for a unit step function applied to a sampler is

$$X(z) = 1 + \frac{1}{z} + \frac{1}{z^2} + \cdots = \frac{z}{z - 1} \tag{15.49}$$

The preceding series is convergent for $|z| > 1$. As was mentioned with

regard to Laplace transforms in Chap. 5, the fact that a transform is convergent for a certain range of values suffices to verify its existence.

Because $z/(z-1)$ is the z transform for a unit step function, this is also the z transform corresponding to the Laplace transform $1/s$.

For a delayed step function e^{-Ts}/s,

$$X(z) = 0 + \frac{1}{z} + \frac{1}{z^2} + \cdots = \frac{1}{z}\frac{z}{z-1} = \frac{1}{z-1} \tag{15.50}$$

Thus, the effect of a delaying factor is to multiply the basic transform by $1/z$.

For an exponential e^{-at} applied to a sampler, the z transform is

$$X(z) = 1 + \frac{e^{-aT}}{z} + \left(\frac{e^{-aT}}{z}\right)^2 + \cdots = \frac{z}{z - e^{-aT}} \tag{15.51}$$

The preceding z transform for e^{-at} is also the z transform corresponding to the Laplace transform $1/(s+a)$. In Table 15.1 is given a partial listing of z transforms and corresponding Laplace transforms.

TABLE 15.1. z TRANSFORMS

Description	Time function	Laplace transform	z transform
Unit impulse	$u_1(t)$	1	1
Unit step	$u(t)$	$\frac{1}{s}$	$\frac{z}{z-1}$
Ramp	t	$\frac{1}{s^2}$	$\frac{Tz}{(z-1)^2}$
Quadratic	$\frac{t^2}{2}$	$\frac{1}{s^3}$	$\frac{T^2z(z+1)}{2(z-1)^3}$
Exponential	e^{-at}	$\frac{1}{s+a}$	$\frac{z}{z-e^{-aT}}$
Exponential times $\cos \frac{\pi}{T}t$	$e^{-at}\cos \frac{\pi}{T}t$	.....	$\frac{z}{z+e^{-aT}}$
Constant raised to power t	$a^{t/T}$	$\frac{1}{s-(1/T)na}$	$\frac{z}{z-a}$ $(a>0)$
Constant raised to power t times $\cos \frac{\pi}{T}t$	$a^{t/T}\cos \frac{\pi}{T}t$	.....	$\frac{z}{z+a}$ $(a>0)$
Sinusoidal	$\sin \omega t$	$\frac{\omega}{s^2+\omega^2}$	$\frac{z \sin \omega T}{z^2 - 2z\cos \omega T + 1}$
Cosine	$\cos \omega t$	$\frac{s}{s^2+\omega^2}$	$\frac{z^2 - z\cos \omega T}{z^2 - 2z\cos \omega T + 1}$
Multiplication by e^{-at}	$e^{-at}f(t)$	$F(s+a)$	$F(e^{+aT}+z)$
Delay by time nT	$f(t-nT)$	$e^{-nTs}F(s)$	$\frac{1}{z^n}F(z)$

If a first-order system of the form $1/(s + a)$ follows a holding circuit, the z transform is obtained as follows:

$$G(s) = \frac{1 - e^{-Ts}}{s(s + a)}$$
$$= \frac{1}{a}\left(\frac{1}{s} - \frac{1}{s + a}\right)(1 - e^{-Ts}) \tag{15.52}$$

The z transform is

$$G(z) = \frac{1}{a}\left(\frac{z}{z - 1} - \frac{z}{z - e^{-aT}}\right)\left(1 - \frac{1}{z}\right)$$
$$= \frac{1}{a}\frac{1 - e^{-aT}}{z - e^{-aT}} \tag{15.53}$$

The over-all equation of operation in terms of z transforms for the system shown in Fig. 15.13*b* is obtained as follows,

$$C^*(s) = E^*(s)G^*(s)$$

whence

$$C(z) = E(z)G(z) \tag{15.54}$$

The substitution of the z representation for the corresponding starred (*) terms in Eq. (15.43) gives

$$E(z) = \frac{R(z)}{1 + GH(z)} \tag{15.55}$$

From Eqs. (15.54) and (15.55), the over-all relationship is

$$C(z) = \frac{G(z)R(z)}{1 + GH(z)} \tag{15.56}$$

Illustrative Example. Let it be desired to determine the unit step-function response for the system shown in Fig. 15.13*b*, in which

$$G(s) = \frac{1 - e^{-Ts}}{s}\frac{K_1}{s} = \frac{K_1}{s^2}(1 - e^{-Ts})$$

and $H(s) = 1$ for unity feedback.

SOLUTION. For this system $G(z) = GH(z)$, which is

$$G(z) = \frac{K_1Tz}{(z - 1)^2}\left(1 - \frac{1}{z}\right) = \frac{K_1T}{z - 1} \tag{15.57}$$

where from Table 15.1 the transform for a ramp function $1/s^2$ is $Tz/(z - 1)^2$. The over-all z transform for $C(z)$ is

$$C(z) = \frac{K_1T}{z - (1 - K_1T)}\frac{z}{z - 1} = \frac{z}{z - 1} - \frac{z}{z - (1 - K_1T)} \tag{15.58}$$

For $1 - K_1T > 0$, the inverse transform of the preceding expression is

$$c(t) = 1 - (1 - K_1T)^{t/T} \tag{15.59}$$

Thus the system is stable if $0 < 1 - K_1T < 1$. When $1 - K_1T < 0$ or $K_1T - 1 > 0$, the inverse transformation of Eq. (15.58) is

$$c(t) = 1 - (K_1T - 1)^{t/T} \cos \frac{\pi t}{T} \tag{15.60}$$

The preceding response is oscillatory. For stability, it is necessary that $0 < K_1T - 1 < 1$.

The application of z transforms is thus seen to be analogous to that of Laplace transforms.

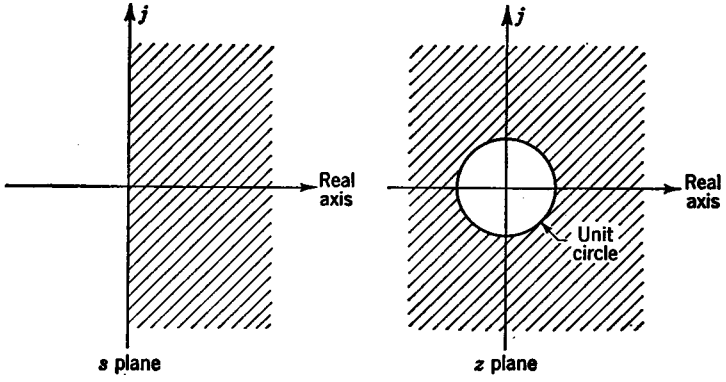


FIG. 15.14. Stability regions for s plane and z plane.

Characteristic Function. In Chap. 6, it was found that a system is unstable if any zeros of the characteristic function are in the right half plane. The right half of the s plane may be designated by $\sigma + j\omega$, in which $\sigma > 0$. The corresponding portion of the z plane is obtained by noting that

$$z = e^{sT} = e^{\sigma T} e^{j\omega T}$$

or

$$|z| = |e^{\sigma T}|\,|e^{j\omega T}| = e^{\sigma T} \tag{15.61}$$

From the preceding result, it follows that, for $\sigma > 0$, then $|z| > 1$. As illustrated in Fig. 15.14, the right half of the s plane corresponds to a value of z outside the unit circle. Thus, for stability, all the zeros of the z-transformed characteristic function must lie within the unit circle. For the preceding example, the characteristic function is

$$D_{G(z)}D_{H(z)} + N_{G(z)}N_{H(z)} = z - (1 - K_1T) \tag{15.62}$$

This is a first-order equation whose zero is

$$z_1 = 1 - K_1T \tag{15.63}$$

For stability, it follows that

$$|z_1| = |1 - K_1T| < 1 \tag{15.64}$$

This is the result previously obtained from an examination of the equation for the time response.

The root-locus methods may also be extended to sampled-data systems. For example, the characteristic function for this system may be written in the form

$$z + K = 0 \tag{15.65}$$

where $K = K_1T - 1$.

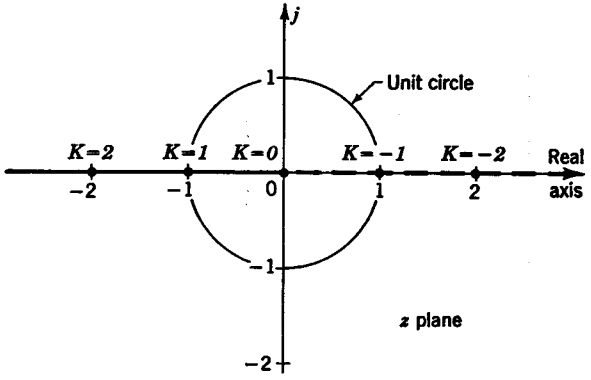


FIG. 15.15. Root-locus plot on z plane.

The solid line of Fig. 15.15 is the root-locus plot for positive values of K, and the dashed line is the corresponding plot for negative values of K. In applying the angle condition for negative values of K the resulting angle must be $0° \pm 360°k$ rather than $180° \pm 360°k$, where

$$k = 0, 1, 2, 3, \ldots$$

Frequency Response. In applying frequency-response methods to sampled-data systems, the zeros of the characteristic function are the zeros of the function $1 + GH^*(s) = 1 + GH(z)$. To illustrate the application of frequency-response techniques to sampled-data systems, let it be desired to determine the value of K_1T so that the preceding system will have an M_m of 1.4. The function $GH^*(s)$ is

$$GH^*(s) = GH(z) = \frac{K_1T}{z - 1}$$

or

$$\frac{GH^*(s)}{K_1T} = \frac{1}{z - 1} \tag{15.66}$$

The path of values for z is the unit circle ($z = e^{j\beta}$) shown in Fig. 15.16*a*.

The vector drawn from the $+1$ point to the unit circle is the vector $z - 1$. The corresponding $GH^*(s)$ plot is shown in Fig. 15.16*b*, whence the required value of K_1T is

$$K_1T = \frac{1}{0.8} = 1.25 \tag{15.67}$$

Thus, in general, the methods previously developed for linear systems may be extended to sampled-data systems.

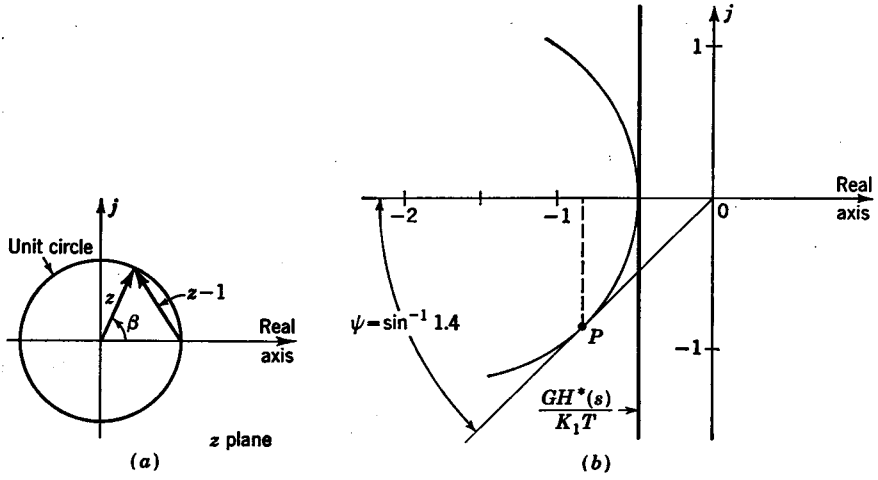


FIG. 15.16. (*a*) Path of values for z; (*b*) frequency response $\dfrac{GH^*(s)}{K_1T} = \dfrac{1}{z-1}$.

15.6. Phase-plane Techniques. The phase-plane method is basically a graphical procedure for determining the transient response of a second-order system[1–4] which is of the general form

$$\ddot{y} + f_1(y,\dot{y})\dot{y} + f_2(y,\dot{y})y = C \tag{15.68}$$

where $f_1(y,\dot{y})$ = a function of y and $\dot{y}$
$f_2(y,\dot{y})$ = another function of $\dot{y}$ and y
C = a constant

For the case in which $f_1(y,\dot{y})$ and $f_2(y,\dot{y})$ are both constant, Eq. (15.68) reduces to a linear differential equation with constant coefficients.

[1] Kochenburger, *op. cit.*
[2] D'azzo and Houpis, *op. cit.*
[3] R. L. Cosgriff, "Nonlinear Control Systems," McGraw-Hill Book Company, Inc., New York, 1958.
[4] W. J. Cunningham, "Introduction to Nonlinear Analysis," McGraw-Hill Book Company, Inc., New York, 1958.

By making the substitution that $v = \dot{y} = dy/dt$, Eq. (15.68) may be reduced to a first-order differential equation as follows:

$$\ddot{y} = \frac{dv}{dt} = \frac{dy}{dt}\frac{dv}{dy} = v\frac{dv}{dy} \tag{15.69}$$

Thus,
$$v\frac{dv}{dy} + f_1(y,v)v + f_2(y,v)y = C$$

or
$$\frac{dv}{dy} = -f_1(y,v) - \frac{f_2(y,v)y}{v} + \frac{C}{v} \tag{15.70}$$

To illustrate the basic principles involved in the phase-plane method, the mass-spring system shown in Fig. 15.17 will first be investigated.

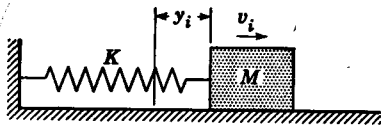


FIG. 15.17. Mass-spring system.

The equation of motion for this system is

$$M\ddot{y} + Ky = 0 \tag{15.71}$$

By letting $dy/dt = v$, so that $\ddot{y} = v(dv/dy)$,

$$Mv\frac{dv}{dy} = -Ky \tag{15.72}$$

The solution of the preceding first-order differential equation is

$$\int_{v_i}^{v} v\,dv = -\frac{K}{M}\int_{y_i}^{y} y\,dy \tag{15.73}$$

or
$$v^2 + \frac{K}{M}y^2 = v_i^2 + \frac{K}{M}y_i^2 = K_1 \tag{15.74}$$

where K_1 is a constant, y_i is the initial displacement of the mass, and v is the initial velocity.

Plotting Eq. (15.74) in the y, v coordinate system would yield an ellipse. Because any ellipse may be plotted as a circle by changing the scale factor, it is apparent from Eq. (15.74) that the desired coordinates are $\sqrt{(K/M)}\,y$ and v.

A phase-plane trajectory is a plot of all corresponding values of y and v for a given system. In particular, Eq. (15.74) is the equation for the phase-plane trajectory shown in Fig. 15.18*a*. For different values of the constant K_1 in Eq. (15.74) different trajectories are obtained as shown in Fig. 15.18*b*. This family of trajectories is called a phase

portrait. The particular trajectory describing the operation of the system is determined by the initial operating point $(\sqrt{K/M}\, y_i, v_i)$.

Because $y = \int v\, dt$, for positive values of v the position y increases and for negative velocities the value of y decreases. Thus, corresponding values of y and v proceed in a clockwise direction along the phase trajectories, as indicated by the arrows in Fig. 15.18. From Eq. (15.69), it follows that the slope of the phase trajectories is $\dfrac{dv}{dy} = \dfrac{1}{v}\dfrac{d^2y}{dt^2}$. Because v is zero when a trajectory crosses the y axis, then $dv/dy = \infty$. Thus, each trajectory crosses the y axis at right angles.

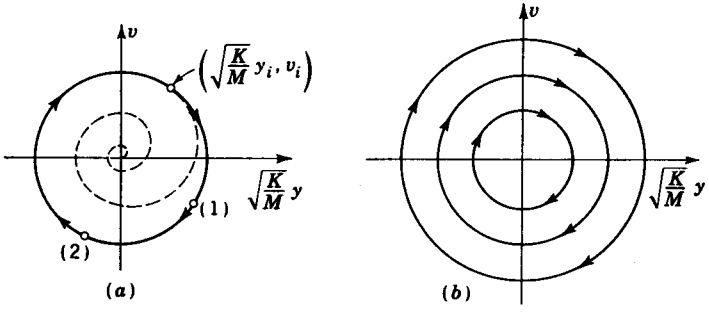


FIG. 15.18. Phase-plane trajectories for mass-spring system.

The time t required for the operating point to move from any station (1) on a trajectory to a second station (2), as indicated in Fig. 15.18*a*, may be computed as follows,

$$dt = \frac{dy}{v}$$

or
$$t_2 - t_1 = \int_{y_1}^{y_2} \frac{dy}{\sqrt{K_1 - (K/M)y^2}} \tag{15.75}$$

The trajectory for any conservative system in which there is no damping to dissipate energy must be a closed path. If there were damping in the system, the path of operation would be a spiral in toward the origin, as is indicated by the dashed line of Fig. 15.18*a*.

Coulomb Friction. When there is coulomb friction existing between the mass M of Fig. 15.17 and the surface over which it is sliding, the equation of motion is

$$M\ddot{y} + Ky = -\mu Mg \qquad v > 0 \tag{15.76}$$
$$M\ddot{y} + Ky = \mu Mg \qquad v < 0 \tag{15.77}$$

where μ is the coefficient of friction.

By making the substitution $y_r = y + \mu Mg/K$, Eq. (15.76) becomes

$$M\ddot{y}_r + Ky_r = 0 \tag{15.78}$$

It is seen that $\dot{y}_r = \dot{y} = v$ and $\ddot{y}_r = \ddot{y}$. The preceding equation is similar to Eq. (15.71), and thus its solution is

$$v^2 + \frac{Ky_r^2}{M} = v^2 + \frac{K}{M}\left(y + \frac{\mu Mg}{K}\right)^2 = K_2 \tag{15.79}$$

The phase portrait for the preceding expression is a family of circles of radius $\sqrt{K_2}$ whose center is at the point $y = -\mu Mg/K$. Because Eq. (15.76) is valid for $v > 0$, only semicircles above the y axis as shown in

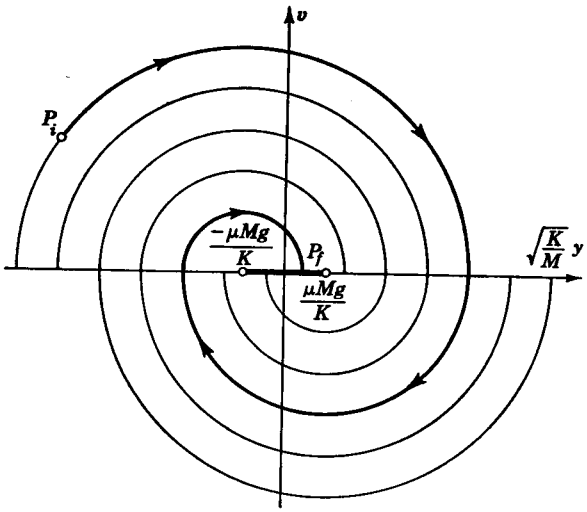


FIG. 15.19. Phase portrait for mass-spring system with friction.

Fig. 15.19 may be obtained from this expression. For the case in which $v < 0$, the substitution $y_L = y - \mu Mg/K$ is made in Eq. (15.77). This yields

$$M\ddot{y}_L + Ky_L = 0 \tag{15.80}$$

The solution of the preceding expression is

$$v^2 + \frac{Ky_L^2}{M} = v^2 + \frac{K}{M}\left(y - \frac{\mu Mg}{K}\right)^2 = K_3 \tag{15.81}$$

Equation (15.81) yields the family of semicircles drawn below the y axis of Fig. 15.19.

If the initial position and velocity of the mass are y_i, v_i as indicated by P_i in Fig. 15.19, the path of motion will be that of the heavy line. The final position at which the mass comes to rest is indicated by P_f. Because of coulomb friction the at-rest position is not at the origin of the phase portrait.

Graphical Determination of Trajectories. For the general case of a nonlinear system, it is not possible to obtain a mathematical expression for the trajectories. When this is so, the trajectories may be constructed by a graphical interpretation of Eq. (15.70), i.e.,

$$\frac{dv}{dy} = F(y,v) = m \tag{15.82}$$

where $m = F(y,v) = -f_1(y,v) - f_2(y,v)(y/v) + C/v$ is the tangent to the trajectory at the point (y,v). For a given slope m, Eq. (15.82)

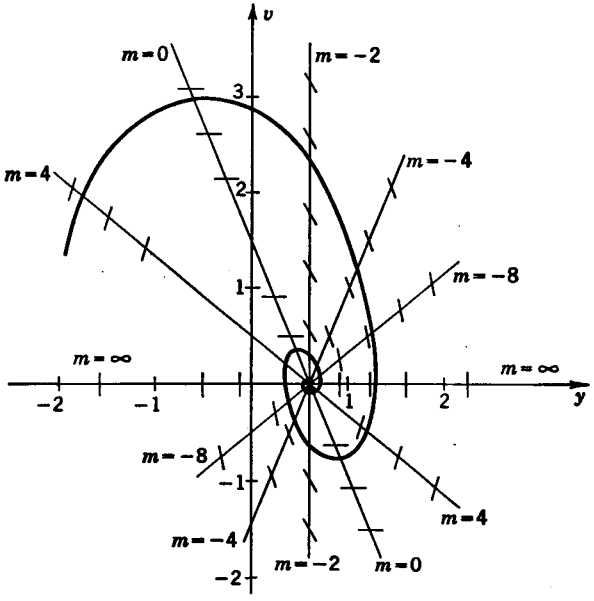


FIG. 15.20. Construction of phase portrait from isoclines.

describes a line such that every trajectory has the same slope m as it crosses this line. Such a line is called an isocline. In Fig. 15.20 is shown a typical family of isoclines, which are constructed as follows: Consider the equation

$$\ddot{y} + 2\dot{y} + 5y = 3 \tag{15.83}$$

where $f_1(y,v) = 2$, $f_2(y,v) = 5$, and $C = 3$. The equation for the isoclines is

$$m = -2 - 5\frac{y}{v} + \frac{3}{v}$$

or

$$v = \frac{-5y}{m+2} + \frac{3}{m+2} \tag{15.84}$$

As was illustrated by this example, the isoclines are straight lines when $f_1(y,v)$ and $f_2(y,v)$ are constant.

APPENDIX I

CORRELATION BETWEEN LAPLACE TRANSFORM, FOURIER SERIES, AND FOURIER INTEGRAL

A greater understanding of the Laplace transform $F(s)$ of a time function $f(t)$ may be obtained by examining the similarities which exist between Laplace transforms and the more familiar Fourier series and Fourier integral.

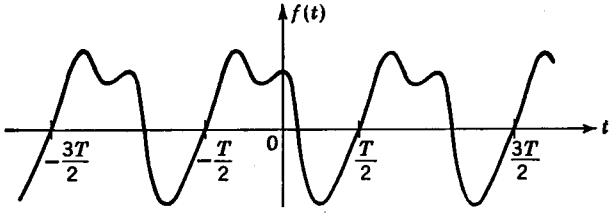


FIG. I.1. Periodic function.

Fourier Series. A periodic function as shown in Fig. I.1 may be represented by the series

$$f(t) = A_0 + \sum_{n=1}^{\infty} (A_n \cos n\omega_0 t + B_n \sin n\omega_0 t) \tag{I.1}$$

where $\omega_0 = 2\pi/T$, in which T is the period. The constants A_0, A_n, and B_n are evaluated as follows: Integration of each term in Eq. (I.1) over a complete period causes each term in the summation on the right-hand side to vanish. Thus

$$\int_{-T/2}^{T/2} f(t)\, dt = A_0 \int_{-T/2}^{T/2} dt + 0 = A_0 T$$

or

$$A_0 = \frac{1}{T} \int_{-T/2}^{T/2} f(t)\, dt \tag{I.2}$$

The value of A_0 is seen to be equal to the average value of the function over a period.

To evaluate A_n each term of Eq. (I.1) is multiplied by $\cos m\omega_0 t$ and

then integrated over a period. Because

$$\int_{-T/2}^{+T/2} \cos n\omega_0 t \cos m\omega_0 t \, dt = \begin{cases} 0 & \text{for } m \neq n \\ \frac{T}{2} & \text{for } m = n \end{cases} \tag{I.3}$$

and

$$\int_{-T/2}^{+T/2} \sin n\omega_0 t \cos m\omega_0 t \, dt = 0 \tag{I.4}$$

it follows that

$$A_n = \frac{2}{T} \int_{-T/2}^{+T/2} f(t) \cos n\omega_0 t \, dt \tag{I.5}$$

Similarly, multiplication of each term of Eq. (I.1) by $\sin m\omega_0 t$ and integration over the period yields the following result for B_n

$$B_n = \frac{2}{T} \int_{-T/2}^{+T/2} f(t) \sin n\omega_0 t \, dt \tag{I.6}$$

Equation (I.1) may be telescoped into a more convenient form by using Eqs. (5.37) and (5.38) to express the cosine and sine in exponential form.

$$A_n \cos n\omega_0 t = \frac{A_n}{2} (e^{jn\omega_0 t} + e^{-jn\omega_0 t})$$

and

$$B_n \sin n\omega_0 t = -j \frac{B_n}{2} (e^{jn\omega_0 t} - e^{-jn\omega_0 t})$$

Thus

$$f(t) = A_0 + \frac{1}{2} \sum_{n=1}^{\infty} (A_n - jB_n) e^{jn\omega_0 t} + (A_n + jB_n) e^{-jn\omega_0 t} \tag{I.7}$$

By also writing Eqs. (I.5) and (I.6) in exponential form, it can be seen that

$$A_n - jB_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t)(\cos n\omega_0 t - j \sin n\omega_0 t) \, dt = \frac{2}{T} \int_{-T/2}^{T/2} f(t) e^{-jn\omega_0 t} \, dt \tag{I.8}$$

and

$$A_n + jB_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t)(\cos n\omega_0 t + j \sin n\omega_0 t) \, dt = \frac{2}{T} \int_{-T/2}^{T/2} f(t) e^{jn\omega_0 t} \, dt \tag{I.9}$$

Substitution of the preceding results into Eq. (I.7) gives

$$f(t) = A_0 + \frac{1}{T} \sum_{n=1}^{\infty} e^{jn\omega_0 t} \int_{-T/2}^{+T/2} f(t) e^{-jn\omega_0 t} \, dt + \frac{1}{T} \sum_{n=1}^{\infty} e^{-jn\omega_0 t} \int_{-T/2}^{+T/2} f(t) e^{jn\omega_0 t} \, dt \tag{I.10}$$

By noting that the last summation is unaltered by changing the sign of n, the Fourier series becomes

$$f(t) = A_0 + \sum_{n=1}^{\infty} \frac{e^{jn\omega_0 t}}{T} \int_{-T/2}^{+T/2} f(t)e^{-jn\omega_0 t}\,dt + \sum_{n=-1}^{-\infty} \frac{e^{jn\omega_0 t}}{T} \int_{-T/2}^{+T/2} f(t)e^{-jn\omega_0 t}\,dt \quad \text{(I.11)}$$

Because the value of the summation for $n = 0$ is A_0,

$$f(t) = \sum_{n=-\infty}^{\infty} \frac{e^{jn\omega_0 t}}{T} \int_{-T/2}^{+T/2} f(t)e^{-jn\omega_0 t}\,dt \quad \text{(I.12)}$$

Equation (I.12) is frequently written in the form

$$f(t) = \sum_{n=-\infty}^{\infty} C_n e^{jn\omega_0 t} \quad \text{(I.13)}$$

where

$$C_n = \frac{1}{T} \int_{-T/2}^{+T/2} f(t)e^{-jn\omega_0 t}\,dt$$

Fourier Integral. As the period T becomes infinite, the Fourier series expression given by Eq. (I.12) is

$$f(t) = \lim_{T\to\infty} \left[\sum_{n=-\infty}^{\infty} \frac{e^{jn\omega_0 t}}{T} \int_{-T/2}^{+T/2} f(t)e^{-jn\omega_0 t}\,dt \right] \quad \text{(I.14)}$$

By letting $\omega = n\omega_0$ and

$$\frac{\Delta\omega}{2\pi} = \lim_{T\to\infty} \frac{n\omega_0 - (n-1)\omega_0}{2\pi} = \lim_{T\to\infty} \frac{1}{T}$$

Eq. (I.14) becomes

$$f(t) = \lim_{\substack{\Delta\omega\to 0 \\ T\to\infty}} \left[\frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{j\omega t}\,\Delta\omega \int_{-T/2}^{+T/2} f(t)e^{-j\omega t}\,dt \right] \quad \text{(I.15)}$$

The limit of Eq. (I.15) is the Fourier integral

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} \left[\int_{-\infty}^{\infty} f(t)e^{-j\omega t}\,dt \right] d\omega \quad \text{(I.16)}$$

The Fourier integral is frequently expressed by the Fourier transform

pair

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(j\omega)e^{j\omega t}\,d\omega \tag{I.17}$$

$$F(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\,dt \tag{I.18}$$

Equation (I.18) is referred to as the direct Fourier transformation, and Eq. (I.17) is the inverse Fourier transformation.

For most physical problems, all it is desired to know is the solution for $t > 0$. Thus, only the initial conditions and information concerning $f(t)$ for $t > 0$ need be considered, in which case the lower limit of integration in Eq. (I.18) may be taken as zero.

To illustrate the use of the Fourier transform, consider the function

$$f(t) = e^{at} \qquad t \geq 0 \tag{I.19}$$

Application of the direct Fourier transform gives

$$F(j\omega) = \int_0^{\infty} e^{at}e^{-j\omega t}\,dt = \frac{e^{at}e^{-j\omega t}}{a - j\omega}\bigg|_0^{\infty} \tag{I.20}$$

If the exponent a is less than zero, the preceding expression becomes

$$F(j\omega) = 0 - \frac{1}{a - j\omega} = \frac{1}{j\omega - a} \qquad a < 0 \tag{I.21}$$

However, if the exponent a is positive, e^{at} becomes infinite when evaluated at $t = \infty$ and thus $F(j\omega)$ diverges.

Laplace Transform. To extend the usefulness of the Fourier transform so that it is applicable to divergent functions, a converging factor $e^{-\sigma t}$ is introduced. Thus, the general transform equation is

$$\begin{aligned} F(\sigma,j\omega) &= \int_0^{\infty} f(t)e^{-\sigma t}e^{-j\omega t}\,dt \\ &= \int_0^{\infty} f(t)e^{-(\sigma+j\omega)t}\,dt \end{aligned} \tag{I.22}$$

The transform for the time function given by Eq. (I.19) is

$$\begin{aligned} F(\sigma,j\omega) &= \frac{e^{(a-\sigma-j\omega)t}}{a - \sigma - j\omega}\bigg|_0^{\infty} \\ &= \frac{1}{(\sigma + j\omega) - a} \qquad \sigma > a \end{aligned} \tag{I.23}$$

The preceding expression is seen to converge when σ is greater than a. To ensure convergence of the Fourier transform, it was necessary that $\int_0^{\infty} |f(t)|\,dt < \infty$. However, the transform indicated by Eq. (I.22) converges when $\int_0^{\infty} |f(t)|e^{-\sigma t}\,dt < \infty$ for some finite σ.

The substitution of $s = \sigma + j\omega$ and $F(s) = F(\sigma,j\omega)$ into Eq. (I.22) yields the Laplace transform equation.

$$F(s) = \int_0^\infty f(t)e^{-st}\,dt \tag{I.24}$$

It is necessary only that some finite value of σ exists such that $\int_0^\infty |f(t)|e^{-\sigma t}\,dt < \infty$ in order to verify the existence of the transform indicated by Eq. (I.24). For most functions $f(t)$ encountered in engineering work, the transform $F(s)$ is convergent. It should also be noted that, to solve differential equations by Laplace transforms, it is not necessary to determine the value or values of σ over which $F(s)$ is convergent. It suffices to know that such a value or values of σ exist.

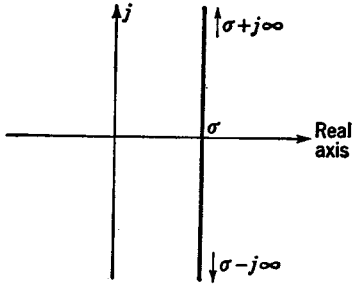


FIG. I.2. Vertical line $s = \sigma + j\omega$.

In effect, Eq. (I.24) is the result of substituting s for $j\omega$ and $F(s)$ for $F(j\omega)$ in Eq. (I.18). The use of these same substitutions in Eq. (I.17) yields the inverse transform, i.e.,

$$f(t) = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} F(s)e^{st}\,ds \tag{I.25}$$

The new limits of integration are obtained by noting that, when $\omega = \pm\infty$, then $s = \sigma \pm j\omega = \sigma \pm j\infty$. Equation (I.25) is a line integral for which the path of integration is a vertical line which is displaced a distance σ from the imaginary axis, as is shown in Fig. I.2. For convergence, it is necessary that σ be such that all the values of s which make $F(s)$ infinite [i.e., poles of $F(s)$] lie to the right of the vertical line shown in Fig. I.2.

APPENDIX II

RESPONSE OF SYSTEM TO AN ARBITRARY INPUT

For some input functions $x(t)$, such as an arbitrary input, the transform $X(s)$ is not easily obtained or is a very complicated expression. For such cases, it is desired to be able to determine the transient response without knowing the value of $X(s)$.

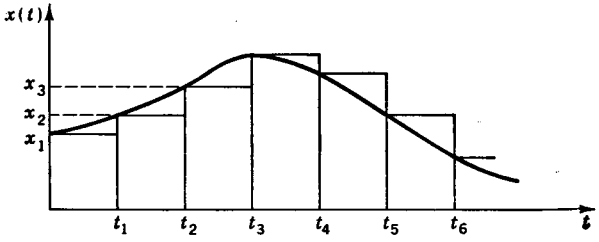


FIG. II.1. Arbitrary function $x(t)$.

The arbitrary input $x(t)$ shown in Fig. II.1 could be approximated by a sum of pulses, i.e.,

$$x(t) \approx x_1[u(t) - u(t - t_1)] + x_2[u(t - t_1) - u(t - t_2)] + \cdots + x_n[u(t - t_{n-1}) - u(t - t_n)] \quad \text{(II.1)}$$

For this case, the response of the system $y(t)$ at time t is obtained by adding the individual responses due to each pulse, as is shown graphically in Fig. II.2; i.e.,

$$y(t) = y_1(t) + y_2(t) + \cdots \quad \text{(II.2)}$$

where $y_1(t)$ is the response due to the first pulse, etc. With the preceding method, accuracy is increased by using more pulses of a smaller width to approximate the input. In a similar manner, it is possible to approximate the input by a series of step changes.

A considerable saving in computational time and effort is realized by the use of the convolution integral. In addition, the convolution-integral method yields an exact rather than an approximate solution. It should first be noted that the general transformed expression for $Y(s)$ is

$$Y(s) = \frac{L_m(s)X(s)}{L_n(s)} + \frac{I(s)}{L_n(s)} = W(s)X(s) + \frac{I(s)}{L_n(s)} \quad \text{(II.3)}$$

where $W(s) = L_m(s)/L_n(s)$.

The inverse transformation of Eq. (II.3) yields the desired time response

$$y(t) = \mathfrak{L}^{-1}[W(s)X(s)] + \mathfrak{L}^{-1}\left[\frac{I(s)}{L_n(s)}\right] \tag{II.4}$$

The last term of the preceding expression is evaluated directly from a knowledge of the initial conditions. For an arbitrary input for which

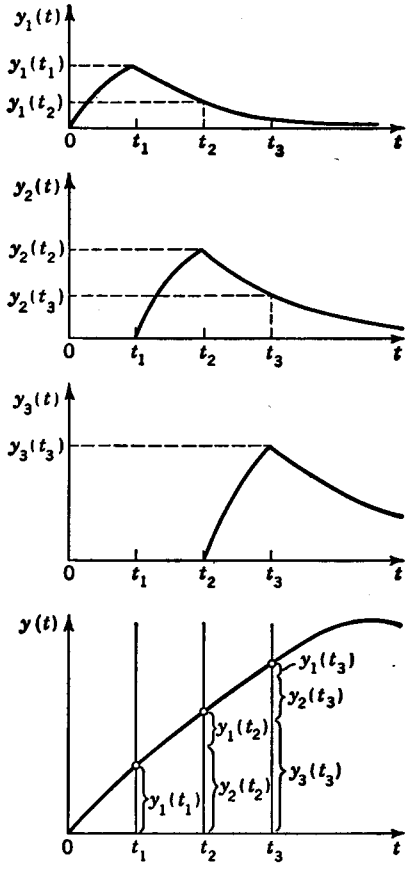


FIG. II.2. Response at time t_n.

$X(s)$ is not known, the first term may be determined by application of the following convolution integral,

$$\mathfrak{L}^{-1}[W(s)X(s)] = \int_0^\infty w(\lambda)x(t-\lambda)\,d\lambda \tag{II.5}$$

where $w(\lambda) = \mathfrak{L}^{-1}\{[L_m(s)/L_n(s)](1)\} = \mathfrak{L}^{-1}[W(s)]$ is the unit impulse response of the system and may be computed directly since $W(s)$ is known.

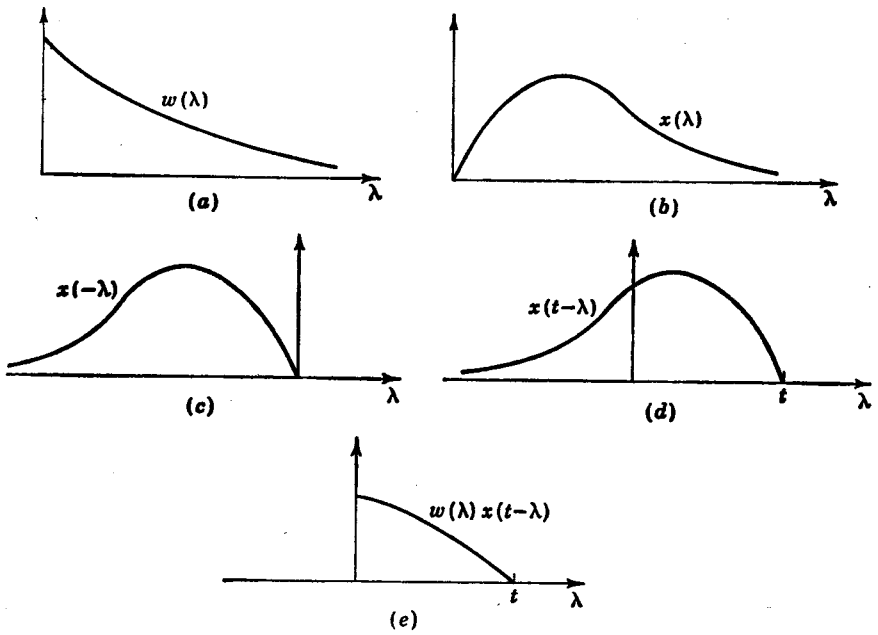


FIG. II.3. Physical interpretation of convolution integral.

It is necessary to introduce the symbol λ to distinguish the two time terms which appear in Eq. (II.5). A plot of the impulse response $w(t)$ versus t would be identical to a plot of $w(\lambda)$ versus λ.

The convolution integral given by Eq. (II.5) is verified as follows: First take the Laplace transform of the right-hand side of Eq. (II.5).

$$\mathcal{L}\left[\int_0^\infty w(\lambda)x(t-\lambda)\,d\lambda\right] = \int_0^\infty \left[\int_0^\infty w(\lambda)x(t-\lambda)\,d\lambda\right] e^{-st}\,dt$$
$$= \int_0^\infty w(\lambda)\left[\int_0^\infty x(t-\lambda)e^{-st}\,dt\right] d\lambda \quad \text{(II.6)}$$

Application of the real-translation theorem to the integral in brackets in Eq. (II.6) gives

$$\int_0^\infty x(t-\lambda)e^{-st}\,dt = X(s)e^{-s\lambda} \quad \text{(II.7)}$$

Substitution of this result into Eq. (II.6) yields

$$\mathcal{L}\left[\int_0^\infty w(\lambda)x(t-\lambda)\,d\lambda\right] = \left[\int_0^\infty w(\lambda)e^{-s\lambda}\,d\lambda\right]X(s)$$
$$= W(s)X(s) \quad \text{(II.8)}$$

Taking the inverse transform of both sides of the preceding expression yields the result given by Eq. (II.5). A direct physical interpretation of

the convolution integral is shown in Fig. II.3. The response of the system $w(\lambda)$ to a unit impulse is shown in Fig. II.3*a*. The arbitrary input is drawn in Fig. II.3*b*. The plot of $x(\lambda)$ versus λ is identical to the graph of the input function $x(t)$ versus t. The input is reflected about the

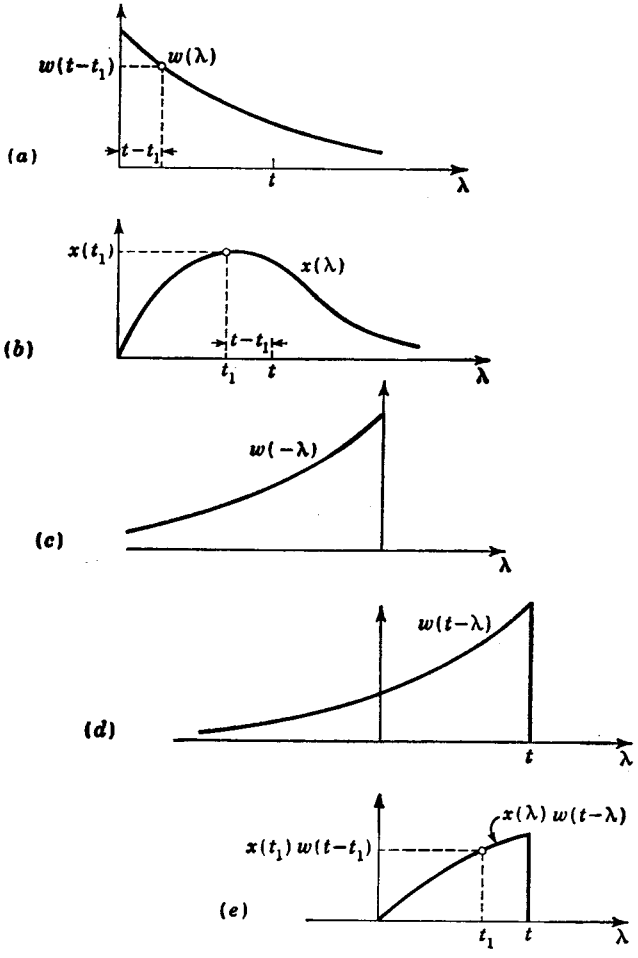


FIG. II. 4. Physical interpretation of the alternative form of the convolution integral.

$\lambda = 0$ axis to yield $x(-\lambda)$ in Fig. II.3*c*. Translating Fig. II.3*c* by t sec yields the $x(t - \lambda)$ plot shown in Fig. II.3*d*. Multiplication of corresponding values of $w(\lambda)$ and $x(t - \lambda)$ yields the curve of $w(\lambda)x(t - \lambda)$ as shown in Fig. II.3*e*. The desired response $y(t)$ at time t is equal to the area under this latter graph. From Fig. II.3*e*, it is apparent that the value of the integral given by Eq. (II.5) is zero for values of λ greater

than t. Thus, the upper limit of integration may be taken as t rather than infinity, i.e.,

$$\mathfrak{L}^{-1}[W(s)X(s)] = \int_0^t w(\lambda)x(t-\lambda)\,d\lambda \tag{II.9}$$

The complete time-response expression is

$$y(t) = \int_0^t w(\lambda)x(t-\lambda)\,d\lambda + \mathfrak{L}^{-1}\left[\frac{I(s)}{L_n(s)}\right] \tag{II.10}$$

By employing techniques similar to those used in deriving Eq. (II.5), it may also be shown that

$$\mathfrak{L}^{-1}[W(s)X(s)] = \int_0^t x(\lambda)w(t-\lambda)\,d\lambda \tag{II.11}$$

where, as before, $w(\lambda)$ is the unit impulse response and $x(\lambda)$ the arbitrary input. The graphical interpretation of the preceding expression is shown in Fig. II.4. It is interesting to consider the portion of the total response $y(t)$ due to the portion of the input which occurs at time t_1, as shown in Fig. II.4*b*. In effect the abscissa $x(t_1)$ is multiplied by the value of $w(t-t_1)$ to give $x(t_1)w(t-t_1)$. Thus, the multiplication factor for each portion of the input depends on the value of $w(\lambda)$ at $\lambda = t - t_1$. The impulse response $w(\lambda)$ is also called the *weighting function because it indicates how much the input applied $t - t_1$ sec in the past has decayed.*

APPENDIX III

OBTAINING THE FREQUENCY RESPONSE FROM THE TRANSIENT RESPONSE

Several techniques are available[1] for determining the frequency response of a system when the transient response is known. The method now to be described was developed by Guillemin.[2] This method possesses the advantage that the accuracy obtained with only a few terms (computational efforts thus being minimized) is better than that obtained by other techniques.

Guillemin's procedure is based primarily upon a graphical interpretation of the following equation:

$$Y(j\omega) = \frac{1}{(j\omega)^n} \int_0^\infty y^{(n)}(t) e^{-j\omega t}\, dt \tag{III.1}$$

The preceding expression is derived as follows: The transform for the nth derivative of a function $y(t)$ for which the initial conditions are zero is given by

$$\mathcal{L}[y^{(n)}(t)] = \int_0^\infty y^{(n)}(t) e^{-st}\, dt = s^n Y(s)$$

or

$$Y(s) = \frac{1}{s^n} \int_0^\infty y^{(n)}(t) e^{-st}\, dt \tag{III.2}$$

The substitution of $j\omega$ for s in the preceding expression yields the result given by Eq. (III.1). By comparison of Eq. (III.1) and Eq. (I.18) it follows that differentiation of $y(t)$ with respect to time corresponds to multiplication of $Y(j\omega)$ by $j\omega$.

The application of Eq. (III.1) for obtaining the frequency response is next demonstrated. In Fig. III.1*a* is shown a typical function $y(t)$.

[1] R. C. Seaman, Jr., B. P. Blasingame, and G. C. Clementson, The Pulse Method for Determination of Aircraft Performance, *Aeronaut. Sci.*, vol. 17, no. 1, pp. 22–38, January, 1950.

[2] E. A. Guillemin, Computational Techniques Which Simplify the Correlation between Steady-state and Transient Responses of Filters and Other Networks, *Proc. Natl. Electronics Conf.*, 1953, vol. 9, 1954.

The straight-line approximation to this function is indicated by $y^*(t)$. The first derivative $y^{*(1)}(t)$ shown in Fig. III.1*b* is seen to be a series of steps. The height of each step is equal to the slope of the corresponding portion of $y^*(t)$. The second derivative $y^{*(2)}(t)$ yields the train of impulses shown in Fig. III.1*c*. The area of each impulse is equal to the vertical distance between steps of Fig. III.1*b*.

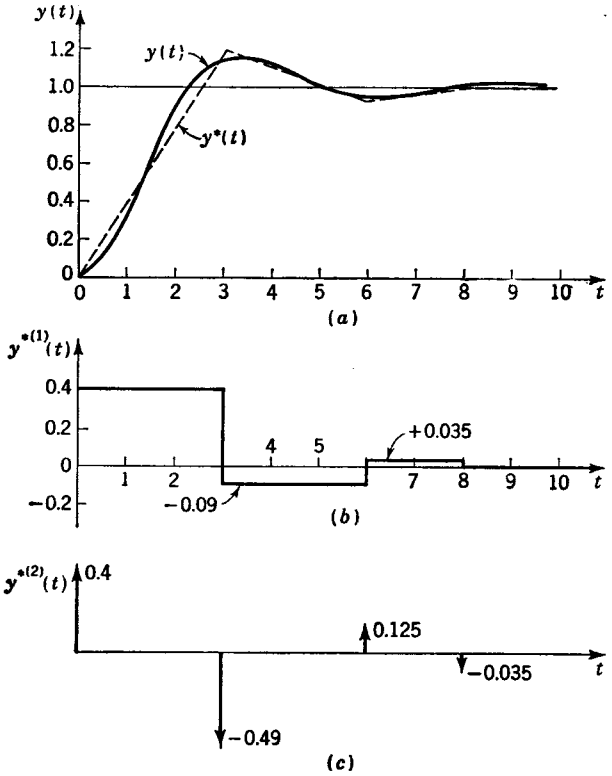


FIG. III.1. Differentiation of $y^*(t)$ to obtain a train of impulses.

The second derivative $y^{*(2)}(t)$ may be written in the form

$$y^{*(2)}(t) = \sum_{k=1}^{\nu} a_k u_1(t - t_k) \tag{III.3}$$

where a_k is the area of the kth impulse, $u_1(t - t_k)$ is the symbolic designation for a unit impulse which occurs at time t_k, and ν is the total number of impulses.

Substituting Eq. (III.3) into Eq. (III.1) and noting that $n = 2$ gives

$$Y^*(j\omega) = \frac{1}{(j\omega)^2} \int_0^\infty \sum_{k=1}^{\nu} a_k u_1(t - t_k) e^{-j\omega t}\, dt$$

$$= \frac{1}{(j\omega)^2} \sum_{k=1}^{\nu} a_k e^{-j\omega t_k} \qquad \text{(III.4)}$$

Application of Eq. (III.4) to the function shown in Fig. III.1 gives

$$Y^*(j\omega) = \frac{1}{(j\omega)^2} (0.40 - 0.49e^{-j3\omega} + 0.125e^{-j6\omega} - 0.035e^{-j8\omega}) \qquad \text{(III.5)}$$

For a given input $x(t)$ and corresponding output $y(t)$, the frequency response is $G^*(j\omega) = Y^*(j\omega)/X^*(j\omega)$. For example, if the response $y(t)$ is given by the function $y(t)$ of Fig. III.1 and $x(t)$ is a unit step function, $x^{*(1)}(t)$ is a unit impulse occurring at $t = 0$. Thus, $X^*(j\omega) = 1/j\omega$, and

$$G^*(j\omega) = \frac{Y^*(j\omega)}{X^*(j\omega)} = \frac{1}{j\omega} (0.40 - 0.49e^{-j3\omega} + 0.125e^{-j6\omega} - 0.035e^{-j8\omega}) \qquad \text{(III.6)}$$

Improved accuracy is obtained by determining $y^{(1)}(t)$ exactly and then approximating this derivative by straight lines rather than the original function $y(t)$. In effect, the function $y(t)$ is now being approximated by a series of parabolas. For this case, the second derivative $y^{*(2)}(t)$ is a series of steps, and the third derivative is a train of impulses. The approximation $Y^*(j\omega)$ now becomes

$$Y^*(j\omega) = \frac{1}{(j\omega)^3} \sum_{k=1}^{\nu} a_k e^{-j\omega t_k} \qquad \text{(III.7)}$$

APPENDIX IV

OBTAINING THE TRANSIENT RESPONSE FROM THE FREQUENCY RESPONSE

Essentially the same approach described in Appendix III may be used to determine the transient response when the frequency response of a system is known. The general operational representation for a differential equation is

$$y(t) = G(p)x(t) \tag{IV.1}$$

For the case in which the initial conditions are zero and the input $x(t)$ is a unit impulse, the transform is

$$Y(s) = G(s) \tag{IV.2}$$

The inverse transformation of Eq. (IV.2) is

$$y(t) = w(t) = \frac{1}{2\pi j}\int_{\sigma-j\omega}^{\sigma+j\omega} G(s)e^{st}\,ds \tag{IV.3}$$

where $w(t)$ is the symbol for the impulse response. The substitution of $j\omega$ for s in Eq. (IV.3) gives

$$w(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(j\omega)e^{j\omega t}\,d\omega \tag{IV.4}$$

Integrating the right-hand side of the preceding expression by parts, and letting $u = G(j\omega)/2\pi$ and $dv = e^{j\omega t}\,d\omega$ gives

$$w(t) = \frac{G(j\omega)e^{j\omega t}}{2\pi jt}\Bigg|_{-\infty}^{\infty} + \frac{1}{2\pi\,(-jt)}\int_{-\infty}^{\infty} G^{(1)}(j\omega)e^{j\omega t}\,d\omega \tag{IV.5}$$

For any realizable function, $G(j\omega)$ goes to zero for infinite values of ω. Thus, the first term on the right-hand side of the preceding expression vanishes. Further integration by parts yields the following general expression for $w(t)$:

$$w(t) = \frac{1}{2\pi(-jt)^n}\int_{-\infty}^{\infty} G^{(n)}(j\omega)e^{j\omega t}\,d\omega \tag{IV.6}$$

By separating $G^{(n)}(j\omega)$ into its real part $G_R{}^{(n)}(j\omega)$ and its imaginary part $G_I{}^{(n)}(j\omega)$ and similarly by writing $e^{j\omega t}$ in its rectangular form $\cos \omega t + j \sin \omega t$, the preceding expression becomes

$$w(t) = \frac{1}{2\pi(-jt)^n} \int_{-\infty}^{\infty} [G_R{}^{(n)}(j\omega) \cos \omega t - G_I{}^{(n)}(j\omega) \sin \omega t]\, d\omega + \frac{j}{2\pi(-jt)^n} \int_{-\infty}^{\infty} [G_R{}^{(n)}(j\omega) \sin \omega t + G_I{}^{(n)}(j\omega) \cos \omega t]\, d\omega \quad \text{(IV.7)}$$

Examination of the coefficient $j/(-jt)^n$ in front of the second integral

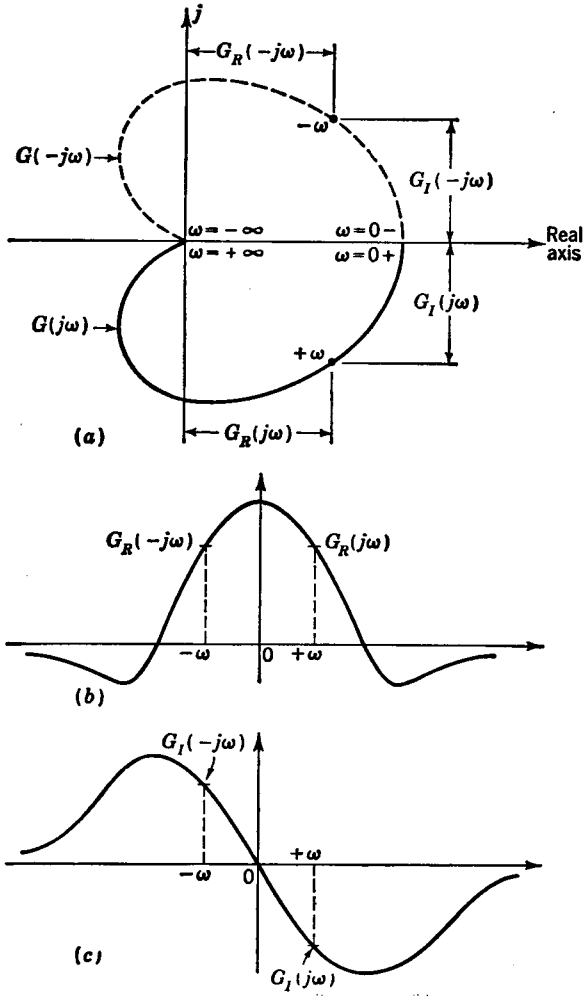


FIG. IV.1. (a) Polar plot $G(j\omega)$; (b) plot of $G_R(j\omega)$ versus ω; (c) plot of $G_I(j\omega)$ versus ω.

shows that for even values of $n(n = 2, 4, 6, \ldots)$ the second integral is imaginary. In order that $w(t)$ be real, the second integral must vanish for even values of n. Similarly, for odd values of n the first integral is imaginary and therefore must vanish.

The preceding result could also be ascertained as follows: In Fig. IV.1*a* is shown a typical polar plot of $G(j\omega)$. For any value of ω, $G_R(j\omega) = G_R(-j\omega)$ so that the real part $G_R(j\omega)$ is an even function of ω, as shown in Fig. IV.1*b*. However, from Fig. IV.1*a* it is to be noted that $G_I(j\omega) = -G_I(-j\omega)$, and thus the imaginary part $G_I(j\omega)$ is an odd function of ω, as shown in Fig. IV.1*c*. Differentiation of an even function yields an odd function, while differentiation of an odd function yields an even function. Thus, the nth derivative of $G_R(j\omega)$ is an even function when n is even and an odd function when n is odd. Similarly, the nth derivative of $G_I(j\omega)$ is an odd function when n is even and an even function when n is odd. In addition, the product of two even functions or the product of two odd functions gives an even function, while the product of an even function and an odd function is an odd function. Thus, since $\cos \omega t$ is an even function of ω and $\sin \omega t$ is odd, it follows that for even values of n the first integrand of Eq. (IV.7) is an even function, while the second integrand is odd. The integral of an odd function from $-\infty$ to $+\infty$ is zero so that the second term of Eq. (IV.7) vanishes for even values of n. Similarly, for odd values of n, the first integrand is an odd function and thus vanishes after integration.

Because the first integral of Eq. (IV.7) is an even function when n is even,

$$w(t) = \frac{1}{\pi(-jt)^n} \int_0^\infty [G_R^{(n)}(j\omega) \cos \omega t - G_I^{(n)}(j\omega) \sin \omega t]\, d\omega \qquad (n \text{ even}) \tag{IV.8}$$

The first term in the integrand of Eq. (IV.8) is an even function of *time*, while the second is an odd function of *time*. In order that $w(t)$ be zero for negative values of time, the two components must be equal and opposite for negative values of t and hence equal for $t > 0$. Thus for even values of n and $t > 0$

$$w(t) = \frac{-2}{\pi(-jt)^n} \int_0^\infty G_I^{(n)}(j\omega) \sin \omega t\, d\omega \tag{IV.9}$$

$$w(t) = \frac{2}{\pi(-jt)^n} \int_0^\infty G_R^{(n)}(j\omega) \cos \omega t\, d\omega \tag{IV.10}$$

By applying similar reasoning to the second integral of Eq. (IV.7), it may

be shown that for odd values of n and $t > 0$

$$w(t) = \frac{2j}{\pi(-jt)^n} \int_0^\infty G_I^{(n)}(j\omega) \cos \omega t \, d\omega \qquad \text{(IV.11)}$$

$$w(t) = \frac{2j}{\pi(-jt)^n} \int_0^\infty G_R^{(n)}(j\omega) \sin \omega t \, d\omega \qquad \text{(IV.12)}$$

When the nth derivative is a train of impulses, Eqs. (IV.10) and (IV.12) become

$$w(t) = \frac{2}{\pi} \frac{(-1)^{n/2}}{t^n} \sum_{k=1}^{\nu} a_k \cos \omega_k t \qquad n \text{ even} \qquad \text{(IV.13)}$$

$$w(t) = \frac{2}{\pi} \frac{(-1)^{(n+1)/2}}{t^n} \sum_{k=1}^{\nu} a_k \sin \omega_k t \qquad n \text{ odd} \qquad \text{(IV.14)}$$

where a_k is the area of the kth pulse and ω_k is the angular velocity at which it occurs.

The procedure to use in applying Eq. (IV.13) or (IV.14) parallels that described for Eq. (III.4) or (III.7). A set of equations similar to Eqs. (IV.13) and (IV.14) can be obtained from Eqs. (IV.9) and (IV.11), in which the imaginary part $G_I(j\omega)$ is employed. Thus, one may work with either the real or the imaginary part of $G(j\omega)$ in order to find the impulse response. After the impulse response has been determined, the convolution technique presented in Appendix II may be used to find the response to any arbitrary input.

PROBLEMS

Chapter 2

2.1 Perform the indicated operations for each case given below. Use for $f(t)$ the function $f(t) = t^2$. Determine the constants of integration if the initial conditions are $x(0) = 7$ and $px(0) = 4$.

(*a*) $x(t) = \frac{1}{p}[pf(t)]$ (*b*) $x(t) = \frac{1}{p}[p^2f(t)]$

(*c*) $x(t) = \frac{1}{p^2}[pf(t)]$ (*d*) $x(t) = \frac{1}{p^2}[p^2f(t)]$

2.2. For each of the mechanical systems shown in Fig. P 2.2,
(*a*) Determine the equation which relates f and x.
(*b*) Determine the equation which relates f and y.
(*c*) Determine the equation which relates x and y.

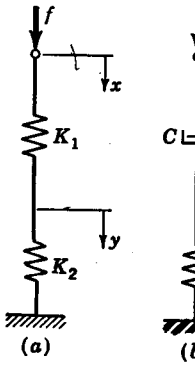


FIG. P 2.2. Translational mechanical systems.

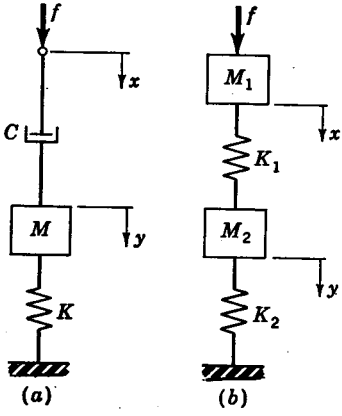


FIG. P 2.3. Translational mechanical systems with a mass.

2.3. For each of the mechanical systems shown in Fig. P 2.3, construct the equivalent grounded-chair representation and
(*a*) Determine the equation which relates f and x.
(*b*) Determine the equation which relates f and y.
(*c*) Determine the equation which relates x and y.

2.4. For the electrical networks shown in Fig. P 2.4*a* and *b*,
(*a*) Determine the equation which relates E_1 and I.
(*b*) Determine the equation which relates E_1 and E_2.

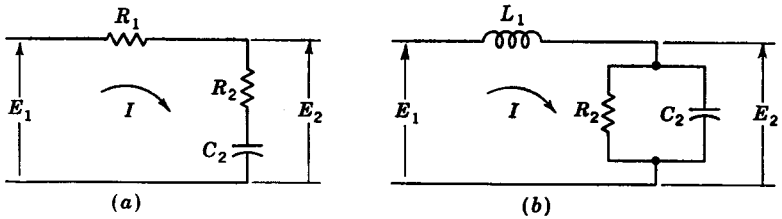


FIG. P 2.4. Electrical networks.

2.5. A schematic diagram of an accelerometer for measuring the linear acceleration d^2x/dt^2 is shown in Fig. P 2.5. Determine the operational form for the differential equation which relates y (the change in the position of the mass relative to the frame) to the acceleration p^2x of the frame.

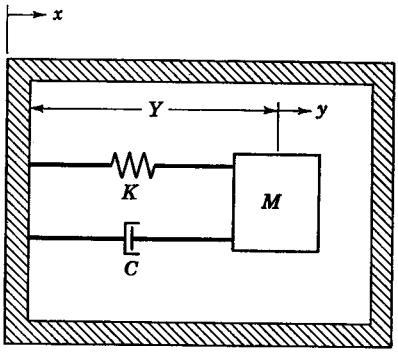


FIG. P 2.5. Accelerometer.

2.6. For each of the mechanical systems shown in Fig. P 2.6, construct the grounded-chair representation and determine the equation relating f and x.

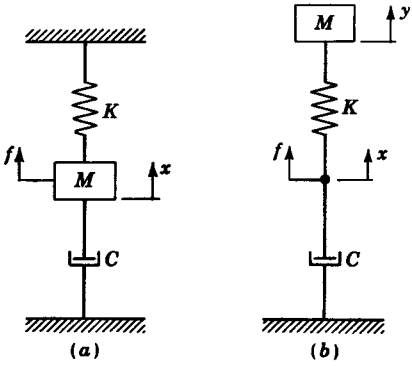


FIG. P 2.6. Mechanical systems.

2.7. Two mechanical vibration absorbers are shown in Fig. P 2.7. Construct the grounded-chair representation and determine the equation relating f and x for each case.

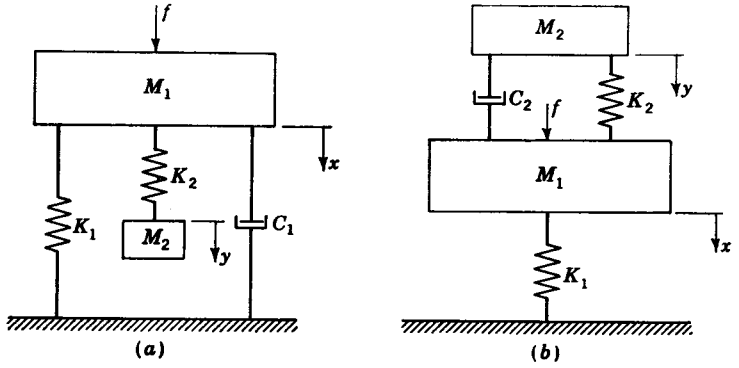


FIG. P 2.7. Mechanical vibration absorbers.

2.8. Construct the electrical analogy for the mechanical system shown in Fig. P 2.2*c* by using

(*a*) Analogous force-voltage terms. Determine the equation relating E and I of this electrical system and compare this equation with that relating f and x of Fig. P 2.2*c*.

(*b*) Analogous force-current terms. Determine the equation relating E and I of this electrical system, and compare this equation with that relating x and f of Fig. P 2.2*c*.

2.9. Construct the mechanical analogy for the electrical system shown in Fig. P 2.4*b* by using

(*a*) Analogous force-voltage terms. Determine the equation relating f and x of this mechanical system, and compare this equation with that relating E_1 and I of Fig. P 2.4*b*.

(*b*) Analogous force-current terms. Determine the equation relating x and f of this mechanical system, and compare this equation with that relating E_1 and I of Fig. P 2.4*b*.

2.10. Determine the equation for the torsional spring rate of the stepped shaft shown in Fig. P 2.10. Are the individual spring rates for the left and right portion of the stepped shaft added by the parallel or by the series law?

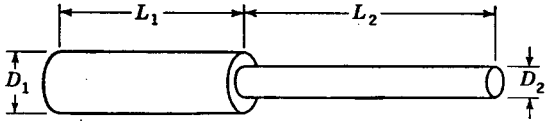


FIG. P 2.10. Stepped shaft.

2.11. (*a*) The torque T which is applied to the shaft shown in Fig. P 2.11*a* is transmitted directly to the load, which consists of inertia $Jp^2\phi$, viscous friction $C_v p\phi$, and external load torque T_L. Determine the equation relating T and the position ϕ of the load.

(*b*) Figure P 2.11*b* shows the force-torque analog for the system of Fig. P 2.11*a*. Compare the equation relating f and x of this analog with that obtained for T and ϕ in part *a*.

(*c*) In Fig. P 2.11*c* the two shafts are connected by gears of diameter D_1 and D_2, respectively. Determine the equation relating T and the position ϕ of the load. (Neglect inertia of the gears.)

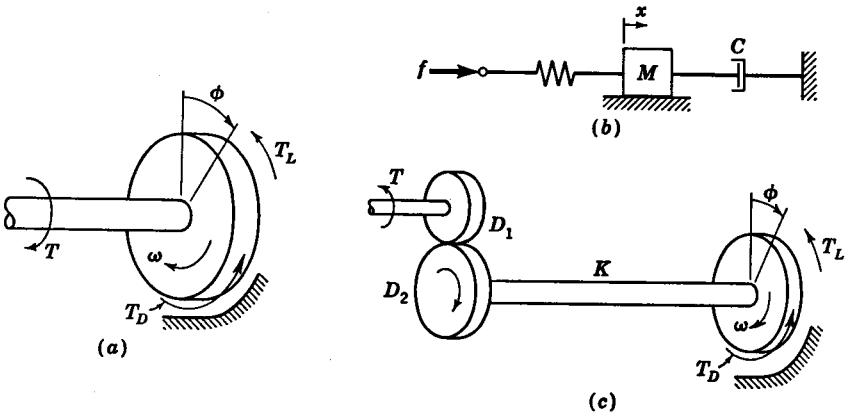


FIG. P 2.11. Torsional system.

2.12. (*a*) In Fig. P 2.12*a* is shown the stable element of a stable platform for an inertial-guidance system. This stable element consists primarily of two disks which are connected by a flexible post. The torque transmitted through the post is $K_s(\phi_2 - \phi_1)$, where ϕ_1 and ϕ_2 designate the angular positions of the upper and lower disks, respectively. Write the expression for the summation of torques acting on each disk, and then eliminate the parameter ϕ_2 in order to determine the operational form of the differential equation relating the motor torque T and the position ϕ_1. Assume that the viscous damping in the bearing which supports the post is negligibly small.

(*b*) In Fig. P 2.12*b* is shown the force-torque analog for the rotational system shown in Fig. P 2.12*a*. For the mechanical system shown in Fig. P 2.12*b*, determine the

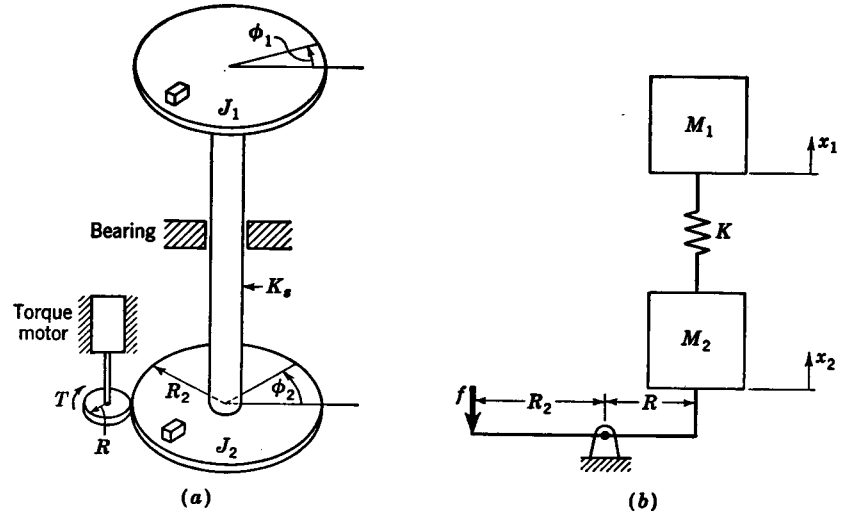


FIG. P 2.12. Stable element for an inertial-guidance system.

equation relating f and x_1. Compare this with the equation relating T and ϕ_1 that was obtained in part a.

(c) Same as Prob. 2.12a except that the torque motor is connected to the upper table rather than the lower table. Determine the equation relating T and ϕ_1.

2.13. For the lever shown in Fig. P 2.13a and b, the variation in the applied force is f, and the variation in spring position is x. (The horizontal line represents the reference position of the lever.) For each system,

(a) Determine the equation relating f and x.

(b) Determine the relationship between t and ϕ (where $t = fL_f$ is the variation in applied torque).

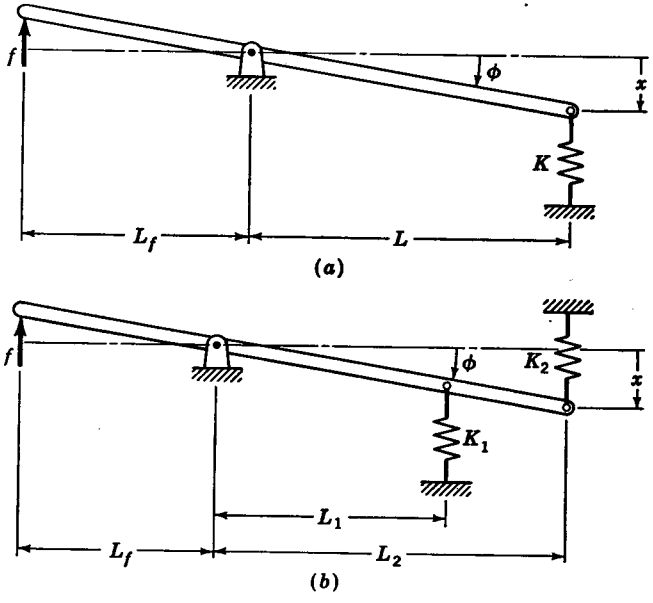


FIG. P 2.13. Torsional system composed of translational elements.

2.14. The lever system shown in Fig. P 2.14 is drawn in its reference position. The variation in spring position is designated by x. The variation in applied force is designated by f. (f and x are zero at the reference position.)

(a) Determine the equation relating f and x.

(b) Determine the relationship between t and ϕ.

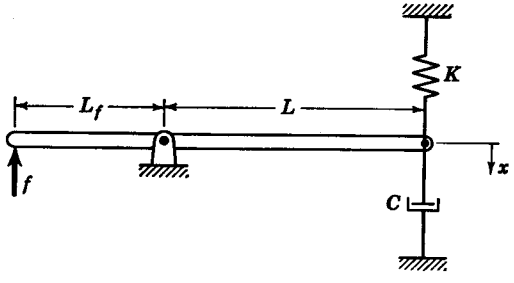


FIG. P 2.14. Lever system.

2.15. For the lever system shown in Fig. P 2.15,
(*a*) Determine the equation relating f and x.
(*b*) Determine the relationship between t and ϕ.

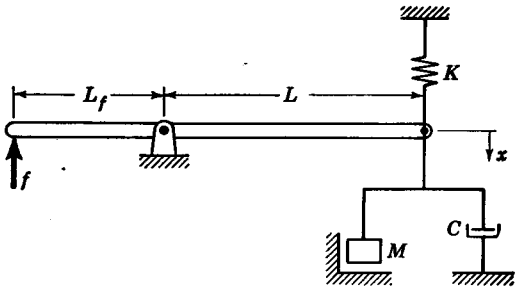


FIG. P 2.15. Lever system.

2.16. Linearize each of the following expressions:
(*a*) $Y = \sin KX$
(*b*) $Y = \tan KX$
(*c*) $Y = e^{KX}$
What is the resulting linearized expression in each case when $X_i = 0$? When $X_i = 1$?

2.17. Linearize the following expressions:

$$(a)\ Z = 0.1X^2 + \frac{100}{Y} \qquad (b)\ Z = X(2Y + 5)$$

If $Y_i = 10$ and $X_i = 20$, what per cent error results in each of the above cases when $y = +1$ and $x = -2$?

2.18. The volume V of a sphere is $V = \frac{4}{3}\pi R^3$. Determine the equation for the linear approximation to V. If $R_i = 10$ in., what per cent error results by using this approximation for V when $R = 11$? What is the per cent error for $R = 9$?

2.19. The equation for the flow of an incompressible fluid through a sharp-edged orifice is

$$Q = C_c A \sqrt{\frac{2g}{\rho}(P_1 - P_2)}$$

where Q = rate of flow, in.^{3}/sec
C_c = coefficient of discharge (unitless)
A = area of orifice, in.^{2}
$P_1 - P_2$ = pressure drop across orifice, psi
ρ = density, lb/in.^{3}
Determine the linear approximation to the preceding equation for the case in which both the area A and pressure drop $P_1 - P_2$ vary slightly from the initial values of A_i and $(P_1 - P_2)_i$.

2.20. Typical operating curves for a d-c motor are shown in Fig. P 2.20. These are curves of torque T versus operating speed N for constant values of voltage E applied to the motor. These curves are a plot of the function $N = F(T,E)$. Effect a linear approximation for N. Evaluate the partial derivatives in this approximation when T_i is 1 in.-lb and E_i is 16 volts.

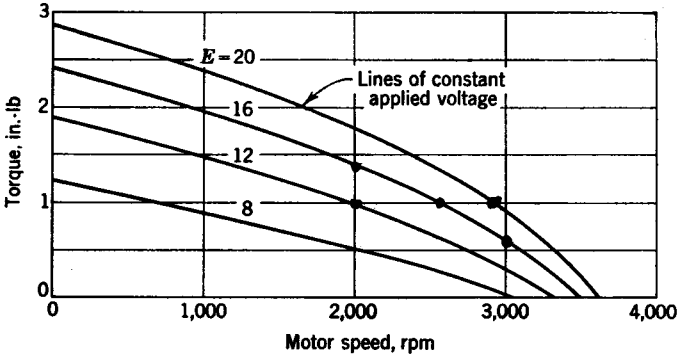


FIG. P 2.20. Operating curves for a d-c motor.

Chapter 3

3.1. Determine the over-all relationship between x and y for the hydraulic servomotor which is shown in Fig. 3.1 for the case where $a \neq b$. Construct the block diagram for this system.

3.2. In Fig. P 3.2 is shown a hydraulic servomotor which is similar to the power-amplifying device used in power steering units. A movement in the x direction of the valve is seen to open passage 1 to the supply pressure, which in turn causes the big piston to move to the right. Because the sleeve is directly connected to this piston, the sleeve also moves to the right to close off flow from the valve. Determine the block diagram relating the input position x to the output y. Identify the time constant.

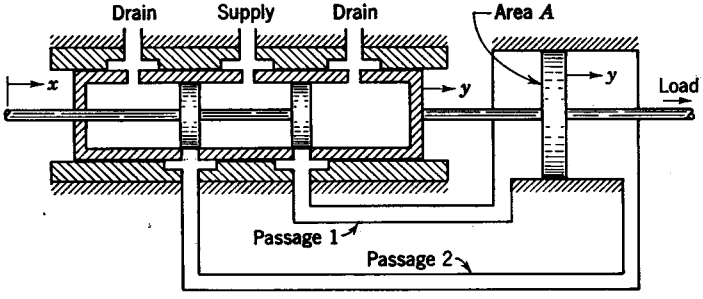


FIG. P 3.2. Hydraulic servomotor.

3.3. In Fig. P 3.3 is shown a modification of the hydraulic power amplifier discussed in Prob. 3.2. Determine the block-diagram representation for this device in which x is the input and y the output. Note that the position of the sleeve is $[a/(a+b)]y$.

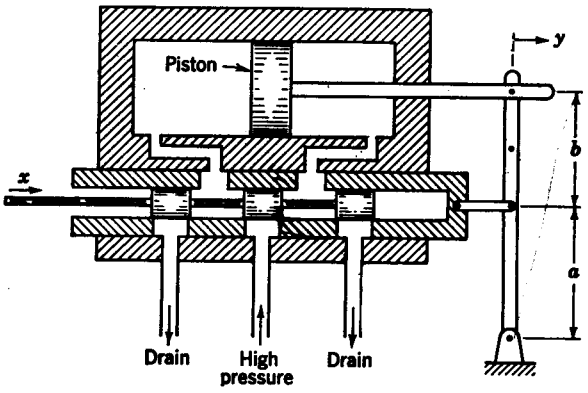


FIG. P 3.3. Hydraulic power amplifier.

3.4. For the hydraulic amplifier shown in Fig. P 3.4, determine the block diagram for the walking-beam linkage and also the block diagrams relating e to y and y to w. Combine these diagrams to determine the over-all block-diagram representation for the system.

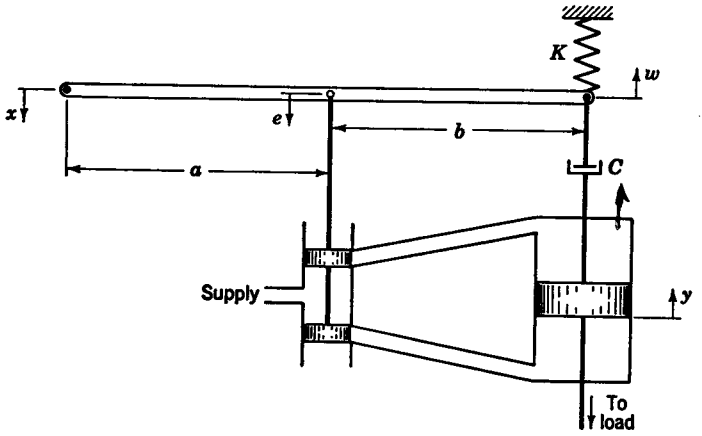


FIG. P 3.4. Hydraulic amplifier.

3.5. In Fig. P 3.5 is shown a tension-regulating apparatus such as is used in the paper industry. To ensure uniform winding, it is necessary to maintain a constant tension F_c as the sheet is being wound on the wind-up roll. To increase the tension in the paper, the tension control lever is raised. This raises the torque control arm of the motor, which increases the torque T_m applied by the motor to the wind-up roll. The

change in torque provided by the motor is $t_m = K_m e/(1 + \tau p)$. For the wind-up roll, it follows that $F_c = T_m/R$, where R is the radius of the wheel. Determine the overall block diagram relating a variation f_r of the reference or desired tension to a variation of the controlled tension f_c.

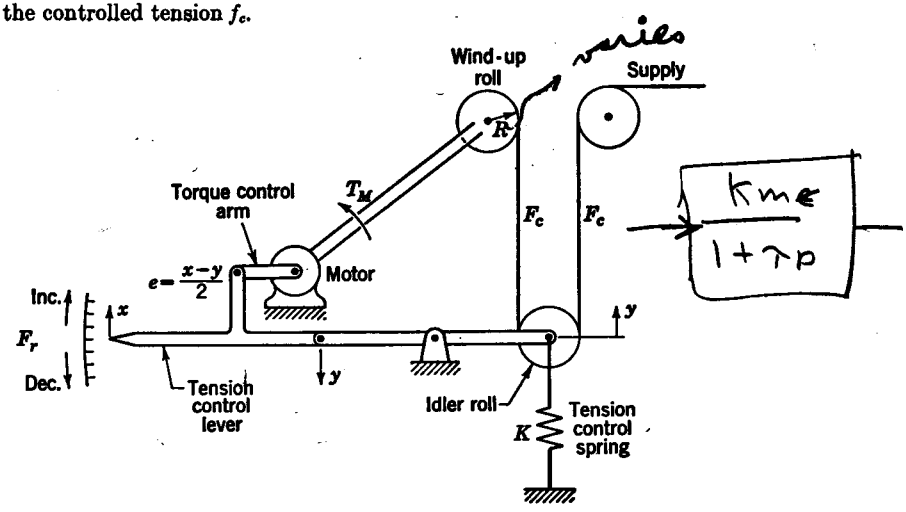


FIG. P 3.5. Tension regulator.

3.6. In Fig. P 3.6 is shown an electrical speed control system. The input potentiometer provides a reference input voltage E_r which is proportional to the desired speed N_{in} ($E_r = K_r N_{in}$). A voltage signal E_c which is proportional to the controlled output speed N_o is provided by the tachometer ($E_c = K_c N_o$). The error $E_r - E_c$ is amplified by an electronic amplifier whose output is $E_f = K_a(E_r - E_c)$. The voltage E_f is

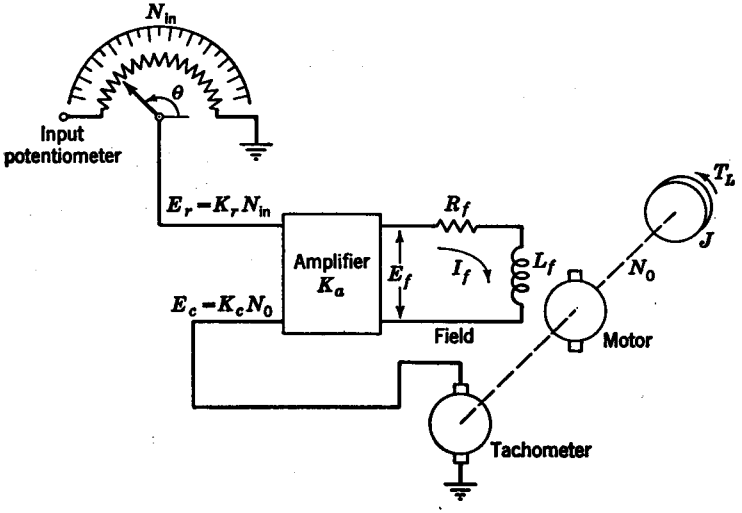


FIG. P 3.6. Electrical speed control system.

applied to the field of a field-controlled d-c motor. The torque exerted on the shaft by the motor (air-gap torque) is proportional to the field current, that is, $T = KI_f$. Determine the over-all block diagram for this speed control system for the case in which the load torque consists of an inertia $Jp^2\theta = Jp(2\pi/60)N_o$ and an external torque T_L.

3.7. In Fig. P 3.7 is shown a liquid-level controller. To raise the level of the fluid, the control lever is moved up (i.e., position z is raised). This raises the valve (position

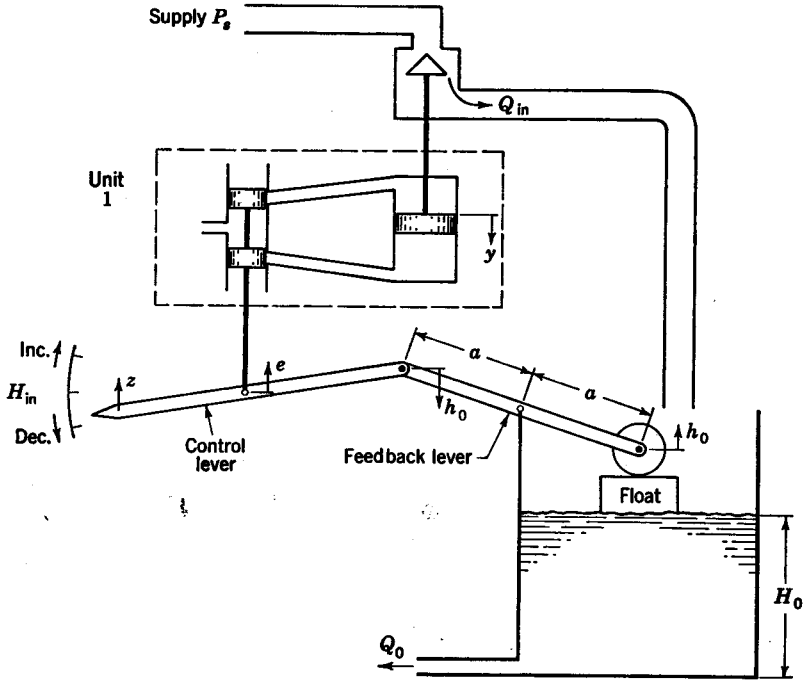


FIG. P 3.7. Level control system.

e), which increases y, thereby admitting more flow Q_{in}. The flow Q_{in} is a function of the flow valve opening Y and the supply pressure P_s. The change in volume of liquid in the tank is the time integral $(q_{in} - q_o)/p$, which is equal to the cross-section area of the tank A_T times the change in level h_o. The flow out Q_o is seen to depend upon the pressure head H_o. Determine the over-all block diagram for this controller.

3.8. (*a*) The same as Prob. 3.7 except that positions e and y are connected by a link to convert the hydraulic integrator to a hydraulic servomotor shown in Fig. P 3.8*a*. That is, unit 2 in Fig. P 3.8*a* is substituted for unit 1 in Fig. P 3.7.

(*b*) The same as Prob. 3.7 except that positions e and y are connected by unit 3 as shown in Fig. P 3.8*b*. This is in effect replacing the hydraulic integrator by a spring-damper combination.

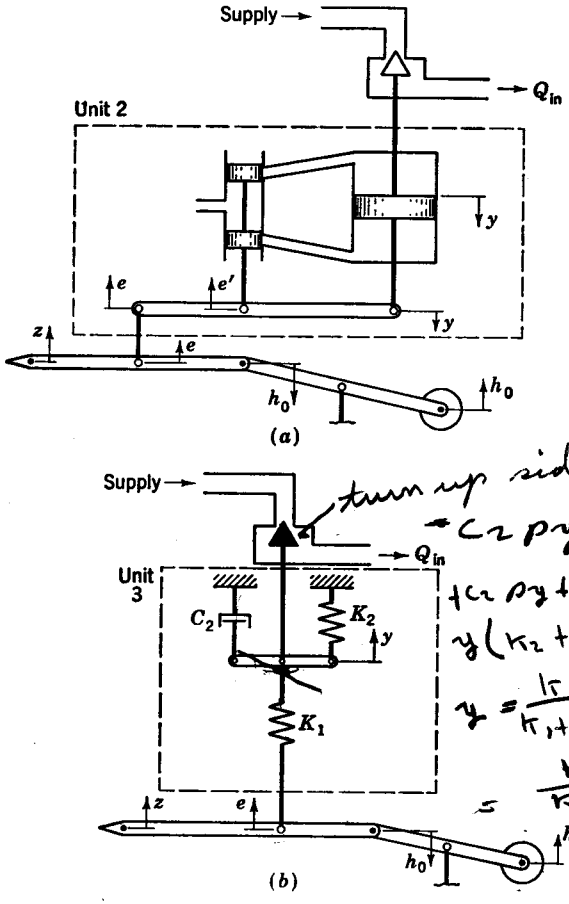


FIG. P 3.8. Alternative control units.

3.9. A system for controlling flow is shown in Fig. P 3.9. Increasing the desired flow setting increases the compression on spring K_1, which causes x and position e of the balanced valve to move up. This in turn causes the flow valve to move down, which increases the flow. The amount of flow out is measured by a venturi-type flowmeter so that the pressure drop $P_1 - P_2$ is a function of Q_o. The diaphragm prevents leakage from the high pressure P_1 to the low pressure P_2, but it permits motion, just as a piston would. The effective area of the diaphragm is A_d. The flow Q_o is seen

to be a function of the flow-valve opening Y and the supply pressure P_s. Determine the over-all block-diagram representation for this system.

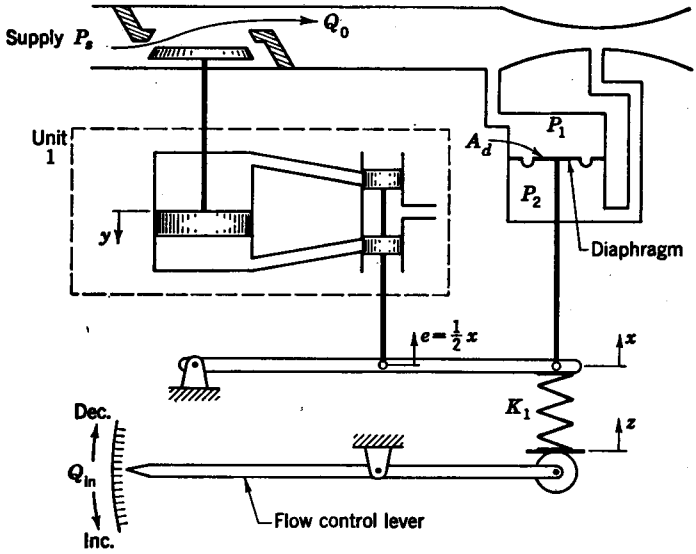


FIG. P 3.9. Flow control system.

3.10. (*a*) Same as Prob. 3.9 except that e actuates the valve through unit 2, shown in Fig. P 3.8*a*, rather than unit 1, shown in Fig. P 3.9.

(*b*) Same as Prob. 3.9 except that e actuates the valve through unit 3, shown in Fig. P 3.8*b*, rather than unit 1, shown in Fig. P 3.9.

3.11. In Fig. 3.10 is shown a general block-diagram representation for a feedback control system. The operators $G_1(p)$, $G_2(p)$, and $H(p)$ may be written in the form

$$G_1(p) = \frac{N_{G_1}}{D_{G_1}} \qquad G_2(p) = \frac{N_{G_2}}{D_{G_2}} \qquad H(p) = \frac{N_H}{D_H}$$

where N represents the numerator and D the denominator of each term. Show that

$$c(t) = \frac{N_{G_1}N_{G_2}D_H r(t) + N_{G_2}D_{G_1}D_H d(t)}{N_{G_1}N_{G_2}N_H + D_{G_2}D_{G_1}D_H}$$

Note that the coefficient for $r(t)$ is the product of the numerator terms $N_{G_1}N_{G_2}$ from $r(t)$ to the output and the denominator term D_H from the output back to $r(t)$. Similarly the coefficient for $d(t)$ is the product of the numerator term N_{G_2} from $d(t)$ to the output and the denominator terms $D_H D_{G_1}$ from the output to $d(t)$. From Fig. 3.17 it is to be noted that $N_{G_1} = K_1$, $D_{G_1} = 1 + \tau_1 p$, $r(t) = C_2 K_s n_{in}$, $d(t) = -C_3 t_L$, etc. By application of the preceding general expression, verify the result given by Eq. (3.41).

3.12. For the control system shown in Fig. P 3.12, determine the differential equation of operation for each of the following cases:

(*a*) $G_1(p) = \dfrac{K_1}{1 + \tau_1 p}$ (*b*) $G_1(p) = \dfrac{K_1}{p(1 + \tau_1 p)}$

(*c*) $G_1(p) = \dfrac{K_1}{p} + \dfrac{K_1}{1 + \tau_1 p}$

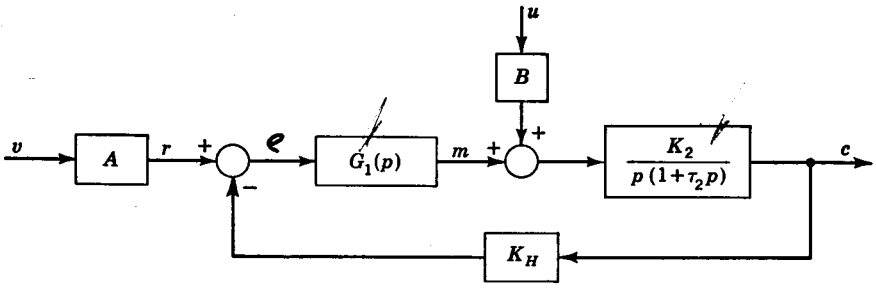


FIG. P 3.12. Block diagram for control system.

Chapter 4

4.1. For the control system shown in Fig. P 3.12, determine the steady-state equation relating v, u, and c for each of the following cases:

(*a*) $G_1(p) = \dfrac{K_1}{1+\tau_1 p}$ (*b*) $G_1(p) = \dfrac{K_1}{p(1+\tau_1 p)}$

(*c*) $G_1(p) = \dfrac{K_1}{p} + \dfrac{K_1}{1+\tau_1 p}$

4.2. The steady-state operating curves for a proportional-type temperature control system are shown in Fig. P 4.2.

(*a*) Determine the equation for steady-state operation about point A.

(*b*) If this were an open-loop rather than a closed-loop system, what would be the steady-state equation of operation?

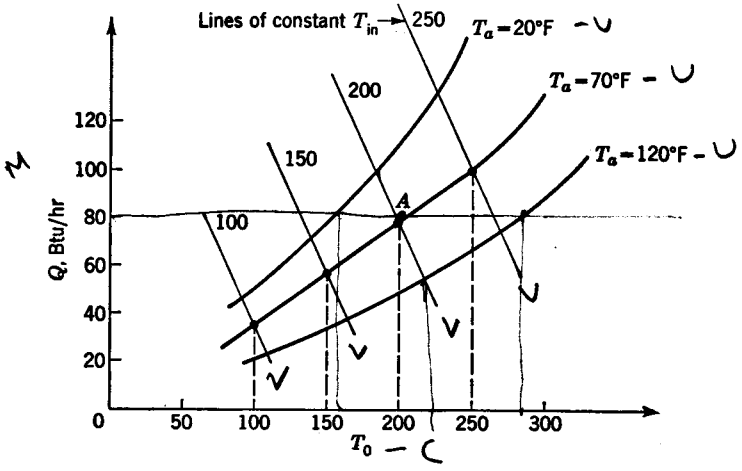


FIG. P 4.2. Steady-state curves for a temperature control system.

4.3. The individual constants in Eq. (4.3) may be evaluated from the steady-state operating curves. From Fig. 4.1*b*, it follows that, for the system to be controlled,

$$(m + Bu)K_{G_1} = c$$

(*a*) Show that

$$\frac{1}{K_{G_2}} = \frac{\partial M}{\partial C}\bigg|_U \qquad \text{and} \qquad B = -\frac{\partial M}{\partial U}\bigg|_C$$

Similarly from Fig. 4.1*b* it follows that for the controller

$$(Av - K_H c)K_{G_1} = m$$

(*b*) Show that

$$K_H K_{G_1} = -\frac{\partial M}{\partial C}\bigg|_V \qquad AK_{G_1} = \frac{\partial M}{\partial V}\bigg|_C \qquad \frac{K_H}{A} = \frac{\partial V}{\partial C}\bigg|_M$$

(*c*) For the case in which $A = 0.5$, evaluate the individual constants K_{G_1}, K_{G_2} K_H, and B in the steady-state equation that was obtained in Prob. 4.2*a*.

4.4. For the illustrative example discussed in Sec. 4.3, determine the speed error when $N_{in} = 4{,}000$ and $T_L = 100$. If the slope of the droop line $\partial M/\partial C\big|_V = -K_H K_{G_1}$ is increased by a factor of 5, what is the new error?

4.5. When the system to be controlled has an integrating element such as to make K_{G_2} infinite, the steady-state operating curves for the system become horizontal straight lines, as shown in Fig. P 4.5. The steady-state operating curves for a remote-control positioning device are shown in Fig. P 4.5. The controlled shaft position is designated by θ_o and the set position by θ_{in}. Variations are to be indicated by $\Delta\theta$. For this remote-control positioning system, the external disturbance is a load torque T_L, and the manipulated variable is a motor torque T_m.

(*a*) Determine the steady-state equation of operation relating $\Delta\theta_o$, $\Delta\theta_{in}$, and t_L for operation about point A.

(*b*) To decrease the error caused by variations in the external disturbance by a factor of 5, what should be the new slope of the droop line?

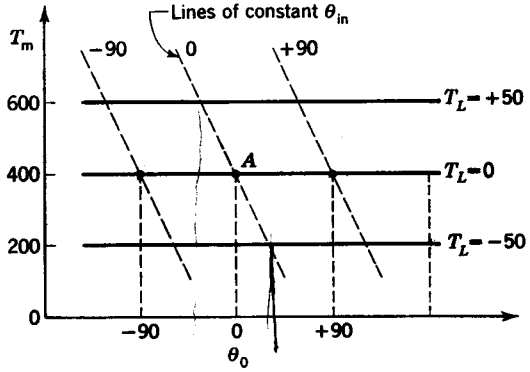


FIG. P 4.5. Steady-state curves for a remote-control positioning system.

4.6. In Fig. P 4.6 is shown the steady-state block diagram for a system which is subjected to two external disturbances. Determine the steady-state equation relating c, v, u_1, and u_2. For which of the following cases will the steady-state operation be independent of variations in u_1? For which cases is it independent of variations in u_2? NOTE: Steady-state operation is independent of K_{G_3}.

(*a*) $K_{G_1} = \infty$ K_{G_2} and K_H are finite

(*b*) $K_{G_2} = \infty$ K_{G_1} and K_H are finite

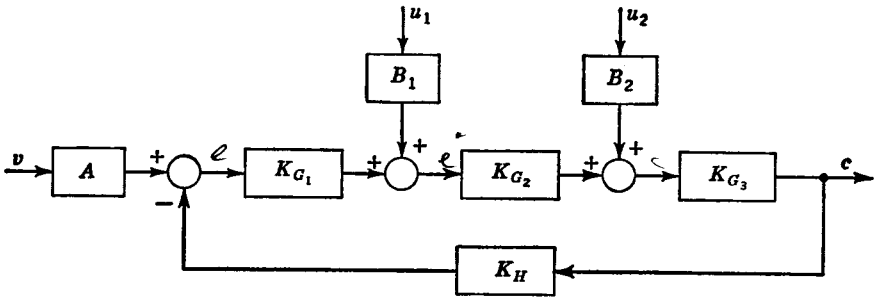
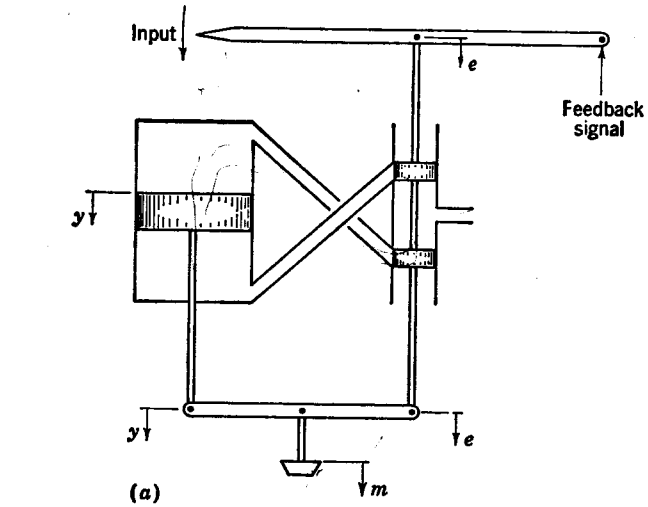


FIG. P 4.6. System subjected to two external disturbances.

4.7. Identify the mode of control for the control elements shown in units 1, 2, and 3 of Figs. P 3.7 and P 3.8*a* and *b*.

4.8. For each of the control elements shown in Fig. P 4.8*a* and *b*, determine the



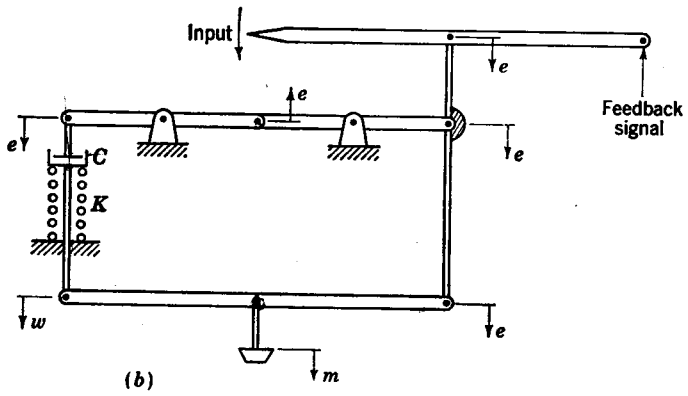


FIG. P 4.8. Control elements.

operational form of the differential equation relating the actuating signal e and the output m.* Identify the mode of control for each case.

4.9. A reproducing shaper is shown in Fig. P 4.9. The position y of the duplicating cutter is seen to follow the position x of the master cutter. Determine the mode of operation of the shaper. What modifications would be necessary to convert this to a proportional plus integral type of controller?

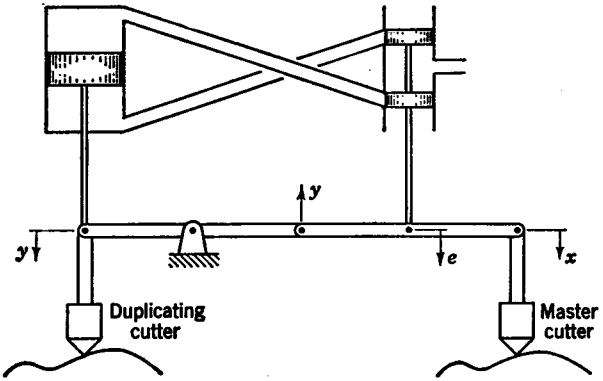


FIG. P 4.9. Reproducing shaper.

Chapter 5

5.1. Determine the general solution for the first-order differential equation

$$(p - r)y(t) = x(t)$$

5.2. Expand the following operators by use of partial-fraction expansion techniques:

(a) $\dfrac{L_m(p)}{L_n(p)} = \dfrac{6}{(p+1)(p+4)}$

(b) $\dfrac{L_m(p)}{L_n(p)} = \dfrac{2p+3}{p(p+3)}$

(c) $\dfrac{L_m(p)}{L_n(p)} = \dfrac{p^2+6p+15}{(p+1)(p+3)(p+6)}$

(d) $\dfrac{L_m(p)}{L_n(p)} = \dfrac{18p+30}{p(p+2)(p+5)}$

5.3. Determine the partial-fraction expansion for the following operators:

(a) $\dfrac{L_m(p)}{L_n(p)} = \dfrac{p+3}{p(p+1)^2}$

(b) $\dfrac{L_m(p)}{L_n(p)} = \dfrac{3}{p^2(p+3)^2}$

(c) $\dfrac{L_m(p)}{L_n(p)} = \dfrac{56p+85}{p^3(p+2)(p+5)}$

5.4. By use of classical techniques, determine the solution of the following differential equations for the case in which $x(t) = h$, which is a constant:

(a) $y(t) = \dfrac{4}{(p+1)(p+3)}x(t)$ $\quad y(0) = py(0) = 0$

(b) $y(t) = \dfrac{p+5}{(p+1)(p+3)}x(t)$ $\quad y(0) = py(0) = 0$

(c) $y(t) = \dfrac{3p+5}{(p+1)^2(p+3)}x(t)$ $\quad y(0) = py(0) = p^2y(0) = 0$

5.5. Same as Prob. 5.4 except that $x(t) = e^{-t}$.

5.6. The "ramp function" shown in Fig. P 5.6 is seen to have a constant slope equal to a. Determine the Laplace transform for this ramp function, whose equation is $x(t) = at$.

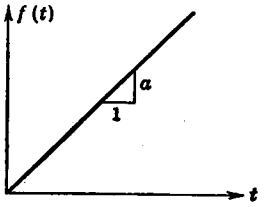


FIG. P 5.6. Ramp function.

5.7. Determine the Laplace transform for

(a) The function $x(t) = at^2$

(b) The function $x(t) = \cos \omega t$

5.8. Obtain the transform $Y(s)$ for each of the differential equations given in Prob. 5.4 in which $x(t) = h$ and $x(0) = h$ at time $t = 0$. Perform a partial-fraction expansion, and invert to yield the response $y(t)$. [It is to be noted that in classical methods the input function $x(t)$ must be continuous. Thus, to check the result of Prob. 5.4, $x(0)$ must be taken as h. With Laplace transforms a discontinuity can exist at $t = 0$.]

5.9. Same as Prob. 5.8 except that $x(t) = e^{-t}$ and $x(0) = 1$ at time $t = 0$.

5.10. Determine the differential equation of operation for the system shown in Fig. P 5.10 for each of the following cases:

(a) $G_1(p) = \dfrac{1}{p}$ (b) $G_1(p) = \dfrac{7.5}{1 + 0.4p}$

A step change of height h_c occurs in the input $x(t)$ at time $t = 0$ such that for $t > 0$, $x(t) = h_c$.

By use of Laplace transforms, determine the response $y(t)$ for the case in which the system is initially at a steady-state operating condition and in addition $x(0) = y(0) = 0$.

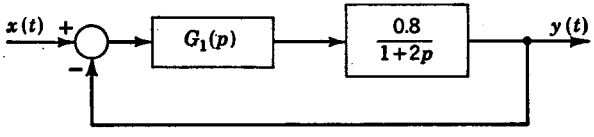


FIG. P 5.10. Control system.

5.11. Same as Prob. 5.10 except that initially $y(0) = 3$.

NOTE: The corresponding value of $x(0)$ may be determined from the equation for steady-state operation.

5.12. Same as Prob. 5.10 except that the step change occurs at time $t_0 = 4$.

5.13. By use of Laplace transforms, determine the response $y(t)$ for the following differential equations for the case in which all the initial conditions are zero [i.e., the

system is initially at a steady-state operating point and in addition $x(0) = y(0) = 0$]. The input $x(t)$ is an impulse of area k which occurs at time $t = 0$.

(a) $y(t) = \dfrac{2p + 1}{(p + 2)(p + 5)}\,x(t)$

(b) $y(t) = \dfrac{2p + 3}{p(p + 2)(p + 3)}\,x(t)$

(c) $y(t) = \dfrac{p^2 + 6p + 15}{(p + 1)(p + 3)(p + 6)}\,x(t)$

5.14. Same as Prob. 5.13 except that $y(0) = 3$.

NOTE: The corresponding value of $\dot{x}(0)$ may be determined from steady-state operation conditions.

5.15. Same as Prob. 5.13 except that the impulse excitation occurs at $t_0 = 4$.

5.16. Determine the final value $y(\infty)$ for each equation given in Prob. 5.13 by

(a) Application of the final-value theorem

(b) Substitution of $t = \infty$ into each of the resulting response equations of Prob. 5.13

5.17. Determine the initial value $y(0+)$ for each equation given in Prob. 5.13 by

(a) Application of the initial-value theorem

(b) Substitution of $t = 0+$ into each of the resulting response equations of Prob. 5.13

5.18. Determine the solution of Eq. (5.52) for the case in which the input is that shown in Fig. 5.10*b* to *d*.

Chapter 6

6.1. Determine the time response $y(t)$ for each of the following transformed equations:

(a) $Y(s) = \dfrac{2s + 12}{s(s + 3)(s + 4)}$

(b) $Y(s) = \dfrac{s^2 + 3s + 8}{(s + 2)(s^2 + 5s + 4)}$

(c) $Y(s) = \dfrac{2s + 3}{s^2(s + 3)}$

6.2. Determine the time response $y(t)$ for each of the following transformed equations:

(a) $Y(s) = \dfrac{10}{s(s^2 + 2s + 5)}$

(b) $Y(s) = \dfrac{2s^2 + 3s + 4}{(s + 3)(s^2 + 2s + 10)}$

(c) $Y(s) = \dfrac{2s + 8}{(s + 2)(s^2 + 4)}$

(d) $Y(s) = \dfrac{20(s + 5)}{s(s^2 + 4)(s^2 + 6s + 25)}$

6.3. The characteristic function for a control system is known to be

$$L_n(p) = (p + 5)(p^2 + 4p + 13)$$

Determine the general form of the equation for the response $y(t)$ of this system when the input excitation $x(t)$ is

(a) An impulse occurring at $t = 0$

(b) An impulse occurring at $t = 5$

(c) A step function occurring at $t = 0$

(d) A step function occurring at $t = 5$

Is the general form of these response expressions affected by the initial conditions?

6.4. The zeros of the characteristic function $L_n(p)$ for a control system are plotted in Fig. P 6.4. Determine the general form of the equation describing the response

$y(t)$ of this system when the input excitation $x(t)$ is
(*a*) An impulse which occurs at $t = 0$
(*b*) An impulse which occurs at $t = 5$
(*c*) A step function which occurs at $t = 0$
(*d*) A step function which occurs at $t = 5$

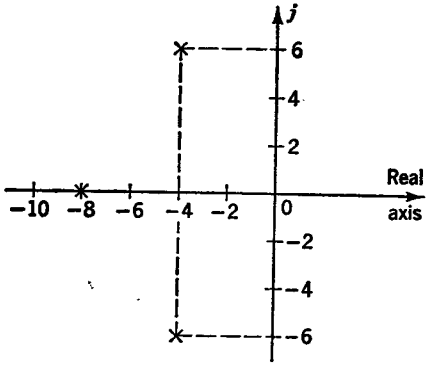


FIG. P 6.4. Plot of zeros of $L_n(p)$.

6.5. Determine the characteristic function for the control system shown in Fig. P 6.5. Use block-diagram algebra to move the constants A and B into the main loop. Does this affect the characteristic function?

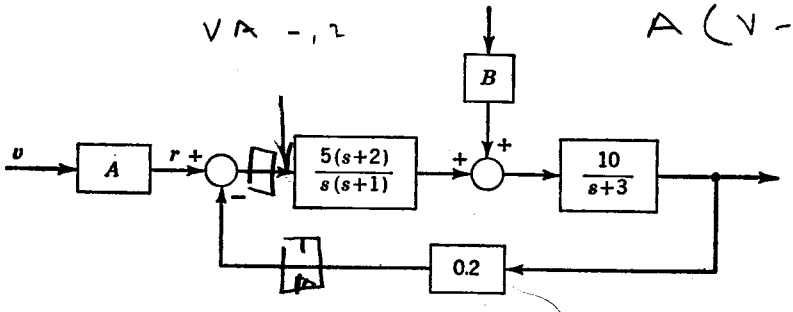


FIG. P 6.5. Control system.

6.6. For each of the characteristic functions given below:
(*a*) Determine the number of zeros that lie on or to the right of the imaginary axis.

$$s^3 + 2s^2 + 5s + 24 \qquad s^4 + 2s^3 + 6s^2 + 2s + 5$$

(*b*) Determine the number of zeros that have a real part greater than or equal to -4.

6.7. Equation (6.65) is the functional relationship which must be satisfied to have all the roots of a cubic equation lie to the left of the imaginary axis (i.e., have a real part less than 0). Derive a similar relationship to determine whether or not all the roots have a real part less than -2.

6.8. For the quartic expression given below, determine the functional relationships which must be satisfied to have all the roots lie to the left of the imaginary axis.

$$b_4s^4 + b_3s^3 + b_2s^2 + b_1s + b_0$$

Chapter 7

7.1. The root-locus plot for the system of Fig. 7.1*a* is given in Fig. 7.1*b*. Determine the response equation $c(t)$ for the case in which $r(t)$ is a unit step function and $K = 8$. All the initial conditions are zero.

7.2. The root-locus plot for the system of Fig. 7.2*a* is given in Fig. 7.2*b*. Determine the response $c(t)$ for the case in which $r(t)$ is a unit impulse and $K = 17$. All the initial conditions are zero.

7.3. Sketch the root-locus plot for the system shown in Fig. P 7.3. Determine the value K to yield a damping ratio of 0.5.

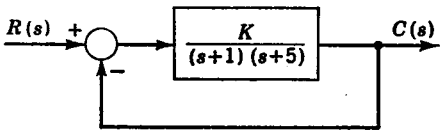


FIG. P 7.3. Control system.

7.4. Sketch the root-locus plot for the system shown in Fig. P 7.4 for each of the following cases:

(*a*) $G_1(s) = \dfrac{K_1}{s}$ (*b*) $G_1(s) = \dfrac{K_1}{s+4}$

For each case, determine the value of $K_1K_2K_3$ to yield a damping ratio of 0.5 for the dominant roots (i.e., the ones located nearest the imaginary axis).

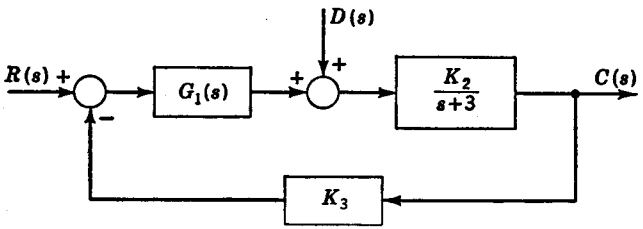


FIG. P 7.4. Control system.

7.5. Sketch the root-locus plot for the system shown in Fig. P 7.5 for each of the following cases:

(*a*) $H(s) = 1$ (*b*) $H(s) = s + 1$

For each case, determine the value of K to yield a damping ratio of 0.5 for the dominant roots. Comment on the effect of adding derivative action in the feedback path (i.e., case *b*).

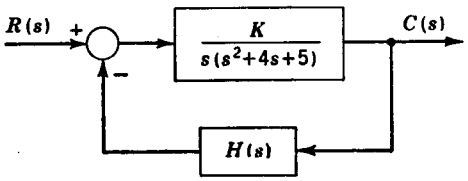


FIG. P 7.5. Control system.

7.6. Sketch the root-locus plot for each of the two systems shown in Fig. P 7.6. Determine the value of K at which each system becomes unstable. Comment on the effect of adding an integrating element as is done in case b.

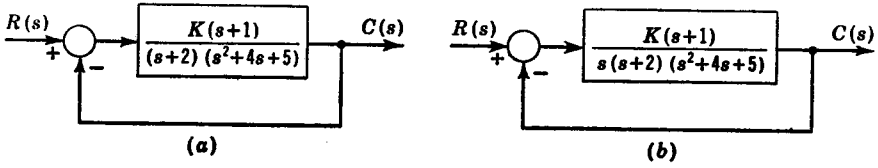


FIG. P 7.6. Control systems.

7.7. Sketch the root-locus plot for each of the following characteristic functions:

(a) $s(s^2 + 16s + 25) + K = 0$ (b) $s(s^2 + 16s + 25) + K(s + 2) = 0$
(c) $s(s^2 + 16s + 25) + K(s + 2)(s + 4) = 0$

7.8. Sketch the root-locus plot for each of the following characteristic functions:

(a) $s(s + 2) + K(s + 4) = 0$ (b) $s^2(s + 2) + K(s + 4) = 0$
(c) $s^2 + 8s + 20 + (s + 2)K = 0$ (d) $s(s^2 + 8s + 20) + K = 0$

7.9. The characteristic equation for Fig. 7.2 is given in Eq. (7.6). If a zero is added to the system, the characteristic equation becomes

$$s(s + 4)(s + 6) + (s + 2)K = 0$$

Sketch the root-locus plot for this new system.

7.10. Sketch the root-locus plot for each of the characteristic equations given below:

(a) $s^2(s + 8) + K = 0$ (b) $s^2(s + 8) + (s + 2)K = 0$

Chapter 8

8.1. Derive the equation of operation for each circuit shown in Fig. 8.5.

8.2. Determine the computer diagram for the following first-order equation:

$$(0.5p + 1)y = x(t)$$

The initial condition is $y(0) = 2$ ft, and the maximum expected values are $x(t)_m = 10$ lb, $y_m = 5$ ft, and $\dot{y}_m = 20$ ft/sec.

8.3. (a) Determine the computer diagram for the following third-order differential equation.

$$(p^3 + 2p^2 + 5p + 10)y = x(t)$$

The initial conditions are $y(0) = 5$ ft, $\dot{y}(0) = 0$, and $\ddot{y}(0) = 10$ ft/sec^2. The maximum expected values are $x(t)_m = 25$ lb, $y_m = 10$ ft, $\dot{y}_m = 20$ ft/sec, $\ddot{y}_m = 50$ ft/sec^2, and $\dddot{y}_m = 100$ ft/sec^3.

(b) Suppose that in part (a) it is found that $e_{3m} = k_3\dddot{y}_m = 25$ volts rather than 100 volts. What is the actual value of $\dot{y}_m$? Revise the computer diagram so that e_{3_m} will be 100 volts.

8.4. Determine the differential equation that is being solved by the computer diagram shown in Fig. P 8.4.

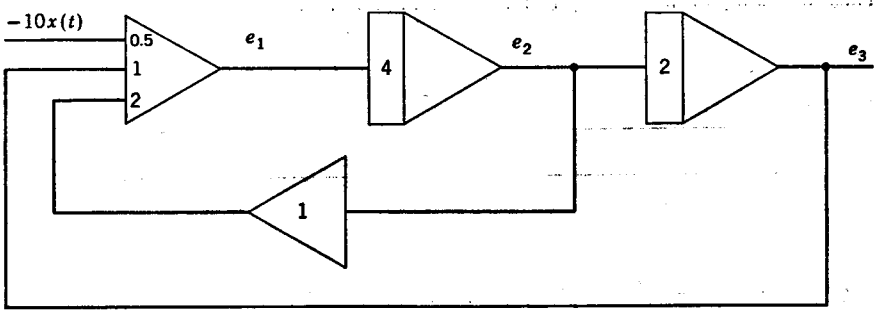


FIG. P 8.4. Computer diagram.

8.5. Let it be desired to speed up the computer solution for Prob. 8.2 by a factor of 5. Determine the computer diagram for this case.

8.6. Let it be desired to slow down the computer solution for Prob. 8.3 by a factor of 2. Determine the computer diagram for this case.

8.7. Derive the equation of operation for each of the amplifier circuits shown in Fig. P 8.7.

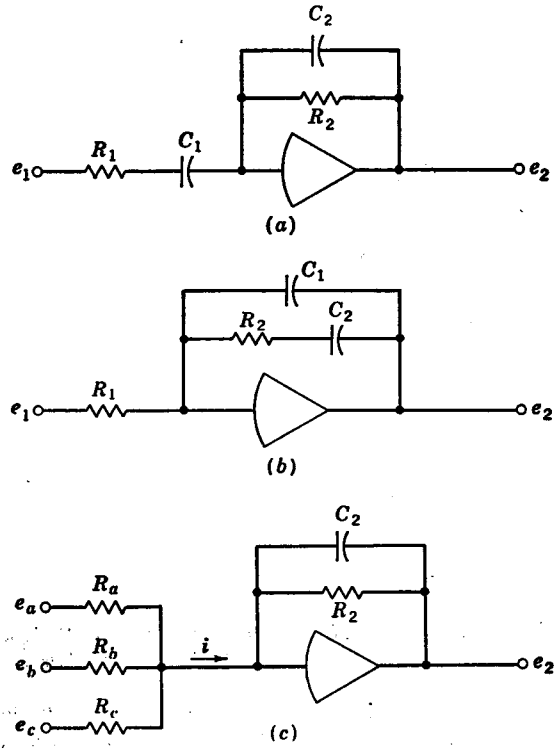


FIG. P 8.7. Amplifier circuits.

8.8. Set up the computer diagram for simulating the system shown in Fig. P 8.8 for the case in which $r(t)_m = 20$, $c(t)_m = 10$, $K = 20$, and $\tau = 0.5$. (Use a 1-μf capacitor.)

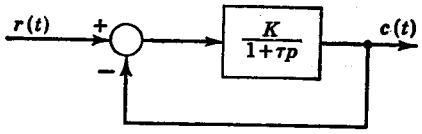


FIG. P 8.8. Control system.

8.9. Determine the computer diagram for simulating the control system shown in Fig. P 8.9. The maximum expected values are $r(t)_m = 5$, $c(t)_m = 10$, $d(t)_m = 25$, and $m(t)_m = 20$. The values of the constants are $K = 10$, $K_2 = 20$, $\tau_1 = 0.2$, and $\tau_2 = 1.0$. (Use 1-μf capacitors.)

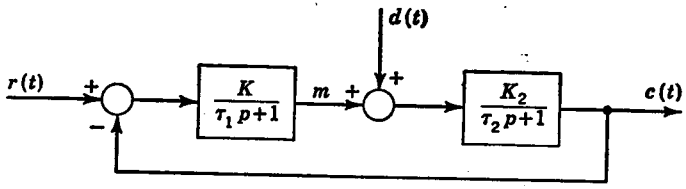


FIG. P 8.9. Control system.

8.10. Determine the block-diagram representation for the control system which is being simulated by the analog computer shown in Fig. P 8.10.

HINT: First write the voltage expression.

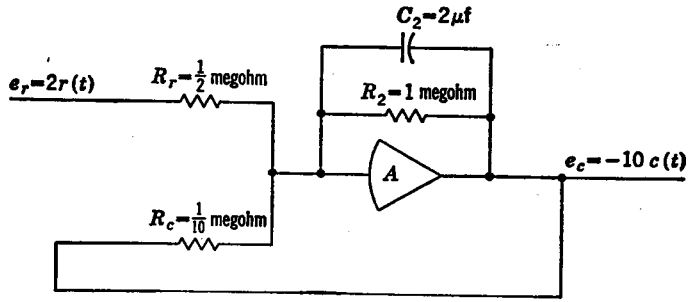


FIG. P 8.10. Computer diagram.

Chapter 9

9.1. Each of the mechanical systems shown in Fig. P 9.1 is excited sinusoidally by a force $f = f_0 \sin \omega t$. For each system, determine

(*a*) The equation for the amplitude ratio y_0/f_0

(*b*) The equation for the phase shift ϕ

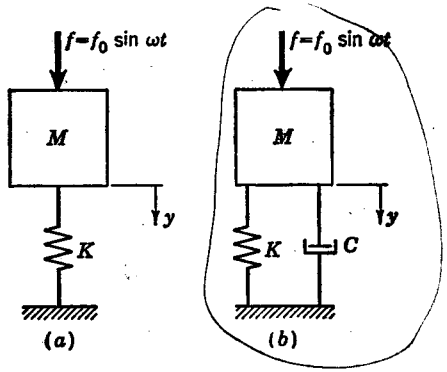


FIG. P 9.1. Mechanical systems.

9.2. Each of the mechanical systems shown in Fig. P 9.2 is excited sinusoidally by a motion of the support $x = x_0 \sin \omega t$. For each system, determine

(*a*) The equation for the amplitude ratio y_0/x_0

(*b*) The equation for the phase shift ϕ

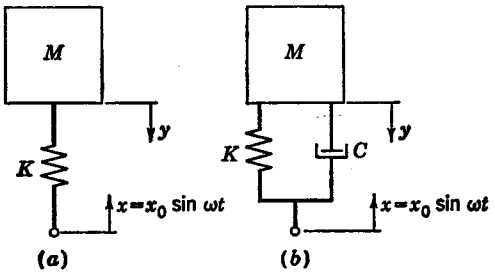


FIG. P 9.2. Mechanical systems.

9.3. For each of the functions given below, evaluate $|G(j\omega)|$ and $\phi = \angle G(j\omega)$ for $\omega = 4(1/0.25) = 16$, $\omega = 2(1/0.25) = 8$, $\omega = 1/0.25 = 4$, $\omega = \frac{1}{2}(1/0.25) = 2$, and $\omega = (\frac{1}{4})(1/0.25) = 1$. [Note that values of ω in the vicinity of the break frequency (1/0.25) yield the most significant information.] Construct the exact log-magnitude diagrams for each function, and sketch in the asymptotes.

(*a*) $G(j\omega) = \dfrac{10}{1 + 0.25j\omega}$

(*b*) $G(j\omega) = \dfrac{10}{j\omega(1 + 0.25j\omega)}$

(*c*) $G(j\omega) = \dfrac{10}{(1 + 0.25j\omega)^2}$

9.4. For the quadratic given below, evaluate $|G(j\omega)|$ and $\phi = \angle G(j\omega)$ for values of $\omega/\omega_n = 4$, $\omega/\omega_n = 2$, $\omega/\omega_n = 1$, $\omega/\omega_n = \frac{1}{2}$, and $\omega/\omega_n = \frac{1}{4}$ and for $\zeta = 0.4$. Repeat for the case in which $\zeta = 0.1$. Construct the exact log-magnitude diagram, and sketch in the asymptotes.

$$G(j\omega) = \frac{10}{1 - (\omega/\omega_n)^2 + j2\zeta(\omega/\omega_n)}$$

9.5. Construct the polar plot for each of the $G(j\omega)$ functions given in Probs. 9.3 and 9.4.

9.6. The asymptotes of the log-magnitude diagram for two $G(j\omega)$ functions are shown in Fig. P 9.6. For each case, the value of ϕ is $-270°$ at very high frequencies. Determine the equation for $G(j\omega)$, and evaluate the gain K for each case.

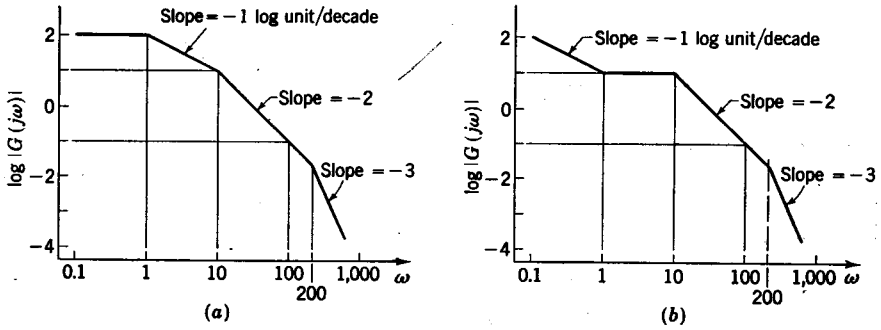


FIG. P 9.6. Log-magnitude diagrams.

9.7. For the system shown in Fig. P 9.7, the frequency-response curves for $G_1(j\omega)$ and $G_2(j\omega)$ were determined experimentally. For both $G_1(j\omega)$ and $G_2(j\omega)$ the phase angle at very high frequencies is $\phi_1 = \phi_2 = -90°$. Construct the log-magnitude plot for $G(j\omega) = G_1(j\omega)G_2(j\omega)$. Determine the equation for $G(j\omega)$, and evaluate the gain K.

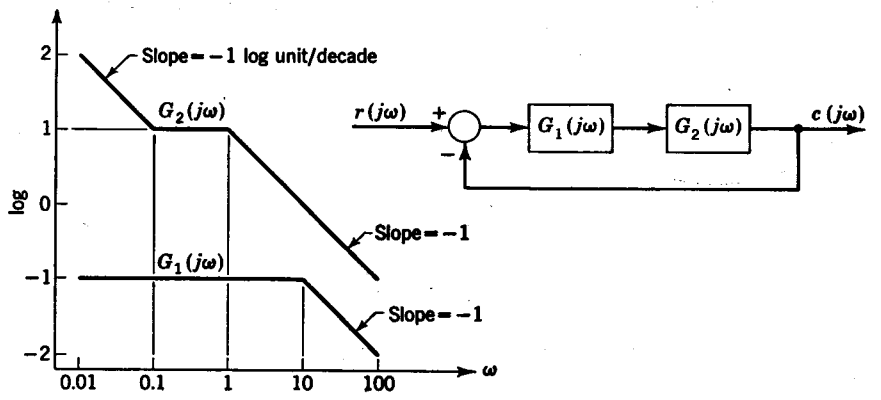


FIG. P 9.7. Log-magnitude diagrams.

9.8. Convert each of the systems shown in Fig. P 9.8 to equivalent unity-feedback systems.

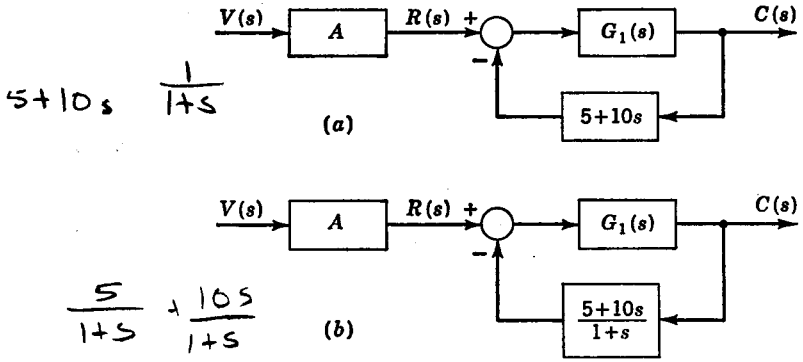


FIG. P 9.8. Block diagrams.

9.9. The polar plots of $G(j\omega)$ for two unity-feedback systems are shown in Fig. P 9.9. For each system, determine whether the response would be better approximated by a first- or a second-order system and also the corresponding value of τ or ζ to be used.

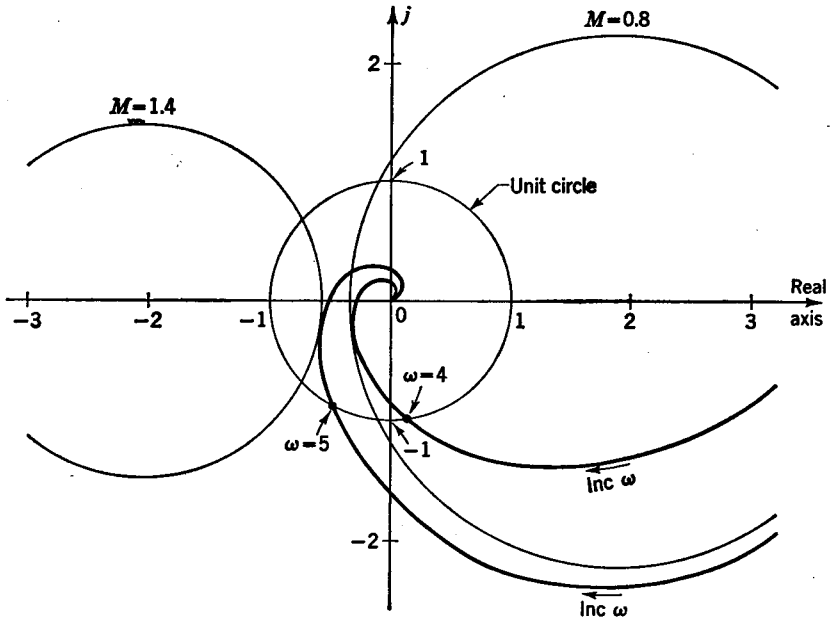


FIG. P 9.9. Polar plots.

9.10. By use of polar plots, determine the value of K to yield an M_m of 1.4 for each of the following unity-feedback systems:

(a) $G(s) = \dfrac{K}{s(1 + 0.25s)}$ (b) $G(s) = \dfrac{K}{s(0.25s^2 + 0.40s + 1)}$

(c) $G(s) = \dfrac{K}{(1 + 0.25s)(0.25s^2 + 0.40s + 1)}$

9.11. Same as Prob. 9.10, but use log-modulus plots (i.e., Nichol's plots) to determine the value of K to yield an M_m of 1.4.

Chapter 10

10.1. For each of the $G(s)H(s)$ plots shown in Fig. P 10.1, the path of values for s is the same as that shown in Fig. 10.2. For each plot, determine the number of zeros of $G(s)H(s)$ which are located in the right half plane ($P = 0$ in all cases).

For each stable system, determine the factor by which the gain should be changed so that the system will just become unstable.

For each unstable system, determine the factor by which the gain should be changed so that the system will just become stable.

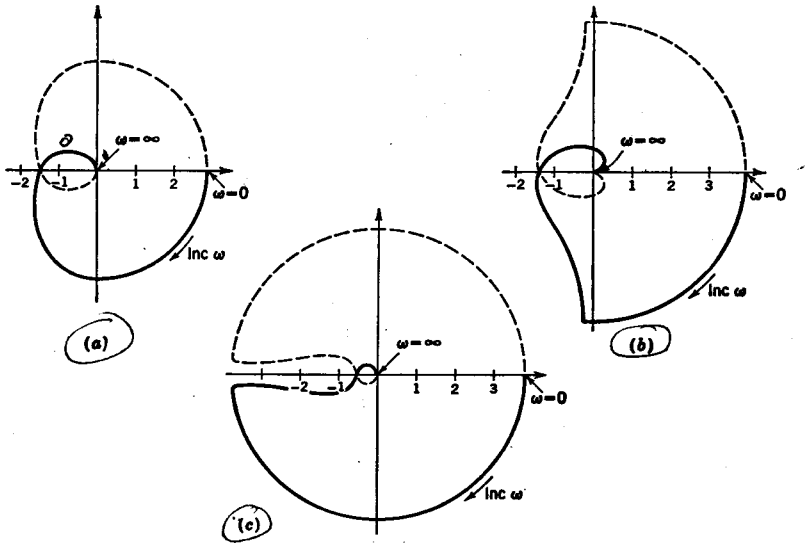


FIG. P 10.1. $G(s)H(s)$ plots.

10.2. For each of the following unity-feedback systems, sketch the complete $G(s)$ plot, and determine the number of zeros of the characteristic function that lie in the right half plane:

(a) $G(s) = \dfrac{10}{(1 + 0.25s)(0.25s^2 + 0.40s + 1)}$

(b) $G(s) = \dfrac{10}{s(1 + 0.25s)}$

(c) $G(s) = \dfrac{10}{s^2(1 + 0.25s)}$ (d) $G(s) = \dfrac{10(1 + s)}{s^2(1 + 0.25s)}$

10.3. Plot the log-magnitude diagram for the system given in Prob. 10.2(*a*). For this system, determine

(*a*) The gain margin and the phase margin

(*b*) The factor by which the gain should be changed to yield a gain margin of 5

(*c*) The factor by which the gain should be changed to yield a phase margin of 40°

10.4. (*a*) For the system $G(s)H(s) = \omega_n^2/(s^2 + 2\zeta\omega_n s + \omega_n^2)$, the phase angle $\phi = \angle G(j\omega)H(j\omega)$ is

$$\phi = -\tan^{-1}\frac{2\zeta(\omega/\omega_n)}{1 - (\omega/\omega_n)^2}$$

When $|G(j\omega)H(j\omega)| = 1$, then show that $\omega/\omega_n = \sqrt{2 - 4\zeta^2}$. Substitute this value of ω/ω_n into the preceding expression for ϕ, and then determine the phase margin ($\gamma = 180° + \phi$) as a function of ζ for this type 0($n = 0$) system.

(*b*) Same as part (*a*) except for the type 1 ($n = 1$) system $G(s)H(s) = \dfrac{K}{s(s + 1/\tau)} = \omega_n^2/s(s + 2\zeta\omega_n)$.

10.5. The substitution of $s + \sigma$ for s has the effect of shifting the imaginary axis a distance σ. Thus, to determine the number of zeros of the characteristic function that have a real part greater than -4, substitute $s - 4$ for s in the function given in Prob. 10.2*b*, and determine the number of zeros of the characteristic function which have a real part greater than -4.

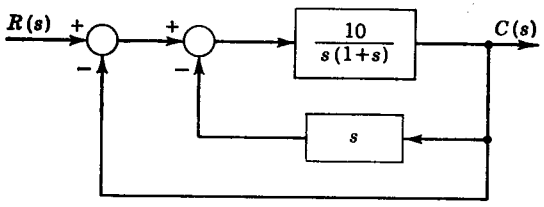
10.6. When the angle $\phi_c + \phi$ in Fig. 10.13 is $-180°$, the corresponding value of $\log |G_cG|$ is -0.5, and thus $|G_cG| = 0.316$. The corresponding gain margin is $K_m = 1/0.316 = 3.16$. If both τ_1 and τ_2 of the lag compensator of Fig. 10.13 are halved, then the equation for $G_c(j\omega)$ becomes $(1 + j\omega)/(1 + 10j\omega)$. This has the effect of moving the curves for $G_c(j\omega)$ to the right, which alters the resultant $G_c(j\omega)G(j\omega)$ plots. Determine the gain margin when this new lag compensator is used.

10.7. Determine the phase margin for the system whose log-magnitude diagrams are given by the $G_c(j\omega)G(j\omega)$ curves of Fig. 10.13. If the original compensator is replaced by the new one described in Prob. 10.6 in which both τ_1 and τ_2 are halved, what will be the new phase margin?

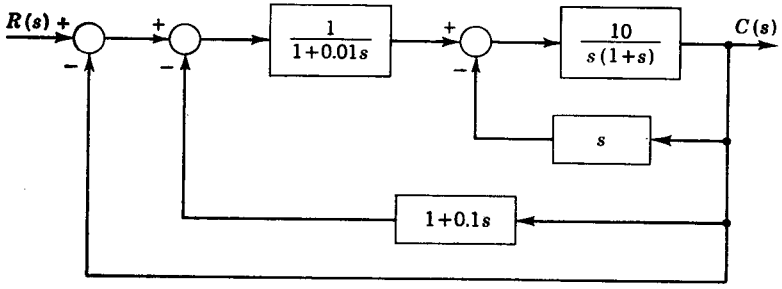
10.8. Determine the gain margin K_m for the system whose log-magnitude diagrams are given by the $G_c(j\omega)G(j\omega)$ curves of Fig. 10.18. If both τ_1 and τ_2 for the lead compensator are doubled, then the equation for $G_c(j\omega)$ becomes $(1 + 2j\omega)/(1 + 0.2j\omega)$. This then shifts the curves for $G_c(j\omega)$ to the left. Determine the gain margin when this new lead compensator is used.

10.9. Determine the phase margin for the system whose log-magnitude diagrams are given by the $G_c(j\omega)G(j\omega)$ curves of Fig. 10.18. If the original compensator is replaced by the new one described in Prob. 10.8 in which both τ_1 and τ_2 are doubled, what will be the new phase margin?

10.10. Construct the approximate log-magnitude plots for each of the two systems shown in Fig. P 10.10. For each system, write the equation for the open-loop transfer function corresponding to the asymptotes. To obtain the exact transfer function for each system, use block-diagram algebra to eliminate the minor feedback loops. Compare the exact and approximate transfer functions.



(a)



(b)

FIG. P 10.10. Block diagrams.

10.11. Same as Prob. 9.10, but use inverse polar plots to determine the value of the gain K to yield an M_m of 1.4.

10.12. Use inverse polar plots to determine the value of β such that the system of Fig. P 10.12 will have an M_m of 1.4.

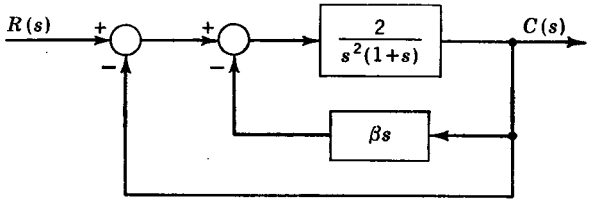


FIG. P 10.12. Control system.

10.13. For the system shown in Fig. P 10.13, determine the value of K_1 to yield an M_m of 1.4.

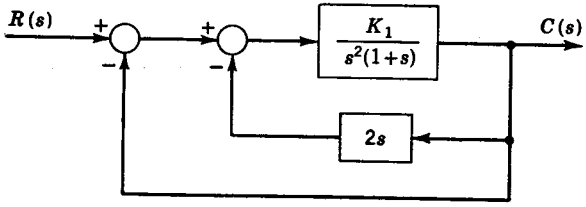


FIG. P 10.13. Control system.

Chapter 11

11.1. The ideal torque required to produce a pressure rise $P_o - P_{in}$ across a pump is equal to the product of the pressure rise and the ideal pump displacement D_i per radian of pump rotation. The torque efficiency of a pump is the ratio of the ideal torque to the actual torque T required to drive the pump $[\eta_T = (P_o - P_{in})D_i/T]$. The volumetric efficiency η_v is the ratio of actual net rate of flow Q discharged per second to the ideal discharge $D_i\omega$, where ω is the angular velocity of the pump. Show that the over-all pump efficiency is equal to the product of the torque efficiency and volumetric efficiency of the pump.

11.2. A hydraulic damper, or dashpot, is shown in Fig. P 11.2. The force F causes a pressure drop ΔP across the piston so that $F = A\,\Delta P$, where A is the net piston area. The leakage flow Q between the piston and cylinder is given by the equation

$$Q = \frac{\pi}{12}\frac{Dd^3\,\Delta P}{L\mu}$$

where Q = rate of flow, in.^{3}/sec
D = diameter of piston, in.
d = mean clearance = ½ difference between cylinder and piston diameters, in.
L = length of piston, in.
ΔP = pressure drop across piston, psi
μ = absolute viscosity, reyns (lb-sec/in.^{2})

The velocity of the piston is $dx/dt = Q/A$. Show that the equation of operation for this damper may be written in the form $F = C(dx/dt)$. What is the resultant expression for the coefficient of viscous damping C?

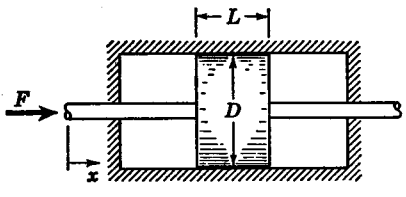


FIG. P 11.2. Hydraulic damper.

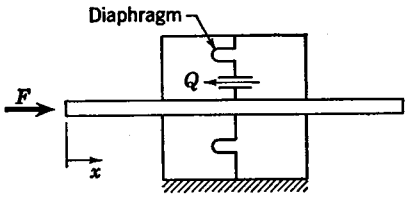


FIG. P 11.3. Hydraulic damper

11.3. A modification of the viscous damper of Fig. P 11.2 is shown in Fig. P 11.3. As in Prob. 11.2, the force F is resisted by the pressure drop ΔP, which acts on the diaphragm. Because the diaphragm prevents leakage, the flow Q must travel through the viscous restriction. The equation for the flow through a viscous restriction is given by Eq. (11.8). The velocity of the diaphragm is $dx/dt = Q/A$. Show that the equation of operation for this damper may be written in the form $F = C(dx/dt)$. Determine the resulting expression for the coefficient of viscous damping C.

11.4. For the flow control system shown in Fig. 11.10*a*, the net flow Q going through the throttle valve is equal to the flow Q_1 supplied by the pump minus the flow Q_d which is bypassed through the differential-pressure-regulating valve to drain (that is, $Q = Q_1 - Q_d$). Because the pump supply Q_1 is constant, for variations about some reference operating condition it follows that $q = -q_d$. The flow through the throttle valve is given by the equation $Q = C_d A_t \sqrt{P_1 - P_2}$. Linearization gives $q = C_1 a_t + C_2(p_1 - p_2)$. Similarly, the amount of flow which is bypassed is given by the expression $Q_d = K_d X \sqrt{P_1}$, where X is the opening of the valve plunger. Linearization gives $q_d = C_4 x + C_4 p_1$. The force balance for the valve plunger is $(P_1 - P_2)A =$

$Mp^2X + KX + F_c$, so that linearization gives $(p_1 - p_2)A = (Mp^2 + K)x$. Combining the preceding linearized expressions yields the block diagram shown in Fig. P 11.4*a*.

Because P_2 rather than P_1 is the independent variable which acts as the external disturbance, the block diagram of Fig. P 11.4*a* should be modified as shown in Fig. P 11.4*b*.

By means of block-diagram algebra, eliminate the minor feedback loop of Fig. P 11.4*b* and determine the resulting characteristic equation for this system.

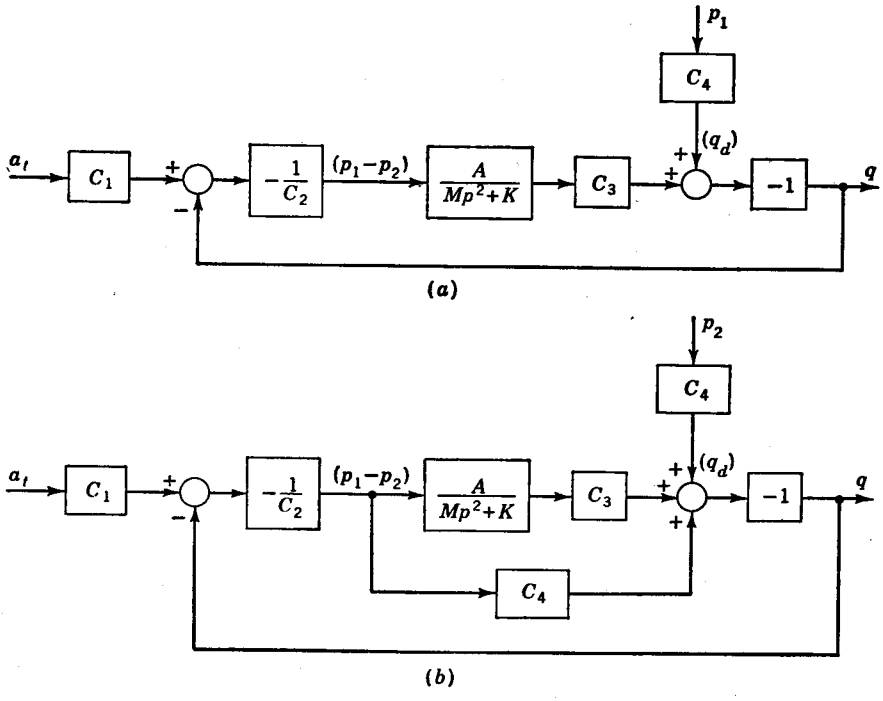


FIG. P 11.4. Block diagram for flow control system.

11.5. For Prob. 11.4, suppose that a viscous damper is inserted as shown in Fig. 11.10*b*. The equation for the flow through this damper is given by Eq. (11.9), and the force balance for the plunger is given by Eq. (11.10). Determine the resulting characteristic equation for this system.

11.6. For the flow control system of Fig. 11.11*b*, the flow $Q = K_d X \sqrt{P_s - P_1}$ goes from the supply line through the unloading valve. This same flow passes through the throttle valve so that $Q = C_d A_t \sqrt{P_1 - P_2}$. Linearization of the preceding expressions gives $q = C_1 x - C_2 p_1$ and $q = C_3 a_t + C_4(p_1 - p_2)$. The force balance for the plunger is $(P_1 - P_2)A = Mp^2X + KX + F_c$; thus for small variations $(p_1 - p_2)A = (Mp^2 + K)x$. Construct the block diagram for this system in which a_t is the input, q the output, and p_2 the external disturbance. What is the resulting characteristic equation?

11.7. Same as Prob. 11.6, except that a viscous damper is inserted between the spring-loaded side of the plunger and the downstream side of the throttle valve.

11.8. Determine the block-diagram representation for the hydraulic servomotor of Prob. 3.1 in which the operation of the four-way valve and cylinder is given by Eq. (11.18) rather than Eq. (2.56).

11.9. Determine the block-diagram representation for each of the hydraulic power amplifiers shown in Fig. P 11.9*a* and *b*. The spring K in Fig. P 11.9*b* represents flexibility in the linkage between the power output and the load. Determine the characteristic equation for each system.

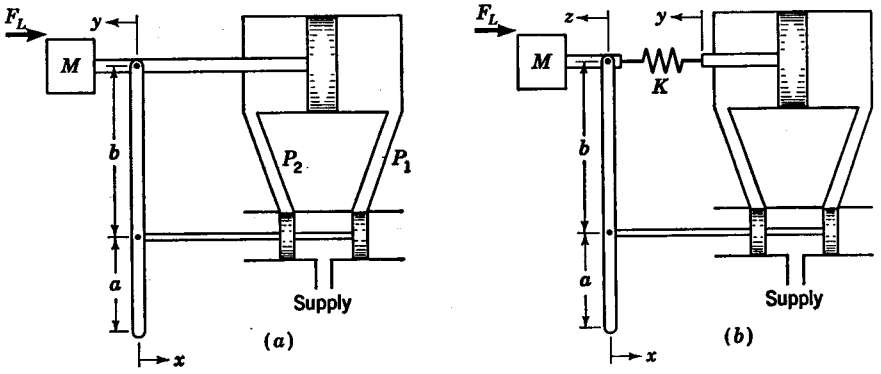


FIG. P 11.9. Hydraulic power amplifier.

11.10. A hydraulic servo system used to control the traverse feed of a machine tool is shown in Fig. P 11.10. Each angular position of the cam corresponds to a desired reference position y_r such that $z = k_r y_r$. For steady-state operation, the valve must be line on line, in which case $x = 0$. Because of the 5:1 lever ratio, it follows that $y_c = 5z = 5(k_r y_r)$. Thus, the value of k_r should be 0.2 in order to have $y_c = y_r$ during steady-state operation. The load on the piston is that due to the tool reactive force on the cross-slide mechanism.

Operation of the valve and cylinder is described by Eq. (11.24). Complete the block diagram for this system. What are the necessary conditions such that there will be no steady-state error due to variations in the external load?

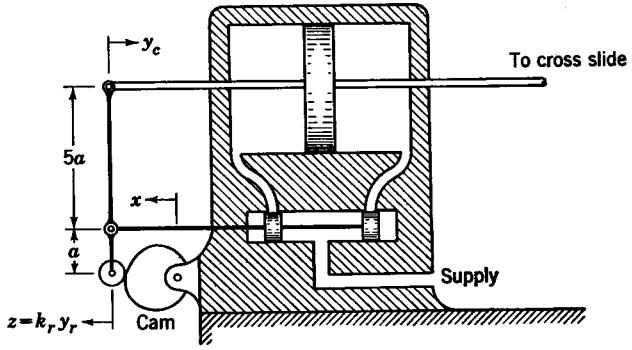


FIG. P 11.10. Machine tool control.

11.11. In Fig. P 11.11 is shown a machine tool with a tape-controlled traverse feed. (The longitudinal and rotational axes may also be tape-controlled if so desired.) The tape reader converts the signal on the punched tape to an electrical voltage $e_r = k_r y_r$ which is proportional to the desired traverse position y_r. A voltage signal $e_c = k_c y_c$ which is proportional to the actual position is fed back and compared with the reference signal. The error signal goes to the coils of a solenoid valve which give a magnetic force ($f_m = k_m e$) in proportion to the error. This force is resisted by a spring such that $f_m = Kx$. The load on the piston is due to the tool reactive force acting on the cross-slide mechanism. Thus, operation of the valve and cylinder may be described by Eq. (11.24). Complete the block diagram for this system. What are the necessary conditions such that there will be no steady-state error due to variations in the external load?

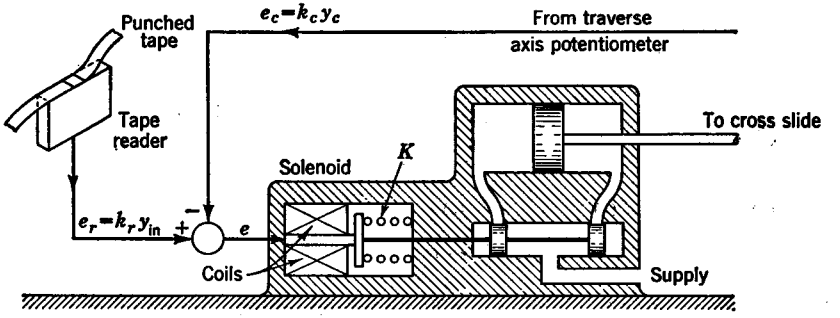


FIG. P 11.11. Tape-controlled machine tool.

11.12. The fluid power source for a hydraulic motor may be a variable-deliver pump (i.e., hydraulic transmission) or a servo valve as shown in Fig. P 11.12. The net flow delivered to the motor is

$$Q_n = Q_1 - Q_{C_1} - Q_L = D_m \dot{\theta}_m \quad (a)$$

where $Q_{C_1} = \frac{V}{B}\frac{dP_1}{dt} = K_c p P_1$ is the equivalent compressibility flow in the P_1 pressure line and $Q_L = K_L(P_1 - P_2) = (K_L/D_m)T_m$ is the leakage flow through the motor.

The return flow Q_2 is the net flow Q_n minus the equivalent compressibility flow Q_{C_2} of the P_2 pressure line

$$Q_2 = Q_n - Q_{C_2} = D_m \dot{\theta}_m - K_c p P_2 \quad (b)$$

The torque developed by the motor is

$$T_m = D_m(P_1 - P_2) \quad (c)$$

Write the flow equations for Q_1 and Q_2 as a function of the respective pressure drops, across the valve and X. Linearize Eqs. (*a*) to (*c*), and then eliminate p_1 and p_2 to

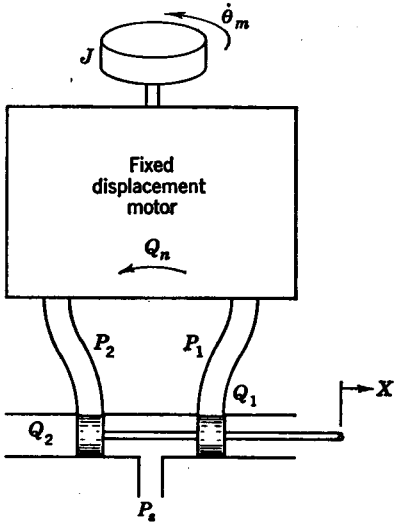


FIG. P 11.12. Hydraulic motor with servo valve.

obtain the functional relationship between $\Delta\theta_m$, t_m, and x. Construct the block diagram for the case in which the load consists of an inertia $Jp^2\theta_m$ and external torque T_L.

Chapter 12

12.1. Derive the equation relating p_o and z for each of the pneumatic control elements shown in Fig. P 12.1*a* to *d*. Explain the significance of each partial derivative which occurs in these equations.

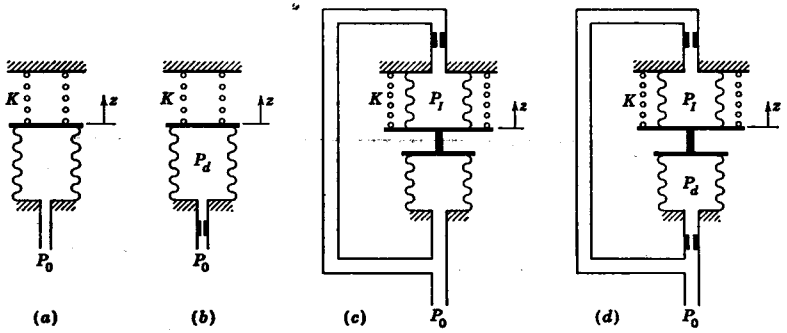


FIG. P 12.1. Pneumatic control elements.

12.2. In Fig. P 12.2 is shown a flapper amplifier in which the controlled pressure is P_o. Determine the over-all block diagram which results when this pressure P_o is connected to each of the pneumatic control elements shown in Fig. P 12.1*a* to *d*. The position z is fed back to the walking-beam linkage as indicated. Note that for this over-all block diagram the input is the error e and the output is the controlled pressure p_o. (The time constant for the flapper-type amplifier may be considered negligible.) Identify the mode of control for each case.

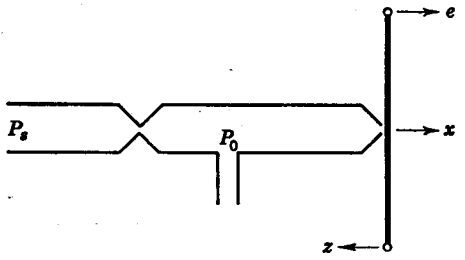


FIG. P 12.2. Pneumatic flapper amplifier.

12.3. The block diagram of a pneumatic position control system is shown in Fig. P 12.3. The controlled position is θ_c, and the reference position is θ_r. Construct the resulting root-locus plot for each of the following:

(*a*) Proportional controller $p_o = K_c e$
(*b*) Proportional plus derivative controller $p_o = K_c(1 + s)e$
(*c*) Proportional plus integral controller $p_o = K_c(1 + 1/s)e$
(*d*) Proportional plus derivative plus integral controller $p_o = K_c(1 + s + 1/s)e$

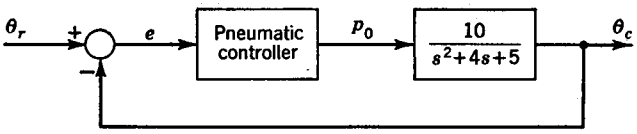


FIG. P 12.3. Position control system.

12.4. Same as Prob. 12.3*b*, but use

(*a*) $p_o = K_c(10 + s)e$ (*b*) $p_o = K_c(0.1 + s)e$

12.5. Same as Prob. 12.3*c*, but use

(*a*) $p_o = K_c(1 + 10/s)e$ (*b*) $p_o = K_c(1 + 0.1/s)e$

12.6. For the pneumatic flapper amplifier shown in Fig. 11.22*a* plot a curve of P_2 versus X. The diameter of the fixed orifice is 0.05 in., the diameter of the flapper opening is 0.20 in., and the supply pressure is 100 psia. Determine K_1 at $P_2 = 30$, 50, and 70 psia.

Chapter 13

13.1. A field-controlled d-c motor is shown in Fig. P 13.1. The motor drives the load through a gearbox so that $\omega_c = n\omega$, where n is the gear ratio, ω is the motor speed, and ω_c is the speed of the load (i.e., the controlled speed). The output shaft is connected to a tachometer, which produces a voltage proportional to the controlled speed ($E_c = K_c\omega_c$). An electronic amplifier is used to amplify the error signal by a factor K_a, that is, $E_f = K_a(E_r - E_c)$. Complete the over-all block-diagram representation for this system. What is the characteristic equation for this system? Sketch the root-locus plot for the case in which the time constant τ_f is 0.1 sec and the viscous friction C_v is negligible.

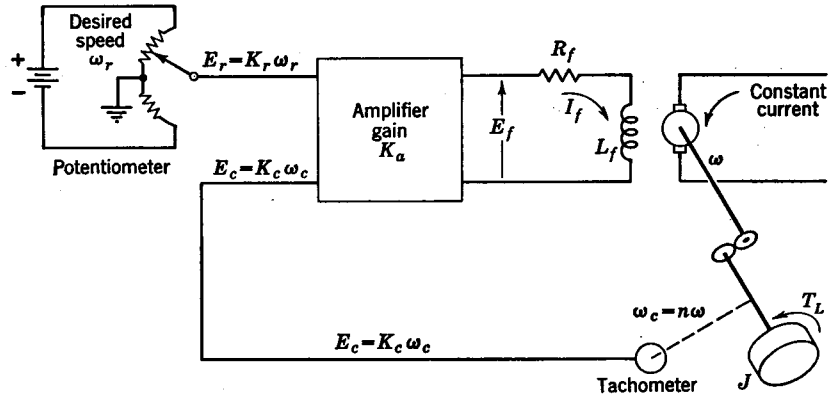


FIG. P 13.1. Field-controlled d-c motor.

13.2. (*a*) In Fig. P 13.2 is shown a generator which is used as a voltage amplifier. The prime mover drives the generator at a constant speed. Determine the equation of operation for the amplification ratio E_2/E_1.

(*b*) Derive Eq. (13.25).

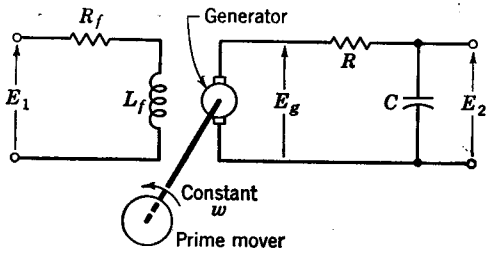


FIG. P 13.2. Voltage amplifier.

13.3. Obtain the over-all block-diagram representation for the positioning servomechanism shown in Fig. 13.11. What is the characteristic equation for this system? Sketch the resulting root-locus plot for the case in which $\tau_{fg} = 0.5$, $\tau_a = 0.1$, and C_v is negligible.

13.4. A Wheatstone bridge as shown in Fig. P 13.4 is commonly used as a comparator for electrical systems. The resistance R_r of the input potentiometer varies in proportion to the reference input. Similarly the resistance R_c is varied in proportion to the controlled variable. The supply voltage E_s and resistances R_1 and R_2 are maintained constant. Determine the equation for the voltage E_e which is a measure of the error.

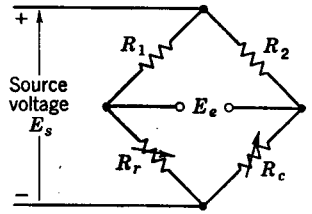


FIG. P 13.4. Wheatstone bridge.

13.5. The gain K of a high-gain amplifier is affected by variations in the tube characteristics due to temperature changes, variations in the supply voltages, etc. Also, slight irregularities in the input signal become greatly magnified at the output (i.e., noise). The use of feedback around an amplifier as illustrated in Fig. P 13.5 tends to make the amplifier insensitive to these changing influences. For this feedback amplifier, the total resistance of the potentiometer across the output is R, and the resistance between the wiper arm and point B is βR. Thus, the potential between point A and B is βE_2, which is fed back. The input to the amplifier is $E_{in} = E_1 + \beta E_2$, and the output is $E_2 = KE_{in}$. Determine the equation of operation for this feedback amplifier.

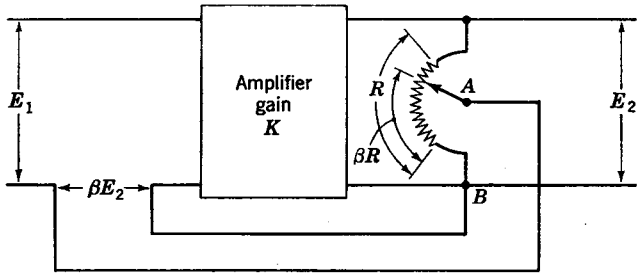


FIG. P 13.5. Feedback amplifier.

13.6. When the amplifier shown in Fig. P 13.5 is operated with the recommended supply voltage, the voltage amplification (without feedback) is 30 per stage. For three stages of amplification, $K = -27{,}000$ at full supply voltage. The negative sign occurs because of the odd number of stages of amplification. When the supply voltage drops 20 per cent, the amplification without feedback is reduced to $K = -20{,}000$.

Determine the voltage amplification E_2/E_1 under the two preceding conditions for this amplifier with feedback in which $\beta = 0.001$. What is the resulting per cent change in over-all amplification?

13.7. Determine the over-all voltage amplification E_2/E_1 for the feedback amplifier shown in Fig. P 13.7.

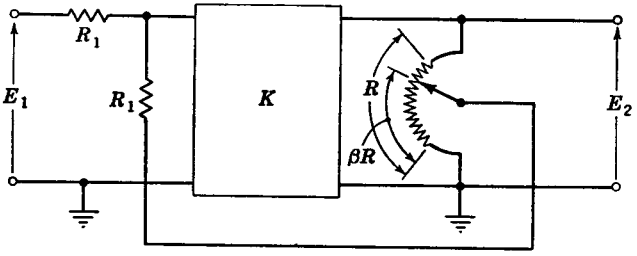


FIG. P 13.7. Feedback amplifier.

13.8. Typical values for a transistor are base resistance $r_b = 1{,}000$ ohms, emitter resistance $r_e = 25$ ohms, collector resistance $r_c = 2$ megohms, and $\alpha = 0.975$.

(*a*) Determine the corresponding values for h_{ie}, h_{fe}, h_{re}, and h_{oe}.

(*b*) Determine each value of G_i and G_v in Table 13.1 for the case in which $R_L = 20{,}000$ ohms.

13.9. Show that for a grounded-base transistor circuit

$$h_{ib} = \frac{h_{ie}}{1 + h_{fe}}$$

$$h_{rb} = \frac{h_{ie}h_{oe}}{1 + h_{fe}} - h_{re}$$

$$h_{fb} = -\alpha = -\frac{h_{fe}}{1 + h_{fe}}$$

$$h_{ob} = \frac{h_{oe}}{1 + h_{fe}}$$

(To obtain these relationships, it is necessary to utilize the approximations that $h_{re} \ll 1$ and $h_{ie}h_{oe} \ll h_{fe}$.)

13.10. Use Eqs. (13.81) to (13.84) to express the values of G_v and G_i in Table 13.1 as a function of the equivalent circuit resistances and α.

13.11. Verify the equations for G_v and G_i given in Table 13.1.

13.12. In working with the four-terminal network of Fig. 13.24*d* it is convenient to use the term $Z_i = v_1/i_1$ and the term $Z_o = v_2/i_2$. Determine equations for Z_i and Z_o in terms of the grounded-emitter h parameters for a:

a Grounded-emitter connection

b Grounded-collector connection

c Grounded-base connection

(Note that $v_1 = R_e i_1$.)

Chapter 14

14.1. In Fig. P 2.5 is shown a linear accelerometer in which y is the motion of the mass relative to the frame and x is the motion of the frame. In the differential equation relating x and y, replace the acceleration term d^2x/dt^2 by a_x (where a_x is the acceleration of the frame). Obtain the resulting transfer function for $Y(s)/A(s)$. Note that because the input is an acceleration the output displacement y will be proportional to the acceleration. Explain why it is desirable to have ω_n very high and ζ between 0.4 and 0.7.

HINT: Sketch the log-magnitude diagram.

14.2. A single-degree-of-freedom gyroscope is shown in Fig. P 14.2. A rotation of the frame about the z axis causes the gyrowheel to precess about the y axis. This

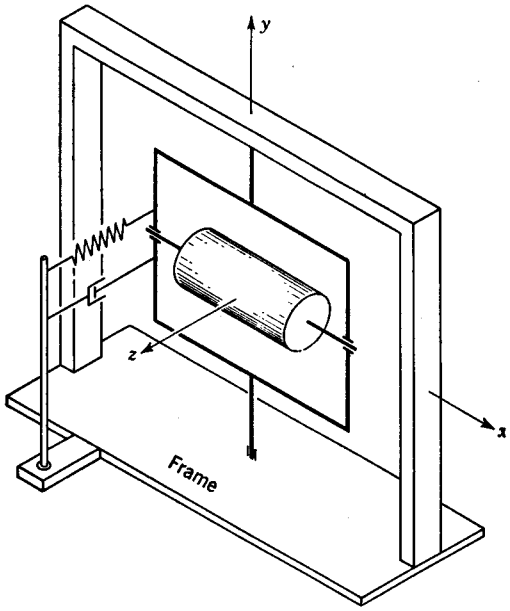


FIG. P 14.2. Single-degree-of-freedom gyroscope.

precession is resisted by a spring and damper which provide an opposing torque such that $T_y = -(K + Cp)\theta_y$. With the aid of Eq. (14.6), determine the equation relating the input motion θ_z to the output θ_y (neglect inertia J_y and damping C_y).

When the opposing torque is provided by the spring only (i.e., no damper), a rate gyro results. What is the equation of operation for this rate gyro? When the resisting torque is provided by the damper only (i.e., the spring K is removed), the device becomes an integrating rate gyro. Determine the resulting equation of operation.

14.3. In Fig. P 14.3 is shown the application of a one-degree-of-freedom gyroscope to provide a stable reference plane. Because of the bearings, motion of the ship is not transmitted to the platform. However, extraneous torques T_{zs} may be transmitted to the platform because of friction in the bearings, inertia forces, etc. Such external disturbing torques tend to rotate the platform about the z axis away from its

reference orientation. A rotation of the platform from its reference position causes the gyroscope to precess about the y axis. An electrical pick-off then detects the error.

For an integrating rate gyroscope which is sensitive to torque about its z axis, it follows that $\theta_y \approx (J_s\omega_s/C)\theta_z$. The electrical pick-off sends a signal to the motor to provide a corrective torque such that $T_m = K_m(\theta_r - \theta_y)$, where θ_r is the desired reference position. From Eq. (14.7) it follows that

$$T_z = T_m - T_{ze} = J_z p^2\theta_z + C_z p\theta_z + J_s\omega_s p\theta_y$$

where T_{ze} is the external disturbing torques tending to rotate the platform from its reference position.

Determine the over-all block-diagram representation for this system in which θ_r is the reference input and θ_z the output.

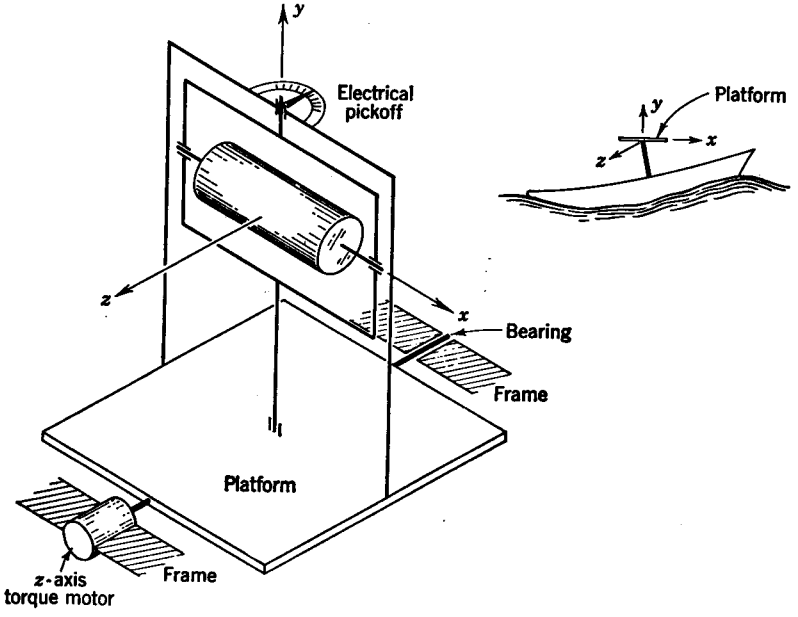


FIG. P 14.3. Stable reference plane.

14.4. Sketch the root-locus plot for the characteristic equation of the system of Prob. 14.3. Assume that the viscous friction $C_z p\theta_z$ is negligibly small.

Add series lead compensation of the form $(1 + \tau_1 s)/(1 + \tau_2 s)$, where $\tau_1 > \tau_2$, to this system. Sketch the resulting root-locus plot. Does lead compensation improve the general stability? Could stability be improved by the use of lag series compensation?

14.5. By the addition of another gimbal and an x-axis torque motor, it is possible to make the platform of Fig. P 14.3 insensitive to rotations about the x axis as well as being insensitive to rotations about the z axis. The gyroscope of Fig. P 14.3 detects rotation about the z axis only. To detect rotations about the x axis, it is necessary to mount another gyroscope on the platform. Make a sketch of the resulting system.

Determine the block-diagram representation in which θ_x is the output and θ_r is the reference input.

14.6. In Fig. P 14.6 is shown the block-diagram representation of an autopilot system for controlling the pitch angle θ_c of an airplane. Determine the required value of the gain K such that the resulting system will have an M_m of 1.4.

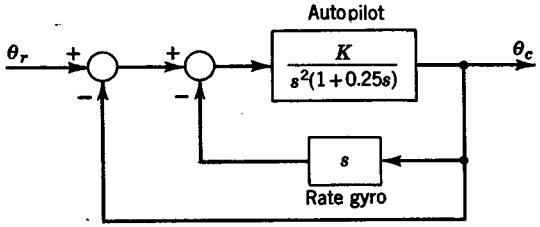


FIG. P 14.6. Autopilot system.

Chapter 15

15.1. The characteristics of a saturation type of nonlinear element are shown in Fig. P 15.1. This element is linear for $-S/2 < x < S/2$ (in this region $y = kx$). The maximum attainable value for y is $kS/2$. Determine the equation for the describing function N.

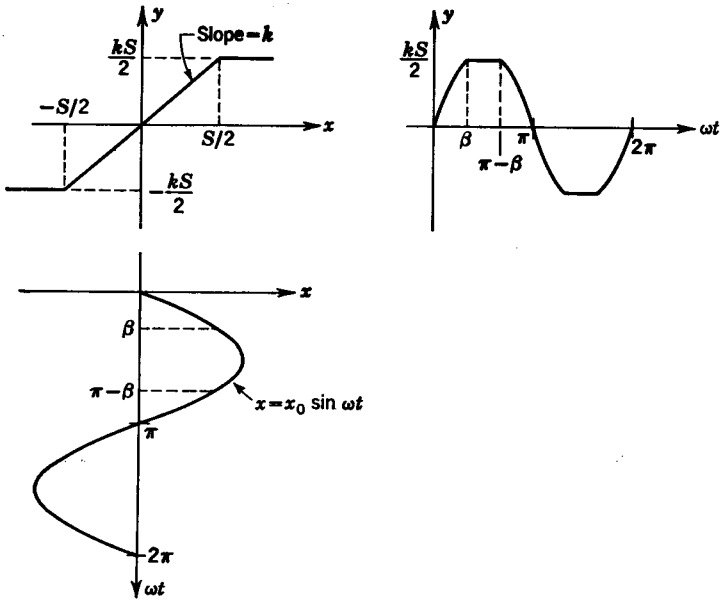


FIG. P 15.1. Saturation-type nonlinear element.

15.2. In applying the root-locus method to the system in Fig. P 15.2, it follows that the angle condition is

$$\measuredangle s + \measuredangle(s + 4) + \measuredangle(s + 6) - \measuredangle N = -180° \pm k360°$$

When the angle of N is zero, as is the case when the nonlinear element does not introduce a phase shift, the root-locus plot is the same as that for the basic system. From the magnitude condition, it follows that

$$|s|\,|s+4|\,|s+6| = |K|\,|N|$$

Thus, the value of the gain along the root-locus plot is KN. The basic root-locus plot for this system is given in Fig. 7.2. For $K = 60$, determine the value of N at which the system becomes unstable.

For the nonlinear element represented by Fig. 15.5, determine the value of x_0/D at which the system becomes unstable for the case in which $k = 5$.

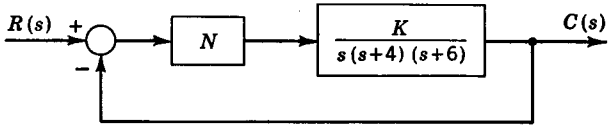


FIG. P 15.2. Control system.

15.3. Same as Prob. 15.2, except that the nonlinear element is represented by Fig. 15.1. Determine the value of Y_0 at which the system becomes unstable.

15.4. Determine the expression for $C(z)/R(z)$ for each of the systems shown in in Fig. P 15.4 in which the sampler is located at different points in the system.

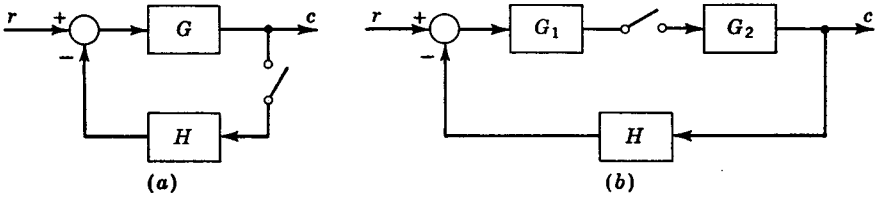


FIG. P 15.4. Sampled-data systems.

15.5. Determine the closed-loop z transfer function for the system shown in Fig. 15.13*b* for the case in which $G(s) = \dfrac{1 - e^{-Ts}}{s}\dfrac{1}{1+\tau s}$ and $H(s) = 1$. Identify the characteristic function of the closed-loop z transfer function, and sketch the root-locus plot. When will the system become unstable?

15.6. Determine the general equation for the phase trajectory for each of the following:

(*a*) $\ddot{x} + C\dot{x} = 0$ (*b*) $\ddot{x} + Kx^3 = 0$
(*c*) $\ddot{x} + K \sin x = 0$

Sketch the trajectories for each of the preceding cases.

15.7. Determine the equation of the isoclines for each of the following:

(*a*) $\ddot{x} + C\dot{x} + Kx = 0$ (*b*) $\ddot{x} + C\dot{x} + Kx^3 = 0$
(*c*) $\ddot{x} + C|\dot{x}|\dot{x} + Kx = 0$

Sketch the trajectories for each of the preceding cases.

INDEX