



FEEDBACK

FEEDBACK

FRED D. WALDHAUER

**Bell Telephone Laboratories
Holmdel, New Jersey**



A Wiley-Interscience Publication

JOHN WILEY & SONS

New York · Chichester · Brisbane · Toronto · Singapore

Copyright © 1982 Bell Telephone Laboratories, Incorporated
Published by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work
beyond that permitted by Sections 107 or 108 of the
1976 United States Copyright Act without the permission
of the copyright owner is unlawful. Requests for
permission or further information should be addressed to
the Permissions Department, John Wiley & Sons, Inc.

Library of Congress Cataloging in Publication Data:

Waldhauer, Fred D.

Feedback.

“A Wiley-Interscience publication.”

Includes index.

1. Feedback (Electronics) I. Title.

TK7835.W29 621.3815'35 81-13104

ISBN 0-471-05319-8 AACR2

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To
Ruth Waina Waldhauer

Preface

Feedback has been one of the more fascinating concepts of technology for centuries, from sixteenth century furnaces that controlled their own temperature to contemporary theories of social interaction. The intuitive understanding of feedback systems at the most elementary level has been made more difficult than necessary by an “endless chain of dependencies” that seems to arise whenever we attempt to analyze a feedback system.

In a system of three interacting things, for example, when thing A affects thing B and thing B , in turn, affects thing C , we feel ourselves to be on firm ground in our understanding of the system, even if we do not know all the details of the interactions between A and B or between B and C . We have a mental picture that follows a cause-and-effect path sequentially from A to B to C .

If C turns around and affects A , however, our mental picture of the interactions is no longer so clear. By introducing *feedback* from C to A , we establish an endless chain of dependencies. The mathematics of the process is well established, but the *schema*, or mental picture, is more complex than it need be. At this point we become involved with the mathematical analysis of the whole process to make sure that we have accounted for everything. We run the risk of getting bogged down in mathematical detail and losing sight of what we are trying to accomplish.

This book adopts a basic change in outlook that greatly simplifies feedback analysis and design. It allows us to retain a clear mental picture of the interactions. In the view developed in this book, we assume that the system

output at C is known. (It must, after all, satisfy some design specification, for example.) Then one part of A is known—the part that comes directly from C through a feedback path. But if the output at C is known, we can infer the input at B from the characteristics of the connecting path from B to C . If we know the input at B , we can similarly infer the contribution to A implied by the input at B (through the characteristics of the connecting path from A to B).

Finally, we *add* the two contributions to A to find the total input to the system. We can thus find the loss of the system—the input divided by the output. No endless chain of dependencies arises, and our mental picture is one of sequential reasoning through the two paths from output to input, in this case from C to A .

To express the distinction between the new theory and the old, we use the term “anticausal analysis” to describe the direction of analysis from output to input.

By applying this change in point of view to many practical areas of circuit analysis and design, we show (1) how it can be used in studying feedback systems and (2) how it is applied to the problems of circuit design. One of the chief benefits of the new approach is that we obtain a traceable path from the initial, rough design approximations to the final, exact analysis and design.

Most of the examples in this book come from electrical circuits, where I have had most of my experience. Examples from audio frequency design to designs of microwave integrated circuits are employed; a uniform approach is adopted over the whole range.

Knuth has said that “the enjoyment of the tools one works with is, of course, an essential ingredient of successful work.”* An object of this book is to provide the reader with an enjoyable set of tools for designing feedback systems. I hope that it will also kindle interest in circuit theory and design.

For readers who would like to apply the methods developed here directly to obtain individual designs of their own, or to check the designs given in the book, 31 programs are given in three appendices. They are written for the Hewlett-Packard HP 41C or 41CV calculator and cover most aspects of the material in the book.

Among these programs is one that synthesizes feedback systems for a prescribed performance. Another converts the HP 41C calculator into a “two-port network calculator.” Included are the four basic functions of addition, subtraction, multiplication and the matrix inverse, as well as lead interchange operations (e.g., conversion from common emitter to common base or common collector), all available at the touch of a button. This “calculator within a calculator” is itself programmable, and means are provided for converting numerical results into network properties, including loss, input and output impedances, and sensitivities as functions of frequency.

These programs were originally written as an aid to the author to assure himself that the approaches taken could be expressed algorithmically. I believe

*Quoted by J. A. Ball in his preface to *Algorithms for RPN Calculators*, Wiley, New York, 1978.

that they have turned out to be more generally useful as teaching tools in themselves. To avail oneself of this feature, a calculator must be acquired (with printer and card reader). Alternatively, the programs of interest can be rewritten for the computer in the reader's own operating system and language.

This book is intended for upper-division undergraduate students of electrical engineering and for professionals who have an interest in designing feedback systems and circuits. It arose from notes written for an in-hours two semester course that I taught at Bell Laboratories. After finishing the book, the reader should be able to design feedback systems in a very direct way, with confidence in the sensitivities of the important design specifications to the devices and components used. I hope that the reader will also be motivated to do original work in this area.

The book is intended for either individual or classroom study at several levels of reader involvement. A good overview of the subject can be obtained by reading the book and following the mathematical developments. To become adept at applying the methods in actual circuit design, the reader should complete the homework problems. Further study is facilitated by the fully documented calculator programs in the appendices.

ORGANIZATION OF THE BOOK

The subject matter is separated into three hierarchical levels: (1) the system level, (2) the circuit level, and (3) the device level. In the interests of clarity in both thought and programs, interaction between hierarchical levels has been restricted to adjacent levels to the fullest extent possible. The book is divided into two parts. Part 1 concerns the relationship between system and circuit levels, and Part 2 concerns the interactions between circuit and device levels. In Part 2 the system considerations of Part 1 are also included.

Separation into hierarchical levels is helpful in breaking the design process into manageable pieces, particularly in the design of monolithic integrated circuits. It is also invaluable in rational programming of the design on the calculator or computer.

FRED D. WALDHUER

*Holmdel, New Jersey
January 1982*

Acknowledgments

I would like first to acknowledge a considerable debt to David Thomas, Executive Director, Transmission Systems Division at Holmdel, whose encouragement, support, and patience made this book possible.

For technical help and help of a more general sort, I express gratitude to M. R. Aaron, Jack Sipress, and many other colleagues—more than is appropriate to mention in a short preface. For other sources who have meant much to me, I thank J. F. McGee of Bell Northern, and an early mentor, W. R. Ayres, who encouraged me in the direction finally taken in this book.

David Favin reviewed the entire manuscript and made invaluable suggestions and comments. I am indebted to several anonymous reviewers at Bell Laboratories whose inputs have significantly improved the content of the book. Gary Baldwin and George Moschytz reviewed an early version of the manuscript, and their help is gratefully acknowledged.

Carol Pellom put the entire book into the UNIX operating system. I am grateful to her for making the process painless, pleasant, and graceful.

F. D. W.

Contents

PART 1 SYSTEMS AND CIRCUITS

Chapter 1 Feedback Amplifiers: An Alternate Foundation. The canonical feedback diagram and equation under the conventional formulation of the feedback problem is given. This is contrasted with the alternate formulation to be used in this book. Examples of feedback in simple nonlinear and frequency dependent systems are discussed, and sensitivities are defined and evaluated. **3**

Chapter 2 Polynomials of Loss: Various Descriptions of Polynomials. This chapter contains a budget of techniques from applied mathematics relevant to feedback systems. These constitute the essentials of the mathematics needed under the new formulation. **43**

Chapter 3 Elements of Feedback Synthesis: A Case Study. A specific three-stage feedback amplifier design is undertaken. The sensitivities of the loss (or gain) to the components and device parameters are developed. An initial look at statistical design is made from a sensitivity point of view. A general synthesis method is developed for single-input–single-output feedback systems. **93**

Chapter 4 Signal Flow Graphs of Polynomials, Rational Functions, and Circuits. The results of Chapters 1 to 3 are consolidated using signal flow

graphs and their related sequential matrices, defined here. Several practical examples of the new design approach are developed from the field of active filters and equalizers. This chapter concludes with a discussion of loop gain, which is clearly defined in the context of signal flow graphs, and the definition of feedback, for which we must jump out of the signal flow graph system. **124**

Chapter 5 Signal Delay in Feedback Systems. Delay in signal paths represents the fundamental limitation on the application of feedback. Delay arises from the transport of electrical carriers in a transistor, for example. This chapter develops the techniques needed to deal with delay: they are applied to the case study example of Chapter 3, completing its practical realization in a 300 MHz, 40 dB amplifier. Delay is also the basic limitation on use of quantized feedback in digital systems. The design of a quantized feedback system for digital regenerative repeaters is given. **162**

PART 2: CIRCUITS AND DEVICES

Chapter 6 Two-Port Analysis of Circuits and Devices. The relationships among various descriptions of two-port networks are examined, with emphasis on the general or $ABCD$ parameters. The $ABCD$ parameters give an anticausal description of the two-port network and are appropriate for the new formulation of the feedback problem. Parameters are classified in two ways: they are either mixed-epoch (h , z , y , g , or S) or separate-epoch ($ABCD$ or F) parameters and are defined for lumped (h , z , y , g , or $ABCD$) or distributed (S and F) circuits. **211**

Chapter 7 Feedback Analysis of the Bipolar Transistor. The principles of the new formulation of feedback apply to physical devices such as transistors. This chapter develops the principles of operation of the transistor as a feedback system under the new formulation. Transistor characterization is simplified, and a useful equivalent circuit is given. **233**

Chapter 8 Two-Port Feedback Analysis. Network properties such as loss and input and output impedances and sensitivities are developed from the $ABCD$ parameters of the network. Each parameter can be expressed as a polynomial in frequency; the polynomial coefficients themselves constitute “handles” for securing desired circuit performance. **273**

Chapter 9 Analog Integrated Circuit Design: Feedback and Feedforward. This chapter develops methods for obtaining the $ABCD$ parameters of arbitrary networks from the characteristics of the active and passive components used in them, using the principles of anticausal analysis. Circuit combinations useful in integrated circuit design are analyzed. **310**

Chapter 10 Output-Stage Design. Anticausal analysis of output stages, such as that used in integrated circuits is advantageous because, by starting the analysis at the output, the drive requirement for the output stage in the presence of nonlinearities is placed directly in evidence. The effects of emitter junction and avalanche nonlinearities are included. Temperature effects in transistors are discussed, with the band-gap reference and Class B biasing for given performance with temperature as examples. **344**

Chapter 11 Noise and Input Stages. Noise in circuits has been treated for two decades under the alternate formulation discussed in this book. In this chapter, thermal, shot, and $1/f$ noise in devices is calculated and represented under the standard treatment. The effects of feedback and feedforward on noise are also discussed, including the noise performance of widely used integrated circuit configurations. **384**

Chapter 12 Differential and Operational Amplifiers. This concluding chapter extends the previous results of Part 2 to three-ports, using differential amplifiers as a vehicle. A general null reference matrix is defined for the three-port and is extendable to general multiport networks. The relation between differential and operational amplifiers is given. **430**

Appendix A. Programs For Manipulating Polynomials **473**

Appendix B. Feedback Analysis and Synthesis Programs **529**

Appendix C. Two-Ports, Transistors, and ABCD Matrices **591**

Index **645**

FEEDBACK

Part 1

Systems and Circuits

Chapter 1

Feedback Amplifiers: An Alternate Foundation

Modern feedback theory may be said to have begun on the Lackawanna Ferry between Hoboken, New Jersey and Manhattan on the morning of August 2, 1927. Harold Black was a passenger on his way to work at Bell Laboratories, where he had been working for some six years on the problem of reducing distortion in amplifiers to be used in repeaters for telephone transmission. On a blank space in his copy of *The New York Times*, he drew the diagram and wrote the equation shown in Fig. 1.1.^{1,2} The diagram has become a commonplace in fields far removed from telephone transmission, appearing in books and journals on control theory, system theory, biology, cybernetics, sociology, and economics. The diagram and the equation represent the canonical view of feedback.

Pinpointing the beginning of feedback theory at this event is arbitrary, perhaps, since Maxwell had analyzed what we recognize as a feedback system—the flyball governor—some 60 years earlier.³ This analysis was based on inventions that preceded it over several centuries, including furnace regulators of Cornelius Drebbel from the sixteenth century, windmill regulators of Mead and others, and steam engines of James Watt in the eighteenth and nineteenth centuries.⁴

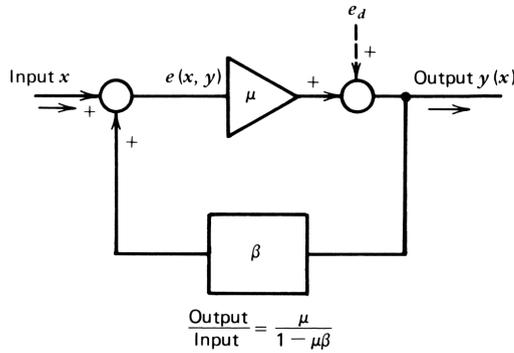


Figure 1.1. Black's feedback diagram and equation.

Nevertheless, Black's diagram and equations were central because they established a *language* with which to talk about feedback systems. This language was later picked up and used in many other disciplines, after Nyquist and Bode had contributed their mathematical insights to problems of amplifier design in the face of inherent instability.^{5,6} In this and the following chapters, we investigate an alternative interpretation of the set of facts represented by Black's diagram and equation. We begin by briefly reviewing feedback under the aspect of the canonical theory introduced by Black.

1.1 CANONICAL FEEDBACK DIAGRAM AND EQUATION

Black was seeking a way of reducing distortion in electronic amplifiers to be used as repeaters for telephone transmission, where even small amounts of distortion would build up to unacceptable levels in many tandem repeaters. To see how the circuit represented by the system diagram in Fig. 1.1 does this, we now develop Black's equation from the diagram. A source signal is applied to the input of an electronic amplifier or *active path* that amplifies it by a factor μ and presents it to the output. A fraction β of the output signal is fed back to the input of the amplifier through a *feedback path* and is of polarity appropriate to reduce the active path input signal. The reason that this reduces distortion in the amplifier is that the undistorted portion of the output signal almost cancels the signal from the source, but the distorted component is not canceled. Its presence at the active-path input tends to cancel the distortion in the active path: it may be regarded as a corrective predistortion applied to the input of the amplifier.

To make these ideas quantitative, we derive Black's equation from the diagram in Fig. 1.1. The output y is related to the input x by solution of the following simultaneous equations:

$$y = \mu e \tag{1.1-1}$$

$$e = x + \beta y \tag{1.1-2}$$

Substituting (1.1-2) in (1.1-1), we obtain Black's equation

$$y = \frac{\mu}{1 - \mu\beta} x \tag{1.1-3}$$

This has been called the *fundamental formula of control theory*.⁷ The quantity $\mu\beta$ is the *loop gain*, and $1 - \mu\beta$ is the *return difference*, so called because if the loop is broken (e. g. at e), and 1 V is applied at the right side of the break, the signal returned to the input is $\mu\beta$, and the difference between this signal and the originating 1 V is $1 - \mu\beta$.

A note on signs is in order. For the closed-loop gain to be stable, it is necessary (but not sufficient) for the sign of either μ or β to be negative. We take up the question of stability in later chapters.

The benefits of feedback are considerable. To see the effect of feedback on distortion, we add a distortion generator e_d in series with the output of the active path, as shown in Fig. 1.1. This generator represents a distortion signal generated in the amplifier. Thus eq. (1.1-1) becomes

$$y = \mu e + e_d \tag{1.1-4}$$

Solving this simultaneously with eq. (1.1-2), we have

$$y = \frac{\mu}{1 - \mu\beta} x + \frac{1}{1 - \mu\beta} e_d \tag{1.1-5}$$

The distortion is reduced by the factor $1 - \mu\beta$. For a substantial reduction in distortion, therefore, the magnitude of $1 - \mu\beta$ must be large: factors of 30th – 100 are common. The beneficial effects of feedback are seen to come from the denominator of the gain expression $1 - \mu\beta$, the return difference.

Another benefit of feedback important to Black's repeaters is the stabilization of gain. An accumulation of gain deviations in many tandem repeaters could lead to overload for an increase in gain and to reduction in signal : noise ratio for a reduction in gain. Bode defined the term "sensitivity" to describe the ratio of the per unit variation in closed-loop gain $K = y/x$ to a small per unit variation in μ :

$$S_\mu^K = \frac{dK/K}{d\mu/\mu} = \frac{d \ln K}{d \ln \mu} \tag{1.1-6}$$

From this equation we can find the sensitivity of the closed-loop gain K to the active-path gain μ :

$$\begin{aligned} S_\mu^K &= \frac{dK}{d\mu} \cdot \frac{\mu}{K} = \frac{1 - \mu\beta + \mu\beta}{(1 - \mu\beta)^2} \cdot \frac{\mu(1 - \mu\beta)}{\mu} \\ &= \frac{1}{1 - \mu\beta} \end{aligned} \tag{1.1-7}$$

This equation says that a 1% variation in μ will cause a $1/(1-\mu\beta)$ percent variation of closed-loop gain. For the canonical diagram in Fig. 1.1, the sensitivity is simply the reciprocal of the return difference. The sensitivity can be found for any parameter in an amplifier. The sensitivity of K to β for the fundamental equation is

$$S_{\beta}^K = \frac{\mu\beta}{1-\mu\beta} \quad (1.1-8)$$

If $\mu = -1000$ and $\beta = 0.1$, for example, the sensitivity to variations in μ is $1/101$, and sensitivity to β is $100/101$. The basic assumption is that the value of β is well controlled (e.g., a ratio of resistors), so that high sensitivity to β is tolerable whereas the value of μ is much less well controlled.

The effect of feedback on noise and other unwanted disturbances is most easily calculated by referring the noise to the input of the amplifier; this is common practice for characterizing and specifying noise. Noise originating internally in the active path is represented in Fig. 1.2. The active path has been

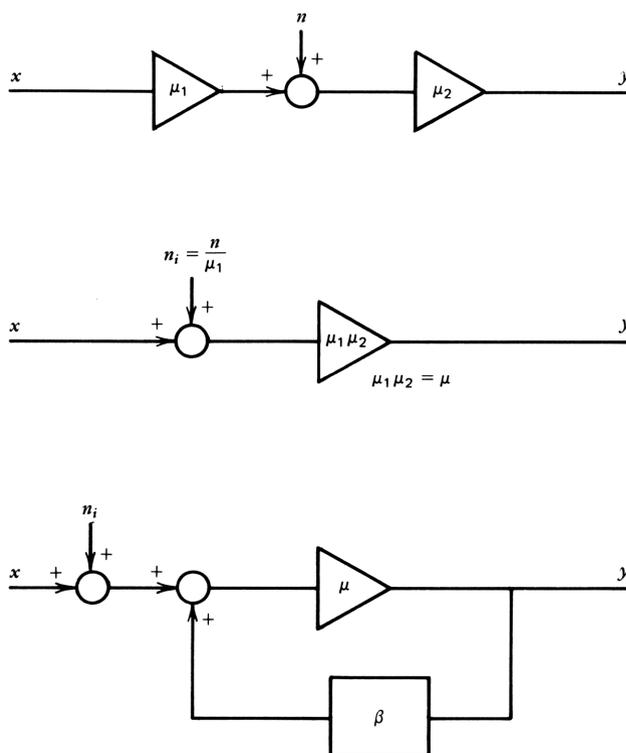


Figure 1.2. By representing all noise sources in the amplifier by an equivalent noise source at the input, noise may be removed from the feedback loop.

split into two (noiseless) portions μ_1 and μ_2 , and a noise source is added between them. This is equivalent to the second diagram, in which the noise source has been divided by μ_1 and moved to the input of the active path. Any other noise sources in the amplifier may be similarly treated, so that the equivalent input noise source n_i will serve to represent them all. To compare the noise performance of the amplifier with and without feedback, we can write

$$y = \mu\beta y + \mu x + \mu n_i \quad (1.1-9)$$

$$y = \frac{\mu x + \mu n_i}{1 - \mu\beta} \quad (1.1-10)$$

Feedback reduces the gain and the noise by the same amount at the output. Thus the signal/noise ratio, *keeping the input signal constant*, is unchanged by the feedback. If noise n_o is injected at the *output*, we may find $n_i = n_o/\mu$ and use it in eq. (1.1-10) to find that the noise is reduced by the factor $1 - \mu\beta$ and that the signal/noise ratio is improved by this factor.

For the fundamental feedback diagram, the benefits of feedback can be summarized to include reduction of distortion and active-path gain variation by a factor of $1 - \mu\beta$, and an improvement in signal/noise ratio (*for given output signal level*) of the same factor. At the input, the signal must also rise by $1 - \mu\beta$ to maintain the given output, so that the improvement in signal/noise ratio comes from the increased input signal.

We have defined two concepts for the canonical diagram that require further discussion: (1) return difference, the denominator of eq. (1.1-3); and (2) sensitivity, in eq. (1.1-6).

Bode made these two concepts precise for general circuits, not just for the canonical diagram due to Black. He chose return difference as the primary concept because, as he said, it “most nearly agrees with the usual conception of feedback.”⁶ In the following section we introduce an alternate formulation of the problem in which return difference disappears but in which sensitivity retains the general meaning given to it by Bode.

The benefits of feedback are not attainable without some cost. First, the gain is reduced by the factor $1 - \mu\beta$ so that additional active-path gain must be provided. A more important and fundamental limitation arises because of *bandwidth limitations* in the active path and *signal propagation delay* around the feedback loop. These effects can cause unsatisfactory dynamic behavior such as ringing and overshoot of the output signal, and even instability. Much of what follows in this and later chapters is concerned with these fundamental limitations and the optimization of performance in their presence. The means by which this is done in this book is considerably simplified by taking an approach that is quite different from the one taken above. We begin the study of the new approach in the following section. In Section 1.3 we discuss bandwidth limitations in the active path. Propagation delay is studied later in this chapter and in Chapter 5.

1.2 AN ALTERNATE FOUNDATION FOR FEEDBACK THEORY

Harold Black wrote his equation on his copy of *The New York Times* as the circuit *gain* of the feedback amplifier. He could as easily have expressed his result as circuit *loss*, merely the reciprocal formulation of eq. (1.1-3):

$$x = \left(\frac{1}{\mu} - \beta \right) y \quad (1.2-1)$$

in which the input x and the active-path input e are expressed in terms of the output y . The negative sign of β arises because feedback signals were defined as adding to e in the previous section; now they are seen to subtract from $e(y) = (1/\mu)y$. Fig. 1.3 contrasts the summation of signals under the conventional and reciprocal formulations. The quantity $(1/\mu) - \beta$ is the *loss ratio*. Loss was used to express the characteristics of transmission lines, to which his repeaters were to be applied, so that the concept of loss as the reciprocal of gain would not have been strange. The loss of a repeater amplifier would have had to be a number less than one. If Black had expressed his result in this way the development of feedback theory might well have taken a different direction. This book builds feedback theory from this alternate point of view. It is shown later that the description of feedback can thereby be simplified substantially.⁸

One conceptual problem with the reciprocal equation concerns the common-sense notion of causality. When we write an equation that says that the input x depends on the output y , we express a mathematical relationship: x is *functionally dependent* on y . We know, on the other hand, that x *causes* y . Most equations that we write in engineering and science are expressed in cause-and-effect form, in which the effect is expressed as functionally dependent on the cause. No doubt this is why Black wrote his equation in the way he did. The output y depends on the input x ; thus it seems “natural” to write the equation

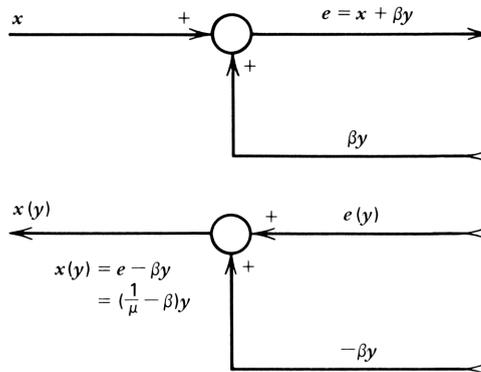


Figure 1.3. Canonical feedback diagram under the reciprocal formulation.

as $y=f(x)$. In this way the mathematical description follows the causal description. When the mathematical description proceeds in an “anticausal” direction, it seems unnatural.

Consider Black’s situation when he developed feedback, however. He knew the output he wanted. It was to be an undistorted signal. After many years of effort, he finally found how to modify the input signal to obtain the desired output. In this sense the reciprocal equation can be read as the answer to the question regarding what input signal is needed to give the required output signal. The reciprocal equation can be considered conceptually as a “designer’s equation.” Although it takes some getting used to, the reciprocal formulation is as intuitively satisfying as the canonical one.

How should we interpret eq. (1.2-1)? The loss ratio of the equation is simply the sum of two components—the loss of the active path and the β loss. The loss of the active path is the reciprocal of μ and is the loss ratio when the β path is set to zero, that is, when the feedback is removed. Likewise, β is the loss ratio when $1/\mu$ is set to zero, that is when the loss of the active path vanishes, or when the gain goes to infinity. The equation contains no denominator; thus the return difference as defined previously is unity.

Let us repeat the gain stability and distortion calculations of the previous section for the reciprocal formulation. Although there is no return ratio or return difference, the physical quantities representing the performance of the amplifier must remain the same. Denoting the ratios $x/y=1/K=L$, the loss ratio, we rewrite eq. (1.2-1) as

$$L = \frac{1}{\mu} - \beta \quad (1.2-2)$$

Applying the sensitivity definition of eq. (1.1-6) to this equation, we find that $dL/d(1/\mu)=1$, so that

$$\begin{aligned} S_{1/\mu}^L &= 1 \cdot \frac{1/\mu}{L} = \frac{1/\mu}{(1/\mu) - \beta} \\ &= \frac{1}{1 - \mu\beta} \end{aligned} \quad (1.2-3)$$

Therefore, the sensitivity of L to $1/\mu$ is the same as that of K to μ , as we would expect since the loss equation is merely a different description of the same physical situation.

There is one important difference, however. The sensitivity under the reciprocal formulation *is also applicable for large changes in the parameter* $1/\mu$ since for a change $\Delta(1/\mu)$, we have

$$\begin{aligned} L &= \frac{1}{\mu} - \beta \\ \Delta L &= \Delta \frac{1}{\mu} \end{aligned}$$

Hence

$$\frac{\Delta L}{L} = \frac{\Delta(1/\mu)}{(1/\mu) - \beta}$$

so that

$$\frac{\Delta L/L}{\Delta(1/\mu)/(1/\mu)} = S_{1/\mu}^L = \frac{1/\mu}{(1/\mu) - \beta} \quad (1.2-4)$$

where $S_{1/\mu}^L$ is now the sensitivity for large parameter changes *and is not a function of the parameter change* (see Problem 1). Suppose, for example, that $1/\mu$ is 0.01 and β is -0.01 , so that $L=0.11$. The sensitivity of L to $1/\mu$ is $1/11$, or 0.0909. Then, if $1/\mu$ vanishes, that is, changes by -100% (a large change, indeed), the percentage change in L would be -9.09% , bringing it down to 0.1, or $-\beta$.

Sum Rule for Sensitivities

The property of the sensitivity of the sum of several elements to one of those elements will be important in later work. If $\Sigma = a + b + c$, then $\partial \Sigma / \partial a = a$, so that $S_a^\Sigma = a / (a + b + c) = a / \Sigma$. We call this the *sum rule* for the sensitivity of a sum of elements to one of these elements. *Furthermore, it is true for a large per unit change in element a.*

As an example, consider the series combination of two resistors, one of 10 Ω , and the other of 90 Ω . The sensitivity of the total resistance to the 10 Ω resistance is $10/100$, or 0.1. If the 10 Ω resistance increases by 10% or 1 Ω , the total resistance increases by 1%. Similarly, the sensitivity of the total to the 90 Ω resistor is 0.9. If this increases by 10%, or 9 Ω , the total increases by $0.9 \times 10\%$, or 9%. In feedback theory based on the alternate, reciprocal foundation, this particularly simple interpretation of sensitivity applies fairly generally because loss expressions tend to be sums.

Returning to the loss equation [eq. (1.2-2)], the sensitivity loss with respect to β can be written directly:

$$S_\beta^L = \frac{-\beta}{(1/\mu) - \beta} \quad (1.2-5)$$

This is the negative of the sensitivity of K to β given in eq. (1.1-8). Although this expression is also true for large changes in β , the sensitivity of K to β given in eq. (1.1-8) is not. The fact that they are equivalent for small parameter changes is evident from the definition of sensitivity given in eq. (1.1-6) and because $\ln u = -\ln(1/u)$, so that

$$S_v^u = -S_v^{1/u} = -S_{1/v}^u = S_{1/v}^{1/u} \quad (1.2-6)$$

which says, for example, that a 1% change in v produces the same change in u as a -1% change in $1/v$.

Distortion Analysis

The analysis of distortion reduction by feedback is different in character under the reciprocal formulation and substantially simpler. It does, however, require a considerable reorientation in thinking about the problem. Instead of a pure sine-wave input producing a distorted output, for example, we now reverse the picture to find out what predistortion is required on the *input* signal to produce a *pure sine-wave output*. Where the input and output are related by a nonlinear function, either formulation can be used. The latter one, however, has not been used extensively and may be unfamiliar. Under the reciprocal formulation, the output signal is the independent variable, and the distortion-producing nonlinearity can be related directly to it. As we see later, this gives a more intuitively satisfying description of the effect of feedback on nonlinearity.

Let us assume that the input–output relationship of the active path can be represented by a power series, in which the input e is instantaneously *dependent* on the *output* y

$$e(y) = \alpha_0 + \alpha_1 y + \alpha_2 y^2 + \alpha_3 y^3 \quad (1.2-7)$$

This equation states that the error signal input $e(y)$ can be regarded as the sum of a direct current (dc) term a_0 , a linear term $a_1 y$, a parabolic term $a_2 y^2$, and a cubic term $a_3 y^3$. Each term is plotted separately in the first column in Fig. 1.4, using (arbitrarily chosen) values of the a coefficients: $a_0 = -0.01$; $a_1 = 0.02$; $a_2 = 0.02$; and $a_3 = 0.05$. The sum of the components is shown at the bottom of the column. For small-output signals, the total is fairly linear, but as the signal increases, the nonlinearity rises rapidly.

We now can find the effect of adding negative feedback on this nonlinearity. As before, the negative feedback path is assumed to be linear and to add no dc term. We take the amount of negative feedback (again an arbitrary choice for illustration) to be $-\beta = 0.06$, or three times as large as the linear term. The feedback signal adds to the linear term (and to the total) by (1.2-1) but does not add to the dc, the parabolic, or the cubic terms. This is shown in the second column in Fig. 1.4. The totals for each of the terms are shown in the third column in Fig. 1.4, which shows only the linear term increased by the feedback. The bottom row shows the sum taken vertically: in the first column is the nonlinear output–input function of the nonlinear active path; in the second, the linear feedback; and in the third, the total of the two.

Clearly, feedback has improved the linearity, and it has done this by increasing the linear term of the series at the input, leaving the distortion products and the dc offset unaffected. We can obtain a *quantitative* measure of the improvement by differentiating eq. (1.2-7):

$$\frac{de}{dy} = \alpha_1 + 2\alpha_2 y + 3\alpha_3 y^2 \quad (1.2-8)$$

We can *normalize* the coefficients to the linear term as follows:

$$\frac{de}{dy} = \alpha_1 \left(1 + \frac{2\alpha_2 y}{\alpha_1} + \frac{3\alpha_3 y^2}{\alpha_1} \right) \quad (1.2-9)$$

in which the second term in the brackets is a measure of the parabolic (or

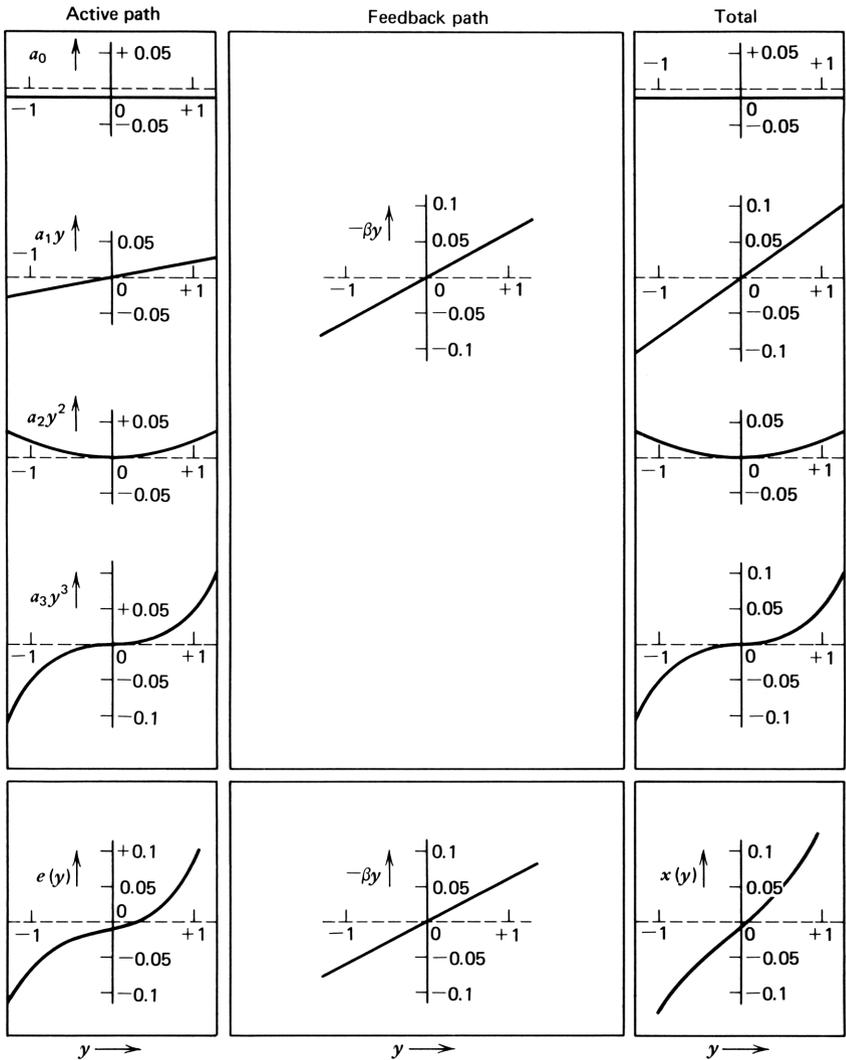


Figure 1.4. Static input-output characteristics for a nonlinear feedback system. The first four rows give the dc offset, the linear component, and the parabolic and cubic components; the fifth row gives the total. The first column is the contribution to the input by the active path; the second is that of the feedback path; and the third is the total of the two.

quadratic) distortion and the third term is similarly a measure of the cubic distortion. Higher-order nonlinearities are absent, but their presence would not affect the discussion. When we add the feedback signal $-\beta y$ to the right side of eq. (1.2-7), we obtain $x(y)$, the input signal with feedback. When this modified equation is differentiated, we obtain

$$\frac{dx}{dy} = (\alpha_1 - \beta) + 2\alpha_2 y + 3\alpha_3 y^2 \quad (1.2-10)$$

When this expression is normalized to its linear term, we obtain

$$\frac{dx}{dy} = (\alpha_1 - \beta) \left(1 + \frac{2\alpha_2 y}{\alpha_1 - \beta} + \frac{3\alpha_3 y^2}{\alpha_1 - \beta} \right) \quad (1.2-11)$$

Comparing the second and third terms in the brackets with the similar terms of eq. (1.2-9), we see that both parabolic and cubic measures of distortion have been reduced by the factor $\alpha_1/(\alpha_1 - \beta)$. (Note that for negative feedback, either α_1 or β must be negative, but not both.) But α_1 is just $1/\mu$, the small-signal value of the loss, so that the individual distortion components are each multiplied by

$$\frac{\alpha_1}{\alpha_1 - \beta} = \frac{1}{1 - \mu\beta} \quad (1.2-12)$$

as obtained in the previous section.

Figure 1.4 clarifies several points that were obscure under the conventional formulation. First, for a given output (note that the output is plotted as the *abscissa* in the curves in Fig. 1.4), addition of linear feedback *does not change the nonlinear components* at the input. It does change the linear component of input, however, and this input tends to swamp out the nonlinear components. Hence the nonlinear components are reduced *relative to the linear component*. In this sense the output-input characteristic with feedback, shown in the lower right graph, is more linear than that without feedback. Second, the *dc offset* a_0 is unchanged by the application of feedback.

It is useful to think of the undistorted output signal as the desired signal and the nonlinear terms in the input signal as corrective predistortion signals needed (along with the undistorted component of input signal) to obtain the desired output signal. Since the output is undistorted, feedback can only increase the undistorted input signal required while leaving the predistortion components unchanged.

The situation depicted in Fig. 1.4 corresponds to negative feedback, in which the slope of $x(y)$ is increased when feedback is applied; that is, the loss increases, or the gain is reduced. Suppose that we now change the sign of the feedback and let $-\beta = -0.06$, as shown in Fig. 1.5, leaving the active path unchanged. This corresponds to positive or regenerative feedback and for the illustrative numbers chosen leads to an $x(y)$ whose slope becomes negative

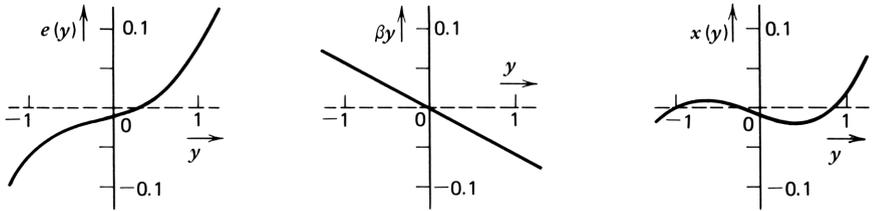


Figure 1.5. Positive feedback: (a) the active-path contribution; (b) the feedback contribution; (c) the total.

over a region near the origin. In this example y is a multivalued function of x for the functional dependencies of the conventional formulation. Under the reciprocal formulation, however, x is a single-valued function of y . The type of curve shown in the right-hand graph in Fig. 1.5 is encountered in various unstable circuits such as flip-flops, multivibrators, and oscillators. In this book we are concerned primarily with negative feedback.

The discussion to this point has sought to point out the advantages in clarity that are obtainable simply by taking the output signal of a feedback system as the independent variable and finding the input signal that is required to obtain that output. Another aspect of this simplification is that equations written under the reciprocal formulation are more easily solved. The reason for this is that where nonlinearities exist, they are more directly related to the output rather than the input signal, so that we can find an explicit expression for the input in terms of the output. The reverse is generally not the case. The circuit example described in Example 1 will serve to illustrate this point.

Example 1. Consider the common emitter transistor circuit in Fig. 1.6, in which an external emitter resistor has been added to improve the linearity. The small-signal input voltage may be approximated by applying Ohm's law:

$$dV_b = (r_e + R) dI_c \quad (1.2-13)$$

where r_e is the (incremental) emitter resistance, given by $r_e = kT/qI_c$; $I_c = I_C + i_c$, the dc quiescent value and the signal current, respectively; and kT/q is 0.026 V

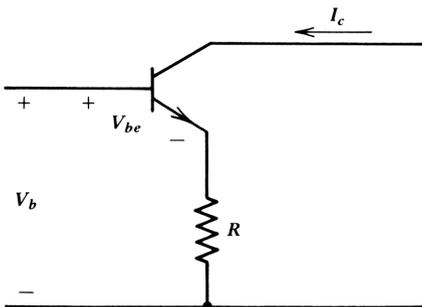


Figure 1.6. Common emitter transistor with emitter resistor feedback for linearization.

at room temperature, so that $r_e = 26 \Omega$ at $I_c = 1 \text{ mA}$. In this equation r_e may be interpreted as the reciprocal of the active-path gain, which is a *transconductance*. The value R may be interpreted as $-\beta$, so that eq. (1.2-13) corresponds to eq. (1.2-1). Thus we can integrate (1.2-13) after the appropriate substitutions to yield

$$\begin{aligned} V_b &= \frac{kT}{q} \int \frac{di_c}{I_c + i_c} + R \int di_c \\ &= \frac{kT}{q} \ln \left(1 + \frac{i_c}{I_c} \right) + Ri_c + V_B \end{aligned} \quad (1.2-14)$$

where the constant V_B is the dc base-to-common voltage consisting of V_{BE} , the dc bias voltage from base to emitter, and RI_C , the dc drop across R . The dc bias voltage from base to emitter V_{BE} is about 0.7 V at 1 mA for silicon transistors. For values of i_c smaller than the quiescent value, we can expand the ln function in a power series:

$$V_b = \left(\frac{kT}{qI_C} + R \right) i_c + \frac{kT}{q} \left(\frac{1}{2} \gamma^2 + \frac{1}{3} \gamma^3 + \frac{1}{4} \gamma^4 + \dots \right) \quad (1.2-15)$$

where $\gamma = i_c/I_C$. The first term is linear in i_c . The second term represents the distortion. Just as in Fig. 1.4, the linear term is increased by the feedback and the distortion terms are unaffected by it. The feedback signal is the voltage across R arising from the output current flowing through it. By the sum rule, the distortion terms have been reduced by the factor $r_e/(r_e + R)$. The sensitivity of V_b to r_e has been reduced by the same factor.

Although it is easy to visualize these relations in the reciprocal formulation, it is not as obvious how they can be put in the form of Black's canonical feedback equation and diagram. However, if we set $g_m = 1/r_e$, calling g_m *transconductance*, we can write

$$dI_c = \frac{g_m}{1 + g_m R} dV_b \quad (1.2-16)$$

From this the canonical diagram in Fig. 1.1 can be drawn with $\mu = g_m$ and $-\beta = 1/R$. Note that *the canonical diagram is not a circuit diagram, but rather a diagram relating signal variables*. We discuss this later when we study more general diagrams of this type—signal flow graphs. In eq. (1.2-14) $g_m = qI_c/kT$, so the equation can be written

$$dI_c = \frac{qI_c}{kT + qI_c R} dV_b \quad (1.2-17)$$

In this equation I_c and V_b both appear on the right side; the separation of

variables that made eq. (1.2-13) easy to integrate is not obtained under the conventional formulation.

1.3 FEEDBACK AROUND A FREQUENCY-DEPENDENT ACTIVE PATH

The study of feedback systems primarily involves their *dynamic* behavior, which can be treated by differential equations in the time domain, or by frequency domain methods. We study dynamic behavior in the frequency domain in Chapters 2 and 3. This section introduces the subject by considering feedback around an active path that has a simple frequency response cutoff, one in which the high-frequency gain is inversely proportional to frequency. A single transistor or an *internally compensated* operational amplifier (such as type 741) have high-frequency characteristics of this type. We are interested here in developing and comparing the mathematical descriptions of such an amplifier under the canonical and the reciprocal formulations.

We assume that the low-frequency active-path gain is μ_0 and that its behavior as a function of the angular frequency ω is given by

$$\mu = \frac{\mu_0}{1 + j\mu_0\tau_1\omega} \quad (1.3-1)$$

where τ_1 remains to be defined. Since μ is a complex quantity, two numbers are required to represent it at any given frequency. We could represent it by its real and imaginary parts, for example. Perhaps the most familiar and useful way to represent μ is by its magnitude and phase. We can write

$$\mu = |\mu|e^{j\theta} \quad (1.3-2)$$

Taking natural logarithms, we obtain

$$\ln \mu = \ln|\mu| + j\theta \quad (1.3-3)$$

Expressed in this way, we have as our two numbers the log of magnitude (expressed in nepers) and the phase (expressed in radians). As a matter of convention, we change the measure of magnitude from nepers to decibels: the magnitude is $20\log_{10}|\mu|$. We also change the measure of phase from radians to degrees by multiplying by $180/\pi$. A plot of these two quantities as a function of frequency on a log scale is called a *Bode diagram* or *plot*.

Figure 1.7 shows a Bode diagram of μ . At low frequencies the magnitude is μ_0 and the phase is zero. If μ were to have a phase reversal, the phase would be 180° without affecting the magnitude. At high frequencies the magnitude falls at 6 dB per octave (20 dB per decade) in a straight-line relationship. It crosses

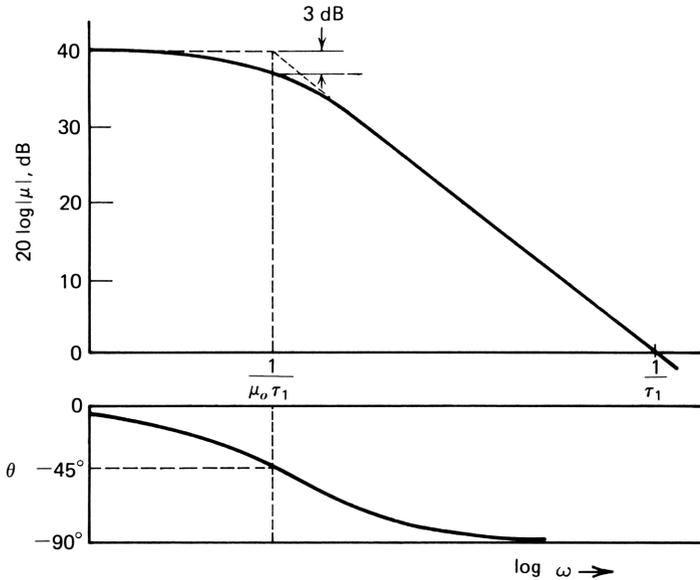


Figure 1.7. Bode diagram of an active path including a simple frequency cutoff at $\omega = 1/\mu_0\tau_1$.

the unity gain (0 dB) ordinate at an angular frequency of $\omega = 1/\tau_1$, the unity gain frequency. At a frequency of $\omega = 1/\mu_0\tau_1$, the denominator of μ is $j + j1$, so that the magnitude is less than the low-frequency value by a factor of $\sqrt{2}$, or 3 dB. This is called a *corner* frequency and, because it is concave downward, is a *downward* corner frequency. The dashed lines represent the *asymptotes* of the response.

The pocket calculator is useful for sketching Bode diagrams on semilogarithmic graph paper; such sketches are useful aids to understanding and are part of the language of feedback systems. In this simple case the real and imaginary parts of the denominator are converted to polar coordinates, and the magnitude is divided into the numerator. The \log_{10} of the result is taken and multiplied by 20 to obtain the magnitude ordinate. The phase is just the negative of the phase of the denominator.

We now wish to find the effect of connecting a feedback network to the active path defined by eq. (1.3-1). From the canonical equation, we have

$$\begin{aligned}
 K &= \frac{\mu}{1 - \mu\beta} \\
 &= \frac{\frac{\mu_0}{1 + j\mu_0\tau_1\omega}}{1 - \frac{\mu_0\beta}{1 + j\mu_0\tau_1\omega}}
 \end{aligned}
 \tag{1.3-4}$$

For example, if we take $\beta = -0.1$, $\mu_0 = 100$, and $\tau_1 = 1.0$, we have

$$\mu = \frac{100}{1 + j100\omega} \quad (1.3-5)$$

$$\mu\beta = \frac{-10}{1 + j100\omega} \quad (1.3-6)$$

$$1 - \mu\beta = \frac{11 + j100\omega}{1 + j100\omega} \quad (1.3-7)$$

and, finally,

$$K = \frac{100}{11 + j100\omega} \quad (1.3-8)$$

Bode diagrams for these four quantities are plotted in Fig. 1.8.

At low frequencies $\mu = 100$, so that $20\log|\mu| = 40$ dB. At $\omega = 0.01$, it has a downward corner frequency above which it begins to fall with frequency at a rate of 20 dB per decade. The phase is $\tan^{-1}(\tau_1\omega\mu_0)$. The curve of the loop gain $\mu\beta$ with $\beta = -0.1$ is of the same shape in magnitude, moved down by 20 dB, and the same shape in phase, moved 180° because of the negative sign of β . The $\mu\beta$ curve crosses the 0 dB line (unity gain) at $\omega = 0.1$. This is termed the *loop-gain crossover frequency*. The return difference $1 - \mu\beta$ has a magnitude of 11 (20.8 dB) at low frequencies, and a downward corner at $\omega = 0.01$, the same as $\mu\beta$. At high frequencies, however, it must equal unity (0 dB) so that it has an upward corner at $\omega = 0.11$. The phase of the return difference is zero at both very low and very high frequencies and reaches a minimum of -57° at $\omega = 0.033$. Finally, the closed-loop gain K is μ divided by $1 - \mu\beta$; on the dB magnitude scale it is the difference between the μ and return difference curves. The phase of K is also the difference between the μ and $1 - \mu\beta$ curves.

This completes the description under the conventional formulation. In addition to the forward-path gain and the feedback loss, we investigated their product (the loop gain) and the sum of unity plus the loop gain (the return difference) to determine the closed-loop gain. Hence five quantities must be evaluated as functions of frequency. Four of them are shown in the Bode plot; the fifth, β , is a constant here and was not shown, but it, too, would generally be a function of frequency. Let us now contrast this description with that under the reciprocal formulation, in which we find the closed-loop *loss* rather than gain.

The loss of the active path is found by taking the reciprocal of eq. (1.3-1):

$$\frac{1}{\mu} = \frac{1}{\mu_0} + j\tau_1\omega$$

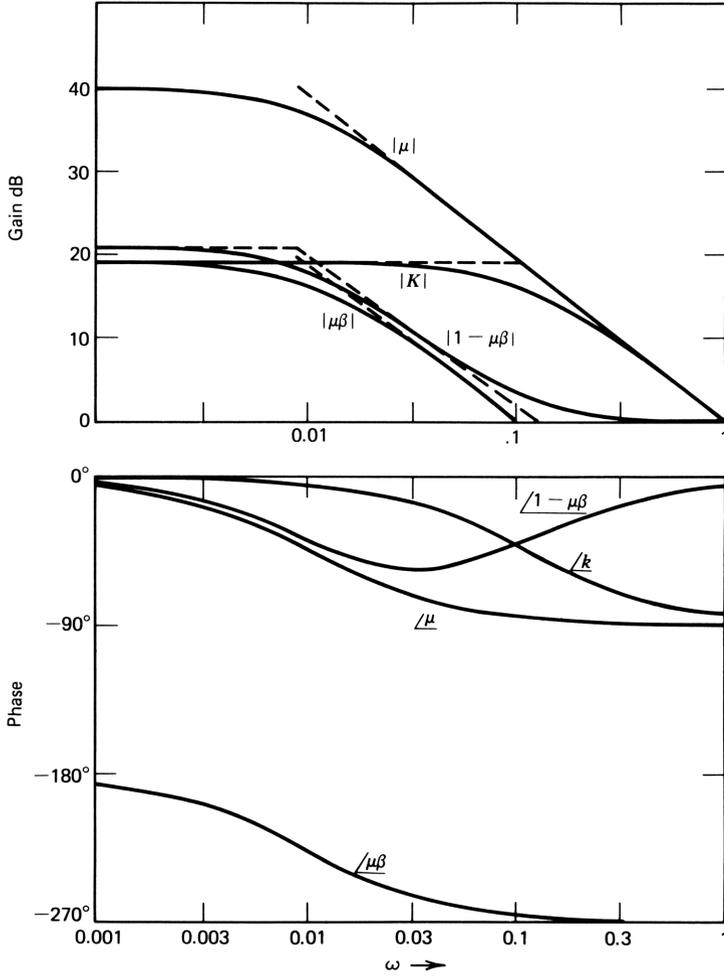


Figure 1.8. Bode diagrams for μ , $\mu\beta$, $1-\mu\beta$, and K for a feedback system incorporating a simple cutoff in the active path.

For convenience, we let $1/\mu_0 = a_0$ and $1/\mu = L_a$, so that

$$L_a = a_0 + j\tau_1\omega \tag{1.3-9}$$

Next, we find the closed-loop loss L by adding the feedback contribution to the input:

$$L = a_0 - \beta + j\tau_1\omega \tag{1.3-10}$$

Taking the values used in Example 1, namely, $\beta = -0.1$, $a_0 = 0.01$ ($\mu_0 = 100$),

and $\tau_1 = 1.0$, we have

$$\begin{aligned} L(j\omega) &= 0.01 + 0.1 + j\omega \\ &= 0.11 + j\omega \end{aligned} \quad (1.3-11)$$

Contrast this result with that of (1.3-8). Under the reciprocal formulation, loop gain and return difference are spurious quantities; we need consider only the sum of the active path contribution to the input L_a and the feedback contribution β .

Bode diagrams of the three quantities of interest, shown in Fig. 1.9, are more easily interpreted than those in Fig. 1.8. The feedback simply adds to a_0 , thereby raising the frequency at which the magnitude of the frequency-sensitive term equals that of the constant (dc) term. The shapes of the curves for L_a and L are identical; the magnitude curve is shifted up and to the right by the

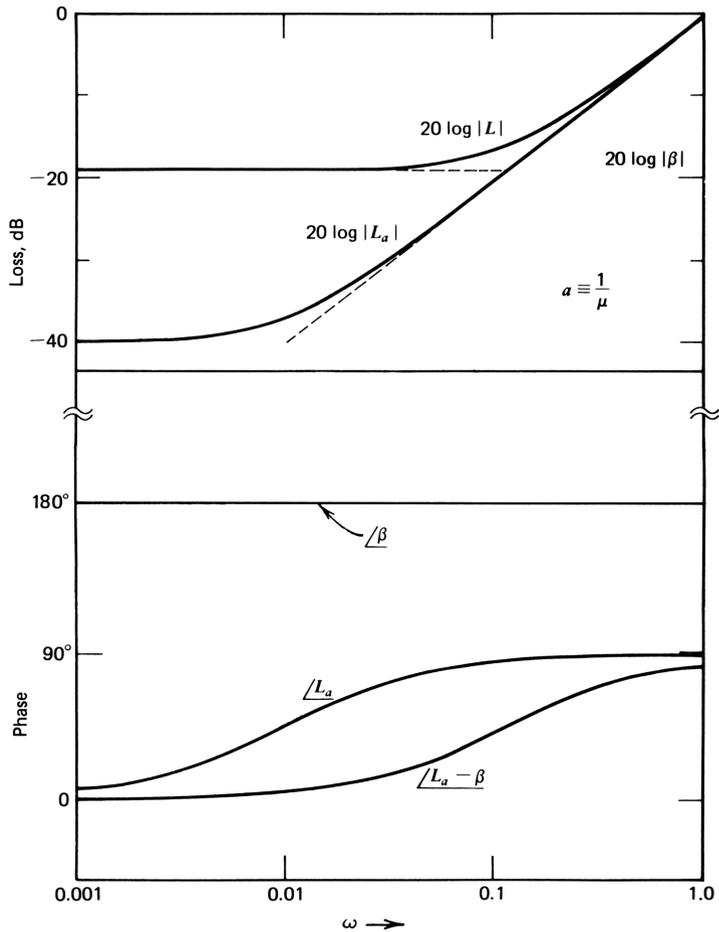


Figure 1.9. Bode diagrams for L_a , β , and L for the feedback system described under the reciprocal formulation.

feedback, and the phase curve is simply moved to the right. Seen in the light of the reciprocal formulation, the conventional view is awkward. Stated in a more positive way, the reciprocal formulation simplifies the description of feedback processes in the frequency domain, and in a way similar to the one that we found for distortion in the previous section. In Chapter 4 we make the mathematical nature of this simplification more precise.

We can interpret this simplification in terms of the canonical block diagram in the following way. The expression for the frequency cutoff of the active path given in eq. (1.3-1) bears a striking resemblance to the fundamental feedback equation (1.1-3); μ_0 corresponds to μ , and $j\omega\tau_1$ corresponds to $-\beta$. Just as we eliminated the denominator in the fundamental feedback equation by use of reciprocal formulation, we eliminate the denominator of the active-path expression when we use *its* reciprocal. This suggests that we can view the forward-path gain in eq. (1.3-1) and Fig. 1.10a as a canonical feedback

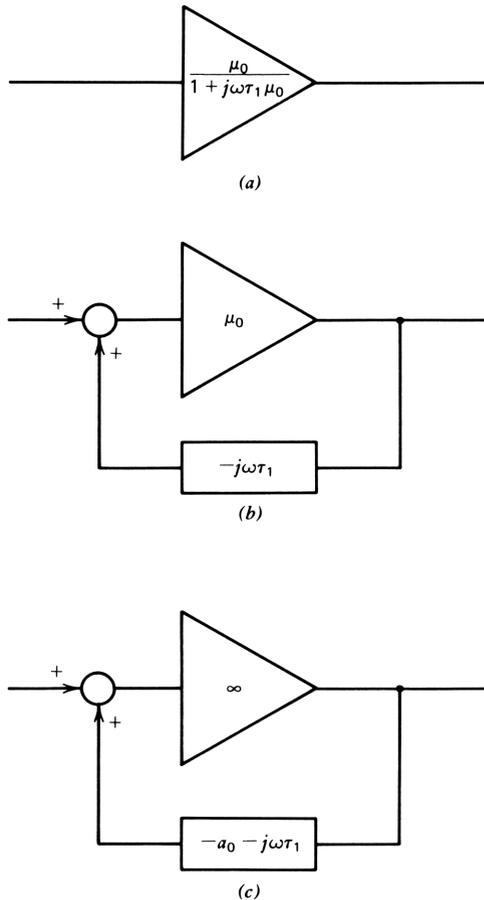


Figure 1.10. Alternate, equivalent representations of a frequency-dependent active path.

structure (shown in Fig. 1.10*b*). Accordingly, we can eliminate the denominator from the expression by using the reciprocal formulation, as we have done before for μ_0 .

Figure 1.10*c* goes one step further. The remaining active-path element μ_0 is *also* placed in the feedback path. It appears as a feedback element $-1/\mu_0 = -a_0$ around an infinite gain amplifier or an amplifier whose loss is zero. With that refinement, we see that the loss $1/\mu = L_a$ of the active path is given directly by eq. (1.3-9). Furthermore, $-\beta$ can be added to the feedback path in Fig. 1.10*c* to obtain the loss of eq. (1.3-10). It is clear that, mathematically, we can assign physical effects in the amplifier to either the feedback or the forward paths according to our choice. As we proceed, it will become apparent that this is also physically true. The concept of a zero loss amplifier is a valuable one widely used in working with operational amplifier circuits. We generalize this result for two- and three-port networks later.

To recapitulate, Fig. 1.10*a* represents the conventional view of the active path, with the path gain $\mu_0/(1+j\omega\tau_1\mu_0)$ multiplying the input signal to obtain the output. Figure 1.10*c*, on the other hand, represents the active path under the reciprocal formulation, with the input signal equal to the negative of the path gain of the “feedback” path. Since input e of the zero-loss amplifier must be zero for any finite output, input signal x is forced to cancel the signal from the output exactly. Clearly, the two representations are equivalent, but with the functional dependencies reversed.

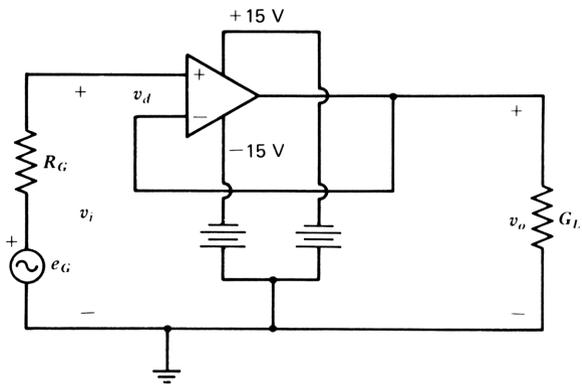
Example 2. The purpose of this example is to move the discussion from the abstract canonical feedback diagrams in Fig. 1.10 to a more physical representation of feedback structures. Consider the circuit in Fig. 1.11*a*, in which an operational amplifier is connected as a unity gain, noninverting amplifier. The negative differential input is connected directly to the output. Differential input voltage v_d is amplified by the factor μ : $v_o = \mu v_d$, where $v_d = v_i - v_o = v_i - \mu v_d$. Hence

$$v_d = \frac{1}{1+\mu} v_i \quad (1.3-12)$$

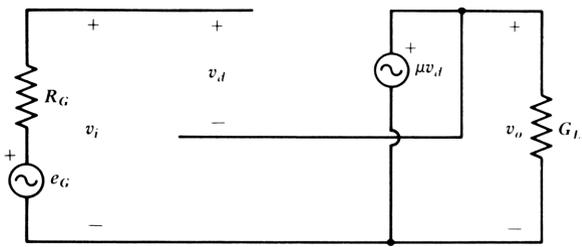
and

$$v_o = \frac{\mu}{1+\mu} v_i \quad (1.3-13)$$

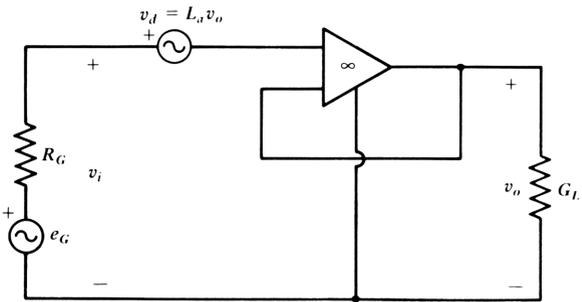
This is in the form of the canonical equation, with $\beta = -1$. Figure 1.11*b* gives an equivalent circuit in which a dependent generator μv_d represents the forward or active path. The input admittance is small and has been ignored in this simple model. The operational amplifier is represented solely by the voltage-controlled voltage source and is described causally since the controlled source both *depends* on v_d and is *caused* by it. The expressions “depends on v_d ”



(a)



(b)



(c)

Figure 1.11. (a) Unity gain amplifier using an operational amplifier; (b) conventional dependent-generator equivalent circuit; (c) equivalent circuit using a generator dependent on the output voltage, connected to the input of a zero-loss amplifier.

and “caused by v_d ” are often taken to mean the same thing, so that the distinction between the two concepts needs clarification.

The same circuit, this time described under the reciprocal formulation, is shown in Fig. 1.11c and makes the distinction clear. In this description the active path is described by an ideal amplifier that has a dependent generator connected in series with its *input* lead. The ideal amplifier is defined as having zero input current and voltage for any finite output current and voltage. No numbers are needed to describe it; its function is to bring both the voltage and current at its input to zero. This circuit element is discussed further in Section 1.7 and later chapters. A nonideal operational amplifier does have a nonzero input voltage, however, and this is represented in Fig. 1.11c by the dependent generator $v_d = L_a v_o$ connected in series with the ideal amplifier input. This representation of the operational amplifier is equivalent to that in Fig. 1.11b, but the *cause* of the output v_d is represented by a *dependent* generator $L_a v_o$. Hence the functional dependency is set up in an “anticausal” direction, so that whereas v_o is *caused* by v_d , v_d *depends* on v_o .

Over most of its frequency range, the voltage loss ratio of a simple voice frequency operational amplifier is given by

$$\frac{v_d}{v_o} = L_a(s) = a_0 + \tau_1 s \quad (1.3-14)$$

in which the Laplace transform complex frequency variable s replaces $j\omega$. This change is made here for notational convenience; it will take on additional meaning in Chapter 2. The value of a_0 for most operational amplifiers is negligible (typically 10^{-5} , or 100 dB of gain).

The loss of the voltage follower circuit is found by adding the voltages in the input loop in Fig. 1.11c. Thus, assuming negligible input current, we have

$$e_G = v_o + v_d$$

or

$$\frac{e_G}{v_o} = L(s) = 1 + \tau_1 s \quad (1.3-15)$$

which is a loss response with a single upward corner frequency at $\omega = 1/\tau_1$, and unity loss up to this frequency. This expression is not accurate for most operational amplifiers because usually there is significant delay of signals between input and output. The means for handling delays in feedback systems is treated in Chapter 5.

1.4 NONLINEARITIES IN DYNAMIC SYSTEMS

The combination of nonlinear components with frequency-dependent ones has always been troublesome. The reciprocal formulation allows us to deal simply

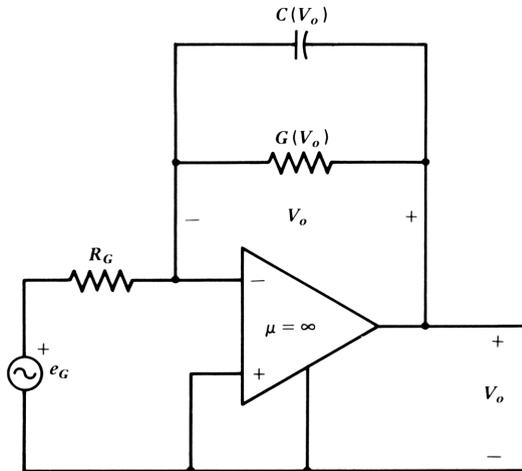


Figure 1.12. Circuit for study of a nonlinear dynamic system. Either G or C may be a nonlinear function of the voltage across it, in this case the output voltage.

with a type of nonlinearity problem that will be important to us later on. The problem concerns Fig. 1.10c, in which either a_0 or τ_1 (or both) may be nonlinear functions of the signal applied to them. In Fig. 1.10c the applied signal voltage is equal to the output signal since the signal voltage at the input is zero.

To make the problem more concrete, we translate this system in Fig. 1.10c to the circuit shown in Fig. 1.12. In this circuit we connect a feedback admittance consisting of a conductance and a capacitance around an infinite gain operational amplifier. This latter is just the amplifier in Fig. 1.11c, in which $v_d=0$. The problem of the previous paragraph can now be translated into circuit terms; that is, we must determine what input voltage is required to obtain a prescribed output voltage when either G or C (or both) are nonlinear functions of the output voltage.

Taking the output voltage as a function of time, the feedback elements will be instantaneous functions of this output voltage, so we must use a time domain description of the signal variables. We may rephrase the question as: "What is the generator voltage *as a function of time* to obtain a prescribed output voltage waveform?" This is a classic problem; one engineering application is that of generating a linear sweep voltage (a *ramp function*) for a cathode ray tube. In this application the output is to be a voltage that linearly increases with time. When this voltage is delivered by a nonlinear circuit, the input waveform must be predistorted to obtain the linear output. The problem is to find the input waveform for a given set of nonlinearities in the amplifier. Most television sets, for example, include such predistortion circuits to correct for nonlinear deflection systems.

We consider first the case of a nonlinear conductance and a linear capacitance. The input voltage to the circuit is expressed as a function of the output voltage v_o by

$$e_G(t) = -R_G \left[\int G(v_o) dv_o \right] - R_G C \frac{dv_o}{dt} \quad (1.4-1)$$

To take the analysis further, we need an expression for the nonlinearity of $G(v_o)$. If we assume a nonlinearity of the type given in eq. (1.2-8),

$$G(v_o) = G_1(1 + 2\alpha_2 v_o + 3\alpha_3 v_o^2) \quad (1.4-2)$$

we have

$$e_G(t) = -R_G \left(G_1 v_o + G_1 \alpha_2 v_o^2 + G_1 \alpha_3 v_o^3 + C \frac{dv_o}{dt} \right) \quad (1.4-3)$$

where we have written dv_o/dt for the time derivative of the output voltage. Thus, for any prescribed output waveform, we obtain the input waveform by adding the four terms of the equation.

As an example, suppose that the specified output is to be a ramp of -1 V/ μ s for a duration of 4μ s, starting at $v_o=0$ and $t=0$. We use a consistent set of units to avoid unnecessary conversions; our fundamental units here are volts, milliamperes, and microseconds, leading to kilohms, millisiemens, and nanofarads. Find the required input waveform if $G_1=0.1$ mmho, $R_G=1$ k Ω , $C=0.1$ nF, and $\alpha_2=\alpha_3=0.1$. Then $v_o=-t$, $v_o^2=t^2$, $v_o^3=-t^3$, and dv_o/dt is a step of -1.0 V. The input waveform is given by

$$e_G(t) = (0.1t - 0.01t^2 + 0.01t^3 + 0.1)u(t) \quad (1.4-4)$$

where $u(t)$ is a unit step. Note that units are consistent: (mmho) (k Ω)=1, and (nF) (k Ω)= μ s. The waveform is plotted in Fig. 1.13.

Next, consider the case of a linear conductance and a nonlinear capacitance in the circuit shown in Fig. 1.12. In this case the current through the capacitance is given by

$$i_C = \frac{dQ}{dt} = \frac{dQ}{dv_o} \cdot \frac{dv_o}{dt} \quad (1.4-5)$$

But since $Q=C(v_o) \cdot v_o$, we obtain

$$\frac{dQ}{dv_o} = \frac{d[C(v_o)v_o]}{dv_o} = C(v_o) + v_o \frac{dC(v_o)}{dv_o} \quad (1.4-6)$$

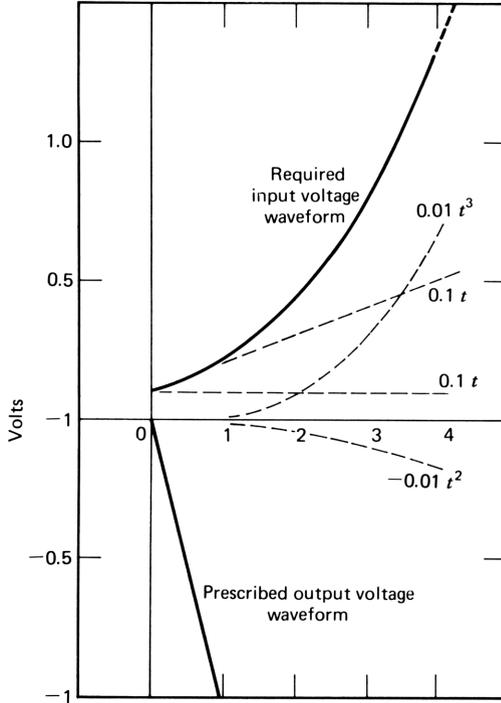


Figure 1.13. Derivation of the input waveform required to obtain a prescribed negative-going ramp at the output.

Thus

$$i_C = \left[C(v_o) + v_o \frac{dC(v_o)}{dv_o} \right] \frac{dv_o}{dt} \tag{1.4-7}$$

Since the current through the (linear) conductance is just Gv_o , we have for the generator voltage

$$e_g(t) = -R_G \left\{ Gv_o + \left[C(v_o) + v_o \frac{dC(v_o)}{dv_o} \right] \frac{dv_o}{dt} \right\} \tag{1.4-8}$$

We cannot go further without knowing the relationship between $C(v_o)$ and v_o . Let us assume as an example that over some range of v_o , $C(v_o)$ can be represented by the function

$$C(v_o) = C_0(1 + kv_o) \tag{1.4-9}$$

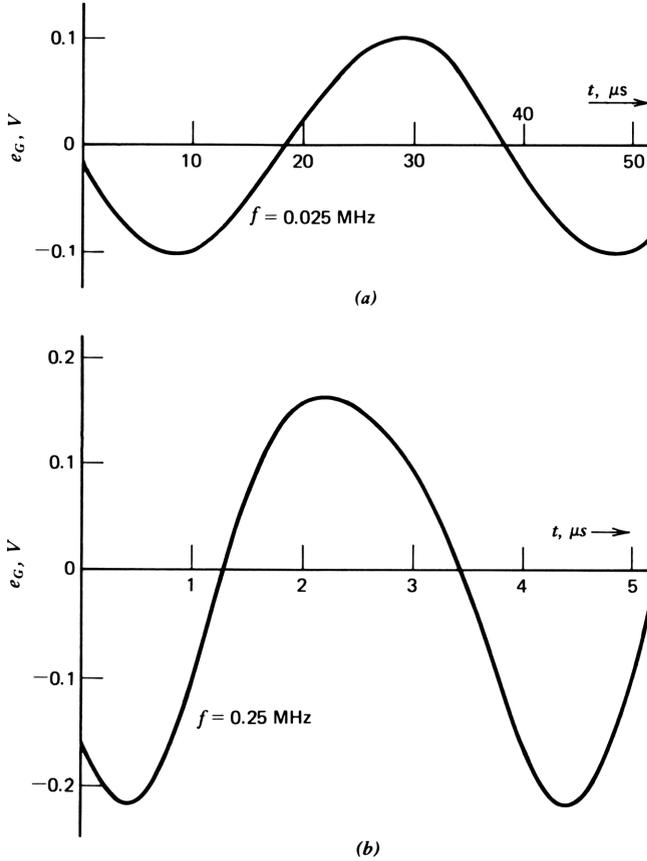


Figure 1.14. Input waveforms at (a) 0.025 MHz and (b) 0.25 MHz to obtain a pure sinusoidal output waveform from the circuit shown in Fig. 1.12.

Then $v_o(dC/dv_o)$ becomes C_0kv_o , so that the input voltage becomes

$$e_G(t) = -R_G \left[Gv_o + C_0(1 + 2kv_o) \frac{dv_o}{dt} \right] \tag{1.4-10}$$

As in the previous example with linear capacitance, if we know v_o at the outset, we can find dv_o/dt so that the input voltage time waveform can be found immediately.

As an example, let $R_G = 1 \text{ k}\Omega$, $G = 0.1 \text{ mmho}$, $C_0 = 0.1 \text{ nF}$. Let us find the required input voltage waveform if the output is to be an undistorted sine wave: $v_o = A \sin \omega t$, with ω in our consistent set of units—megaradians per second). From eq. (1.4-10), we can write the input waveform directly:

$$e_G(t) = -R_G GA \sin \omega t - R_G C_0 A \omega \cos \omega t (1 + 2kA \sin \omega t) \tag{1.4-11}$$

Since $\cos \omega t \sin \omega t = \frac{1}{2} \sin 2\omega t$,

$$e_G(t) = -0.1A \sin \omega t - 0.1A\omega \cos \omega t \\ - 0.1A^2\omega k \sin 2\omega t \quad (1.4-12)$$

The first two terms constitute the linear part of the input signal, and the third term represents the corrective predistortion required to obtain a pure sine-wave output. The latter term rises as the square of the amplitude, whereas the linear terms rise in proportion to the amplitude. Figure 1.14 shows the input waveform for $A=1$, $k=0.2$, at two frequencies, $\omega = \pi/20$ ($f=0.025$ MHz) and $\omega = \pi/2$ ($f=0.25$ MHz).

The preceding analyses were made for nonlinear elements in the feedback network of an ideal operational amplifier. These analyses apply equally well to nonlinearities in the forward path of the canonical diagram if these forward path nonlinearities can be expressed as functions of the output signal variable. The equivalent diagrams in Fig. 1.10 show the translation of circuit parameters back and forth between the feedback and the active paths of the canonical diagram. In the case of the nonlinear conductance, for example, the analysis would be the same if the feedback conductance were absent and if the forward path consisted of a *transresistance* of $R_m = 1/G = 1/G_1(1 + 2\alpha_2 v_o + 3\alpha_3 v_o^2)$. Similarly, the feedback capacitance could be represented in the frequency domain as a forward-path transimpedance, $Z_m = 1/j\omega C(v_o)$. In either case, if there is an output-dependent nonlinearity in the forward path, it is more convenient to transform it to a feedback network, as in Fig. 1.10c, to allow an analysis such as that carried out previously.

1.5 SENSITIVITIES FOR FREQUENCY-DEPENDENT LOSS RATIOS

As in the case of the sum of series resistors, the sensitivity of the sum of several elements Σ to one of those elements a , is simply a/Σ . Since the loss ratio in eq. (1.3-10) is such a sum, we may obtain the sensitivity of L to each of its component parts directly.

Thus for

$$L = a_0 + j\omega\tau_1 - \beta \quad (1.5-1)$$

we can use the sum rule to write

$$S_{a_0}^L = \frac{a_0}{a_0 - \beta + j\omega\tau_1} \quad (1.5-2)$$

The sensitivity of L to a_0 is a complex function of frequency. At zero frequency the result is just that obtained in Section 1.2. But how should we interpret the frequency dependence of the sensitivity and the fact that it is a complex quantity?

For a simple interpretation of sensitivity as a complex function of frequency, consider the series combination of a resistor R and an inductor L . What is the sensitivity of the total series impedance to the resistance? To the inductance? From the sum rule,

$$S_R^Z = \frac{R}{R + j\omega L} \quad (1.5-3)$$

and

$$S_L^Z = \frac{j\omega L}{R + j\omega L} \quad (1.5-4)$$

At dc the sensitivity of the impedance to the resistance is unity, whereas at high frequencies the sensitivity is zero. The opposite is the case for the inductor.

Since the sensitivity is a complex quantity, we need two numbers to specify it—the real and imaginary parts or the magnitude and the phase. Both interpretations are of interest and are developed in the following paragraphs.

Real and Imaginary Parts

Since the form of eqs. (1.5-2) and (1.5-3) is the same, we can develop the interpretation of the real and imaginary parts of the sensitivity for either the impedance or the loss ratio.

Bode plots separately express the magnitude and phase of loss polynomial L . Let us now find the sensitivities of the magnitude and phase of L to a_0 . Expressing L in terms of its magnitude and phase, we obtain

$$L = |L|e^{j\theta} \quad (1.5-5)$$

and taking $\ln L$,

$$\ln L = \ln|L| + j\theta \quad (1.5-6)$$

The definition of sensitivity is given in eq. (1.1-6) as

$$S_{a_0}^L = \frac{d \ln L}{d \ln a_0} \quad (1.5-7)$$

Substituting (1.5-6) into (1.5-7), we obtain

$$S_{a_0}^L = \frac{d \ln|L|}{d \ln a_0} + j \frac{d\theta}{d \ln a_0} \quad (1.5-8)^*$$

*I believe that this substitution and the ensuing result were first formulated by E. J. Angelo, Jr. in 1955 (MRI research report R-449-55, PIB-379, "Design of Feedback Systems").

Thus the real part of the sensitivity of L to a_0 is the sensitivity of the *magnitude* of L to a_0 :

$$\operatorname{Re}\{S_{a_0}^L\} = S_{a_0}^{|L|} \quad (1.5-9)$$

The real part of the sensitivity gives us the per unit change in $|L|$ for a small per unit change in a_0 . Equivalently, it expresses the ratio of the decibel change in $|L|$ to a small decibel change in a_0 . The imaginary part is $d\theta/(da_0/a_0)$, the change in θ for a small per unit change in a_0 . If a_0 were to increase by 1%, for example, $da_0/a_0=0.01$, the change in θ would be

$$d\theta \Big|_{\frac{da_0}{a_0}=0.01} = 0.01 \operatorname{Im}\{S_{a_0}^L\} \quad (1.5-10)$$

The sensitivity of the phase to a_0 , $S_{a_0}^\theta$, is the per unit change in the phase for a small per unit change in a_0 . It is generally not a useful quantity by itself, since we are seldom interested in the per unit change of an angle. Thus we usually use the $\operatorname{Im}\{S_{a_0}^L\}$. This is related to $S_{a_0}^\theta$ as follows:

$$\begin{aligned} \operatorname{Im}\{S_{a_0}^L\} &= \frac{d\theta}{da_0/a_0} \\ &= \theta \frac{d\theta/\theta}{da_0/a_0} \\ &= \theta S_{a_0}^\theta \end{aligned} \quad (1.5-11)$$

We call this quantity the *angular phase sensitivity* to distinguish it from $S_{a_0}^\theta$. It is the measure of phase sensitivity that is usually of interest. The complete expression for the sensitivity in rectangular form is

$$S_{a_0}^L = S_{a_0}^{|L|} + j\theta S_{a_0}^\theta$$

Returning to the example of eq. (1.5-1), we can write

$$S_{a_0}^{|L|} = \operatorname{Re}\left\{ \frac{a_0}{a_0 - \beta + j\tau_1\omega} \right\} \quad (1.5-12)$$

$$\theta S_{a_0}^\theta = \operatorname{Im}\left\{ \frac{a_0}{a_0 - \beta + j\tau_1\omega} \right\} \quad (1.5-13)$$

The magnitude and angular phase sensitivities to β are given by similar expressions in which the numerators are replaced by β . The magnitude and phase sensitivities to τ_1 are given by similar expressions in which the numerator

is replaced by $j\tau_1\omega$. This latter fact relies on the relationship

$$S_{j\tau_1\omega}^L = S_{\tau_1}^L \quad (1.5-14)$$

since at any frequency, $j\omega$ is a constant, so that $d\ln j\omega = 0$. Thus

$$\frac{d\ln L}{d\ln j\omega\tau_1} = \frac{d\ln L}{d\ln \tau_1 + d\ln j\omega} = \frac{d\ln L}{d\ln \tau_1} = S_{\tau_1}^L \quad (1.5-15)$$

The magnitude and angular phase sensitivities for the example, with $a_0 = 0.01$, $\beta = -0.1$, and $\tau_1 = 1$, are shown in Fig. 1.15. The magnitude sensitivities give the percent change in $|L|$ for a 1% change in the parameter. At low frequencies, for example, the sensitivity to a_0 is $0.01/(0.01+0.1) = 0.091$, so that a 10% change in a_0 would cause a 0.91% change in $|L|$, or a 1 dB change in a_0 would cause a 0.091 dB change in $|L|$. Therefore, the sensitivity to a_0 is small and drops with frequency. The sum of the magnitude sensitivities is unity (of course). At low frequencies $|L|$ is mostly sensitive to β and at high frequencies, to τ_1 .

The phase sensitivities are expressed as the imaginary part of $S_x^L = \theta S_x^\theta$, as noted previously, with θ in degrees. In this example, a 10% change in the parameter would give a change in θ equal to 10% of the scale value. (Similarly, a +1 dB change in the parameter value would give a change in θ equal to +12.2% of the scale value, since a +1 dB change is equivalent to a +12.2% change.) The maximum ordinate value for $\theta S_{\tau_1}^\theta$ is 29° . Thus a 10% change in τ_1 will cause a maximum change in phase of 2.9° , at $\omega = 0.11$.

Several sensitivity relationships help to ease the calculation of sensitivities. We encountered one of these relationships in Section 1.2, namely

$$S_x^{1/y} = \frac{d\ln 1/y}{d\ln x} = \frac{-d\ln y}{d\ln x} = -S_x^y \quad (1.5-16)$$

Another relationship we use is

$$S_y^x = \frac{d\ln x}{d\ln y} = \left(\frac{d\ln y}{d\ln x} \right)^{-1} = 1/S_x^y \quad (1.5-17)$$

If k is a constant, we can write

$$S_x^{k \cdot y} = \frac{d\ln k + d\ln y}{d\ln x} = \frac{d\ln y}{d\ln x} = S_x^y \quad (1.5-18)$$

If u and v are functions of x , we can use the alternate definition of sensitivity

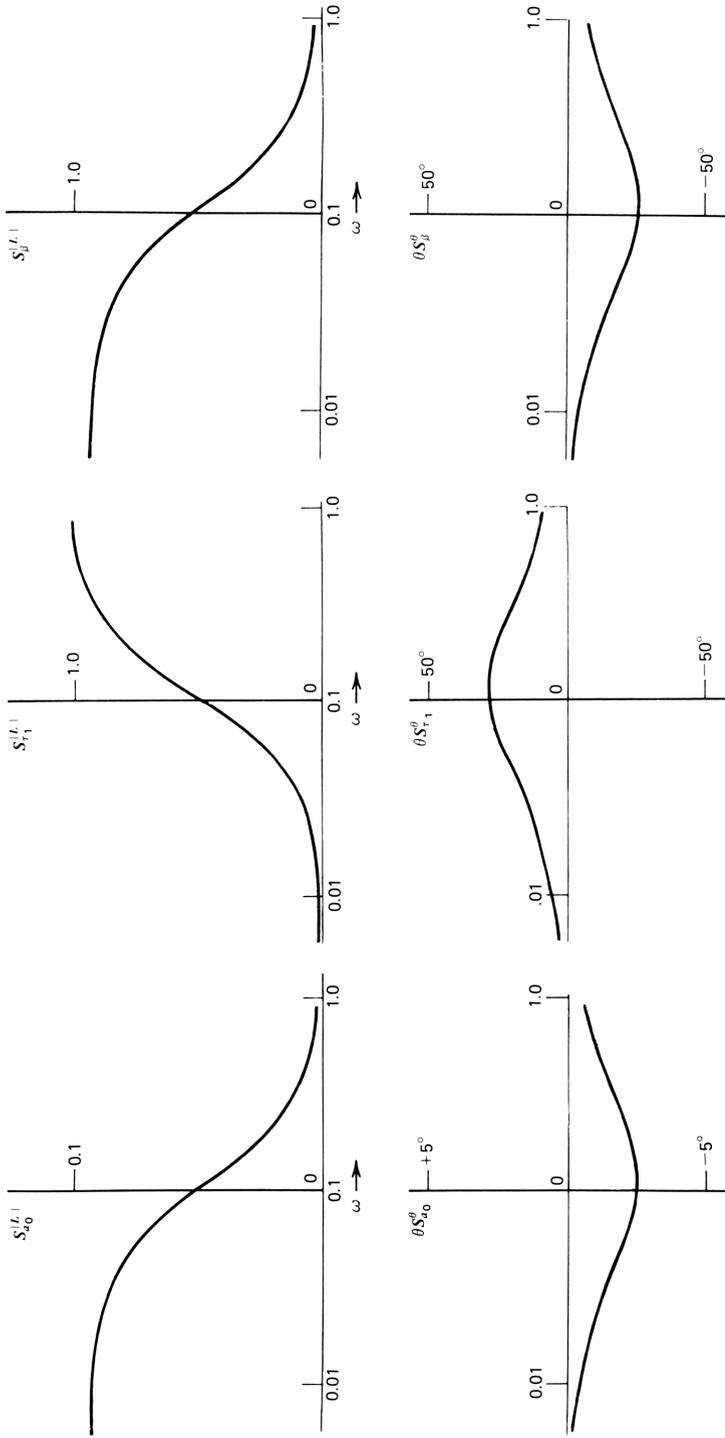


Figure 1.15. Real and imaginary parts of the sensitivities of loss to a_0 , τ_1 , β for a simple binomial loss ratio.

to obtain

$$\begin{aligned}
 S_x^{u+v} &= \frac{d(u+v)/(u+v)}{dx/x} = \frac{1}{u+v} \left(\frac{du}{dx/x} + \frac{dv}{dx/x} \right) \\
 &= \frac{1}{u+v} \left(u \frac{du/u}{dx/x} + v \frac{dv/v}{dx/x} \right) \\
 &= \frac{1}{u+v} (uS_x^u + vS_x^v) \tag{1.5-19}
 \end{aligned}$$

Table 1.1 is a collection of sensitivity relationships that have a twofold purpose: (1) the relationships will be found convenient in working with sensitivities; and (2) perhaps more important, they serve to familiarize us with the concept of sensitivity. In Problem 12 the reader is asked to derive each of these expressions from the definition of sensitivity. Most can be derived in one or two lines.

Table 1.1 Sensitivity Relationships^{9, a}

1	$S_x^{c^y} = S_x^y$	
2	$S_x^{y^n} = nS_x^y$	
3	$S_x^y = 1/S_y^x$	
4	$S_x^y = S_{u_1}^y \cdot S_x^{u_1} + S_{u_2}^y \cdot S_x^{u_2} + \dots$	where $y = y(u_1, u_2, \dots, u_n)$
5	$S_x^y = S_x^{ y } + j\phi_y \cdot S_x^{\phi_y}$	
6	$S_x^{ y } = \text{Re } S_x^y$	
7	$S_x^{\phi_y} = \frac{1}{\phi_y} \text{Im } S_x^y$	
8	$S_x^{u \cdot v} \dots = S_x^u + S_x^v + \dots$	
9	$S_x^{u+v+\dots} = \frac{1}{u+v+\dots} (uS_x^u + vS_x^v + \dots)$	
10	$S_x^y = -S_x^{1/y} = -S_{1/x}^y = S_x^{1/y}$	
11	$S_x^{u/v} = S_x^u - S_x^v$	
12	$S_x^{e^y} = yS_x^y$	
13	$S_x^{\ln y} = \frac{1}{\ln y} \cdot S_x^y$	

^aCoordinates y, u, v are single-valued differential functions of x .

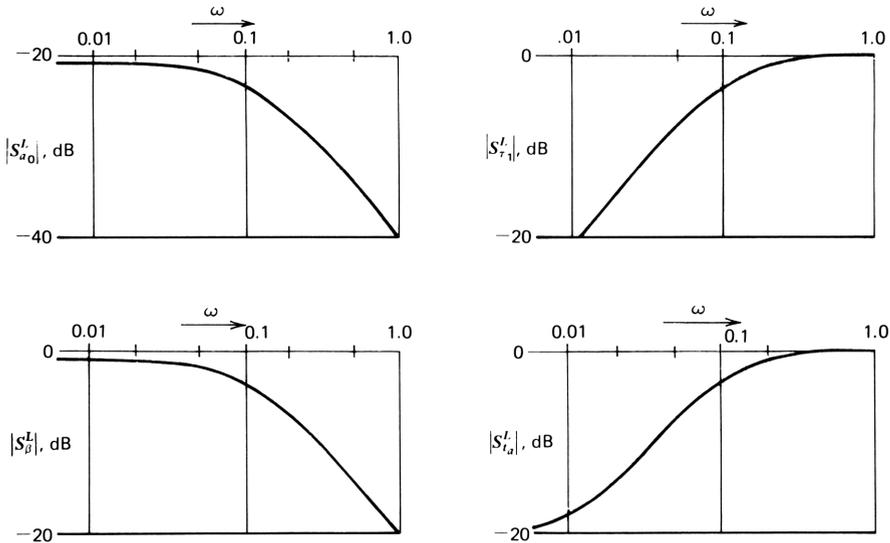


Figure 1.16. Magnitude of the sensitivities of loss to a_0 , τ_1 , and β for the simple binomial loss ratio. The fourth diagram gives the sensitivity of loss to the active-path loss t_a .

Sensitivity Magnitude

The complex quantity sensitivity can be expressed by its magnitude and phase as well as by its real and imaginary parts. This is of particular importance when we are interested in distortion reduction or other benefits of feedback that involve the swamping out of undesirable effects (signals) by the feedback signal. In such cases the phase is not of great importance; we are interested in the ratio of the magnitude of the feedback signal relative to the active-path signal. In Fig. 1.16 the magnitudes (in decibels) of the sensitivities of loss to a_0 , τ_1 , and β corresponding to the example are plotted as a function of frequency. (Note the scale difference between the a_0 sensitivity and the remaining curves.)

These curves are useful in estimating the distortion reduction afforded by feedback. Suppose that a_0 is nonlinear, as discussed in Section 1.2. The magnitude of the sensitivity of loss to a_0 tells us specifically how much reduction we may expect in distortion products from this particular nonlinearity, and similarly from the curves for τ_1 and β . The fourth curve, which is for the sensitivity of loss to $t_a = a_0 + \tau_1 s$, is also included to relate sensitivity to return difference as defined under the conventional formulation.

Relationship Between Sensitivity and Return Difference

Let us now form the sensitivity of the loss to the active-path loss:

$$S_{L_a}^L = \frac{L_a}{L_a - \beta}$$

This is the ratio of the input contribution of the active path to the input to the total input. We now take the reciprocal of this sensitivity:

$$\frac{1}{S_{L_a}^L} = 1 - \mu\beta$$

Hence the reciprocal of the sensitivity is the return difference as found under the conventional formulation. Therefore, the magnitude of the sensitivity is the reciprocal of the magnitude of the return difference for the canonical diagram in Fig. 1.1.

Under the reciprocal formulation there is no return difference or loop gain, but the sensitivity is the same for the two formulations. For this reason, we assert the primacy of the concept of sensitivity rather than that of return difference to describe feedback structures. The two concepts are quite different at a fundamental level. Whereas sensitivity is a physical property of a structure, return difference is not. The latter depends largely on the way we (1) define the signal variables and (2) assign the different parts of the structure to the active path or to the feedback path. We have seen (in Fig. 1.10) that we can assign the total structure to the active path (in Fig. 1.10*a*) or to the feedback path (in Fig. 1.10*d*). As we analyze feedback structures in greater detail in later chapters, we see that this ambiguity of assignment carries through to the most basic levels of analysis so that return difference can be defined only with respect to arbitrary choices. In relation to the physical phenomenon it is to represent, sensitivity may be said to be a “well-formed” parameter, whereas return difference is not.

Sensitivity is a useful concept because it keeps things in proportion. It helps quantify our ideas about what is and what is not important in a system—puts the effect of variations of system parameters on a common basis. The basis here has been the per unit change in the magnitude and phase of the loss ratio, but the concept can be applied, often usefully, to any quantity dependent on several parameters. It can be applied in situations where the component parameters and their variations are only partially known or completely unknown but where the sensitivity can be determined experimentally.

The concept of market elasticity is an example.¹⁰ Here, we are interested in knowing how much the market for a product will decrease when the price is raised. We can sometimes find out by changing the price in a test market area and determining the effect on sales. If we raise the price 10% and if the sales fall off by 5%, we say that the market elasticity is -0.5 . This is just the definition of sensitivity we have been discussing here, but with the difference that the sensitivity is found by experiment without knowing the parameters or their quantitative relations that went into establishing it. Armed with this knowledge, we are in a better position to find the reason for the elasticity and can make a better judgment as to whether to market the product or how to improve it.

1.6 THE REFERENCE CONDITION

A key difference between the conventional approach and that taken under the reciprocal formulation is highlighted by the notion, introduced by Bode, of the *reference condition*. Bode used this idea to clarify a situation that does not arise for Black's simple canonical diagram: when a portion of the input signal leaks through the feedback network to the output, an output signal will exist even when $\mu=0$ (or, as Bode said, "when the tube is dead"). We treat this situation in a later chapter. The point is that a reference condition was defined by setting $\mu=0$.

Under the reciprocal formulation we define a reference condition by setting $L_a=1/\mu=0$. In *this* reference condition the active path becomes a zero-loss (infinite gain) amplifier. Thus in the reference condition, $L=-\beta$. The reference condition for the canonical feedback diagram thereby serves to define β .

This concept will serve to clarify relationships in feedback amplifiers, particularly in what are conventionally called *multiple-loop feedback amplifiers*. It is in this area that Bode's theory exhibits its greatest difficulty. We see later that under the reciprocal formulation the "multiple loops" disappear, so that these structures present no special difficulties and, indeed, do provide an improvement over the single-loop design that Bode envisioned.¹¹

Use of the infinite gain reference condition is at least implicit in a second equation that Black wrote down on his copy of *The New York Times* in 1927:

$$\text{Gain} = \frac{1}{\beta} \left(\frac{\mu\beta}{1-\mu\beta} \right) \quad (1.6-1)$$

in which β emerges as the reciprocal of the gain when the loop gain $\mu\beta$ goes to infinity.

The zero-loss reference condition, or *null reference*, is one of the most useful concepts the reader will encounter in this book. It has been introduced in its single-dimensional form in this chapter, but it is extended to two-port networks in Chapter 6 and in a three-dimensional version in the discussion of differential and operational amplifiers in Chapter 12. This is useful because it allows us to build up transmission functions and characteristics starting out with a clean slate; each physical effect that causes a departure from zero loss (infinite gain) is a portion of the *sum* of all such physical effects. All are added on a common basis so that we can compare their relative effects.

1.7 AN ANALOGY FROM PROJECT MANAGEMENT

A main feature of the alternative foundation for feedback theory is the focus on the output as the *goal*.¹² Making the output the independent variable (or, later, the set of variables) is a bold step, but with ample precedent. Taking the output as the known (or independent) variable presumes that you know where you are headed (even if you do not). It can be likened to the project

management technique known as *PERT* (project evaluation and review technique),¹³ used to make sure that a project keeps on course toward a specific objective.

To manage a project by way of *PERT*, the goal (the output) first must be specified in as much detail as possible and a specific end date set. Then each element that goes into making the final goal is analyzed, particularly with respect to what materials must be present, what labor input is needed, and most crucially, how long that element will take to do. Project evaluation and review technique recognizes that many subprojects will have to come together at points well before completion of the total project, and these are treated in the same way as the whole project. These subprojects *and their goals* provide a main feature of *PERT*: they provide specific milestones that can be identified ahead of time so that when they are reached, the project status can be measured against the original plan. The central idea of *PERT* is to start with the project goal and work backward in time to the specific milestones along the way to identify the subprojects that *must* be done on schedule to ensure timely project completion.

The *PERT* technique consists of measuring the *difference between where the project is and should be* at any given milestones. If it is where it should be, no additional steps apart from those detailed in the original plan need be taken. If a subproject is late, however, two alternatives must be considered: first, if it is not on the original critical path, the situation is reevaluated to find whether it has “gone critical,” in which case the second alternative comes into play. This second alternative, when the milestone is missed, calls for increased effort on the critical subproject or a rescheduling of the whole project.

Although the analogy between this management system and the feedback system under the aspect of the reciprocal formulation is rough, it is useful in two ways: (1) it takes the perhaps unfamiliar idea of analyzing a feedback system by working backward in time and shows how this concept has been used in another field; and (2) *PERT* works if it is applied properly. The same may be said for feedback analysis and design under the reciprocal formulation.

1.9 SUMMARY

A brief review of the problem that led to Black’s invention and canonical feedback equation was given at the beginning of this chapter. This equation forms the foundation of classical feedback theory. An alternate foundation, introduced in Section 1.2, substituted loss (the reciprocal of gain) for gain in the canonical feedback equation. Under this reciprocal formulation of the feedback problem, return difference is eliminated, and analysis of the effects of feedback on gain stability (sensitivity), nonlinearity, and noise is simplified.

In Section 1.3 we applied the reciprocal formulation to an active path with a simple high-frequency cutoff. We found that the reciprocal formulation similarly simplifies the description of the active path itself. We next studied a

nonlinear dynamic system in which the nonlinearity can be expressed as a function of the output signal variable.

In Section 1.5 we developed the sensitivity concept for systems that include frequency dependencies, including high-frequency cutoffs and transport delay. In Section 1.6 we defined a null reference condition for feedback systems: a system is said to be in the reference condition when the loss of the active path is zero (i.e., when the gain is infinite). Finally, in Section 1.7 we drew an analogy between the analysis of systems under the reciprocal formulation with the PERT technique of project management.

PROBLEMS

- 1 To see why the definition of sensitivity given in Section 1.1 is restricted to small per unit variations of μ , assume that μ changes by a sizable amount $\Delta\mu$, and find an expression for

$$S_{\mu}^L = \frac{\Delta K/K}{\Delta\mu/\mu}$$

in which S_{μ}^L is the sensitivity for large variation of the parameter μ . The expression must, of course, be equivalent to eq. (1.1-7) for small changes of μ .

- 2 Let the transfer characteristic of the μ path of the feedback system of Fig. 1.1 be given by the nonlinear relationship

$$y = \mu_1 e + \mu_3 e^3$$

Thus

$$\frac{dy}{de} = \mu_1 + 3\mu_3 e^2$$

If $\mu_1 = 100$, $\mu_3 = 10$, and $\beta = 1.1$, find $K = dy/dx$ and show that the distortion term $e\mu_3 e^2$ is reduced by a factor about equal to the sensitivity.

- 3 In the text we took as a measure of nonlinearity of L the values of the coefficients $2\alpha_2/(\alpha_1 - \beta)$ and $3\alpha_3/(\alpha_1 - \beta)$. Another measure of nonlinearity is the sensitivity of L to the output signal variable y . Applying the definition of sensitivity to eq. (1.2-10), derive an expression for S_y^L . Next, derive $S_y^{1/\mu}$ using eq. (1.2-8). What is the ratio of these sensitivities, and what is the significance of the ratio?
- 4 Consider two resistors in parallel, R_1 and R_2 . Draw a canonical feedback diagram whose output is the voltage across the pair and whose input is the current through the pair. Let $\mu = R_2$. What is β ? Find the sensitivity of the parallel resistance to R_1 . Repeat the procedure, this time letting the output be the current and the input the voltage.

- 5 The active path of a feedback system as illustrated by either Fig. 1.1 or 1.3 is described by the following equation:

$$y = k_1 \ln(e + 1)$$

where the error e is given by

$$e = x + k_2 y$$

Sketch the input-output characteristic for values of $k_1 = 10$ and $k_2 = 0$ and -0.5 . Let y vary from -10 to 10 .

- 6 The output of a bipolar transistor differential amplifier is related to its input voltage (the differential input voltage between their minus and plus input terminals) by the equation

$$y = A \tanh \frac{v_d}{0.052}$$

where v_d is the input differential voltage, y is the output voltage, and A is the small-signal voltage gain. When feedback is connected around this amplifier as shown in Fig. 1.17, we obtain e_G as the sum of the feedback-path and active-path signals:

$$e_G = R_G G_G y + v_d$$

Sketch the static input-output characteristic with and without feedback if $A = 10$ and $R_G G_G = 0.1$. For each case, what is the small-signal ratio for $y = 0$? For $y = 9$ V?

- 7 The sensitivity of loss ratio to large changes in a parameter is called the *large-signal sensitivity* and is

$$S_x^L = \frac{\Delta L/L}{\Delta x/x} \neq \frac{d \ln L}{d \ln x}$$

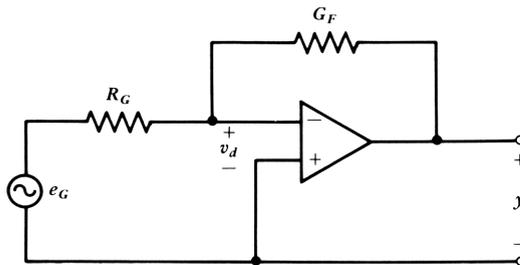


Figure 1.17.

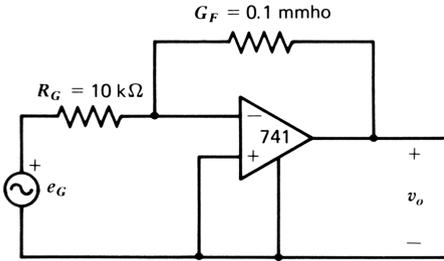


Figure 1.18.

Show that the large-signal sensitivity is equal to

$$S_x^L = 1 - \frac{L_r}{L}$$

where L is the loss ratio with x taking its nominal value and L_r is the loss ratio with x taking a reference value (i.e., the worst-case value of x).

- 8 The circuit in Fig. 1.18a is a unity gain inverting amplifier. For this circuit, use the reciprocal formulation to derive an expression for L in terms of t_α , G_F , and R_G . Put the equation in the form of $L = t'_\alpha - \beta$, where t'_α is the time constant of the operational amplifier modified by the circuit immittances G_F and R_G . With $t_\alpha = -0.15j\omega$, sketch Bode plots of t'_α , β , and L . Define β as the loss of the circuit with $t_\alpha = 0$.
- 9 If the feedback conductance in Problem 8 is replaced by a 100 pF capacitor (0.1 nF), the circuit becomes an integrator (because the output time response is the integral of the input signal with respect to time). Sketch Bode plots of t_α , β , and L for this circuit.
- 10 A silicon diode is connected from the output to the input of an ideal operational amplifier as shown in Fig. 1.19. The conductance of the diode is given by $g_d = qI_d/kT = I_d/0.026$. Find an expression for the input current to the circuit as a function of the output voltage. If the diode voltage is given by $v_d = kT/q \ln I_d/I_s$, and if $I_s = 10^{-12}$ mA, find the constant of integration and plot a graph of input current against output voltage. Ignore diode capacitance.

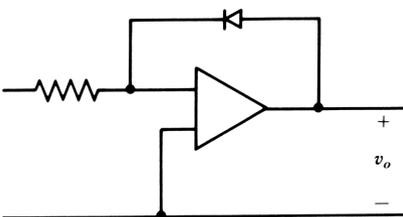


Figure 1.19.

- 11 The capacitance of the (reverse-biased) collector junction of a transistor varies with collector bias voltage according to the relation

$$C_{jc} = \frac{C_{jc0}}{(1 + V_{CB}/\phi_c)^{m_c}}$$

Typical values for ϕ_c and m_c are 0.85 and 0.33 V, respectively. We may model the effect of this capacitance on the input current to the transistor by connecting such a capacitance from the output to the negative input of an ideal operational amplifier, as shown in Fig. 1.19. Obtain an expression for the input current of the model as a function of output voltage.

- 12 Derive each sensitivity expression in Table 1.1.
 13 Use the sum rule to obtain a simpler derivation of eq. (1.5-19). [*Hint*: Use eq. (1.5-7).]

REFERENCES

- 1 H. S. Black, "Translating System," U.S. Patent No. 1,686,792, issued December 1937.
- 2 H. S. Black, "Stabilized Feedback Amplifiers," *BSTJ* **13**, 1 (January 1934).
- 3 J. C. Maxwell, "On Governors," *Proc. R. Soc. London* **16**, 270–283 (1968). Reprinted in Reference 4.
- 4 O. Mayr, *The Origins of Feedback Control*, MIT Press, Cambridge, MA, 1970.
- 5 H. Nyquist, "Regeneration Theory," *BSTJ* **11**, 126–147 (1932).
- 6 H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945.
- 7 D. Berlinski, *On Systems Analysis: An Essay Concerning the Limitations of Some Mathematical Methods in the Social, Political, and Biological Sciences*, MIT Press, Cambridge, MA, 1976.
- 8 F. D. Waldhauer, "Anticausal Analysis of Feedback Amplifiers," *BSTJ* **56** (10), 1337–1386 (1977).
- 9 G. S. Moschytz, *Linear Integrated Networks: Fundamentals*, Van Nostrand Reinhold, New York, 1974, p. 224.
- 10 J. M. Henderson and R. E. Quandt, *Microeconomic Theory*, 2nd ed., McGraw-Hill, New York, 1971, Chapter 2.
- 11 H. W. Bode, "Feedback—The History of an Idea," presented at the Symposium on Active Networks and Feedback Systems, Polytechnic Institute of Brooklyn, April 1960. See p. 12 of the Proceedings. This paper is also reprinted in *Selected papers on Mathematical Trends in Control Theory*, Dover, New York, 1964. Many other good papers appear here.
- 12 R. C. Dorf, *Modern Control Systems*, 2nd ed., Addison-Wesley, Reading, MA 1974, p. 5.
- 13 G. Nadler, *Work Design: A Systems Concept*, Irwin, Homewood, IL, 1970.

Chapter 2

Polynomials of Loss: Various Descriptions of Polynomials

In Section 1.3 we saw that an amplifying device that has a simple high-frequency cutoff can be represented under the reciprocal formulation by a *binomial* in the frequency variable. The loss ratio of the device, in other words, is represented by the sum of a dc or low-frequency constant and a term that is linear in frequency. In this chapter we extend the study of frequency dependencies of the loss ratio to polynomials of higher degree. In the interests of clarity, we restrict the discussion largely to polynomials in the frequency variable, which represent *low-pass* loss ratios.

The chapter gives several computational tools that will be used in the rest of the book and that are available mostly in packaged computer programs.

Section 2.1 analyzes a simple feedback configuration under the reciprocal formulation.

Section 2.2 describes the scaling in frequency and amplitude of loss polynomials, and includes analysis of linear, quadratic, and cubic polynomials.

Section 2.3 describes Newton's method for obtaining the roots of polynomials and may be familiar to some readers. The section also gives a brief review

of Laplace transform theory to relate time domain specifications to the general problem of establishing performance specifications for feedback systems, with emphasis on settling time.

Section 2.4 evaluates polynomials of loss for $s=j\omega$ as they relate to Bode and Nyquist diagrams. Although this may be familiar, the reverse process described in Section 2.5 (viz., finding the polynomial coefficients from the Bode diagram) is less familiar but is useful in feedback system design.

Section 2.6 derives coefficients of various polynomials that may be used as system response specifications. Included are Butterworth, Bessel, and Chebyshev polynomials and a cookbook for finding them. The frequency scaling deserves attention, as it differs from that of some other texts.

Section 2.7 treats a central concern of the book—namely, sensitivities of loss to the polynomial coefficients. This section is essential to later work.

2.1 A FEEDBACK AMPLIFIER

Electronic amplifiers provide us with particularly simple and practical examples of feedback structures because signals are all in electrical form. Figure 2.1a shows a three-stage transistor feedback amplifier. For the purposes of this initial look at feedback amplifiers, we can replace the transistors by the approximate representations in Fig. 2.1b, in which the amplification function is performed by an ideal amplifier and the amplification is degraded from the ideal by a resistance in the common lead of the amplifier. Finite bandwidth comes about by connecting a capacitance in shunt with the input. The model with the ideal amplifier is an anticausal model in that the amplification function is idealized and the departure from the ideal comes about from feedback element r . It is mathematically equivalent to the causal model also shown in Fig. 2.1b, which incorporates a dependent generator.

For the anticausal model, the input voltage and current are easily obtained as a function of the output current:

$$v_{be} = r i_c \quad (2.1-1)$$

$$i_b = r C s i_c = \tau s i_c \quad (2.1-2)$$

in which r and C are intended for the i th stage of the amplifier. The product $rC = \tau$ is the *unity loss time constant*. The reciprocal of the unity loss time constant is the (angular) frequency at which the current loss (or gain) is unity.

For the bipolar transistor, C is the *diffusion capacitance* and is roughly proportional to collector current; r is the emitter resistance and is inversely proportional to collector current. To a first approximation, τ is constant with collector bias current. In this sense τ is a more fundamental constant of the device than is C . We can represent this on the circuit diagram as in Fig. 2.1c, in which an output-current-dependent current generator connected across the

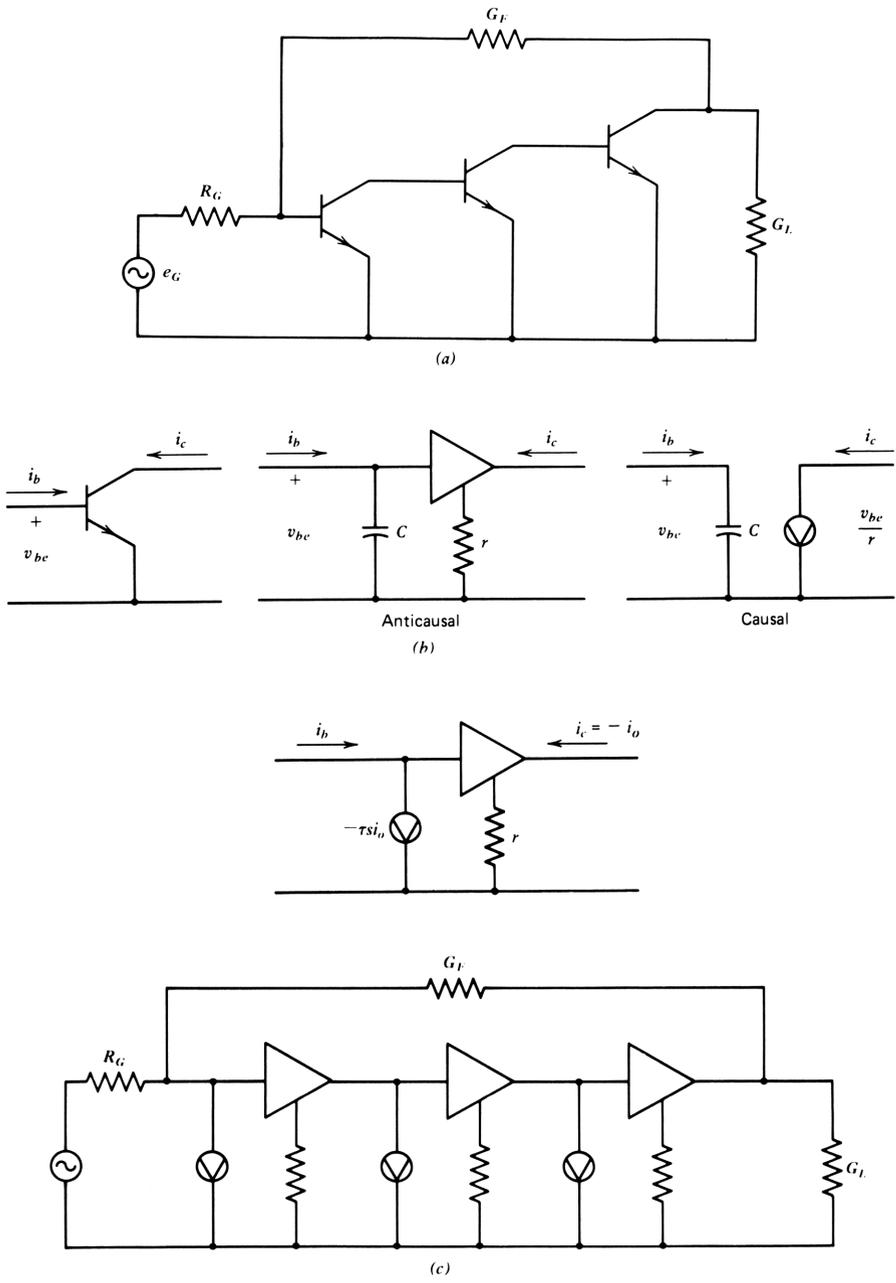


Figure 2.1. Three-stage feedback amplifier and device equivalent circuit for analysis.

input terminals replaces the capacitance. The two representations are equivalent, but the current generator more closely represents the physical situation. These matters are discussed in Chapter 7.

The transistor is incorporated into the three stage amplifier in Fig. 2.1*d*. This circuit can be analyzed by inspection if we enter the problem in an appropriate way, as follows. Assume an output signal voltage of unity. The signal current to be supplied by the third stage is then $G'_L = G_L + G_F$ flowing out of the collector, since the unit output voltage creates a signal current in both G_L and G_F . We ignore the current induced in G_F by the voltage at the input of the amplifier to which the other end of G_F is connected (it will be small if the amplification is sizable). With current G'_L flowing *out* of its collector, the input voltage of the third stage will be $-r_3 G'_L$, and the input current to that stage is $-\tau_3 s G'_L$. This latter is, in turn, the current flowing out of the second stage; thus we can write the input voltage and current for the second stage. We then repeat the process for the first stage; this gives us the input voltage and current to the *active path* of the amplifier:

$$v_{\text{in(active)}} = -r_1 \tau_2 \tau_3 G'_L s^2 \quad (2.1-3)$$

$$i_{\text{in(active)}} = -\tau_1 \tau_2 \tau_3 G'_L s^3 \quad (2.1-4)$$

To the input *current*, we now add the feedback current from the output. Again ignoring the current induced in G_F by v_{in} , the feedback current is $-G_F v_o$, so that with feedback we obtain

$$i_{\text{in}} = -(G_F + \tau_1 \tau_2 \tau_3 G'_L s^3) \quad (2.1-5)$$

The source or generator voltage is obtained from

$$e_G = v_{\text{in}} + i_{\text{in}} R_G \quad (2.1-6)$$

where $v_{\text{in}} = v_{\text{in(active)}}$, so that

$$e_G = -(R_G G_F + r_1 \tau_2 \tau_3 G'_L s^2 + R_G \tau_1 \tau_2 \tau_3 G'_L s^3) \quad (2.1-7)$$

$$= -(a_0 + a_1 s + a_2 s^2 + a_3 s^3) \quad (2.1-8)$$

If we replace the unit output voltage by v_o , we obtain the *loss ratio* e_G/v_o as a *polynomial* in the frequency variable s ; thus $a_0 = R_G G_F$, $a_1 = 0$, $a_2 = r_1 \tau_2 \tau_3 G'_L$, and $a_3 = R_G \tau_1 \tau_2 \tau_3 G'_L$.

When we express the response of the amplifier as loss rather than gain, we find a simple sum of terms rather than the more complex feedback concepts of forward-path gain and feedback loss. "Loop gain" becomes an irrelevant concept in this formulation. The feedback loss simply adds to the active-path loss.

Suppose that the total load conductance G'_L is 10 mS (millisiemens). Further, suppose that $R_G = 1.0 \text{ k}\Omega$, $r_1 = r_2 = r_3 = 0.1 \text{ k}\Omega$, $C_1 = C_2 = C_3 = 10 \text{ pF}$, and $G_F = 0.01 \text{ mS}$. Then $\tau_1 = \tau_2 = \tau_3 = 1.0 \text{ ns}$, and

$$L(s) = -(0.01 + 0.0s + 1.0s^2 + 10s^3) \tag{2.1-9}^*$$

The loss has thus been written as a function of frequency for this amplifier essentially by inspection. This loss function can be shown to be unstable, so that the amplifier is useless. Before going on to make it into a stable structure, however, we should know more about polynomials themselves and what characteristics of them make for desirable loss functions. As we see later, the missing linear coefficient of (2.1-9) guarantees instability, although its presence does not assure stability. Problem 1 shows one way of providing the missing coefficient.

2.2 ALTERNATE DESCRIPTIONS OF POLYNOMIALS

As in the case of the six blind men gathered around an elephant and putting their impressions together to try to describe it, polynomials can take on different descriptions, each of which gives us added insights to their properties. In this and the following section we explore several of these descriptions and their relationships. It is desirable to be able to pass easily back and forth between these descriptions for a full understanding of what is involved.

The “primary description” is the polynomial itself—a sum of terms in power of the frequency variable s :

$$L(s) = \sum_{i=0}^n a_i s^i \tag{2.2-1}$$

when we establish the units in which s is expressed and the values of the a_i , we have specified the polynomial completely. A fully equivalent polynomial is obtained when we scale the polynomial in amplitude by dividing through by the dc term a_0 :

$$L(s) = a_0 \sum_{i=0}^n \frac{a_i}{a_0} s^i \tag{2.2-2}$$

At very high frequencies a Bode plot of the polynomial is asymptotic to a straight line whose slope is $20n \text{ dB}$ per decade of frequency. The frequency at

*Equations involving physical quantities are expressed in this book in a self-consistent set of units to avoid the necessity for unit conversions. In this example, the volt, the milliamper, and the nanosecond are taken as the fundamental set of units, consistent with the picofarad, kilohm, millisiemen, and gigaradian per second units.

which this line intersects the dc value of loss is termed the *asymptotic cutoff frequency*, ω_0 . The polynomial may be scaled in frequency by changing the frequency variable to a new one, $p = s/\omega_0$. Here, p is the frequency normalized to the asymptotic cutoff frequency.

A polynomial scaled to its dc value and asymptotic cutoff frequency can be written

$$L(p) = a_0 \left(1 + \sum_{i=1}^{n-1} b_i p^i + p^n \right) \quad (2.2-3)$$

where

$$p = \frac{s}{\omega_0}$$

By the definition of ω_0 as the frequency at which the high-frequency asymptote intersects the dc value, we can write

$$\omega_0 = \left(\frac{a_0}{a_n} \right)^{1/n}$$

Thus the b_i in (2.2-3) are

$$b_i = \frac{a_i \omega_0^i}{a_0}$$

The advantage of this *normalized* form is that the character of the polynomial is more easily seen. We have made the first and last coefficients in the brackets unity, so that two fewer coefficients need be dealt with. The shape of the Bode plot remains unchanged. Scaling of the dc value shifts the Bode plot up or down, whereas scaling of the frequency shifts it to the left or right. The normalization of (2.2-3) shifts the dc value and cutoff frequency of the polynomial in the brackets both to unity.

In feedback system design, where the degree of the polynomials that we deal with are relatively low, the reduction to the normalized form is significant. In a system described by a cubic polynomial, for example, only two (rather than four) numbers describe the dynamics of the system. The other two numbers are there, of course; they allow us to rescale the solution to the dc loss and cutoff frequency of the original problem.

The loss polynomial of the amplifier of the previous section can be normalized as follows. First, dividing by the dc term, we obtain

$$L(s) = -0.01(1 + 0.0s + 100s^2 + 1000s^3) \quad (2.2-4)$$

setting

$$\omega_0 = \left(\frac{a_0}{a_n} \right)^{1/3} = 0.1 \text{ Grad/s} \quad (2.2-5)$$

we have

$$L(p) = -0.01(1 + 0.0p + p^2 + p^3) \quad (2.2-6)$$

The character of all normalized cubic polynomials such as this is established by the linear and quadratic coefficients b_1 and b_2 .

Binomial Loss Ratios

For a binomial—a polynomial that has only a dc term and a term linear in frequency—the center term in the brackets of (2.2-3) vanishes, and we have

$$L_1(p) = a_0(1 + p) \quad (2.2-7)$$

where it is seen that the binomial in normalized form has no distinguishing features. All have the Bode plot given in Fig. 1.6; variation of a_0 shifts the Bode plot up or down, and variation of the asymptotic cutoff frequency shifts it left or right, with no change in shape. The root of the binomial is real, at $p = -1$, or at $s = -\omega_0$.

Quadratic Loss Ratios

A quadratic in the frequency variable

$$L_2(s) = a_0 + a_1s + a_2s^2 \quad (2.2-8)$$

$$= a_0(1 + b_1p + p^2) \quad (2.2-9)$$

where

$$b_1 = \frac{a_1\omega_0}{a_0}$$

and

$$\omega_0 = \left(\frac{a_0}{a_2} \right)^{1/2}$$

has only one parameter to distinguish it, the value of b_1 ; b_1 is twice the *damping factor* ζ :

$$\zeta = \frac{b_1}{2} \quad (2.2-10)$$

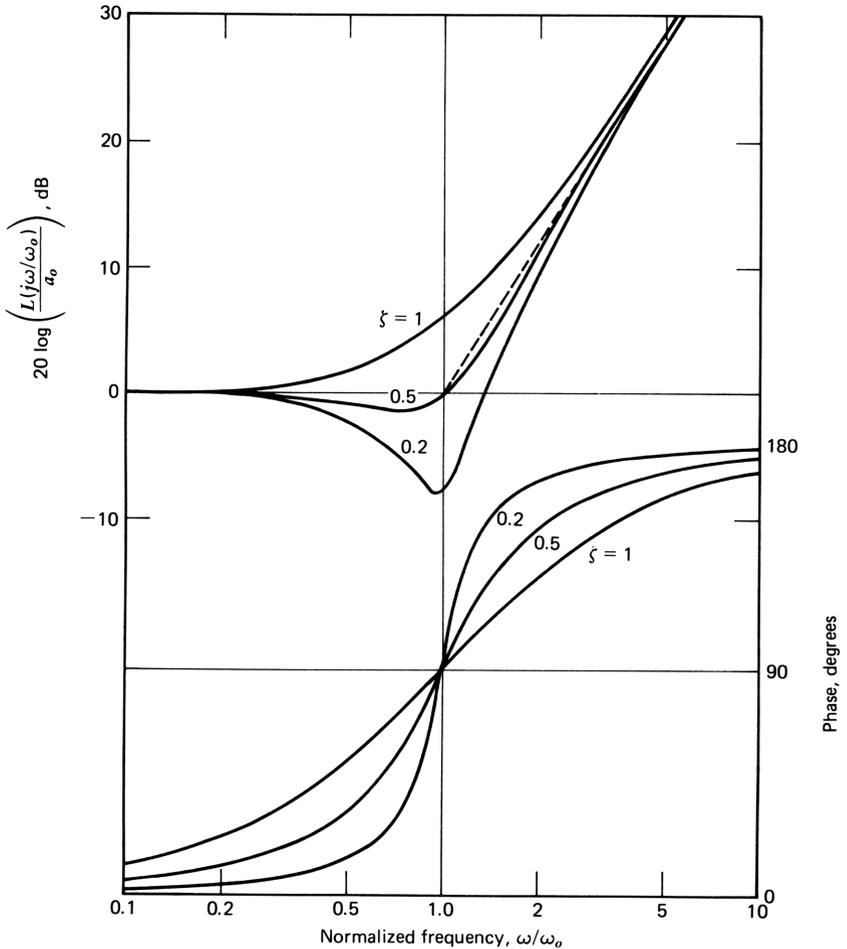


Figure 2.2. Bode plots for quadratic loss ratios.

or the reciprocal of the *quality factor* Q :

$$Q = \frac{1}{b_1} \quad (2.2-11)$$

Bode plots for the quadratic loss ratio are given in Fig. 2.2 for various values of the damping factor. Values of ζ greater than unity have not been included in this plot; where ζ is greater than unity, the quadratic can be factored into two real, linear factors, each of which has the Bode plot in Fig. 2.2. These can then be *added* to obtain the Bode plot of the quadratic. Note, however, that for $\zeta > 1$, the value of ω_0 will in general be different for the two factors, so that the cutoff frequencies will be displaced from one another.

For $\zeta > 1$, root locations for the quadratic loss ratio are found by solving eq. (2.2-9):

$$p_1, p_2 = \zeta \pm j\sqrt{1 - \zeta^2} \tag{2.2-12}$$

since $s = p\omega_0$,

$$s_1, s_2 = -\zeta\omega_0 \pm j\omega_0\sqrt{1 - \zeta^2}$$

In polar form these roots are

$$s_{1,2} = \omega_0 e^{\pm j\theta} \tag{2.2-13}$$

where

$$\theta = \cos^{-1} \zeta$$

the root locations for the quadratic are equidistant from the origin at $|s| = \omega_0$, independent of the damping factor. As the damping factor is varied, the locus of the roots is a circle of radius ω_0 centered on the origin, as shown in Fig. 2.3. The angle of the phasors is independent of ω_0 and depends only on θ . The roots enter the right half of the s plane for negative damping values. As we see in Section 2.5, the significance of any root entering the right half of the s plane is that the physical device represented by the function becomes unstable.

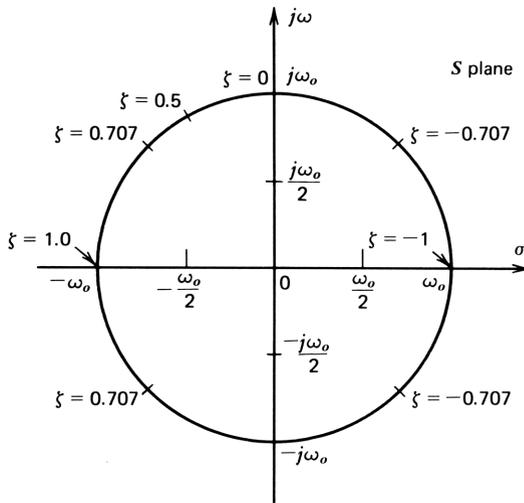


Figure 2.3. Root locations for quadratic loss ratios as a function of the damping.

Cubic Loss Ratios

Loss polynomials of higher degree can be built up as the product of linear and quadratic terms. As an example, we reconsider the cubic polynomial. The normalized cubic polynomial

$$L(p) = 1 + b_1 p + b_2 p^2 + p^3 = 0 \quad (2.2-14)$$

contains one real root and two other roots that may be either real or complex, so we can factor the polynomial as

$$L(p) = (1 + \rho^2 p) \left(1 + \frac{2\xi p}{\rho} + \frac{p^2}{\rho^2} \right) \quad (2.2-15)$$

where ρ is the asymptotic bandwidth of the quadratic factor and ξ/ρ is its damping coefficient. (Note that $1/\rho^2$ is the asymptotic bandwidth of the simple factor.) Multiplying these two factors, we obtain the b coefficients as

$$b_1 = \rho^2 + \frac{2\xi}{\rho} \quad (2.2-16)$$

$$b_2 = \frac{1}{\rho^2} + 2\xi\rho \quad (2.2-17)$$

For positive b coefficients, the necessary and sufficient conditions for stability are that $\xi > 0$, so the quadratic roots remain in the left half plane. Hence

$$2\xi = b_1 \rho - \rho^3 > 0$$

$$2\xi = \frac{b_2}{\rho} - \frac{1}{\rho^3} > 0$$

or

$$b_1 > \rho^2 \quad (2.2-18)$$

$$b_2 > \frac{1}{\rho^2} \quad (2.2-19)$$

Multiplying these two inequalities, we obtain the stability condition as

$$b_1 b_2 > 1 \quad (2.2-20)$$

The stability condition merely tells us the constraints on b_1 and b_2 to avoid oscillation. More stringent restrictions on the b coefficients are required if the response is to conform to a useful performance specification. One possibility is

to restrict the damping coefficient of the quadratic term to be greater than a given value (or to be equal to a specified value). To this end, eqs. (2.2-16) and (2.2-17) have been used to obtain the b coefficients for various values of the damping coefficient and are plotted in Fig. 2.4.

In Fig. 2.4 the stability borderline is a line of -1 slope passing through the point at which $b_1=b_2=1$, corresponding to zero damping. Above and to the right of this line, curves of constant values of quadratic damping from 0.01 to 1 are plotted. The curve for $\zeta=0$ and 1 separate the b_2, b_1 plot into three regions: below and to the left of the $\zeta=0$ line is the region of unstable operation; above and to the right of the $\zeta=1$ curve is a region in which all three cubic roots are real; the region between represents the set of b values for which a pair of the roots is complex but stable, with the third root on the negative real axis. Lines of constant values of ρ are also plotted.

The line for $\rho=1$ is of particular interest. It has a slope of $+1$ and passes through the point $b_1=b_2=1$. It represents the value of ρ for which all three roots lie on the unit circle in the p plane (a circle of radius ω_o in the s plane). It passes through the point 2,2, a polynomial of maximally flat amplitude, to be discussed later, and the point 3,3, at which all three roots are at $p=-1$ on the p plane ($-\omega_o$ on the s plane).

For values of ρ less than unity, the importance of the real root is less: for values of ρ of 0.5 or less, the real root may be ignored for most purposes, as we now show. For this value of ρ , eq. (2.2-15) becomes

$$L(p) = (1 + 0.25p)(1 + 4\zeta p + 4p^2) \quad (2.2-21)$$

Thus the asymptotic bandwidth of the quadratic factor is 0.5, and at this cutoff frequency the effect of the simple, real factor is not greatly different from unity since the real root occurs at $p=-4$, eight times the quadratic cutoff. Taking $p=j\omega$, we can evaluate $L(j\omega)$ as

$$L(j\omega) = (1 + j.25\omega)(1 + j4\zeta\omega - 4\omega^2) \quad (2.2-22)$$

Evaluating this at $\omega=0.5$, the effect of dropping the real factor is to introduce an error of $10 \log(1 + 0.125^2) = 0.067$ dB and a phase error of $\tan^{-1} 0.125 = 7.1^\circ$. Hence the region above the $\rho=0.5$ curve in Fig. 2.4 is one for which quadratic analysis is generally adequate. As p is reduced, the importance of the real root drops rapidly, as ρ^2 . The roots of the remaining quadratic are given directly by the plot in Fig. 2.4 in terms of the b coefficients.

Conversely, as ρ increases above unity, the importance of the real root increases, and that of the quadratic pair decreases. To be able to ignore the quadratic factor, however, it is not sufficient that the quadratic cutoff, or quadratic asymptotic cutoff be at a much higher frequency than the real factor cutoff. It is also necessary that the damping factor be sufficiently high to prevent the loss ratio in the cutoff region from staging a return to low values of

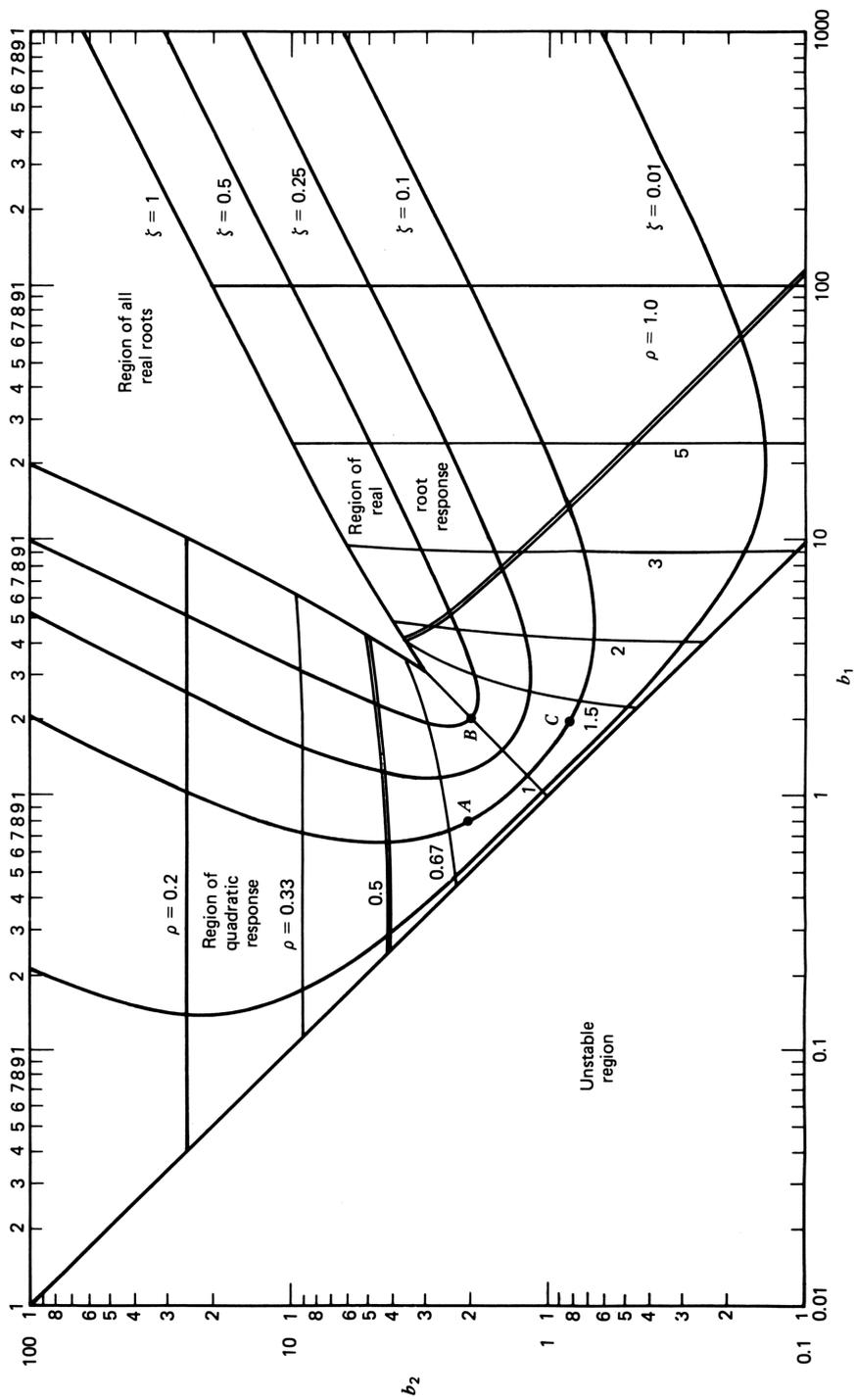


Figure 2.4. Analysis of the cubic polynomial scaled to unit asymptotic bandwidth; plot gives bandwidth and damping of the quadratic factor as a function of the b coefficients.

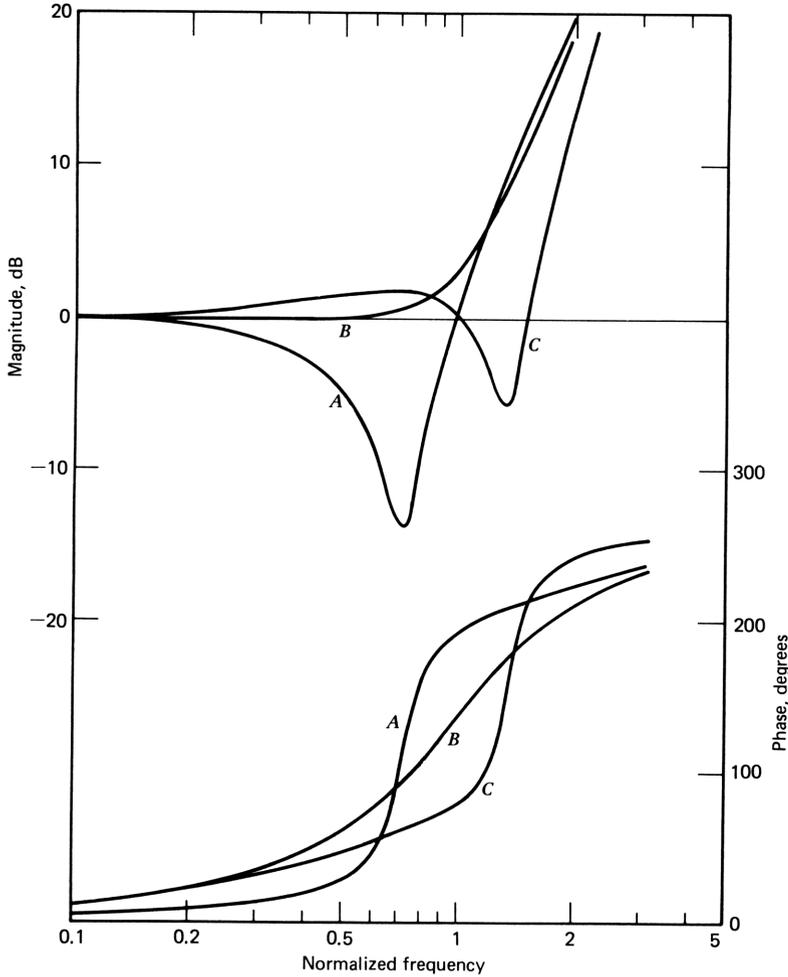


Figure 2.5. Bode plots of three cubic loss ratios.

loss ratio at high frequencies, of the sort shown in curve C in Fig. 2.5. This plot of $L(j\omega)$ is for the values $b_1=2$, and $b_2=0.8$, point C in Fig. 2.4. This corresponds to a damping coefficient of 0.1, which gives a sharp dip in the loss ratio (a peak of gain), which is usually an unsatisfactory response.

For comparison, two other curves are shown in Fig. 2.5. For curve A, the loss ratio is dominated by the quadratic factor. This curve is drawn for $b_1=0.8$ and $b_2=2$, point A in Fig. 2.5. Curve B shows the loss ratio for the maximally flat amplitude case, with $b_1=b_2=2$ (point B). For this case, neither factor is dominant.

Is there a region in Fig. 2.4 that can be said to be controlled by the real root, so that the cubic can be approximated by a simple first-degree cutoff? To be able to ignore the quadratic factor, we need merely to ensure that the quadratic

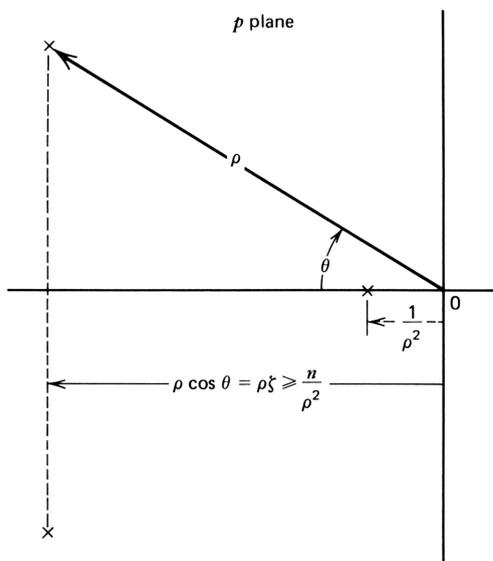


Figure 2.6. Root diagram for determining conditions under which the quadratic roots of a cubic may be ignored.

factor roots lie more deeply in the left half of the p plane than the real root, as depicted in Fig. 2.6. In Fig. 2.6 the real part of the quadratic factor is taken to be n times as large as the real root. Thus we can write

$$\rho \zeta \geq \frac{n}{\rho^2} \quad (2.2-23)$$

Substituting this inequality in eqs. (2.2-16) and (2.2-17), we have

$$b_1 \geq \rho^2 + \frac{2n}{\rho^4} \quad (2.2-24)$$

$$b_2 \geq \frac{2n+1}{\rho^2} \quad (2.2-25)$$

Thus, for values of the b coefficients that satisfy these equations and for $0 < \zeta < 1$, the quadratic roots can often be ignored. For $n=5$, the worst error (at the real root asymptotic cutoff frequency of $0.04\omega_0$) occurs for $\zeta=1$ and is 0.07 dB and 10° . The equation for the borderline case is obtained by substituting equal signs in (2.2-24) and solving for b_1 as a function of b_2 :

$$b_1 = \frac{1+2n}{b_2} + 2n \left(\frac{b_2}{1+2n} \right)^2 \quad (2.2-26)$$

This equation is plotted on Fig. 2.4 as a double line for $n=5$; to the right of this curve, the quadratic factor may be dropped.

For feedback systems only a restricted area of Fig. 2.4 is of interest in the sense of exhibiting cubic response. In this area the roots may be found from the b coefficients by reading ρ and ζ from the plot: these are then used with eq. (2.2-15) to obtain the roots.

Stabilizing the Amplifier in Section 2.1

Readers who have completed Problem 1 will have already found that the addition of a feedback capacitor C_F in parallel with G_F does essentially one thing: it adds a term $-R_G C_F s$ to the loss ratio, thereby converting eq. (2.1-9) to

$$L(s) = -0.01 \left(1 + \frac{R_G C_F s}{0.01} + 100s^2 + 1000s^3 \right) \quad (2.2-27)$$

In normalized form, $p=s/10$. With $R_G=1.0 \text{ k}\Omega$,

$$L(p) = -0.01(1 + 10C_F p + 1.0p^2 + 1.0p^3) \quad (2.2-28)$$

In this equation $b_2=1.0$ for any C_F , so that the locus on Fig. 2.4 as C_F is varied is the horizontal line for $b_2=1.0$. When C_F is zero, the loss is in the unstable region. When $C_F=0.1$, $b_1=1.0$, and the loss is on the border of instability. The damping of the quadratic roots reaches a maximum for $b_1=3$, or for $C_F=0.3 \text{ pF}$, giving

$$L(p)' = -0.01(1 + 3p + p^2 + p^3) \quad (2.2-29)$$

A measure of its margin against instability is the value of damping of the quadratic roots, which is about 0.2, according to Fig. 2.4. This corresponds to a value of θ (in Fig. 2.6) of 78° . This margin is small, only 12° away from instability.

Addition of Load Capacitance to the Amplifier

To begin the study of higher-order polynomials, consider the amplifier described in Section 2.1 driving an additional load of 10 pF of capacitance connected directly across the load. Can we find the loss ratio by inspection again, as we did in the earlier case? By *superposition*, the load capacitance adds new terms to the loss, terms we can find by substituting $C_L s$ for G_L in eq. (2.1-7) and including only terms containing C_L . Thus the *change* in loss caused by capacitance is

$$\Delta e_G = r_1 \tau_2 \tau_3 C_L s^3 + R_G \tau_1 \tau_2 \tau_3 C_L s^4 \quad (2.2-30)$$

and the new equation for the loss of the amplifier is found by adding these two

terms to (2.1-7):

$$\begin{aligned} L(s)' &= r_G G_F + R_G C_F s + r_i \tau_2 \tau_3 G_L' s^2 \\ &\quad + (R_G \tau_1 \tau_2 \tau_3 G_L' + r_i \tau_2 \tau_3 C_L) s^3 \\ &\quad + R_G \tau_1 \tau_2 \tau_3 C_L s^4 \end{aligned} \quad (2.2-31)$$

The loss polynomial is now a quartic. Note the ease with which the expression is obtained through superposition. This facility is lost when the analysis proceeds from input to output.

When $C_L = 10 \text{ pF}$, the added terms are $1.0s^3$ and $10s^4$; thus

$$L'(s) = -(0.01 + 0.3s + 1.0s^2 + 11s^3 + 10s^4) \quad (2.2-32)$$

What are the characteristics of this polynomial? Is it stable? Where do its roots lie? In the following sections we answer these questions for polynomials in general. For the present, we can use normalization to get some idea of the characteristics of the polynomial. To compare it with the cubic for the amplifier without load capacitance, we set $p = 10s$ for comparison purposes and obtain

$$L'(p) = -0.01(1 + 3p + 1.0p^2 + 1.1p^3 + 0.1p^4) \quad (2.2-33)$$

From this equation we see that at the cubic cutoff frequency ($p = 1.0$), the quartic term is about 10% of the magnitude of the quadratic and cubic terms, leading us to believe that the effect of the quartic term is modest. On the other hand, the margin against instability is small, so the small change may be enough to cause instability. To resolve these questions, further study of polynomials is indicated.

Calculator Programs

A good way to consolidate what has been discussed is to commit the relationships to a program for the computer or calculator. By having to tell the machine exactly what is to be done, the concepts become clearer and misunderstandings are corrected. In addition, when later work requires that we make use of these concepts, they are available to us in their most directly useful form, so that we do not have to go back over the material. For these reasons, calculator programs are given in the appendix at the end of the book with explanations of how they have been developed. Readers who are interested in working with the concepts to be developed are urged to obtain either a calculator or time on a computer facility and to follow the development of these programs.

Five programs have been developed in conjunction with the concepts of this section. The first two scale polynomials of any degree up to eighth. A third

builds up polynomials from linear and quadratic factors. The fourth finds the roots of a quadratic equation, and the fifth finds the roots of a cubic. All are explained in Appendix A.

2.3 ROOTS OF POLYNOMIALS BY NUMERICAL ANALYSIS

Roots of polynomials of degree higher than the cubic are usually determined by numerical analysis—a sort of directed trail-and-error approach. An initial guess of the location of a root of a given function $L(s)$ is made; the function is evaluated at this value of s . If $L_1(s)=0$, the guess was right and the root is found. If $L_1(s)\neq 0$, some method is used to improve the guess, and the process is repeated. Many approaches to the problem of improving the guess have been devised. For finding roots of polynomials, where the roots may be complex, Newton’s method is satisfactory. A brief explanation of the method is given here since we find it useful later on. For a complete discussion of the method, see, for example, Chapter 2 of Atkinson’s work,¹ or Conte and de Boor.²

To begin, we assume real roots of $L(s)$ since this is easier to depict graphically. Let the n th guess of the root of $L(s)$ be s_n . For Newton’s method, the $(n + 1)$ th guess is

$$s_{n+1} = s_n - \frac{L(s_n)}{L'(s_n)} \tag{2.3-1}$$

To find the $(n + 1)$ st guess, we must evaluate the function and its derivative at s_n , divide the latter into the former, and subtract the result from s_n . The process is repeated until the magnitude of $L(s)$ falls below a small error ϵ at which point the evaluated root has satisfactory accuracy. The process is depicted in Fig. 2.7, in which the first guess was $s=0$. The tangent to the curve at $s=0$ is drawn; where it intersects the s -axis is given by eq. (2.3-1). This becomes the second guess for s , where a second tangent is drawn to find s_3 ; the process is repeated until $|L(s)|$ is sufficiently small.

For a polynomial, we must evaluate the function

$$L(s) = \sum_{i=0}^n a_i s^i \tag{2.3-2}$$

and its derivative

$$L'(s) = \sum_{i=0}^n i a_i s^{i-1} \tag{2.3-3}$$

a process most easily carried out on a computer or a calculator. A program for implementing Newton’s method for polynomials up to twelfth degree (extendable by adding memory register locations) is given in Appendix A.

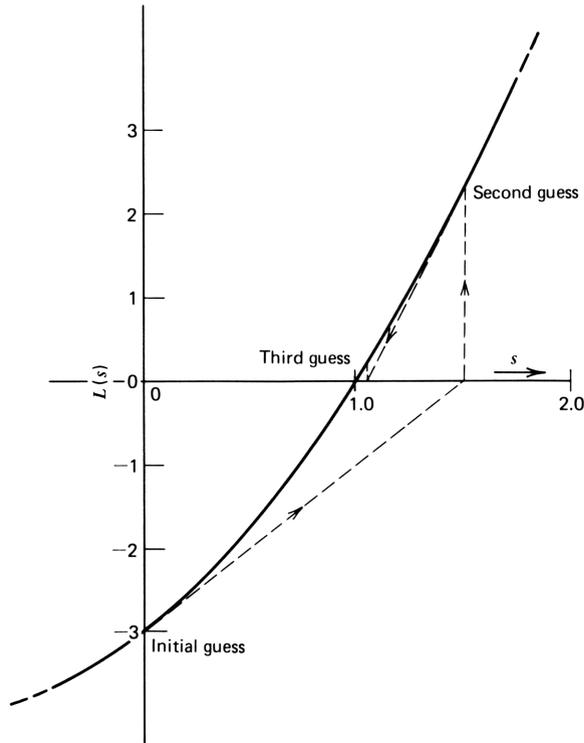


Figure 2.7. Use of Newton's method for finding a real root of $L(s)=0$.

The example in Fig. 2.7 is drawn for the quadratic function

$$L(s) = -3.0 + 2s + s^2 \quad (2.3-4)$$

whose derivative is

$$L'(s) = 2 + 2s \quad (2.3-5)$$

Taking $s=0$ as a first guess, we have $L(0) = -3.0$ and $L'(0) = 2$, so that the second guess for s is 1.5. To obtain the third guess, we evaluate $L(1.5) = 2.25$ and $L'(1.5) = 5$, so that

$$s_3 = 1.5 - \frac{2.25}{5} = 1.05$$

A calculator tape showing the convergence on the root at $s=1$ is shown in Fig. 2.8. When $|L(s)|$ becomes less than $\epsilon = 10^{-8}$, the root is approximated accurately enough, and the polynomial is *deflated*—that is, divided by the root factor. The process is repeated for the remaining polynomial, which in this case

```

                SF 00
                XEQ "ROOTS"

ROOTS

DEG?
                2   RUN
POLY, ASC. ORDER
R20= -3.000E0
R21= 2.000E0
R22= 1.000E0

OK?
                RUN
L=3.00E0
L=2.25E0
L=2.03E-1
L=2.44E-3
L=3.72E-7
L=0.00E0
S=1.0000

POLY, ASC. ORDER
R20= 3.000E0
R21= 1.000E0

L=3.00E0
L=0.00E0
S=-3.0000

END
    
```

Figure 2.8. Calculator tape showing convergence on the real roots of a polynomial. Magnitude of loss $|L|$ is shown on calculator tape.

is just $3 + s$, with the derivative 1.0. Since the deflated polynomial is linear in this example, the root is found immediately. (Why?)

The program automatically determines the starting point, in such a way that it usually finds the smallest roots first for better accuracy. The initial guess should include an imaginary part. The reason for this is that for real s , both $L(s)$ and its derivative are real. Hence if the initial guess is also real, successive guesses will never depart from the real axis according to eq. (2.3-1). Starting out with an imaginary part of s avoids this trap. (A special subroutine in the program does this automatically if the value of s becomes real during the iterations.) For purposes of the example in Fig. 2.8, we overrode the initial guess of $j1.5$, substituting $s=0$.

For an example containing complex roots, let

$$L(s) = 1 + 2s + 2s^2 + s^3 \tag{2.3-6}$$

whose derivative is

$$L'(s) = 2 + 4s + 3s^2 \tag{2.3-7}$$

Let us take $s=j1$ as an initial guess (the program takes $s=0.20+j0.46$ but $j1$ illustrates the process more clearly). Then $L(s)=-1+j1, =\sqrt{2} \angle 135^\circ$ and $L'(s)=-1+j4=4.123 \angle 104.0^\circ$, so that the new value of s to be tried is

$$\begin{aligned} s_2 &= j1 - \frac{1.414 \angle 135^\circ}{4.123 \angle 104.0^\circ} \\ &= -0.2941 + j0.8235 \end{aligned} \quad (2.3-8)$$

thus beginning the process of converging on the root at $-0.5+j0.866$. The calculator tape shown in Fig. 2.9 illustrates the convergence. After the complex root is found, the original polynomial is deflated by the quadratic factor of this root and its complex conjugate, following which the remaining root is found.

The program in Appendix A will usually find roots without difficulty, including finding its own starting point—the initial guess for the root position. Where difficulty is encountered, an interactive mode is provided, allowing for variation of starting point and strategy.

```

                SF 00
                XEQ "ROOTS"

ROOTS

DEG?          3.0000   RUN
POLY, ASC. ORDER
R20= 1.000E0
R21= 2.000E0
R22= 2.000E0
R23= 1.000E0

OK?          RUN
L=1.41E0
L=3.88E-1
L=9.22E-2
L=9.51E-3
L=7.95E-5
L=5.04E-9
S=1.0000+-j120.00

POLY, ASC. ORDER
R20= 1.000E0
R21= 1.000E0

L=1.41E0
L=1.00E-10
L=0.00E0
S=-1.0000

END

```

Figure 2.9. Calculator tape showing convergence for a polynomial with complex roots using Newton's method.

Time Domain Performance³

Factoring loss polynomials has one important application in the study of feedback systems—namely, establishing the performance of the system in the time domain. The stability characteristics of the system are most clearly seen by observing its behavior as a function of time. The output of a system characterized by a loss polynomial $L(s)$ is described in terms of its input signal $X(s)$ by

$$Y(s) = \frac{X(s)}{L(s)} \tag{2.3-9}$$

For the following development, we take $X(s)$ as the Laplace transform of unit step, equal to $1/s$. This will give us the *step response* of the system. Where the loss polynomial is factored, we can write

$$Y(s) = \frac{1}{s(s-\lambda_1)(s-\lambda_2)\cdots(s-\lambda_n)a_n} \tag{2.3-10}$$

where the factor s in the denominator comes from $X(s)$. If all roots of the denominator are distinct (no multiple roots), this equation can be written as the sum of factors by the method of partial fractions:

$$Y(s) = \frac{k_0}{s} + \frac{k_1}{s-\lambda_1} + \frac{k_2}{s-\lambda_2} + \cdots + \frac{k_n}{s-\lambda_n} \tag{2.3-11}$$

$$= \frac{k_0}{s} + \sum_{i=1}^n \frac{k_i}{s-\lambda_i} \tag{2.3-12}$$

where the λ_i may be real or complex. The inverse Laplace transform of this sum is equal to the inverse Laplace transform of each term and is the superposition of a series of time responses; thus

$$y(t) = \sum_{i=0}^n k_i e^{\lambda_i t} \tag{2.3-13}$$

the k_i in either of these equations are termed the *residues*. Where the λ_i are real (and separate) there are three cases of interest: negative, zero, and positive values of λ . Step responses for these three cases are shown in Fig. 2.10 for values of λ of -1 , 0 , and $+1$, corresponding respectively to a decaying exponential, a step, and a growing exponential. The latter case represents an *unstable* time response.

Where the λ_i are complex, they must occur in conjugate pairs, which then can be paired. The residues are also complex conjugate, so we can write the

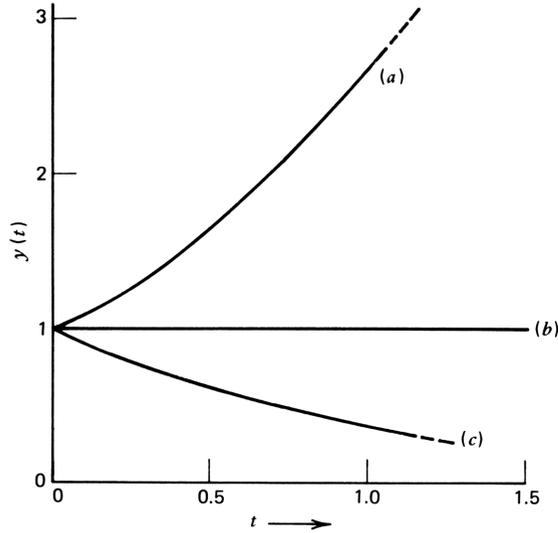


Figure 2.10. Time responses for system with a single real root: (a) positive; (b) zero; (c) negative.

frequency response for each pair, letting $\lambda = \alpha + j\beta$,

$$Y_p(s) = \frac{k_p}{s - \alpha + j\beta} + \frac{k_p^*}{s - \alpha - j\beta} \quad (2.3-14)$$

The corresponding time response is

$$y_p(t) = k_p e^{(\alpha + j\beta)t} + k_p^* e^{(\alpha - j\beta)t} \quad (2.3-15)$$

$$= e^{\alpha t} (k_p e^{j\beta t} + k_p^* e^{-j\beta t}) \quad (2.3-16)$$

Expressing the residue k_p in polar form, $k_p = \rho e^{j\theta}$ and $k_p^* = \rho e^{-j\theta}$; we can write the time response as

$$y_p(t) = \rho e^{\alpha t} [e^{j(\beta t + \theta)} + e^{-j(\beta t + \theta)}] \quad (2.3-17)$$

$$= 2\rho e^{\alpha t} \cos(\beta t + \theta) \quad (2.3-18)$$

For either real or complex roots, the residue at the j th root is found from the equation

$$k_j = \left. \frac{(s - \lambda_j) X(s)}{L(s)} \right|_{s = \lambda_j} \quad (2.3-19)$$

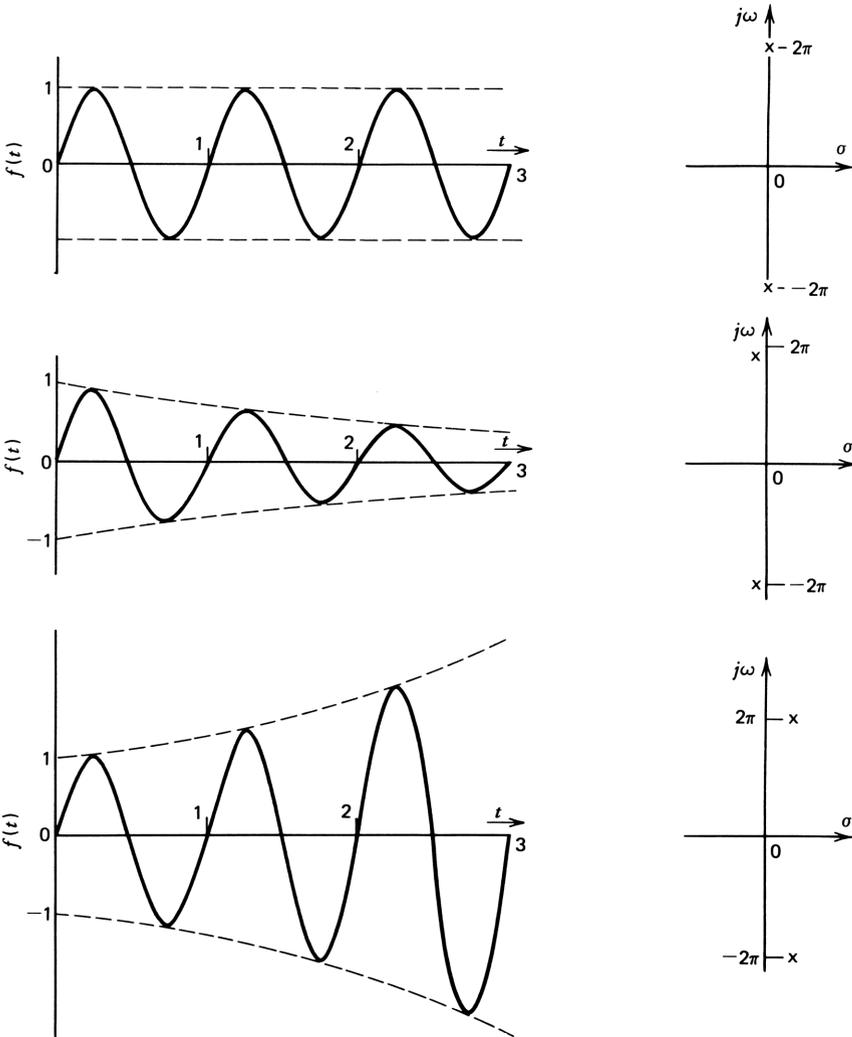


Figure 2.11. Time responses for system with one pair of complex roots whose real part is (a) zero, (b) negative, and (c) positive.

Since $s - \lambda_j$ is a factor of $L(s)$, it must be factored out before the evaluation of the residue is done. With $L(s)$ in factored form, $X(s)$ is divided by each factor $\lambda_j - \lambda_i$, except, of course, $\lambda_i - \lambda_i = 0$.

The step responses for three pairs of complex conjugate root positions are shown in Fig. 2.11 for positive, zero, and negative values of the *real part* of the roots. When the real part is zero, a sinusoidal signal results. Note that for just the single pole pair on the $j\omega$ axis, the angle of the residue is 90° , so that eq.

(2.3-19) gives a sinusoidal response. Negative real parts yield an exponentially decaying sinusoid, whereas positive real parts yield a growing exponential, which is an unstable condition.

Settling Time

The time taken for a system to come to within a small percentage error of its final value is often important. Analog-to-digital converters constitute a typical example. Such converters sample the signal periodically and hold the value of the sample while it is being processed into a digital code. Required accuracies are often uncommonly high. In the case of a coder that is to produce a 14 bit code for each sample, the analog sample is to be held to within a fraction of the smallest step, which is 2^{-14} of the maximum amplitude, about 30 parts per million (ppm).

A feedback amplifier is often called on to amplify the held sample. One measure of its suitability for this purpose is its *settling time*, the time taken for it to settle to within a given fraction (e.g., 30 ppm) of its final value under the excitation of a step input. Where the amplifier loss is described by a polynomial, settling time may be regarded as a property of the polynomial itself. Since time responses die away exponentially, we may expect that (1) settling time is related to the time constant of the exponential decay term— α in the term $k_1 e^{-\alpha t}$ or in the term $k_2 e^{\alpha t} \cos(\beta t + \theta)$ —and (2) only the root(s) nearest the $j\omega$ axis will be involved, as the effect of other roots will have died away earlier in time. The residue at the root in question also affects settling time, but unless it is very small (i.e., unless the root is canceled by a pole of loss), the effect of the residue constant is small. Thus we can write

$$k_j e^{-\alpha_j T_s} = \epsilon \quad (2.3-20)$$

where ϵ is the allowable error after T_s nanoseconds and α_j is the distance from the j axis of the pole (or pole-pair) nearest the axis. The settling time is given by

$$T_s = \frac{-\ln \epsilon + \ln k_j}{\alpha_j} \approx \frac{-\ln \epsilon}{\alpha_j} \quad (2.3-21)$$

As an example, the roots of the amplifier described in Section 2.1 with 0.3 pF added as in Problem 2 are at -0.0361 and at $-0.0319 \pm j0.163$ Grad/s, so that the complex roots dominate the settling time. For this amplifier to settle to within 30 ppm of the final value would take about $[-\ln(30 \times 10^{-6})]/0.0319 = 330$ ns.

The complete step response of this amplifier is shown in Fig. 2.12. The initial slow start of the response comes about because the real root is close to the $j\omega$ axis. The vertical scale has been expanded for $t > 200$ ns to show the settling of the waveform.

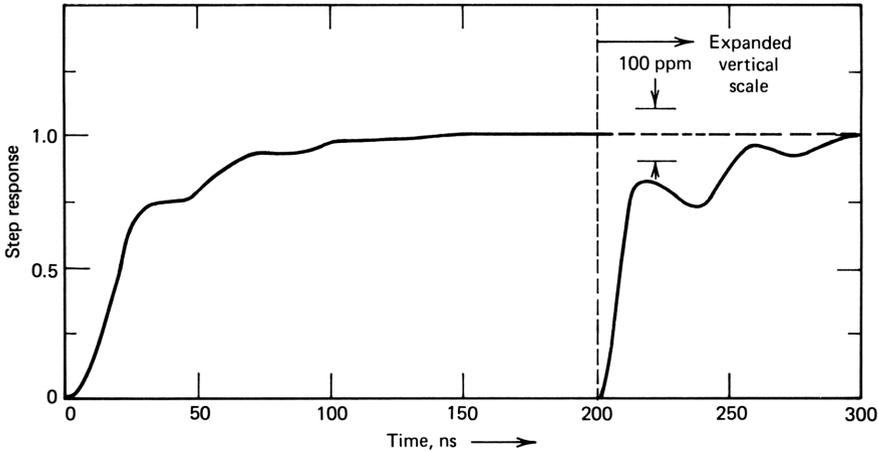


Figure 2.12. Step response, normalized to unity output, for the amplifier in Section 2.1 stabilized by a 0.3 pF feedback capacitance.

2.4 POLYNOMIAL EVALUATION

The most common method of presenting transfer functions is by the Bode plot —magnitude in decibels and phase in degrees versus log frequency for $s=j\omega$. With $s=j\omega$, a simpler method of evaluation is possible, allowing us to evaluate the polynomial on a calculator in less than 40 program steps. The polynomial is split into its even and odd parts, each of which is expressed in nested form. A seventh-degree polynomial, for example, may be written

$$L(j\omega) = a_0 - \omega^2 [a_2 - \omega^2 (a_4 - \omega^2 a_6)] + j\omega \{ a_1 - \omega^2 [a_3 - \omega^2 (a_5 - \omega^2 a_7)] \} \tag{2.4-1}$$

To find the imaginary part of $L(j\omega)$, a_7 is multiplied by $-\omega^2$ and a_5 is added to it; the result is multiplied by $-\omega^2$ and a_3 is added to it; this result is then multiplied by $-\omega^2$ and a_1 is added to it. This result is then multiplied by ω to give the imaginary part. The real part is obtained similarly. To obtain a Bode diagram, the real and imaginary parts are converted to polar form, and the magnitude is converted to decibels. The polynomial evaluation program in Appendix A uses this method.

The program was used to generate the Bode plots in Fig. 2.13 for the amplifier in Section 2.1, both with and without the stabilizing feedback capacitance of 0.3 pF. The Bode plot with $C_F=0$ is characteristic of unstable systems: the phase *decreases* with frequency at a frequency near the right half plane roots. Stable behavior is illustrated by the curve for $C_F=0.3$ pF, which shows rising phase as the loss rises.

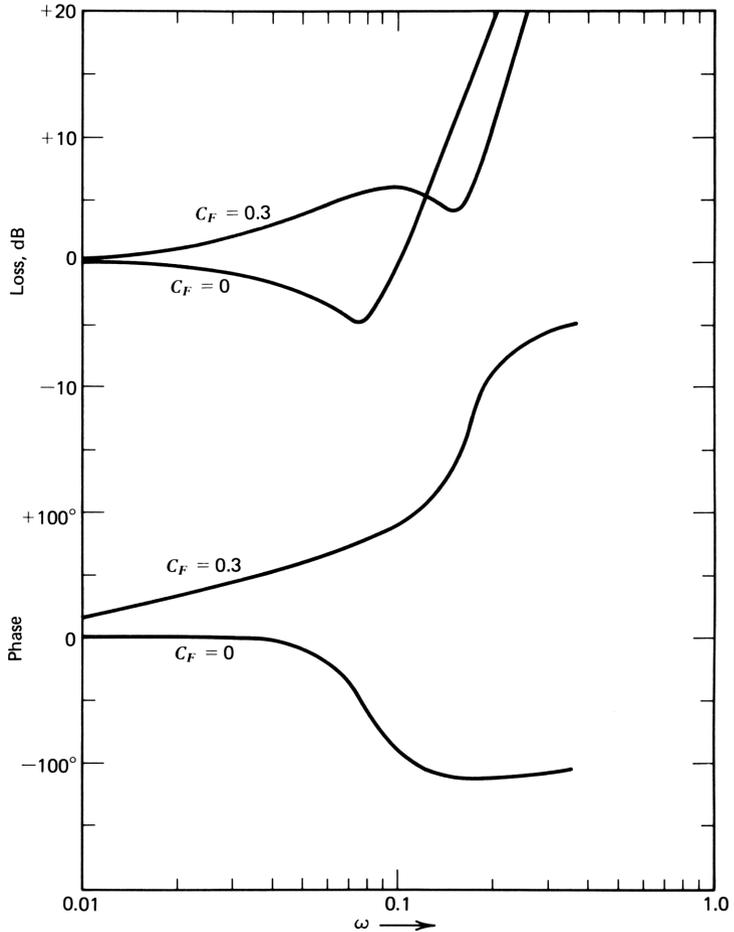


Figure 2.13. Bode plots for amplifier with stable and unstable loss ratios.

The same program can be used to generate different plots of the same information, termed a *Nyquist diagram*, where the real part of $L(j\omega)$ is plotted against the imaginary part at various frequencies. The length of the phasor from the origin to the curve is the magnitude of $L(j\omega)$, and the angle of this phasor with the real axis is the phase. The magnitude is zero at each root of $L(s)$; thus the origin of this diagram corresponds to *all* root locations on the s plane.

It is shown (in the theory of functions of complex variables) that the Nyquist diagram is a conformal mapping of the s plane onto the $L(s)$ plane; the *Nyquist diagram maps the $j\omega$ axis onto the $L(j\omega)$ plane*. The meaning of “conformal” in this context is that angles are preserved in the mapping process — a small square on the s plane will map to a small square on the $L(j\omega)$ plane. Facing north on the $j\omega$ axis in the s plane, the right half plane is to one’s right.

On the Nyquist diagram the map of the right half plane exists to the right of the *map* of the $j\omega$ axis facing in the direction of increasing frequency.

Examples of Nyquist diagrams for linear, quadratic, and cubic loss ratios are given in Fig. 2.14; in each case we have mapped not only the $j\omega$ axis, but also a small pennant arbitrarily placed in the right half plane. The pennant has no significance except to convey a “feel” for the mapping process and the kinds of distortions it introduces in the shapes of things. Consider the linear loss ratio given as

$$L_1(s) = 1 + \tau_1 s \tag{2.4-2}$$

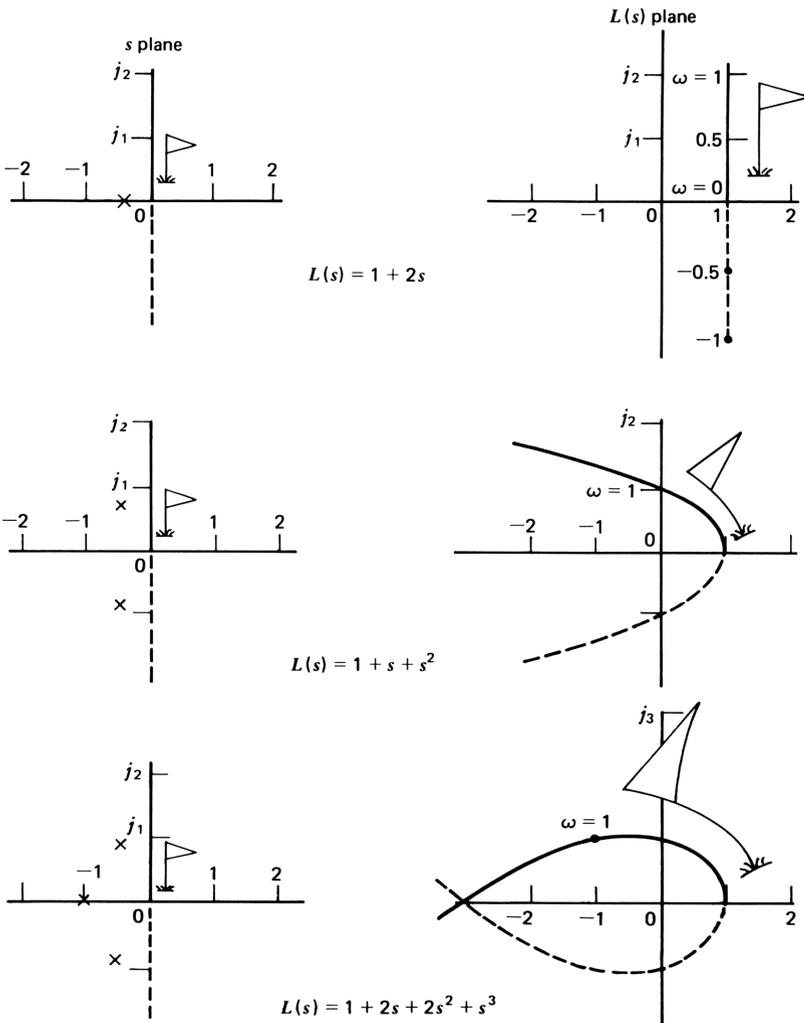


Figure 2.14. Nyquist sketches for linear, quadratic, and cubic stable loss ratios.

Setting $s = \sigma + j\omega$, we find the real and imaginary parts of $L_1(s)$ as

$$L_1(\sigma + j\omega) = 1 + \sigma\tau_1 + j\omega\tau_1 \quad (2.4-3)$$

If we set $\sigma = 0$, $s = j\omega$ so that we have the map, on the $L_1(s)$ plane, of the $j\omega$ axis. There is one root of $L(s)$ at $s = -1/\tau_1 + j0$; this point (of course) maps to the point $0 + j0$ on the $L(s)$ plane: any roots of a loss function or of the characteristic equation map to the point $0 + j0$ for any $L(s)$. The map of $L_1(s)$ looks like the s plane itself, except that the imaginary axis is displaced one unit to the left, and the size is scaled by the factor τ_1 . The pennant is to the right of the imaginary axis and is scaled appropriately. The figure is drawn with the map of the negative $j\omega$ axis shown as a dashed line. As one walks north along the $j\omega$ axis of the s plane, the right half plane is to the right; correspondingly, as one walks along the map of the $j\omega$ axis on the $L(s)$ plane, the map of the $j\omega$ axis is likewise on the right, since angles are preserved.

The map of the quadratic function is drawn in exactly the same way. For small values of ω , the map is similar to the linear case since the $-\omega^2$ term is negligible. As the frequency increases, this term bends the map to the left in the parabolic shape shown. The map of both roots are at $0 + j0$. The pennant is bent and larger than in the linear case. The cubic function includes the $-j\omega^3$ factor, thus causing the map to descend at high frequencies. In all three cases the *origin* is in a region that is a map of the left half plane; all three cases are stable.

An elegant theory of stability based on complex variable theory was given by Nyquist in the early 1930s. It is not needed now because of the ease with which we can find the roots of polynomials on a computer or a calculator. The Nyquist diagram, on the other hand, is of help in understanding and interpreting stability problems that arise in feedback systems, particularly those including transport delays, to be discussed in Chapter 5.

Unstable loss ratios are shown in Fig. 2.15. For the linear case, we take the example

$$L_1(s) = 1 - \tau_1 s \quad (2.4-3)$$

As ω is increased from zero, $L(j\omega)$ moves negatively, so that the right half plane maps to the region to the *left* of $L(j\omega)$; therefore, $L(s)$ has a root in the right half plane, and the system is unstable. The quadratic loss ratio is unstable for the same reason; the presence of the $-\omega^2$ term bends $L(j\omega)$ to the left as before, but the right half plane is to the left of the curve.

The unstable cubic function illustrated in Fig. 2.15 has all positive coefficients, so that the right half plane is to the right of the map near $\omega = 0$; evidently this instability is of a type different from those of the first two cases. Here, the cubic coefficient is large enough to cause the map to intersect the real axis between the origin and the point at $1 + j0$. The region corresponding to the right half plane includes the origin, so that the system is unstable. Only two of

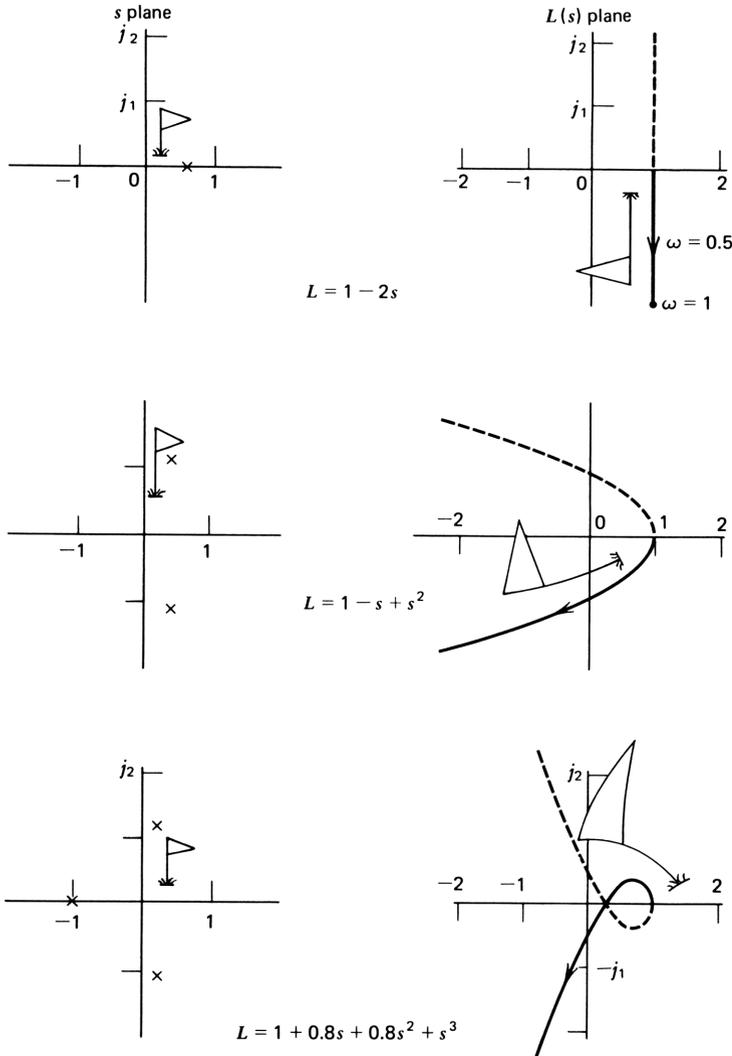


Figure 2.15. Nyquist diagrams of unstable loss ratios.

the three roots of the characteristic equation are in the right half plane. The Nyquist diagram shows that the point at the origin is shaded twice by the right half plane, once for positive frequencies and once for negative frequencies. This superposition of layers of the right half plane is common in more complicated Nyquist diagrams; in the stretching process of mapping, the several layers that may result are called *Riemann surfaces*.

With this background we can replot the information of the Bode plot in Fig. 2.13 as a Nyquist diagram. Figure 2.16 shows Nyquist diagrams for $C_F=0.3$ pF and for 0 pF as well as two other cases. For $C_F=0.3$ pF, the origin is to the

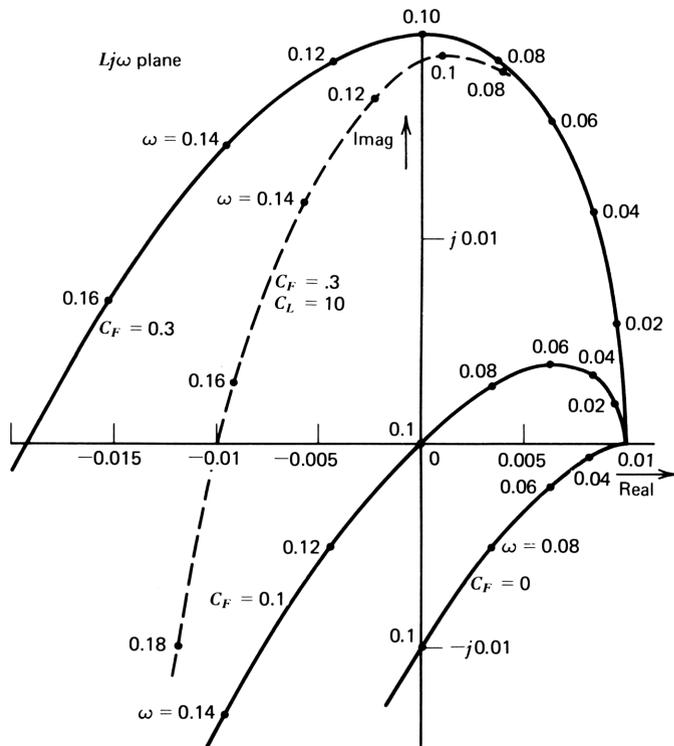


Figure 2.16. Nyquist diagrams amplifier loss with feedback capacitors of 0, 0.1, and 0.3 pF. The dashed line shows the effect of adding 10 pF of load capacitance.

left of the map of the $j\omega$ axis as we proceed in a direction of increasing frequency, so that the system is stable. The angle of the phasor from the origin to the curve increases with increasing frequency. For $C_F=0$, on the other hand, the origin is to the *right* of the curve as we move in the direction of increasing frequency; the system is unstable.

Two other cases are also of interest. When $C_F=0.1$ pF, the map of the $j\omega$ axis passes through the origin of the Nyquist diagram. This indicates that roots of $L(s)$ exist on the $j\omega$ axis as in the top time response in Fig. 2.11. Note that the map of the $j\omega$ axis passes through the origin twice—once for positive frequencies and once for negative (not shown). Hence there are two roots on the $j\omega$ axis. In this case the phase “switches” by 180° as we pass through the origin.

The dashed line in Fig. 2.16 shows the effect of adding load capacitance of 10 pF to the amplifier in Section 2.1. The amplifier is stable since the origin is to the left of the curve, but the loss is smaller than without the load capacitance, since the curve passes closer to the origin. The loss approaches a minimum for $\omega=0.16$ and is about equal to the low-frequency loss.

2.5 EVALUATION OF THE POLYNOMIAL COEFFICIENTS FROM $L(j\omega)$

The inverse process to finding the Bode diagram from the polynomial is to find the polynomial coefficients from values of $L(j\omega)$ that might come from measurements or from a computation. This process is closely allied to the *modeling process* and to the *approximation problem*. In the modeling process we wish to develop a model—a mathematical model first, then a physical model—that matches a set of measurements of frequency response. In particular, we are interested in a polynomial model that we later show can be translated into a physical structure. The existence of such a model is proved in a theorem by Weierstrass.¹

Let $L(j\omega)$ be continuous for $\omega_1 < \omega < \omega_2$, and let $\epsilon > 0$. Then there is a polynomial $P(j\omega)$ for which

$$L(j\omega) - P(j\omega) < \epsilon, \quad \omega_1 < \omega < \omega_2 \quad (2.5-1)$$

A proof is given in the reference. The theorem states that given a polynomial of sufficiently high degree, we can model any continuous frequency response to any desired degree of accuracy.

The approximation problem is the same problem with the frequency response now supplied as a performance specification rather than a measured response. The two problems are similar in that in either case we are given an arbitrary function and wish to find a polynomial approximation to it.

Here, we are interested in the more restricted problem of finding the coefficients of a polynomial of known degree, given a minimum amount of data about its frequency response. With this simpler process under our belts, we shall be able to translate back and forth between the polynomial and its representation on a Bode plot. We shall also have a better understanding of modeling and of the approximation problem.

A simple example illustrates the process. Suppose that we know that a function is a binomial in frequency and that we wish to find a_0 and a_1 where

$$L(s) = a_0 + a_1 s \quad (2.5-2)$$

Then

$$L(j\omega) = a_0 + ja_1\omega \quad (2.5-3)$$

To find a_0 and a_1 , we merely find the real and imaginary parts of $L(j\omega)$ and equate the former to a_0 and the latter to $a_1\omega$. Thus a measurement at a single frequency suffices to obtain the coefficients. Since each measurement of $L(j\omega)$ gives us two numbers (a real and imaginary part), we can obtain two coefficients from each measurement.

To evaluate the coefficients of a quadratic or a cubic polynomial, the value of $L(j\omega)$ at two separate frequencies must be available, whereupon we can

write

$$\operatorname{Re}[L(j\omega_1)] = a_0 - a_2\omega_1^2 \quad (2.5-4)$$

$$\operatorname{Re}[L(j\omega_2)] = a_0 - a_2\omega_2^2 \quad (2.5-5)$$

Solving these equations simultaneously, we obtain

$$a_0 = \frac{1}{\omega_2^2 - \omega_1^2} \{ \omega_2^2 \operatorname{Re}[L(j\omega_1)] - \omega_1^2 \operatorname{Re}[L(j\omega_2)] \} \quad (2.5-6)$$

and

$$a_2 = \frac{1}{\omega_2^2 - \omega_1^2} \{ \operatorname{Re}[L(j\omega_1)] - \operatorname{Re}[L(j\omega_2)] \} \quad (2.5-7)$$

Similar equations for the imaginary parts give a_1 and a_3 since

$$\operatorname{Im}[L(j\omega)] = \omega(a_1 - a_3\omega^2)$$

giving

$$a_1 = \frac{1}{\omega_2^2 - \omega_1^2} \left\{ \frac{\omega_2^2 \operatorname{Im}[L(j\omega_1)]}{\omega_1} - \frac{\omega_1^2 \operatorname{Im}[L(j\omega_2)]}{\omega_2} \right\} \quad (2.5-8)$$

and

$$a_3 = \frac{1}{\omega_2^2 - \omega_1^2} \left\{ \frac{\operatorname{Im}[L(j\omega_1)]}{\omega_1} - \frac{\operatorname{Im}[L(j\omega_2)]}{\omega_2} \right\} \quad (2.5-9)$$

Evaluation of quadratic and cubic coefficients from loss magnitude and phase will be useful in the work ahead; a program for this evaluation is given in Appendix A. Similar procedures are also used for higher-degree polynomials. The case of the quartic or quintic polynomials is given here; the extension to yet higher-degree polynomials will be obvious from this discussion.

The quintic polynomial can be written

$$L(j\omega) = a_0 - a_2\omega^2 + a_4\omega^4 + j\omega(a_1 - a_3\omega^2 + a_5\omega^5) \quad (2.5-10)$$

so that by measuring $L(j\omega)$ at three frequencies, ω_1 , ω_2 , and ω_3 , we can write the matrix equation

$$\begin{bmatrix} 1 & -\omega_2^2 & \omega_1^4 \\ 1 & -\omega_2^2 & \omega_2^4 \\ 1 & -\omega_3^2 & \omega_3^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_2 \\ a_4 \end{bmatrix} = \begin{bmatrix} \operatorname{Re} L(j\omega_1) \\ \operatorname{Re} L(j\omega_2) \\ \operatorname{Re} L(j\omega_3) \end{bmatrix} \quad (2.5-11)$$

A similar equation can be written for the odd coefficients, replacing the real part on the right by the imaginary part divided by ω . We then invert the matrix to obtain the coefficients:

$$\begin{bmatrix} a_0 \\ a_2 \\ a_4 \end{bmatrix} = \begin{bmatrix} 1 & -\omega_1^2 & \omega_1^4 \\ 1 & -\omega_2^2 & \omega_2^4 \\ 1 & -\omega_3^2 & \omega_3^4 \end{bmatrix}^{-1} \begin{bmatrix} \operatorname{Re} L(j\omega_1) \\ \operatorname{Re} L(j\omega_2) \\ \operatorname{Re} L(j\omega_3) \end{bmatrix} \quad (2.5-12)$$

and similarly

$$\begin{bmatrix} a_1 \\ a_3 \\ a_5 \end{bmatrix} = \begin{bmatrix} 1 & -\omega_1^2 & \omega_1^4 \\ 1 & -\omega_2^2 & \omega_2^4 \\ 1 & -\omega_3^2 & \omega_3^4 \end{bmatrix}^{-1} \begin{bmatrix} \frac{\operatorname{Im} L(j\omega_1)}{\omega_1} \\ \frac{\operatorname{Im} L(j\omega_2)}{\omega_2} \\ \frac{\operatorname{Im} L(j\omega_3)}{\omega_3} \end{bmatrix} \quad (2.5-13)$$

Expressed more compactly, this is

$$a_i = F^{-1} \operatorname{Re} L(j\omega_i), \quad i \text{ even} \quad (2.5-14)$$

$$a_i = F^{-1} \frac{\operatorname{Im} L(j\omega_i)}{\omega_i}, \quad i \text{ odd} \quad (2.5-15)$$

where F^{-1} is the inverse frequency matrix of eqs. (2.5-11) and (2.5-12).

In general, an $n \times n$ frequency matrix must be inverted to evaluate the coefficients of a polynomial of degree $2n-1$ or $2n-2$. The frequencies should be chosen appropriately to give reasonable computational accuracy. The lowest frequency will give most information on a_0 and a_1 and should be chosen where the sensitivity of $L(j\omega)$ to these two coefficients is reasonably high—a change in either a_0 or a_1 should change $L(j\omega_1)$ significantly—and similarly for the remaining coefficients. A discussion of these sensitivities is given in Section 2.7. A reasonable choice of frequencies for many useful polynomials is to separate the frequencies (three for a quintic) by an octave or two on either side of the asymptotic cutoff frequency. A program for finding the coefficients of up to a quintic polynomial using this procedure is given in Appendix A.

2.6 POLYNOMIAL PERFORMANCE SPECIFICATIONS: THE SYNTHESIS PROBLEM

The discussion so far has concerned the analysis of polynomials, with application to the loss of feedback systems. In the example in Section 2.1 we found the amplifier to be unstable and took the expedient of adding a feedback

capacitor to stabilize it. We then analyzed the stabilized system. Do we need to “take what we get” and hope to stumble on a satisfactory solution? As we see throughout the remainder of this book, this is not the case. In this section we develop the characteristics of polynomials that give desirable performance for various applications. This is a special case of the more general approximation problem for equalizers and filters where a rational function—the ratio of two polynomials—may be called for. Here, we restrict ourselves to polynomials. We further restrict the discussion to polynomials of the low-pass type, in which the dc coefficient is not zero.

The field of satisfactory low-pass loss polynomials is represented here by three types plus an interpolation between two of them. The first are the Butterworth or maximally flat amplitude (MFA) polynomials; the second the Bessel or maximally flat delay (MFD) polynomials; and the third are the Chebyshev or equiripple polynomials. The interpolation yields a set of transitional polynomials intermediate between the MFA and MFD polynomials.

Butterworth Polynomials

Flat frequency response, by which is meant constancy of the *magnitude* of $L(j\omega)$, is prized in as diverse areas as loudspeakers and frequency-division multiplex telephone transmission systems. In both cases phase response is thought to be less important than flatness of the magnitude function, and for the same reason—the ear is relatively insensitive to phase distortion. The Butterworth polynomials (as well as the Chebyshev polynomials to be discussed later) provide an approximation to flatness within a given band up to the cutoff frequency, beyond which the loss rises at a rate dictated by the degree of the polynomial at $20n$ dB per decade (or $6n$ dB per octave). The squared magnitude of $L(j\omega)$ is given by

$$L(j\omega)^2 = 1 + \omega^{2n} \quad (2.6-1)$$

This function is seen to be unity at $\omega=0$, and its first derivative is

$$\frac{dL(j\omega)^2}{d\omega} = 2n\omega^{2n-1} \quad (2.6-2)$$

It is zero at $\omega=0$. Similarly, all derivatives up to the $(2n-1)$ st are zero at $\omega=0$, which is why (2.6-1) is termed a *maximally flat amplitude function*. The square of the loss magnitude at $\omega=1$ is equal to 2 for any value of n , so that $L(j1)=\sqrt{2}$, or 3 dB.

There are $2n$ roots of eq. (2.6-1) equally spaced on a unit circle, as shown for $n=4$ and $n=5$ in Fig. 2.17; there are no roots on the $j\omega$ axis for any n . Geometric considerations give the root locations as*

$$s_k = \exp\left(j \frac{2k-n+1}{n} \cdot \frac{\pi}{2}\right), \quad k = 1, 2, \dots, 2n \quad (2.6-3)$$

*For a more detailed treatment, see M. Van Valkenburg, *Introduction to Modern Network Synthesis*, Wiley, New York, 1960, p. 373 ff.

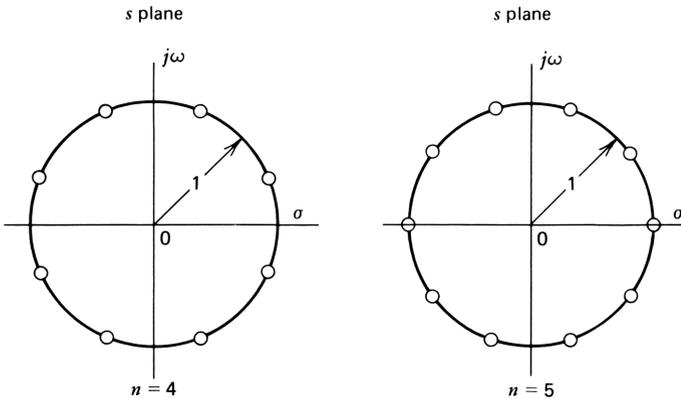


Figure 2.17. Root locations of Butterworth polynomials of fourth and fifth degrees.

It is easy to show that

$$L(s) = L(-s) \tag{2.6-4}$$

or that

$$L(s)^2 = L(s)L(-s) \tag{2.6-5}$$

[Note that magnitude signs are not needed on the right-hand side, since the phase of $L(s)L(-s)$ is zero.] Since we are interested in a stable loss polynomial, we simply associate the right half plane roots with $L(-s)$ and the left half plane roots with $L(s)$ and take the latter for the desired MFA polynomial. The polynomial coefficients are obtained by multiplying the factors corresponding to the n left half plane roots together. A calculator program that does this is given in Appendix A, and the first six MFA polynomials thus obtained are given in Table 2.1.

Bode plots for the first five MFA polynomials are given in Fig. 2.18. As shown previously, the loss at the asymptotic cutoff frequency is 3 dB for any n , and the cutoff slope rises with n . The phase curves show increasing phase distortion as n increases. These and most other phase curves to be presented have been modified to remove linear phase, or constant delay. This makes the departure from linear phase—phase distortion—clearer. Linear phase amounts to a simple time delay and does not change the relative phase of the signal components passing through the amplifier or system. When it is removed, as here, the higher-order delay terms—parabolic, cubic, and so on—are more easily seen. In terms of the normalized polynomial, we obtain

$$L(p) = 1 + b_1 p + b_2 p^2 + \dots + p^n \tag{2.6-6}$$

We plot a phase-reduced polynomial $L_r(p)$:

$$L_r(p) = (1 + b_1 p + b_2 p^2 + \dots + p^n) e^{-b_1 p} \tag{2.6-7}$$

Table 2.1 Normalized MFA, MFD, and Transitional Polynomials

MFA	
1	$1 + p$
2	$1 + 1.41p + p_2$
3	$1 + 2p + 2p^2 + p^3$
4	$1 + 2.61p + 3.41p^2 + 2.61p^3 + p^4$
5	$1 + 3.24p + 5.24p^2 + 5.24p^3 + 3.24p^4 + p^5$
6	$1 + 3.86p + 7.46p^2 + 9.14p^3 + 7.46p^4 + 3.86p^5 + p^6$
MFD	
1	$1 + p$
2	$1 + 1.73p + p^2$
3	$1 + 2.47p + 2.43p^2 + p^3$
4	$1 + 3.20p + 4.39p^2 + 3.12p^3 + p^4$
5	$1 + 3.94p + 6.89p^2 + 6.78p^3 + 3.81p^4 + p^5$
6	$1 + 4.67p + 9.92p^2 + 12.36p^3 + 9.62p^4 + 4.50p^5 + p^6$
Transitional ($m=0.5$)	
1	$1 + p$
2	$1 + 1.56p + p^2$
3	$1 + 2.22p + 2.20p^2 + p^3$
4	$1 + 2.89p + 3.87p^2 + 2.85p^3 + p^4$
5	$1 + 3.57p + 6.01p^2 + 5.96p^3 + 3.51p^4 + p^5$
6	$1 + 4.25p + 8.60p^2 + 10.63p^3 + 8.47p^4 + 4.17p^5 + p^6$

At low frequencies, the phase of the original polynomial is simply $\tan^{-1} b_1 \omega$. This is the value of the phase removed by the exponential, giving zero resulting phase at low frequencies.

Bessel Polynomials

The set of MFD polynomials may be derived from the following relationship for the (normalized) b coefficients⁵:

$$b_k = \frac{(2n-k)!}{(n-k)!k!} \cdot \left(\frac{n!}{2n!} \right)^{(n-k)/n} \quad (2.6-8)$$

where n is the degree of the polynomial. This equation was programmed on the calculator to obtain the normalized coefficients in Table 2.1. Bode diagrams are shown in Fig. 2.19, again with $b_1 \omega$ radians removed from the phase curve. This phase curve is seen to fall off monotonically with frequency, giving rise to

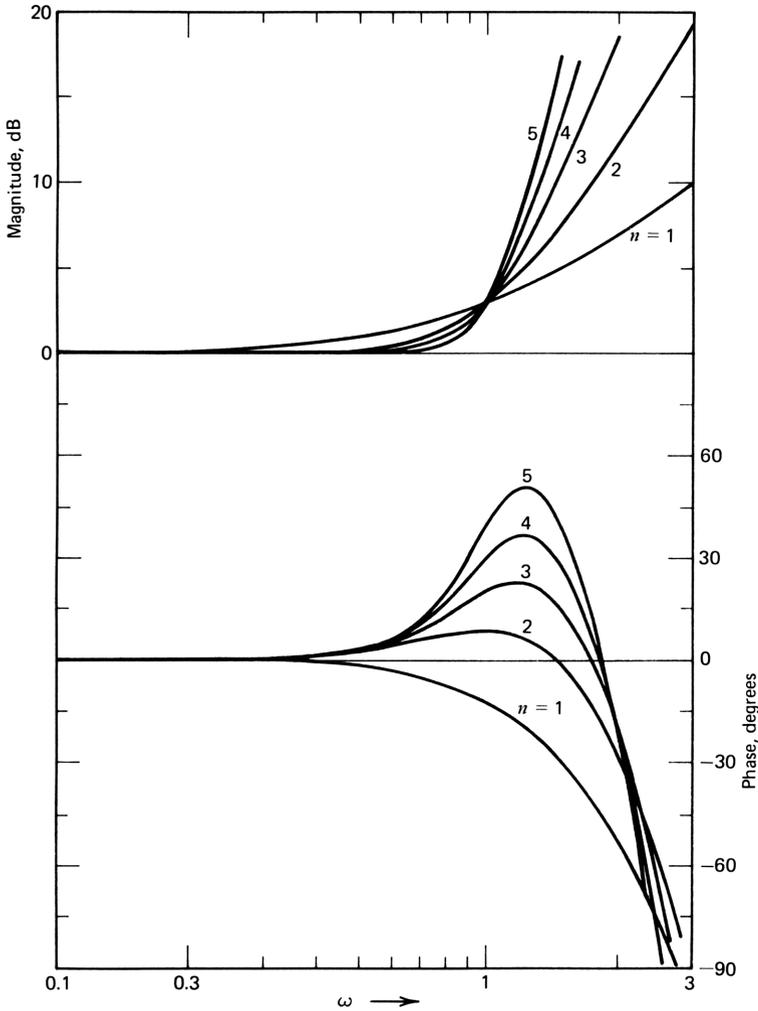


Figure 2.18. Bode diagrams of Butterworth polynomials of first to fifth degree. Flat delay (linear phase) has been removed from the phase curves.

the MFD designation. The delay provided is b_1 units of time and is flat up to the asymptotic cutoff frequency, above which the delay becomes less.

Transitional Polynomials

The MFA and MFD polynomials represent two possible low-pass-system or amplifier specifications, the first appropriate where flatness of loss magnitude is the prime consideration and the second, where flatness of delay is desired. It is convenient to have an intermediate performance specification for cases in which simultaneous requirements are placed on both frequency and transient response.

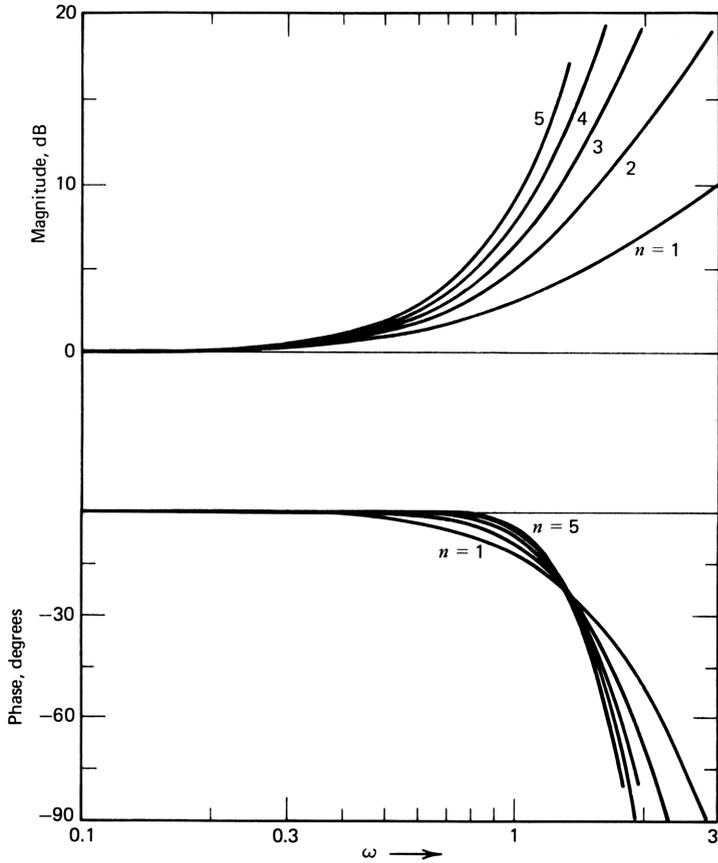


Figure 2.19. Bode diagrams of MFD polynomials with flat delay removed.

A set of *transitional polynomials*, intermediate between the MFA and MFD polynomials, can be generated as follows. Calling the transitional polynomial coefficients b_{kT} , and the MFA and MFD polynomial coefficients b_{kMFA} and b_{kMFD} , we set

$$b_{kT} = b_{kMFD}^m \cdot b_{kMFA}^{(1-m)} \quad (2.6-9)$$

where $0 < m < 1$ is the interpolation factor. For $m=0$, the response is MFA, and for $m=1$, the response is MFD. For $m=0.5$, the polynomial coefficients are the geometric mean between those of the MFA and MFD polynomials. The coefficients for this case are given in Table 2.1.* Bode diagrams are given in Fig. 2.20, again with $b_1\omega$ radians removed from the phase curve.

*The concept of *transitional polynomials* between MFA and MFD responses was introduced by Peless and Murakami.⁵ Their method differs from that given here in that they used the MFA and MFD polynomials in factored form and used the factor m to interpolate between sets of pole positions, interpolating the angle of the root linearly and the magnitude of the root geometrically. Their method gives negligibly different results but involves more computation.

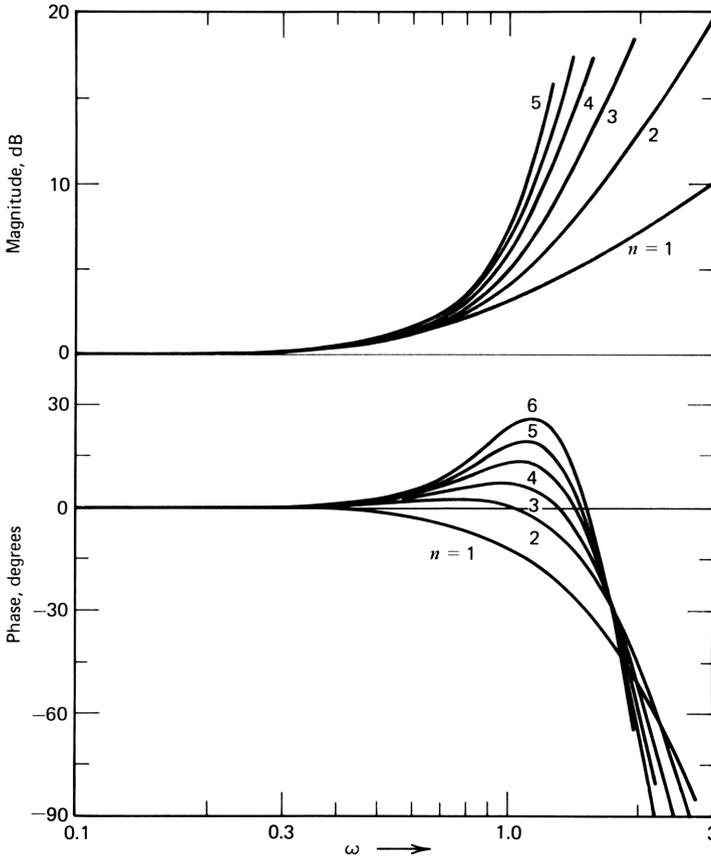


Figure 2.20. Bode diagrams of transitional polynomials with $m=0.5$; flat delay removed.

The step responses of the MFA, MFD, and transitional polynomials are compared in Fig. 2.21. These curves give the *output* time responses for a unit step *input*. By comparing the waveforms of the responses, we can judge which of the three polynomial types will be suitable for a given application; we may wish to use an interpolation factor different from $m=0.5$ given in Fig. 2.21c. The MFA responses exhibit considerable overshoot, over 8% for the third degree (not shown) and 12% for the fourth degree; the corresponding overshoot for the MFD responses is less than 1%. The MFD responses have the property of simultaneously minimizing rise time and bandwidth, a most useful property for systems where bandwidth limitations are needed to reduce noise.

Chebyshev Polynomials

Another set of polynomials that finds considerable use are the Chebyshev polynomials, which focus on providing an accurate match to the magnitude of the loss. By removing the MFA restriction on the derivatives and replacing it by the restriction that the response not deviate by more than a given amount

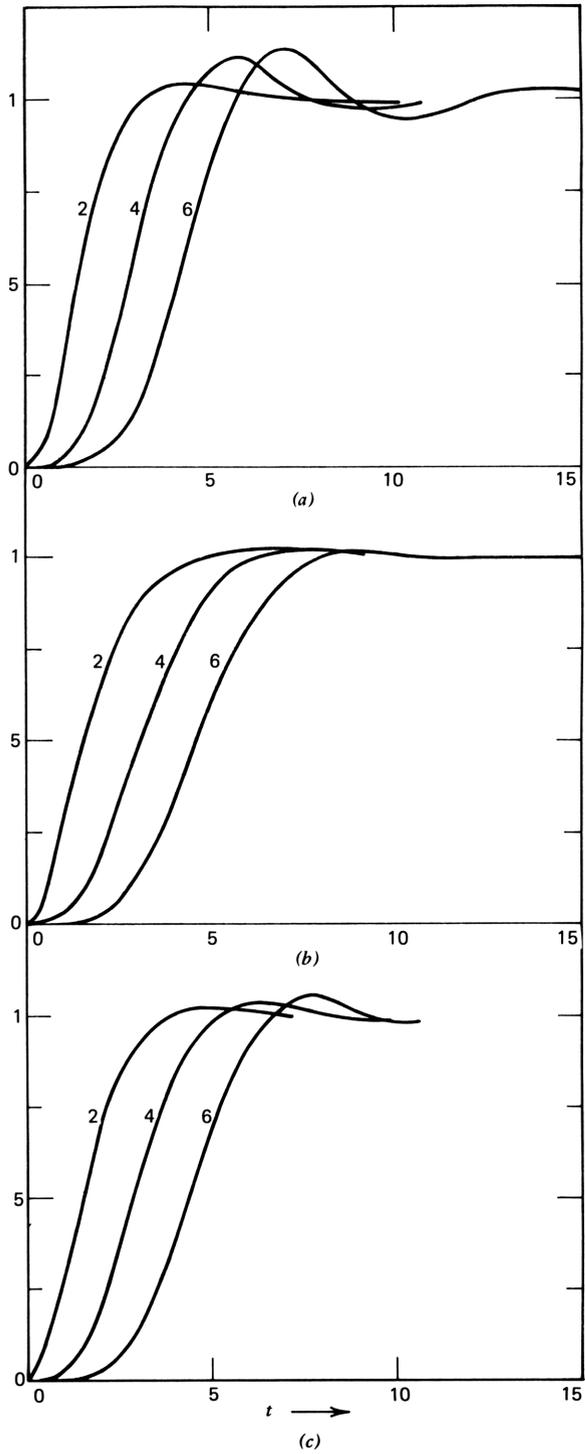


Figure 2.21. Step responses of Butterworth, MFD, and transitional polynomials of unit asymptotic bandwidth.

over a given bandwidth, a larger useful bandwidth may be obtained for a given asymptotic cutoff frequency. Furthermore, a sharper out-of-band cutoff slope can be obtained near the cutoff frequency. Chebyshev polynomials have found use in a wide range of systems where good transient performance is not important.

Where MFA and MFD polynomials are each specified by a single number—the degree of the polynomial (for a given frequency and amplitude normalization)—the Chebyshev polynomials require an additional parameter, the allowable in-band ripple, as illustrated in the magnitude response in Fig. 2.22. The frequency is scaled to the maximum frequency for which the magnitude response is within the specification. As for the MFD polynomials, we give a “cookbook” for rustling up Chebyshev polynomials. For the theory behind it, the reader should consult Van Valkenburg⁴ or Guillemin⁶ (both treatments are insightful).

The procedure for obtaining the Chebyshev polynomial of degree n for a given ripple amplitude r (in decibels) is to define a constant ϵ as follows:

$$\epsilon = (10^{r/10} - 1)^{1/2} \tag{2.6-10}$$

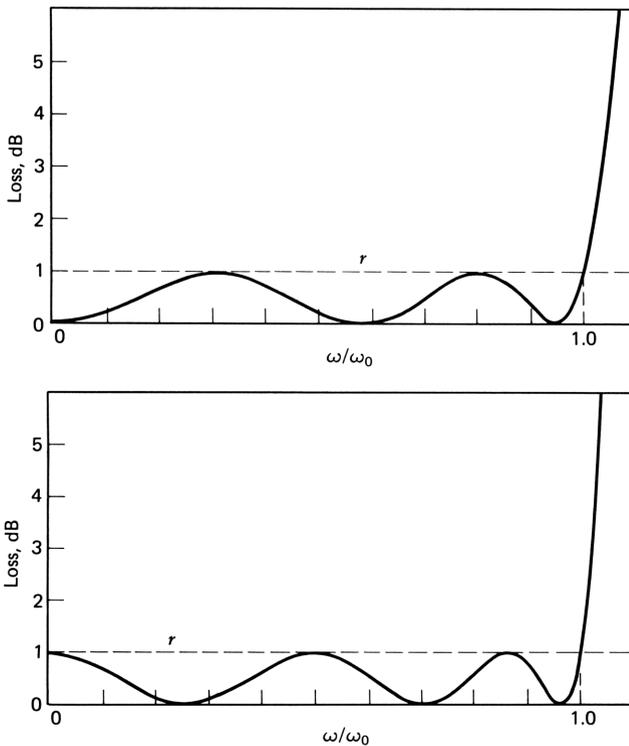


Figure 2.22. Loss (in dB) of Chebyshev polynomials of fifth and sixth degrees. Ripple width is 1 dB.

We then define the value of a second constant β for the polynomial from the relation

$$\beta = \frac{1}{n} \sinh^{-1} \frac{1}{\epsilon} \quad (2.6-11)$$

or alternatively, by trigonometric identity

$$\beta = \frac{1}{n} \ln \left(\frac{1}{\epsilon} + \sqrt{\frac{1}{\epsilon^2} + 1} \right) \quad (2.6-12)$$

Next, we find each root of the n th degree MFA polynomial, converting the root position into rectangular form. We then multiply the imaginary part of the MFA root by $\cosh \beta$ and the real part by $\sinh \beta$. This is the corresponding root of the Chebyshev polynomial whose maximum deviation from flat response is r decibels from unity up to a normalized frequency of unity. Note that the corresponding Butterworth loss polynomial has a magnitude of 1.414, or 3 dB at this (unity) frequency.

As an example, suppose that we desire a ripple amplitude of 0.2 dB up to a frequency of 1 Grad/s for a third-degree Chebyshev polynomial. From (2.6-10), $\epsilon = 0.217$, and from (2.6-11), $\beta = 0.744$, giving $\sinh \beta = 0.815$ and $\cosh \beta = 1.290$.

The cubic Butterworth polynomial has a real root at $\sigma = -1$. For the Chebyshev, this is multiplied by $\sinh \beta$, giving the real root at -0.815 . The real part of the Butterworth complex roots are at $\sigma = -0.5$, so the real parts of the Chebyshev complex roots are at $\sigma = -0.407$. The imaginary part of the Butterworth roots are at $\pm j0.866$; these are multiplied by $\cosh \beta$, giving the imaginary parts of the Chebyshev roots at $\pm j1.117$. With the root positions thus defined, we can multiply the factors to obtain the Chebyshev cubic polynomial with 0.2 dB ripple up to 1 Grad/s:

$$L_{CH}(s) = 1 + 1.804s + 1.415s^2 + 0.8683s^3 \quad (2.6-13)$$

A program that follows this procedure to obtain Chebyshev polynomials of any degree (up to seventeenth) and ripple (in decibels) is given in Appendix A.

The asymptotic bandwidth of this polynomial is slightly wider than that of the Butterworth polynomial used in its construction, by a factor of $(1/0.8684)^{1/3} = 1.048$. We could, of course, normalize the Chebyshev polynomial to unit asymptotic bandwidth:

$$L_{CH}(p) = 1 + 1.891p + 1.554p^2 + p^3 \quad (2.6-14)$$

Thus allowing us to place this polynomial in the diagram in Fig. 2.4. The polynomial in eq. (2.6-13), on the other hand, has known accuracy up to $\omega = 1$; since this was the reason for using the Chebyshev polynomial in the first place, this is the normalization to be used in comparing it with the Butterworth polynomial, for example.

If we renormalize the Butterworth polynomial such that it also has response within 0.2 dB up to unity frequency, it can be compared directly with the Chebyshev polynomial. For the Butterworth, we can solve eq. (2.6-1) to give the frequency at which the loss has increased by 0.2 dB:

$$\omega_{0.2} = (10^{0.2/10} - 1)^{1/2n}$$

For $n=3$, $\omega_{0.2}=0.601$; normalizing the Butterworth to this frequency, we obtain

$$L_B(p) = 1 + 1.202s + 0.7224s^2 + 0.2171s^3 \tag{2.6-15}$$

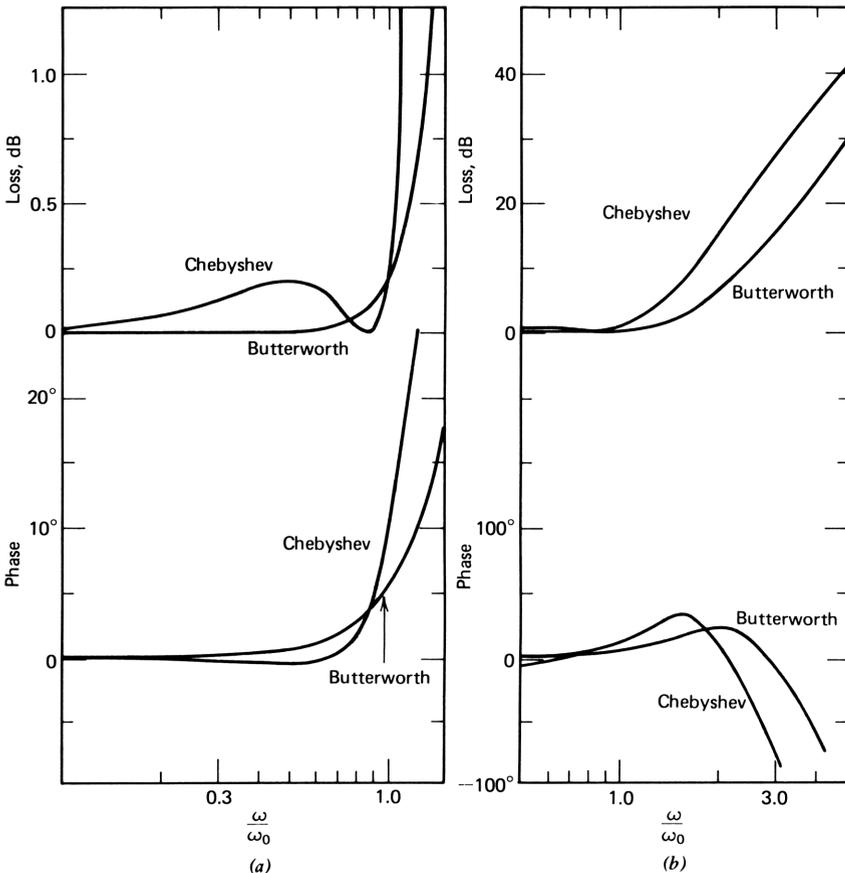


Figure 2.23. Bode diagrams comparing Chebyshev and Butterworth polynomials (each has a maximum in-band error of 0.2 dB): (a) in-band performance; (b) performance in cutoff region.

This can be compared directly with the polynomial in (2.6-13); both have 0.2 dB maximum magnitude error up to the normalized cutoff.

The loss of the Chebyshev and Butterworth functions, plotted in Fig. 2.23, show their relative performance in magnitude and phase both in band in Fig. 2.23*a* and in the cutoff region in Fig. 2.23*b*. The Chebyshev function exhibits its 0.2 dB error at $\omega=0.5$ and 1.0, whereas the Butterworth has an error of this amount only at 1.0. Up to $0.9 \omega_0$, the phase performance of the Chebyshev function is superior to the Butterworth. Out of band, the Chebyshev polynomial has 12 dB more loss (asymptotically) than the Butterworth; in general, for small ϵ , the Chebyshev out-of-band attenuation is $3(n-1)$ times that of the Butterworth. In the frequency domain the Chebyshev polynomial is superior to the Butterworth, in general.

The roots of the Chebyshev and Butterworth cubic polynomials are shown in Fig. 2.24, and the responses to a unit step are given in Fig. 2.25. Here, the Chebyshev polynomial does not do as well. The overshoot is somewhat worse, and the settling time is twice that of the Butterworth function, as seen in the expanded portion of the step response plot. The poorer settling time could have been predicted from the root diagram in Fig. 2.24: the real part of the complex Chebyshev roots is only half that of the Butterworth roots.

There is another aspect to the relative evaluation of polynomials as performance specifications for systems—sensitivity of the loss to the polynomial coefficients, as discussed in Section 2.7.

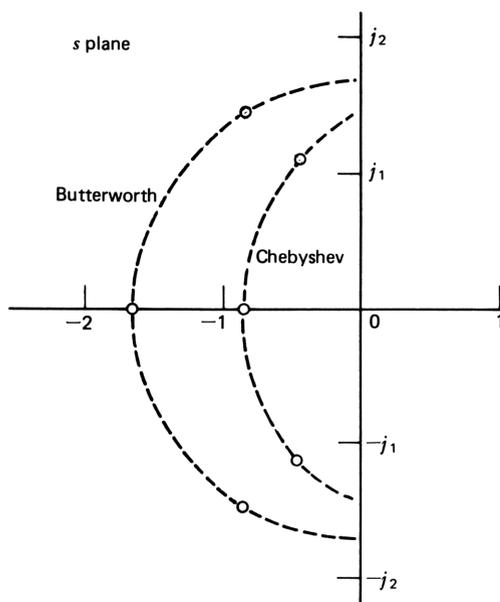


Figure 2.24. Root locations for the Chebyshev and Butterworth polynomials in Fig. 2.23.

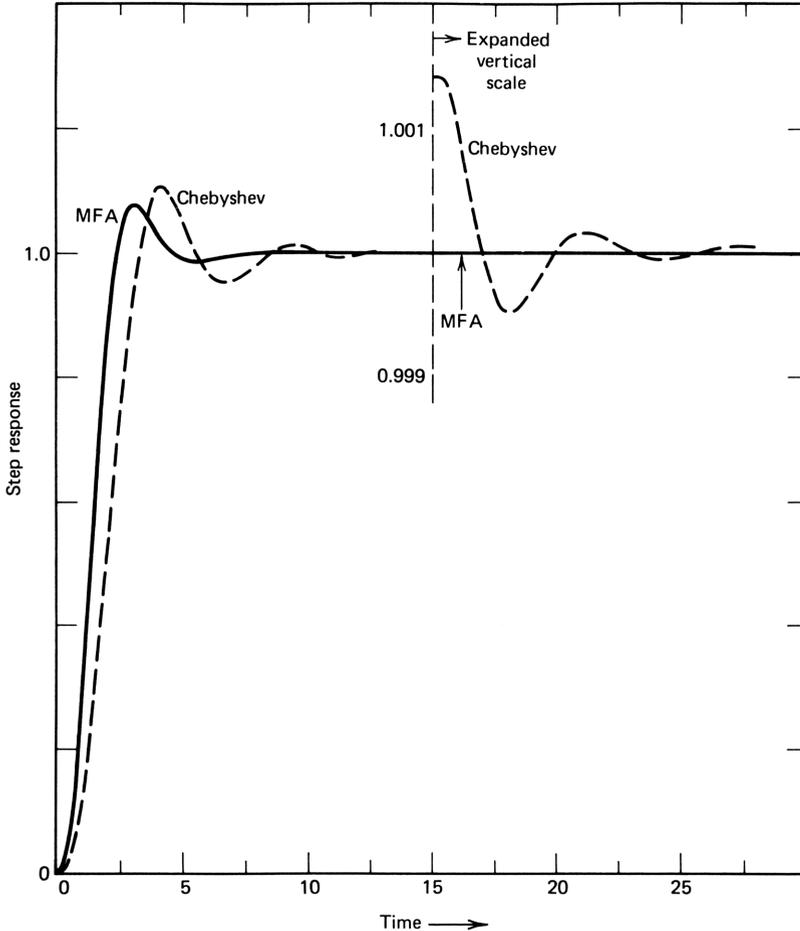


Figure 2.25. Comparison of step responses of Chebyshev and Butterworth polynomials. In the expanded-scale portion of the curves, the Butterworth error is too small to be seen.

2.7 SENSITIVITIES OF LOSS TO THE POLYNOMIAL COEFFICIENTS

We have discussed the relative merits of several polynomial types in their application as characteristic polynomials for feedback systems. Any of them, to be useful, must be realized by hardware whose characteristics may vary from the specified values. We should like to select a polynomial for realization that is insensitive to such variations, one that is “rugged.” We have seen, for example, that the Chebyshev series of polynomials are capable of realizing wide bandwidth when compared with the Butterworth polynomials, but we suspect that such polynomials may be somewhat “touchy” in requiring accurate “tuning.” To quantify the notions of ruggedness and touchiness, we can

define a measure that can be obtained with little more trouble than that involved in the evaluation of the polynomial in magnitude and phase.

The measure is the sensitivity of the loss to the value of each of the polynomial coefficients as a function of frequency. As noted in Chapter 1, sensitivity is a complex number, the *real* part of which is the sensitivity of the loss *magnitude* to the coefficient and the imaginary part of which is a measure of the sensitivity of the phase to the coefficient. Either (or both) can be used depending on which one is important in the application. As a measure of “touchiness,” we adopt the *magnitude of the sensitivity* as a function of frequency since this includes both sensitivities in a single measure.

From the sum rule, the sensitivity of loss to the value of coefficient a_i is given by

$$S_{a_i}^L = \frac{a_i \omega^i}{L(j\omega)} \quad (2.7-1)$$

This is the contribution of the i th term of the polynomial to the total. If we have evaluated $L(j\omega)$ at a given frequency by the method given in Section 2.4, it is a trivial matter to divide it into $a_i \omega^i$ to obtain the sensitivity of $L(j\omega)$ to a_i . As stated in the previous paragraph, we take the magnitude of the sensitivity as a convenient single measure; in some applications it may be desirable to emphasize magnitude or phase sensitivity, simply by converting magnitude and phase of (2.7-1) into rectangular form.

Figure 2.26 compares the loss and the sensitivities of loss to the polynomial coefficients for three standard cubic polynomials, the MFD, the Butterworth, and the Chebyshev with 0.2 dB ripple width. The normalization for these curves is different from that in the previous section. Here we give each of the polynomials a unit asymptotic bandwidth; Table 2.1 gives the polynomials for the MFD and Butterworth cases, and eq. (2.6-14) gives the polynomial for the Chebyshev case. This is the appropriate normalization when bandwidth performance is limited by the devices (transistors) used since the final asymptote will be essentially the same for all three cases.

Table 2.2 compares the performance of these three polynomials on the bases of magnitude and phase of response, sensitivities of loss to the polynomial coefficients, and settling times. In both magnitude and phase, the Chebyshev polynomial gives broader band performance than the other two by a wide margin. At values of ω/ω_0 below 0.6, it also has superior sensitivity performance. The peak sensitivities are higher, however, so we might expect the margin against instability to be poorer than the other two polynomials. Also, its settling times are longer, as noted previously, even with the different normalization used here.

Note that the conclusions here are for the simplified device model introduced in Section 2.1. When more accurate models are developed (later), delay becomes more important, and higher-degree polynomials arise in a more

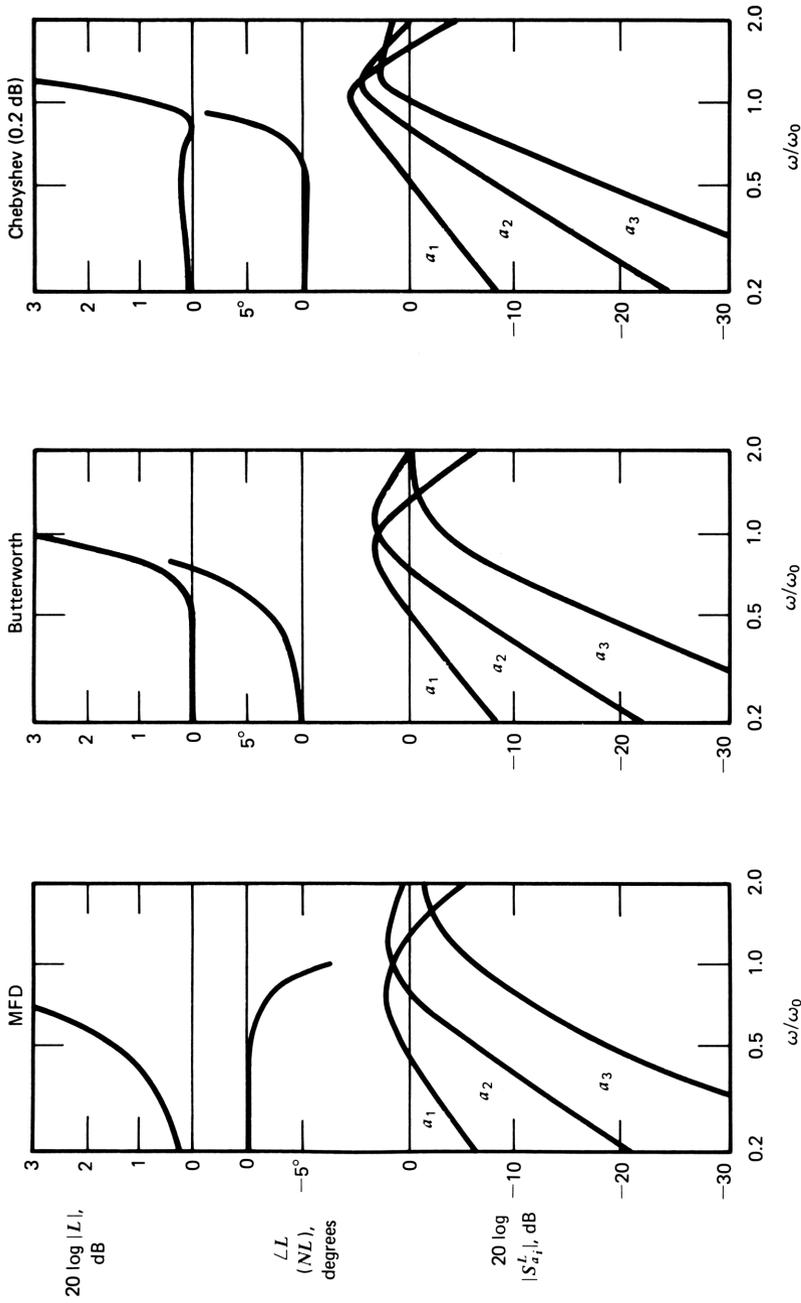


Figure 2.26. Comparison of MFD, Butterworth, and Chebyshev (0.2 dB ripple width) cubic polynomials for equal final asymptotic response. The upper two curves represent loss magnitude and phase; the lower three curves represent the sensitivity magnitudes of loss to a_1 , a_2 , and a_3 .

Table 2.2 Comparison of Cubic MFD, Butterworth, and Chebyshev (0.2 dB Ripple) Polynomials (Same Asymptotic Cutoff)

	MFD	Butterworth	Chebyshev	Unit
Response ± 0.1 dB	0.20	0.60	0.94	ω/ω_0
Phase $\pm 1.0^\circ$	0.65	0.25	0.70	ω/ω_0
Sensitivities ^a				
a_1	0.3	0.6	-0.6	dB
a_2	-6.0	-6.0	-8.5	dB
a_3	-19.0	-18.0	-18.5	dB
Peak sensitivities				
a_1	2.0	3.2	5.2	dB
a_2	2.0	3.2	4.2	dB
a_3	0	0	2.3	dB
Settling times ^b				
To 1%	6	10	11.	ns
To 0.01%	13	18	23	ns

^aAt $\omega/\omega_0=0.5$.^bFor $\omega_0=1$ Grad/s.

complete description. Although this complicates the stability question somewhat, the principle of direct realization of polynomial coefficients still applies.

In Chapter 3 we see how to investigate the sensitivity of loss to each of the components and devices of a feedback system, including the effect of the choice of the loss polynomial.

PROBLEMS

- Find a polynomial expression for the amplifier in Fig. 2.1 in which a capacitor C_F is connected in parallel with the feedback conductance G_F . In finding the current through C_F , ignore the input voltage of the first transistor, and also ignore the shunt loading of C_F on the output stage.
- Suppose that the load conductance of the amplifier in Fig. 2.1 is equal to the reciprocal of the common lead resistance of the third-stage device. What is the input voltage of the third stage for an output voltage of 1.0 V? What is the effect on the input voltage and current of the third stage of paralleling it with n identical transistors? (This is similar in effect to using an output-stage device n times larger than the earlier stages.)
- Given the following polynomials, find (1) whether they are stable, (2) the damping factor of any quadratic roots, and (3) the asymptotic cutoff frequency. Normalize each polynomial.
 - $L(s)=0.1+2s+20s^2+100s^3$
 - $L(s)=0.01+0.4s+20s^2+1250s^3$
 - $L(s)=0.1+0.05s+210s^2+100s^3$

- 4 Sketch a Bode diagram of the loss magnitude and phase of the amplifier of Fig. 2.1 with 0.3 pF capacitor in parallel with G_F .
- 5 Find the settling time to within 100 ppm for the amplifier in Section 2.1 if the feedback capacitance is made 0.2 pF.
- 6 Find the approximate settling time to within 0.01 percent for the amplifier of Section 2.1 with $C_F=0.30$ pF with a load capacitance of 10 pF. Note that this requires you to solve a quartic polynomial for its roots.
- 7 Find the step response for the loss polynomial

$$L(s) = 1 + 2s + 2s^2 + s^3$$

for values of time from zero to 15 ns (with s given in gigaradians per second).

- 8 A doublet is a closely spaced pole and zero. If a doublet is included in the loss of an amplifier, the effect on the step response is to introduce a decaying exponential with a time constant equal to the reciprocal of the doublet frequency. The residue in the pole is small because of the cancellation introduced by the zero. Estimate the effect on the settling time to 0.01% in Problem 7 if the loss is multiplied by the doublet $(1 + 100s)/(1 + 101s)$. (Develop an expression for the time response of the doublet itself, ignoring the other roots of the loss, and find its settling time.)
- 9 Sketch Nyquist diagrams of the amplifier of Section 2.1 in which capacitor C_F is connected in parallel with G_F and where $C_F=0.9$ pF; repeat for $C_F=1.1$ pF. Sketch Bode plots for the two cases.
- 10 The amplifier in Fig. 2.1 with circuit values given in Section 2.1 has an unknown value of feedback capacitance connected in parallel with G_F . The amplifier is measured and found to have the following loss and phase at 0.05 and 0.1 Grad/s:

$$L(j.05) = -38.77 \text{ dB} / \underline{49.4^\circ}$$

$$L(j0.1) = -40.00 \text{ dB} / \underline{90^\circ}$$

Find the value of the feedback capacitance.

- 11 The active path of an amplifier is known to have cubic polynomial response. It is measured at two frequencies: $\omega_1=0.05$, $\omega_2=0.1$ Grad/s, with the following result:

$$L(j.05) = 51.1 \text{ dB} / \underline{153.4^\circ}$$

$$L(j0.1) = -39.0 \text{ dB} / \underline{206.6^\circ}$$

Feedback is to be added to this amplifier with loss equal to -40 dB, constant with frequency. Will the resulting amplifier be stable?

- 12 An amplifier is to be designed to have -30 dB of loss and an amplitude characteristic that is not to deviate from flat loss by more than 0.05 dB up to a frequency of 1.0 Grad/s. Assuming a design roughly like that of Fig. 2.1 in which all three transistors have the same time constant and that bandwidth is controlled by the cubic coefficient $R_G\tau_1\tau_2\tau_3G_L$, find the required values of the transistor time constant if the amplifier is realized by (1) MFA, (2) MFD, and (3) Chebyshev polynomials.

REFERENCES

- 1 K. E. Atkinson, *An Introduction to Numerical Analysis*, Wiley, New York, 1978.
- 2 S. D. Conte and C. de Boor, *Elementary Numerical Analysis: An Algorithmic Approach*, McGraw-Hill, New York, 1972.
- 3 F. F. Kuo, *Network Analysis and Synthesis*, Wiley, New York, 1966.
- 4 M. F. Van Valkenburg, *Introduction to Modern Network Synthesis*, Wiley, New York, 1960.
- 5 Y. Peless and T. Murakami, "Analysis and Synthesis of Transitional Butterworth-Thomson Filters and Bandpass Amplifiers," *RCA Rev.*, 60–94 (March 1957).
- 6 E. A. Guillemin, *Synthesis of Passive Networks*, Wiley, New York, 1957.

Chapter 3

Elements of Feedback Synthesis: A Case Study

Several basic aspects of the new theory are explained in this chapter through a case study concerning the design of a feedback amplifier. The longer-term goal is to enable us to choose a feedback configuration—a circuit configuration—that is uniquely appropriate for a given application and to design it to meet a set of performance specifications. The goal is too general to realize fully, but it should be kept in mind as we try to come close. Several concepts are introduced in this chapter related to feedback synthesis, analysis, and performance variability. The case study provides us with a framework or paradigm that will help to make these concepts intuitive.

3.1 AN AMPLIFIER WITH BUTTERWORTH RESPONSE

We saw in Section 2.1 how the loss polynomial of the simple model used there could be written by inspection by starting at the amplifier output and proceeding to add terms to the voltage and currents as one proceeded toward the

input. The loss expression, from eqs. (2.1-7) and (2.2-27), is

$$\begin{aligned} L(s) &= -(R_G G_F + R_G C_F s + r_1 \tau_2 \tau_3 G_L s^2 + R_G \tau_1 \tau_2 \tau_3 G_L s^3) \\ &= a_0 + a_1 s + a_2 s^2 + a_3 s^3 \\ &= a_0 \left(1 + \frac{a_1}{a_0} s + \frac{a_2}{a_0} s^2 + \frac{a_3}{a_0} s^3 \right) \end{aligned} \quad (3.1-1)$$

Suppose now that we wish the amplifier voltage loss to be given by one of the polynomials in Section 2.6. Then we might write the polynomial

$$L(s) = a_0 \left(1 + \frac{b_1}{\omega_0} s + \frac{b_2}{\omega_0^2} s^2 + \frac{b_3}{\omega_0^3} s^3 \right) \quad (3.1-2)$$

To obtain Butterworth response, we would make $b_1 = b_2 = 2$ and $b_3 = 1$. To synthesize the desired polynomial, we equate coefficients of like degree of eqs. (3.1-1) and (3.1-2) within a set of constraints on the components and device parameters that might take various forms.

Suppose, for example, that the desired polynomial is specified completely, including the dc loss a_0 and the asymptotic bandwidth ω_0 . Equating the cubic coefficients, we can write

$$\frac{a_0 b_3}{\omega_0^3} = a_3 = -R_G \tau_1 \tau_2 \tau_3 G_L \quad (3.1-3)$$

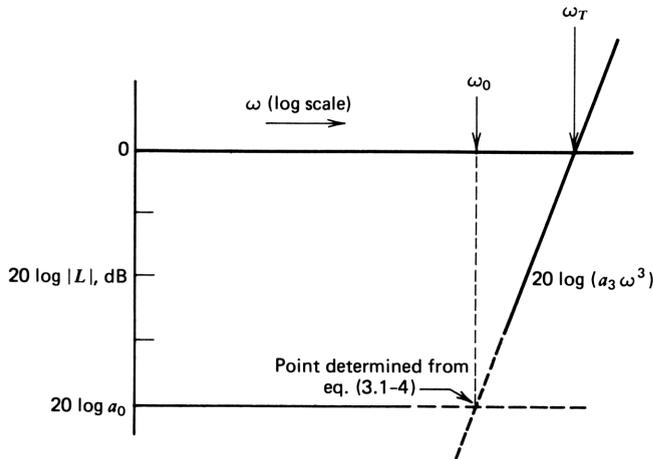


Figure 3.1. Bode diagram for the proposed amplifier showing the relation of the final asymptotic loss to the dc loss; b_3 is unity. $\omega_0 = (a_0 / R_G \tau_1 \tau_2 \tau_3 G_L)^{1/3}$; $\omega_T = (1 / R_G \tau_1 \tau_2 \tau_3 G_L)^{1/3}$.

If terminating immittances R_G and G_L are also specified, the product of the transistor time constants must be selected to give the desired bandwidth. Usually, however, the transistor time constants are known at the outset (at least approximately) so that it is more appropriate to find what dc loss can be achieved for a given set of transistors:

$$a_0 = \frac{a_3 \omega_0^3}{b_3} = - \frac{R_G \tau_1 \tau_2 \tau_3 G_L \omega_0^3}{b_3} \quad (3.1-4)$$

This constraint between a_0 and a_3 is shown graphically in Fig. 3.1. In the following case study we assume that the transistors are given and that the dc loss is to be found using (3.1-4).

To realize Butterworth response, we must set the remaining three coefficients of (3.1-1) equal to those of like degree of (3.1-2), so that $a_3 = a_0 b_3 / \omega_0^3$, or

$$a_0 - \frac{a_3 \omega_0^3}{b_3} = 0 \quad (3.1-5)$$

$$a_1 - \frac{b_1 a_3 \omega_0^2}{b_3} = 0 \quad (3.1-6)$$

$$a_2 - \frac{b_2 a_3 \omega_0}{b_3} = 0 \quad (3.1-7)$$

These four equations—including (3.1-4)—are the basic synthesis equations for designing a feedback amplifier with a prescribed cubic loss ratio. For the circuit in question, these equations can be written in terms of the device parameters and components:

$$-R_G G_F + \frac{R_G \tau_1 \tau_2 \tau_3 G_L \omega_0^3}{b_3} = 0 \quad (3.1-8)$$

$$-R_G C_F + \frac{b_1 R_G \tau_1 \tau_2 \tau_3 G_L \omega_0^2}{b_3} = 0 \quad (3.1-9)$$

$$-r_1 \tau_2 \tau_3 G_L + \frac{b_2 R_G \tau_1 \tau_2 \tau_3 G_L \omega_0}{b_3} = 0 \quad (3.1-10)$$

The second term in each equation is known at the outset since the desired polynomial, the source and load immittances, and the transistor time constants are all given. To synthesize the polynomial, at least one free variable must remain for each coefficient. Terms G_F and C_F are the free variables for a_0 and a_1 , respectively; we must take r_1 as the free variable for a_2 since the other

parameters are specified. Thus the values are

$$G_F = \frac{\tau_1 \tau_2 \tau_3 G_L \omega_0^3}{b_3} \quad (3.1-11)$$

$$C_F = \frac{b_1 \tau_1 \tau_2 \tau_3 G_L \omega_0^2}{b_3} \quad (3.1-12)$$

$$r_1 = \frac{b_2 R_G \tau_1 \omega_0}{b_3} \quad (3.1-13)$$

For purposes of later comparisons, we use the set of numbers given in Section 2.1 for the device parameters and components:

$$\tau_1 = \tau_2 = \tau_3 = 1.0 \text{ ns}$$

$$R_G = 1.0 \text{ k}\Omega$$

$$G_L = 10 \text{ mS}$$

$$\omega_0 = 0.1 \text{ Grad/s}$$

For Butterworth response, we then obtain $G_F = 0.01 \text{ mS}$, $C_F = 0.2 \text{ pF}$, and $r_1 = 0.20 \text{ k}\Omega$.

To obtain a_2 in the preceding design, we took r_1 as a free variable since there were no others. In practice, we may wish to specify r_1 independently (e.g., for best noise performance; thus we might add a component that will give us control of the a_2 parameter for fixed r_1). We might, for example, add a conductance G_3 between the input of the third stage and the common lead (ground) as shown in Fig. 3.2. What is its effect on the loss of the amplifier? Since the system is linear, we can use superposition to analyze the change in loss due to the addition.

The input voltage to the third stage for unit output voltage is $-r_3 G_L$, so the current through G_3 is $-G_3 r_3 G_L$, a dc (constant) term. The second stage in supplying this current has an added input current of $\tau_2 s$ times this amount. The input current of the first stage is $-\tau_1 s$ times the second-stage input current, and the first-stage input voltage is $-r_1$ times that, giving

$$\Delta L(s)|_{G_3} = -r_1 \tau_2 G_3 r_3 G_L s - R_G \tau_1 \tau_2 G_3 r_3 G_L s^2 \quad (3.1-14)$$

By adding this *change* in loss to that given in (3.1-4), we obtain the loss polynomial of the modified amplifier. Its coefficients are

$$a_0 = -R_G G_F \quad (3.1-15)$$

$$a_1 = -R_G C_F - r_1 \tau_2 G_3 r_3 G_L \quad (3.1-16)$$

$$a_2 = -r_1 \tau_2 \tau_3 G_L - R_G \tau_1 \tau_2 G_3 r_3 G_L \quad (3.1-17)$$

$$a_3 = -R_G \tau_1 \tau_2 \tau_3 G_L \quad (3.1-18)$$

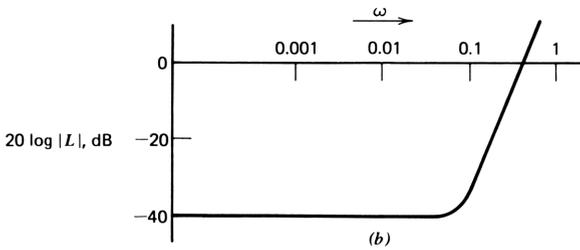
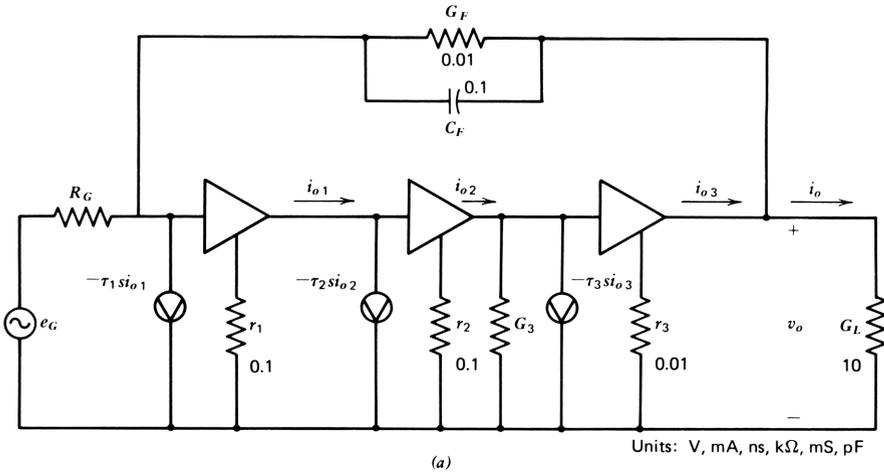


Figure 3.2. An amplifier capable of providing cubic Butterworth response: (a) circuit diagram; (b) loss magnitude versus frequency.

These four equations constitute a complete set of *analysis* equations for the amplifier in Fig. 3.2. By substituting these coefficient values in the synthesis equations (3.1-5) to (3.1-7), we can solve for the values of the free variables G_F , C_F , and G_3 ; thus

$$G_3 = \left(\frac{b_2 \omega_0}{b_3} - \frac{r_1}{R_G \tau_1} \right) \frac{\tau_3}{r_3} \tag{3.1-19}$$

$$G_F = \frac{\tau_1 \tau_2 \tau_3 G_L \omega_0^3}{b_3} \tag{3.1-20}$$

as before, and

$$C_F = \left(\frac{b_1 \tau_1 \tau_3 \omega_0^2}{b_3} - \frac{r_1 G_3 r_3}{R_G} \right) \tau_2 G_L \tag{3.1-21}$$

Note that for a sufficiently high value of r_1 , G_3 may assume a negative value. This is because r_1 causes the a_2 coefficient to be higher than the desired value. In this case r_1 should be reduced (e.g., by increasing the dc current in the first stage for a bipolar transistor). Alternatively, the amplifier may be characterized as a quadratic design; the performance specification polynomial can be changed to reflect this by reducing b_3 , for example, or by increasing ω_0 . (With b_3 reduced to zero, the loss is a quadratic polynomial. Small values of b_3 approximate this.)

Equations (3.1-19) to (3.1-21) constitute a set of design equations for the amplifier, giving the feedback elements explicitly in terms of the other device and component values and the desired polynomial. The design is for an amplifier of given bandwidth for devices that have a given time constant. In this design a_3 contains only the time constants and the source and load immittances, so that the design is of minimum loss (maximum gain) for the devices used and for the polynomial chosen.

With the circuit values given in Fig. 3.2, with $b_1=b_2=2$ for the cubic Butterworth, and with $\omega_0=0.1$ Grad/s, these equations give

$$G_3 = 10.0 \text{ mS}$$

$$C_F = 0.1 \text{ pF}$$

$$G_F = 0.01 \text{ mS}$$

In the discussion that follows we refer to this amplifier design as *Design A*. A Bode diagram for the loss as a function of frequency is shown in Fig. 3.2*b*. The loss polynomial for this design is given by (3.1-1), for which we can now fill in the numbers:

$$L(s)_{\text{Design A}} = 0.01 + 0.2s + 2s^2 + 10s^3 \quad (3.1-22)$$

The design of this feedback amplifier has involved working primarily with the specified polynomial and scaling it properly. Problem 1 deals with the same material for a Chebyshev design with the same initial specifications on bandwidth and on the devices to be used.

The key to the design was to provide means for the independent control of a_2 since we already had control of the other three coefficients. What were the means, exactly? We were seeking a signal that was an undifferentiated version of the output signal so that the cascaded combination of the first two stages would give the desired s^2 term. The input voltage of the third stage is just such a voltage; G_3 converts it to a current to allow the first two stages to work on it. The common lead resistor of the third stage r_3 makes the output current *observable* as a voltage at the input of the third stage. *Control* of a_2 is thereby exercised by selecting G_3 .

If r_3 is nonlinear, or variable, the preceding method is not a good way to secure the desired response. For a nonlinear r_3 , the input voltage of the third

stage will be nonlinearly related to the output, and this nonlinearity will be converted to a current nonlinearity by G_3 and will be passed through to the amplifier input as a required predistortion to obtain an undistorted output, as discussed in Chapter 1. If r_3 varies, the a_1 and a_2 coefficients will vary, reducing response accuracy. A further objection to this method is that the response is load dependent since the part of the input signal arising from G_F and C_F are output *voltage* related and are not dependent on G_L , whereas the parts of the amplifier input signal related to G_3 are directly dependent on G_L .

A better method of providing the desired response is by use of local feedback; one such method is shown in Fig. 3.3, in which a feedback conductance is connected from output to input of the third stage. To see how this works, we make two approximations initially; we remove them later. We assume that the feedback conductance G_2 is negligible compared with the load and that the input voltage of the third stage is negligible compared with its output voltage. With these assumptions, we can immediately write the equation for the change in loss due to G_2 ; the increase in third-stage input current is $-G_2v_0$, leading to an increase of amplifier input current of $-\tau_1\tau_2G_2s^2$ and an increase in input voltage of $-r_1\tau_2G_2s$, so that

$$\Delta L|_{G_2} = -(r_1\tau_2G_2s + R_G\tau_1\tau_2G_2s^2) \tag{3.1-23}$$

For this approximate analysis, the form of the change in loss due to G_2 is the same as that for G_3 , where G_2 replaces $G_2r_3G_L$ in eq. (3.1-14). Since the G_3 design gave $G_3 = 10$ mS, $G_L = 10$ mS, and $r_3 = 0.01$ k Ω , we obtain $G_2 = 1.0$ mS for this design. Next, we turn to an exact analysis of the third stage with G_2 feedback. The amplifier with local feedback around the third stage is called *Design B*.

3.2 FEEDBACK OVER AN INDIVIDUAL STAGE

In this section we take a more detailed view of the third stage introduced in Fig. 3.3. We make an exact analysis of the stage using the simple device model to introduce and define four separate concepts concerning feedback. Later we

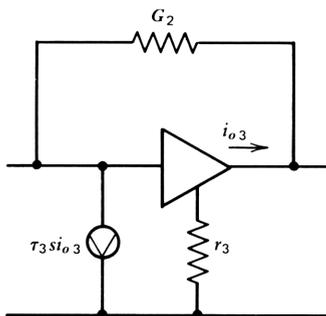


Figure 3.3. An improved third stage for the amplifier shown in Fig. 3.2.

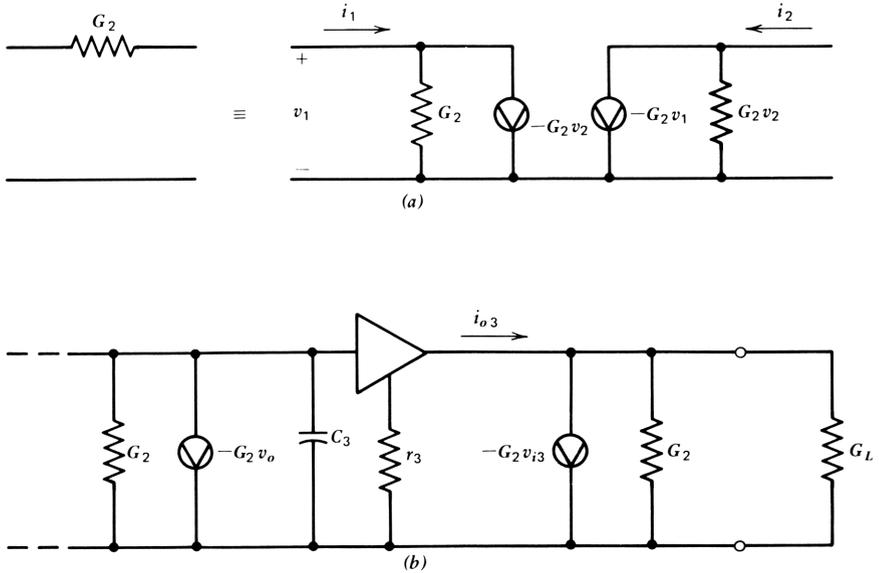


Figure 3.4. Individual feedback stage: (a) an equivalent circuit for the feedback resistor; (b) an equivalent ladder circuit for the stage.

see how this can be applied in general to feedback structures; our purpose here is to show how these four concepts arise in a simple case. The key to defining the four concepts is to describe the feedback network—in this case conductance G_2 —by its two-port parameters. In Fig. 3.4a the feedback conductance is represented by an equivalent network containing four elements—two current generators and two conductances. On first glance this substitution seems an unlikely way to simplify the analysis since we now have four elements rather than one to deal with. As we see later, however, this separation will allow us to define four distinct aspects of feedback.

The feedback resistor in isolation can be represented by its two-port y parameters, defined in the equations

$$\begin{aligned} i_1 &= y_{11}v_1 + y_{12}v_2 \\ i_2 &= y_{21}v_1 + y_{22}v_2 \end{aligned} \tag{3.2-1}$$

In matrix form this is

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{3.2-2}$$

or, stating it more compactly,

$$i_1 = y_{ik}v_k, \quad i = 1, 2; \quad k = 1, 2 \tag{3.2-3}$$

If the conductance of the feedback resistor is G_2 , direct analysis of the circuit gives

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} G_2 & -G_2 \\ -G_2 & G_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{3.2-4}$$

This is the set of equations represented by the equivalent circuit in Fig. 3.4a. One advantage of this representation of the feedback resistor is that it allows us to represent the amplifier in Fig. 3.4a by an *equivalent ladder circuit*, shown in Fig. 3.4b.

The equivalent ladder can be analyzed by starting at the output and working toward the left to the input. The currents and voltages are shown in Fig. 3.4. The advantage of working in this direction is that the analysis proceeds sequentially. With one exception—the G_2v_i generator connected across the output of the device—we never encounter an undefined variable, one whose value has not yet been found. What is more, the one exception can usually be ignored as negligible; when v_i is found, it can be evaluated, and any required correction can be made. In this simple case there is no need for this expedient; we do an exact analysis.

The output current of the device in Fig. 3.4b is given by inspection as

$$i_{o3} = (G_2 + G_L)v_o - G_2v_i \tag{3.2-5}$$

This current flows through the common lead resistance r_3 of the device, giving an input voltage of $-r_3$ times this current; thus

$$v_i = -r_3(G_2 + G_L)v_o + r_3G_2v_i \tag{3.2-6}$$

This can be written

$$v_i = -\frac{r_3(G_2 + G_L)}{1 - G_2r_3}v_o \tag{3.2-7}$$

The input current is comprised of the current through the shunt admittances plus the current contributed by the dependent generator $-G_2v_o$:

$$i_i = (G_2 + C_3s)v_i - G_2v_o \tag{3.2-8}$$

or

$$i_i = -\left(\begin{array}{c} \text{Output loading} \quad \text{Input loading} \\ \downarrow \quad \uparrow \\ \frac{(G_2 + G_L)r_3(G_2 + C_3s)}{1 - G_2r_3} + G_2 \end{array} \right) v_o \tag{3.2-9}$$

Direct feedthrough
Feedback

Equations (3.2-7) and (3.2-9) give a complete description of the linear properties of the device model with feedback. The latter equation can be written

$$i_i = - \left(\frac{1+r_3G_L}{1-r_3G_F} G_F - \frac{G_F+G_L}{1-r_3G_F} \tau_3 s \right) v_o \quad (3.2-10)$$

This equation gives i_i as a binomial in the frequency variable. In this form it is easily incorporated into the equations for the design of the amplifier in Fig. 3.3.

Representing the feedback resistor by its two-port elements—two conductances and two dependent generators—nicely separates the four effects of applying feedback mentioned at the beginning of this section. The conductances represent *input* and *output loading*, and the generators represent signal transfer through the feedback resistor in either direction. In particular, we can identify $y_{11}=G_2$ as the *input loading*, a shunting of the input due to the feedback resistor. Similarly, $y_{22}=G_2$ is the *output loading*. The purpose in applying feedback in the first place is $y_{12}=-G_2$; the G_2v_o generator at the input *augments* the input *current* by an amount proportional to the output *voltage*. Where there is distortion, for example, it is this generator that swamps it out, in the sense of the discussion of Section 1.2. Finally, $y_{21}=G_2$ gives *direct feedthrough* of the input voltage signal, converted to a current by the feedback resistor and added to the current to be supplied by the transistor. As noted, this is usually small. Therefore, feedback conductance G_2 has four roles in the operation of the stage. These are indicated by the arrows in eq. (3.2-9).

The sizes of the four effects are the key to understanding the feedback process. Where feedback is applied to achieve a particular object, one usually thinks (in terms of the preceding example) of what y_{12} will do. It is important to know how the other three aspects affect the outcome of applying feedback. To this end, we evaluate the ratio of input current to output voltage as we add in the four effects of G_2 to the stage represented in Fig. 3.4. The stage response is a binomial in the frequency variable, a_0+a_1s . In Table 3.1 we give the binomial coefficients first for $G_2=0$; in the next four rows we successively add in the effects of y_{12} , y_{11} , y_{22} , and y_{21} . The last row gives the exact response. For the case studied, the input augmentation or feedback accounts for 90% of both a_0 and a_1 ; input loading accounts for about 10% of a_0 ; output loading for about 10% of a_1 and direct feedthrough for about 0.1% of a_0 and 1% of a_1 .

Table 3.1 Successive Approximations to Loss of Eq. (3.2-9)

Approximation	y_{11}	y_{12}	y_{21}	y_{22}	a_0	a_1
No feedback	0	0	0	0	0	1.0
Add input augmentation	0	-1	0	0	1.0	1.0
Add input loading	1	-1	0	0	1.1	1.0
Add output loading	1	-1	0	1	1.11	1.1
Add direct feedthrough	1	-1	-1	1	1.111	1.111

The description of feedback as a combination of four distinct effects corresponding to the four two-port parameters of the feedback network will allow us to develop a general method for the analysis and classification of feedback structures; we discuss this later. Note that the outer feedback network G_F, C_F of either design can be treated in exactly the same way as the *local* feedback around the third stage, with feedback admittance $Y_F = G_F + C_F s$ of the outer loop represented by its four y parameters.

3.3 SYNTHESIS OF DESIGN B

We can find the polynomial coefficients of the amplifier in Fig. 3.5a by straightforward circuit analysis. To do this, we use the equivalent ladder circuit in Fig. 3.5b. The equivalent ladder is drawn just as for the single stage (which is incorporated into it). For the three-stage amplifier, however, the outer feedback circuit is also expressed in terms of its y parameters. The input and output loading ($y_{11} = y_{22} = Y_F$) are placed in shunt with the input and output of the amplifier. The two generators corresponding to $y_{12} = y_{21} = -Y_F$ are also connected at input and output.

To avoid cluttering the equations with nonessentials, we ignore the effects of all but y_{12} for the outer feedback path. When we ignore y_{11} , we ignore G_F of about 0.01 mS relative to the source conductance of 1 mS and C_F of about 0.1 pF relative to $C_1 = 10$ pF; both are 1% corrections. When we ignore y_{22} , y_F is compared with the load conductance, which is 10 times the source conductance, so that the correction is even less. If we ignore y_{22} , we can certainly ignore y_{21} since the current through it is less than that through y_{22} by the ratio v_{i1}/v_o , the voltage loss of the amplifier active path. Hence, the approximation is a good one. When the equations are used by the computer, it is easy to incorporate these small effects.

Circuit analysis to find the polynomial coefficients proceeds as in Section 2.1, starting at the output and proceeding toward the input. With the simplifying assumptions made in the previous paragraph—the same as those made implicitly in the earlier analysis—we can find the equations for the polynomial coefficients by direct substitution in the earlier analysis. Only the third stage has changed. The input current of the third stage of Design A is

$$i_{i3(A)} = -(G_3 r_3 G_L + \tau_3 G_L s) v_o \quad (3.3-1)$$

whereas the input current of the feedback stage is given by eq. (3.2-10):

$$i_{i3(B)} = - \left[\frac{1 + r_3 G_L}{1 - r_3 G_2} G_2 + \frac{G_2 + G_L}{1 - r_3 G_2} \tau_3 s \right] v_o \quad (3.3-2)$$

By substituting the constant and linear terms of (3.3-2) for those of (3.3-1) in eqs. (3.1-15) to (3.1-18), we obtain the polynomial coefficients of the amplifier

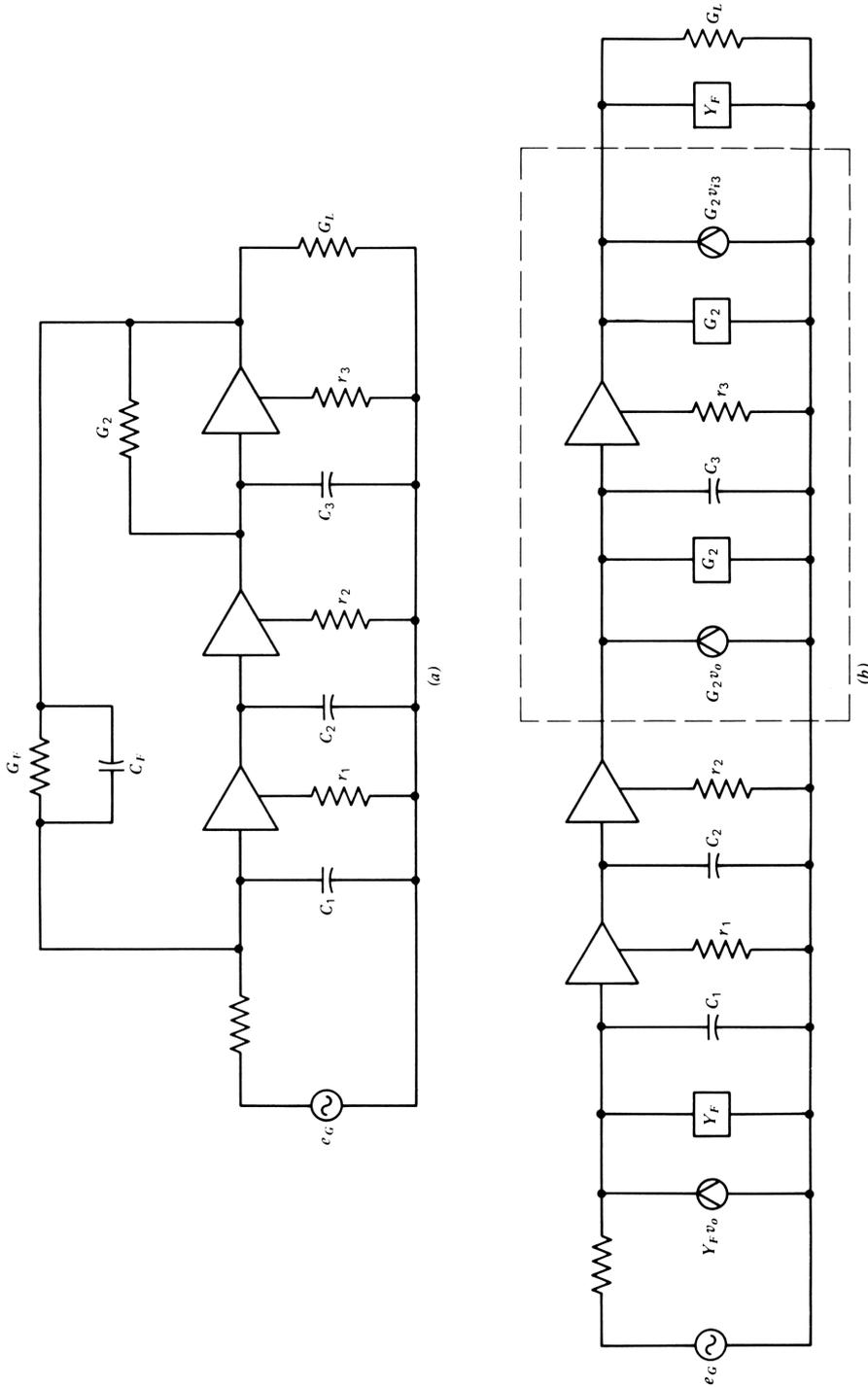


Figure 3.5. Equivalent ladder circuit for Design B.

in Fig. 3.3 directly. Making the approximation that $r_3G_2 \ll 1$, we have

$$a_0 = -R_G G_F \tag{3.3-3}$$

$$a_1 = -R_G C_F - r_1 \tau_2 G_2 (1 + r_3 G_L) \tag{3.3-4}$$

$$a_2 = -r_1 \tau_2 \tau_3 (G_L + G_2) - R_G \tau_1 \tau_2 (1 + r_3 G_L) G_2 \tag{3.3-5}$$

$$a_3 = -R_G \tau_1 \tau_2 \tau_3 (G_L + G_2) \tag{3.3-6}$$

The denominator (due to direct feedthrough) has been ignored since its effect on the coefficients is 1% or less.

The four equations for the coefficients complete the *analysis* of the amplifier in Fig. 3.5a. To *synthesize* the amplifier to realize a prescribed set of coefficients such as the cubic Butterworth, we substitute these four equations into the synthesis equations, (3.1-5) to (3.1-7), and solve them for the values of G_2 , C_F , and G_F . The solution is found by direct substitution if G_2 is found first since $a_2 \neq f(C_F, G_F)$. Although there are many terms, the solution is straightforward; with $b_3 = 1$, we have

$$G_2 = \frac{G_L \tau_3 (b_2 \omega_0 R_G \tau_1 - r_1)}{R_G \tau_1 (1 + r_3 G_L) - \tau_3 (b_2 \omega_0 R_G \tau_1 - r_1)} \tag{3.3-7}$$

Also

$$G_F = \frac{-a_0}{R_G} = \frac{-a_3 \omega_0^3}{R_G} = \tau_1 \tau_2 \tau_3 (G_L + G_2) \omega_0^3 \tag{3.3-8}$$

and finally

$$\begin{aligned} C_F &= \frac{-a_1}{R_G} - \frac{r_1 \tau_2 G_2 (1 + r_3 G_L)}{R_G} \\ &= \frac{b_1 G_F}{\omega_0} - \frac{r_1 \tau_2 G_2 (1 + r_3 G_L)}{R_G} \end{aligned} \tag{3.3-9}$$

With the device parameters and circuit values used for Design A, and again assuming Butterworth response with $\omega_0 = 0.1$ Grad/s, we obtain $G_2 = 1.0$ mS, $G_F = 0.011$ mS, and $C_F = 0.11$ pF, thus completing Design B. The loss polynomial is a little different from that of Design A, because G_2 loads the output stage:

$$L(s)_{\text{Design B}} = -(0.011 + 0.22s + 2.2s^2 + 11s^3) \tag{3.3-10}$$

The two designs now completed give essentially the same performance, at least for nominal values of the components and terminating immittances.

There are significant differences in sensitivity, however. Design B is superior because the a_2 coefficient is realized (in part) by feedback from the output voltage. The following sections clarify and quantify the differences between the two designs.

3.4 SENSITIVITIES OF THE POLYNOMIAL COEFFICIENTS TO THE DEVICES AND COMPONENTS

In cases where device characteristics or circuit component values may vary, we are interested in how such variations affect the polynomial coefficients. As we saw in Chapter 1, it is a simple matter to obtain the sensitivities by the sum rule. For Design B, we use eqs. (3.3-3) to (3.3-6) to find the sensitivities of the polynomial coefficients to the component and device parameter values by inspection. The sensitivity of coefficient a_i to component x is found from the equation for a_i by dividing all terms that include x by the total, which is a_i . Thus the sensitivity of a_0 to G_F or to R_G is unity since G_F and R_G are included in the only term in the equation for a_0 . The sensitivity of a_0 to any other components or device parameters is zero. The sensitivity of a_1 to C_F is less than unity because other terms that do not contain C_F as a factor add to a_1 . Thus

$$S_{C_F}^{a_1} = \frac{-R_G C_F}{-R_G C_F - r_1 \tau_2 G_2 (1 + r_3 G_L)} = \frac{-R_G C_F}{a_1} \quad (3.4-1)$$

Since the amplifier in Fig. 3.3 contains five components including R_G and G_L and six device parameters, there are 11 such sensitivity expressions for each coefficient, or 44 in all. They are all shown in Table 3.2. A similar table could be prepared for Design A. The numerical values of the component sensitivities for both designs are given in Table 3.3. The sensitivities are similar except for the sensitivities to r_3 and G_L .

The values given in Table 3.3 can be obtained as previously or may be obtained by perturbing the value of each component in turn by a small percentage and reanalyzing the circuit by using eqs. (3.3-3) to (3.3-6). Program "AN1" in Appendix B analyzes Design B by using these equations, and program "SCX" finds the sensitivities of coefficients to components by this method.

The array or matrix of sensitivities shown in Table 3.3 is presented to show the dimensions of a complete sensitivity analysis for a simple model of a practical amplifier design. It will also be useful later in finding the statistical variation in the loss from variations of the component values. Where entries are zero, the coefficient of the column in which the zero is found is not affected by the component (or device parameter) in question. One row—that for r_2 —contains only zero entries; had we known this, r_2 could have been omitted from the model. When the model is improved to include other immittances, such as output-to-input feedback capacitance over each device, r_2 will be found to have a small but nonzero effect on performance.

Table 3.2 Sensitivities of Polynomial Coefficients to Device Characteristics and Component Values for Design B

x	$S_x^{a_0}$	$S_x^{a_1}$	$S_x^{a_2}$	$S_x^{a_3}$
Components				
R_G	1	$\frac{-R_G C_F}{a_1}$	$\frac{-R_G \tau_1 \tau_2 (1+r_3 G_L) G_2}{a_2}$	1
G_F	1	0	0	0
C_F	0	$\frac{-R_G C_F}{a_1}$	0	0
G_2	0	$\frac{-r_1 \tau_2 G_2 (1+r_3 G_L)}{a_1}$	$\frac{-G_2 [r_1 \tau_2 \tau_3 + R_G \tau_1 \tau_2 (1+r_3 G_L)]}{a_2}$	$\frac{G_2}{G_L + G_2}$
G_L	0	$\frac{-r_1 \tau_2 G_2 r_3 G_L}{a_1}$	$\frac{-G_L (r_1 \tau_2 \tau_3 + R_G \tau_1 \tau_2 r_3 G_2)}{a_2}$	$\frac{G_L}{G_L + G_2}$
Devices				
τ_1	0	0	$\frac{-R_G \tau_1 \tau_2 (1+r_3 G_L) G_2}{a_2}$	1
r_1	0	$\frac{-r_1 \tau_2 G_2 (1+r_3 G_L)}{a_1}$	$\frac{-r_1 \tau_2 \tau_3 (G_L + G_2)}{a_2}$	0
τ_2	0	$\frac{-r_1 \tau_2 G_2 (1+r_3 G_L)}{a_1}$	1	1
r_2	0	0	0	0
τ_3	0	0	$\frac{-r_1 \tau_2 \tau_3 (G_L + G_2)}{a_2}$	1
r_3	0	$\frac{-r_1 \tau_2 G_2 r_3 G_L}{a_1}$	$\frac{-R_G \tau_1 \tau_2 r_3 G_2 G_L}{a_2}$	0

The sensitivity of the coefficients to r_3 is of particular interest because of the role of r_3 in establishing the nonlinear performance and variability of loss, as noted previously. From Table 3.2, we see that the sensitivities of a_0 and a_3 to r_3 are zero. The sensitivities of a_1 and a_2 to r_3 for Design B are

$$S_{r_3}^{a_1} = \frac{-r_1 \tau_2 G_2 r_3 G_L}{a_1} = 0.045 \quad (3.4-2)$$

$$S_{r_3}^{a_2} = \frac{-R_G \tau_1 \tau_2 r_3 G_2 G_L}{a_2} = 0.045 \quad (3.4-3)$$

Table 3.3 Numerical Values of Sensitivities of Polynomial Coefficients to Component and Device Parameter

x	Design A/Design B			
	$S_x^{a_0}$	$S_x^{a_1}$	$S_x^{a_2}$	$S_x^{a_3}$
Components				
R_G	1	0.5	0.5	1
G_F	1	0	0	0
C_F	0	0.5	0	0
G_3/G_2	0	0.5	0.5/0.55	0/0.09
G_L	0	0.5/0.045	1/0.5	1/0.91
Devices				
τ_1	0	0	0.5	1
r_1	0	0.5	0.5	0
τ_2	0	0.5	1	1
r_2	0	0	0	0
τ_3	0	0	0.5	1
r_3	0	0.5/0.045	0.5/0.045	0

To compare these results with those for Design A, we obtain the sensitivities by inspection of eqs. (3.1-16) and (3.1-17):

$$S_{r_3}^{a_1} = \frac{-r_1\tau_2G_3r_3G_L}{a_1} = 0.5 \quad (3.4-4)$$

$$S_{r_3}^{a_2} = \frac{-R_G\tau_1\tau_2G_3r_3G_L}{a_2} = 0.5 \quad (3.4-5)$$

The effect of a change in r_3 is roughly 10 times worse for Design A than for Design B. If control for G_2 is better than that of r_3 , the response of the latter design will retain its accuracy in the face of changes of r_3 considerably better than will Design A. Nonlinearity of r_3 will cause nonlinearity of the whole amplifier for either design at high frequencies where a_1 and a_2 are important, but Design B will be affected by an order of magnitude less than Design A in this respect.

A calculation of the sensitivity of the coefficients to the load conductance shows a similar desensitization of a_1 and a_2 in the second design. The interpretation here is that the response is less sensitive to load immittance variations for Design B; this is equivalent to saying that the output impedance is lower for Design B. The effect of feedback on impedances is discussed in more detail later on.

To explain these facts by saying that feedback reduces distortion and gain variations is imprecise. Feedback is an elusive notion, although sensitivity is not. Both designs include *local feedback* in the third stage: r_3 provides an input voltage to the third stage that is proportional to output current. In Design A, G_3 converts this to a current that must be provided by the earlier stages. In Design B, on the other hand, G_2 provides this current directly from the output voltage of the amplifier. Feedback is present in either case, but in the first design it was provided by an immittance element r_3 that is subject to variation, whereas this effect is minimized in the second. Note that the residual sensitivity to r_3 in Design B comes about because of the input loading by G_2 on the third stage— y_{11} of the feedback network in the terminology introduced in Section 3.2. It is considerably easier to quantify the effects of feedback through the concept of sensitivity than by attempting to define the notion of feedback quantitatively. “Feedback” in the sense of loop gain has a precise definition for any particular way of looking at a system, but it is subject to the view that we adopt. Sensitivity, on the other hand, is a property of the physical system and does not vary with the method we choose to describe it. This is why we adopt sensitivity as the primary concept of feedback systems. These ideas are developed further in Chapter 4.

3.5 SENSITIVITIES OF LOSS TO INDIVIDUAL COMPONENTS AND DEVICE PARAMETERS

We are only part way to our goal of finding the variation of loss as a function of frequency due to variation of the values of the components and device parameters of a feedback system. In this section we combine loss-to-coefficient sensitivities with coefficient-to-component sensitivities.^{1,2}

In Chapter 2, we found the sensitivities of loss to the coefficients of the loss polynomial and presented them as a series of Bode plots, one for each coefficient. In Section 3.4 we found the sensitivities of each coefficient to each component and device parameter. It now remains to combine these two sets of sensitivities to obtain the sensitivities of the loss to each component and parameter of the system. From Table 1.1 we can write

$$S_{x_j}^L(j\omega) = \sum_{i=0}^n S_{a_i}^L \cdot S_{x_j}^{a_i} \quad (3.5-1)$$

where x_j denotes the value of the j th component or device parameter of the system. To obtain the total sensitivity of $L(j\omega)$ to component x_j , we must account for the effect of component x_j on all the a_i and sum the resulting sensitivities, each of which is a complex number.

This central concept can be illustrated by using the two amplifier designs completed previously. Let x_j be r_3 of Design A, for example. We have calculated that the sensitivities of a_1 and a_2 are both 0.5 for this design, and the loss polynomial is Butterworth; the sensitivities of loss to a_1 and a_2 were

calculated in Chapter 2 for this case, and we can write

$$S_{r_3}^L \text{ Design A} = 0.5S_{a_1}^L + 0.5S_{a_2}^L \quad (3.5-2)$$

where the sum is a phasor sum. (Since the sensitivities of a_0 and a_3 to r_3 are both zero, they do not appear in the sum.) For Design B, the sensitivities of loss to a_1 and a_2 were both 0.045, so the sensitivities of loss to r_3 have the same frequency shape but are less by a factor of $0.045/0.5$, or 21 dB. These calculations were carried out and are plotted as Bode diagrams in Fig. 3.6, using program "SLX" in Appendix B. The curves also tell us just how sensitive each design is to r_3 . At the asymptotic cutoff frequency, for example, distortion caused by nonlinearity of r_3 is not reduced at all in Design A, whereas it is reduced by 21 dB for Design B. (Where we speak of distortion at a single frequency, we mean intermodulation products caused by signals in the immediate vicinity of the frequency.)

Note that the sensitivities of loss to the coefficients are the same in the two equations, (3.5-2) and (3.5-3), since they are related only to the loss polynomial, Butterworth in both cases. Separating those sensitivities dependent on the polynomial and those dependent on the components is a valuable clarification. Stability questions, for example, concern only the coefficient sensitivity, not the component sensitivity; to produce a stable design, one must choose a

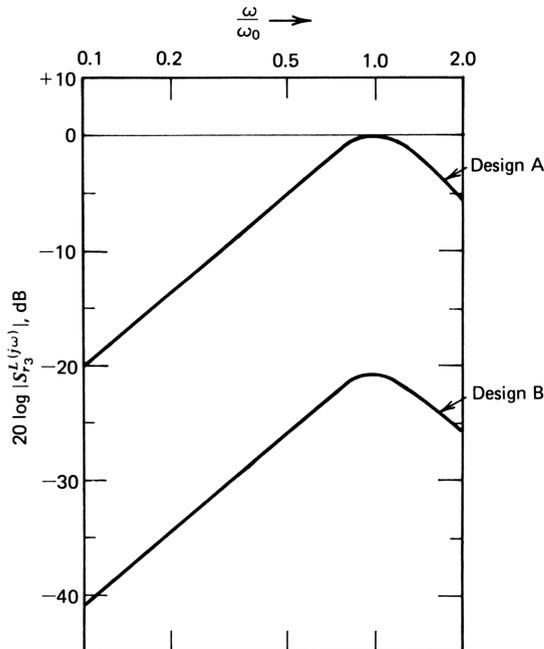


Figure 3.6. Magnitudes of the sensitivities of loss to device parameter r_3 as a function of frequency for Designs A and B.

stable loss polynomial and then ensure that it is realized as planned. The loss polynomial appears in the denominator of the sensitivity function, so that poles of the sensitivity function (zeros of the loss polynomial) must not be allowed to creep into the right half plane.

We have expressed the *magnitude* of the sensitivity in decibels. In this form it is the negative of what is commonly understood as feedback: when one says that there is 20 dB of feedback in an amplifier, he means that the sensitivity of loss to some element is -20 dB. Since we have seen that the sensitivity to each element or component is different in general, it follows that the feedback—in this commonly understood sense—is different for each element. Where the term “feedback” is used, therefore, the element to which it refers must be specified. We have seen in this section how sensitivities may be added to obtain the total sensitivity; the complex components must be added separately to obtain the total.

At this point we should clarify the relationship between the magnitude of sensitivity and its real and imaginary parts, already discussed in Section 1.6. Suppose that we express the loss of an amplifier at some frequency in polar form: with α the attenuation (in nepers) and β the phase in radians, we have

$$L = e^{\alpha + j\beta} \quad (3.5-3)$$

Now let some component x change by Δx , causing a new value of L , say, L' :

$$L + \Delta L = L' = e^{\alpha' + j\beta'} \quad (3.5-4)$$

Figure 3.7a shows these two loss phasors. The change ΔL is also shown. If we divide L' by L , that is, if we normalize to L , we have

$$\frac{L'}{L} = 1 + \frac{\Delta L}{L} = \exp[\alpha' - \alpha + j(\beta' - \beta)] \quad (3.5-5)$$

The phasor diagram for this equation is shown in Fig. 3.7b. If we take the natural log of this equation, for small ΔL , we have

$$\begin{aligned} \ln\left(1 + \frac{\Delta L}{L}\right) &= \alpha' - \alpha + j(\beta' - \beta) \\ &= \frac{\Delta L}{L} \end{aligned} \quad (3.5-6)$$

Clearly, from Fig. 3.7b, the real part, $\alpha' - \alpha$, gives the per unit change in the magnitude of L'/L (or equivalently, the change in nepers) and the imaginary part, $\beta' - \beta$, the change in the angle (in radians) of L'/L . The magnitude of $\Delta L/L$ gives a measure of the effect of Δx on L : we know that the phasor L'/L will terminate on a circle of radius $\Delta L/L$ centered at $+1$, as shown. Thus when we are given the magnitude of $\Delta L/L$, we know only that the resulting

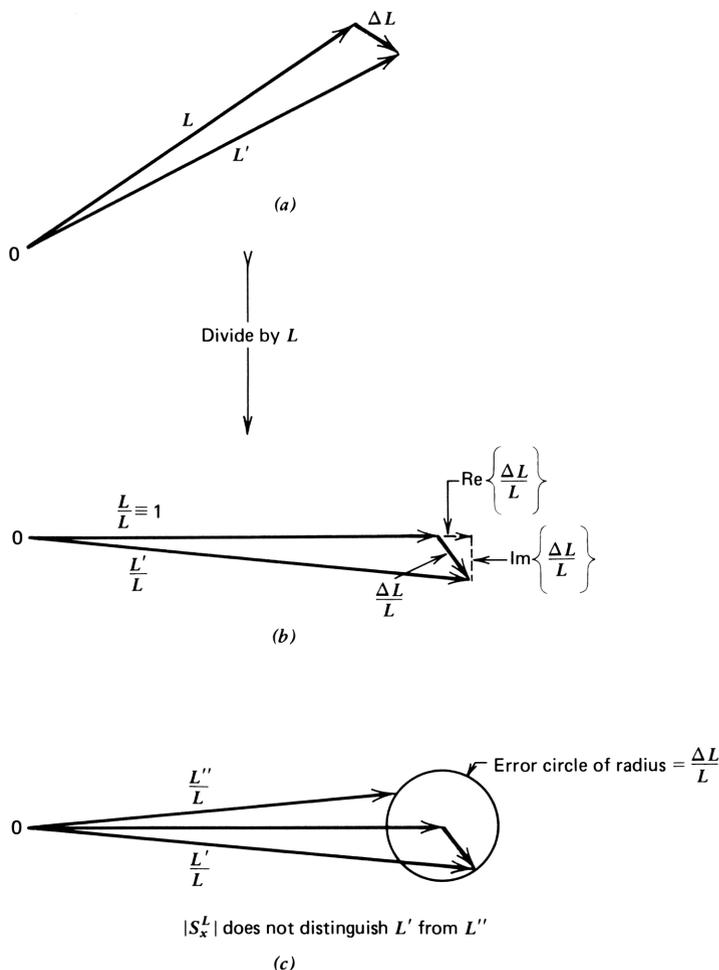


Figure 3.7. Phasor diagrams indicating the relationship between the magnitude and real and imaginary parts of the variation of loss and the sensitivity of loss to component x_j .

loss will be within the bound defined by the circle; we lose the information about the relative effects of Δx on the magnitude and phase of $L(j\omega)$. For determining the effect of feedback on distortion, this is all the information that is needed. The distortion component is proportional to $\Delta L/L$; the improvement in linearity is the reciprocal of this, or the magnitude of feedback. The practice of specifying the magnitude of feedback arose in this way; as a general single measure of the benefits of feedback, it is adequate.

The preceding comments are directed to the *variation* of L from *variation* of component x . The *sensitivity* of L to X is obtained by dividing the variation $\Delta L/L$ by the scalar $\Delta x/x$. The phasor diagram for sensitivity is the same as that for variation of L , expanded by the factor $x/\Delta x$; we concentrate on the portion of the diagram at $+1$ in Fig. 3.7b or 3.7c. The magnitude of the

sensitivity defines a circle whose radius is the reciprocal of the feedback magnitude.

If we are interested in the variation of magnitude (or phase), we should take the real (or imaginary) parts of the sensitivity, as shown in Fig. 3.8. Note that at low frequencies (in the band of interest) the primary effect of r_3 variation is on the phase of L rather than on the magnitude. Comparison of effects expressed in angles with those expressed in magnitudes implies a common basis for the comparison. The basis is that of eq. (3.5-6), in which an in-phase per unit change in L is equivalent to the change in nepers, and a quadrature per unit change in L is equivalent to the change in radians. Hence, for example, a change of 0.1 Np (0.86 dB) is comparable with a change of 0.1 rad (5.7°)

As noted previously, the sensitivity of loss to G_L should be better for Design B than for Design A. In Fig. 3.9 the magnitude of the sensitivities of loss to G_L are plotted for both designs. The equations for the plots are obtained from Table 3.3. Design A is given in eq. (3.5-8), and Design B is given in eq. (3.5-9):

$$S_{G_L}^{L(j\omega)} = 0.5S_{a_1}^L + 1.0S_{a_2}^L + 1.0S_{a_3}^L \tag{3.5-8}$$

$$S_{G_L}^{L(j\omega)} = 0.045S_{a_1}^L + 0.5S_{a_2}^L + 0.91S_{a_3}^L \tag{3.5-9}$$

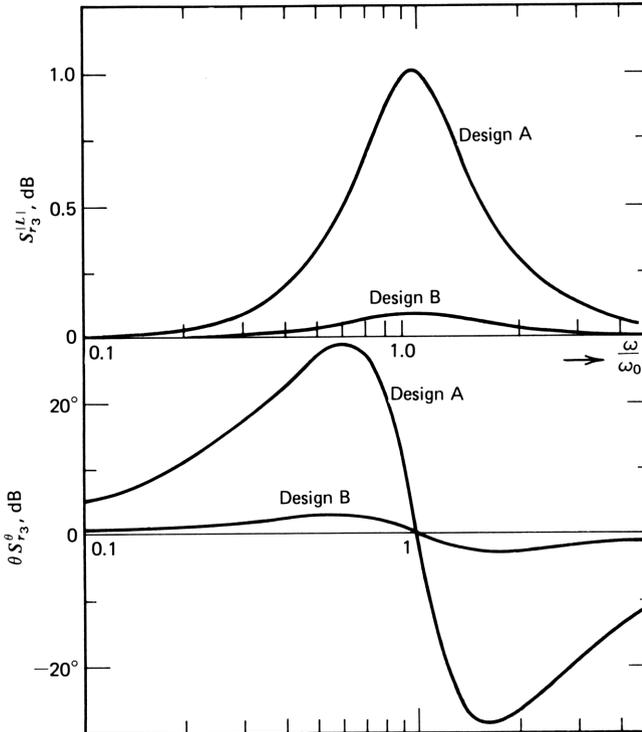


Figure 3.8. Real and imaginary parts of the sensitivities of loss to r_3 for Designs A and B.

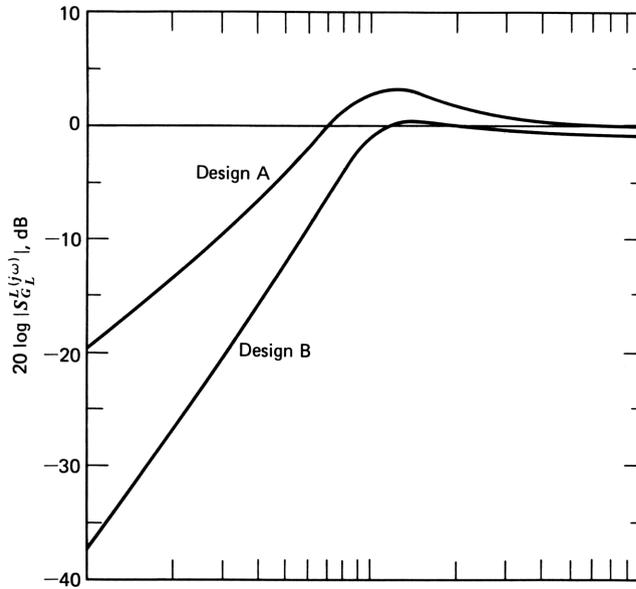


Figure 3.9. Magnitude of the sensitivities (in dB) of loss to G_L for Designs A and B.

3.6 VARIATION OF LOSS AND PHASE FROM COMPONENT AND DEVICE VARIATIONS

The total variation of loss is found by adding the products of the component variations and their respective losses to component sensitivities. For small changes, the variation in magnitude is found by adding the real parts of the product and the variation in phase by adding the imaginary parts:

$$\frac{\Delta L}{L} = \sum_j \frac{\Delta x_j}{x_j} \operatorname{Re} [S_{x_j}^L] \quad (3.6-1)$$

To express the change in decibels, this must be multiplied by $20 \log e = 8.69$ dB/Np:

$$\arg \frac{\Delta L}{L} = \sum_j \frac{\Delta x_j}{x_j} \operatorname{Im} S_{x_j}^L \quad (3.6-2)$$

The variations of the components are usually known only statistically. The per unit component variation $\Delta x_j/x_j$ is a random variable, and we are interested in the statistical properties of the loss, which is the complex sum of several random variables. The components may vary in a deterministic way; the mean value of the variation due to temperature of all the resistors on an integrated circuit chip, for example, may vary together. Superimposed on this may be a

smaller random variation characterized by its variance, or the square root of the variance, the standard deviation. For a sum of normally distributed random variables, the mean value of the sum is the sum of the mean values, and the variance of the sum is the sum of the individual variances. Letting μ_L be the mean value of the variation of $\Delta L/L$ and σ_L the standard deviation of $\Delta L/L$, we can write

$$\mu_L = \sum_j \mu_{x_j} (\text{Re } S_{x_j}^L + j \text{Im } S_{x_j}^L) \tag{3.6-3}$$

where the real part is the variation in the magnitude of $\Delta L/L$ and the imaginary part the variation in phase. Similarly, the standard deviation of $\Delta L/L$ is

$$\sigma_L^2 = \sum_j \sigma_{x_j}^2 \left[(\text{Re } S_{x_j}^L)^2 + (\text{Im } S_{x_j}^L)^2 \right] \tag{3.6-4}$$

For each x_j , the sensitivity in the preceding expressions is the phasor sum of $n+1$ terms for a polynomial of degree n , as given by eq. (3.5-1). This summation is to be done for each of the j components using the coefficient-to-component sensitivity matrix values such as those given for Design B in Table 3.3. Calculator program “STAT” in Appendix B evaluates the mean and standard deviation of the loss and phase as given in (3.6-3) and (3.6-4) for a system that has a arbitrary number of components for which the coefficient-to-component sensitivity matrix values are known (they are found by Program “SCX” and a system analysis program, as described in Section 3.4).

To illustrate the calculation, the simple circuit in Fig. 3.10 is used; the amplifier is considered ideal, so that the loss is given by

$$\begin{aligned} L &= -R(G + Cs) \\ &= a_0 + a_1s \end{aligned} \tag{3.6-5}$$

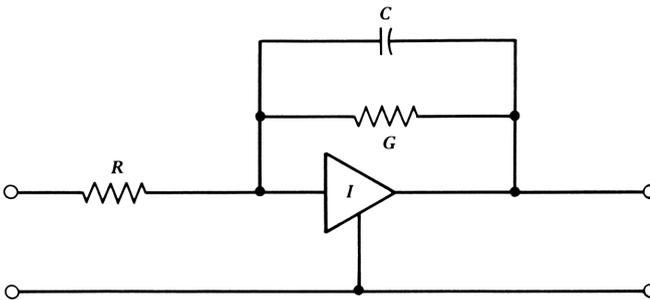


Figure 3.10. Circuit for calculation of mean and standard deviation of loss and phase variation.

The coefficient-to-component sensitivity matrix is written by inspection:

$$S_{x_j}^{a_i} = \begin{bmatrix} S_R^{a_0} & S_R^{a_1} \\ S_G^{a_0} & S_G^{a_1} \\ S_C^{a_0} & S_C^{a_1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (3.6-6)$$

Suppose that $R=G=C=1.0$ and that we wish the mean and standard deviation at an angular frequency of 0.5 (in a consistent set of units). Then the loss-to-coefficient sensitivities are given by

$$S_{a_0}^L = \frac{a_0}{L} = \frac{1}{1+j0.5} = 0.8 - j0.4$$

$$S_{a_1}^L = \frac{j0.5a_1}{L} = \frac{j0.5}{1+j0.5} = 0.2 + j0.4$$

This can be written as a column vector:

$$S_{a_i}^L = \begin{bmatrix} S_{a_0}^L \\ S_{a_1}^L \end{bmatrix} = \begin{bmatrix} 0.8 - j0.4 \\ 0.2 + j0.4 \end{bmatrix} \quad (3.6-7)$$

Premultiplying this vector by the matrix of (3.6-6), we obtain the loss-to-component sensitivities as

$$S_{x_j}^L = \begin{bmatrix} S_R^L \\ S_G^L \\ S_C^L \end{bmatrix} = \begin{bmatrix} 1.0 + j0 \\ 0.8 - j0.4 \\ 0.2 + j0.4 \end{bmatrix} \quad (3.6-8)$$

Now suppose that the mean values of the resistors change by -10% and that the mean value of the capacitance does not change. Suppose also that the standard deviation of the change in both resistors and capacitors is 2%. The mean value of the loss variation is obtained by premultiplying the column vector in (3.6-8) by a row matrix of mean values of the component changes:

$$\mu_L = \begin{bmatrix} -0.1 & 0.1 & 0 \end{bmatrix} \left[\begin{bmatrix} 1.0 \\ 0.8 \\ 0.2 \end{bmatrix} + j \begin{bmatrix} 0 \\ 0.4 \\ 0.4 \end{bmatrix} \right]$$

$$= -0.02 - j0.04$$

in nepers and radians. Converting to decibels and degrees,

$$\mu_L = -0.17 \text{ dB} \quad \text{and} \quad -2.29^\circ$$

This change in mean values could be eliminated by using a capacitor that has a +10% change in mean value to match the resistor change.

To find the standard deviation of the loss variation, we form a row matrix of the variances of the three components and postmultiply it by the squared value of the real and imaginary parts of (3.3-8):

$$\sigma_L^2 = \begin{bmatrix} 0.0004 & 0.0004 & 0.0004 \end{bmatrix} \left[\begin{bmatrix} 1.0 \\ 0.64 \\ 0.04 \end{bmatrix} + j \begin{bmatrix} 0 \\ 0.16 \\ 0.16 \end{bmatrix} \right]$$

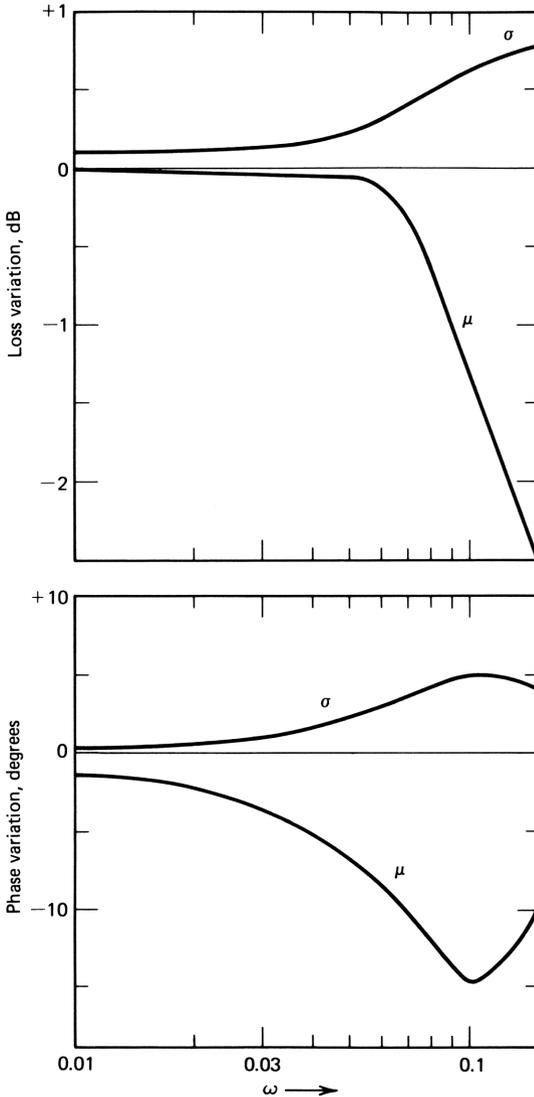


Figure 3.11. Mean and standard deviation of loss and phase for Design B.

from which we obtain

$$\sigma_L = 0.0259 + j0.0113$$

Thus the standard deviation of the loss change is 0.23 dB and 0.65° . A similar calculation can be made at other frequencies to obtain the variation of μ and σ with frequency.

The loss variation (mean and standard deviation) for the Design B amplifier is shown as a function of frequency in Fig. 3.11, calculated in the same way as for the simple three-component circuit. In this calculation the mean value of all resistors were assumed to change by -10% (conductances by $+10\%$) and capacitors by 0% . The mean value of time constants of the transistors were assumed to change by -10% . The standard deviations of the components were all assumed to be 1% ; the device parameter standard deviations were assumed to be 5% . Program "STAT" was used for the calculation; the program and a sample calculation for Design B are given in Appendix B. Thus it is reasonably straightforward to obtain the loss variation as a function of component and device variations.

3.7 SYNTHESIS BY ITERATION OF ANALYSIS: NEWTON'S METHOD

Analysis of circuits can always be done; if the process becomes too tedious by hand calculation, we can turn to the computer for help. Such analysis can provide us with the polynomial coefficients as well, although most present computer circuit analysis programs do this at best indirectly. Such programs can be modified to obtain the coefficients, or they can be found by use of a program such as "RCU" in Appendix A, starting with the loss and phase at several frequencies. The synthesis process—finding a small set of feedback immittances to give us the desired polynomial coefficients—can often be done analytically, but this is unnecessary if we have selected a configuration for which there exists a *dominant element* that controls each coefficient. In the case study, such elements were G_F , C_F , and G_2 or G_3 for a_0 , a_1 , and a_2 , respectively. Intuitively, we know that if we analyze the circuit for its polynomial coefficients and find that one of them is smaller than we desire, we can increase its dominant element. The amount of the increase will be some function of the differences between the desired coefficient and the coefficient obtained from the analysis. If we can approximate this function, we should be able to reduce this difference and try again. The process can be repeated until the difference is negligible. When this procedure is carried out for all coefficients (perhaps simultaneously), we should be able to converge on a design. Thus in Design B if a_1 is too small, increase C_F ; or if a_2 is too small, increase G_2 . When we increase G_2 , however, the effective output loading $G_L + G_2$ increases, and the value of a_3 changes slightly, changing the value of C_F that will be required. This effect is small, however, since the sensitivity of a_0 (or a_3) to G_2 is small—only 0.091 in Table 3.3—so that by going through the process a few

times we should converge on the right numbers for the design. In this section we show how this is to be done and put the intuitive notion on a firmer mathematical basis.

We begin by defining three variables representing the error of the polynomial coefficients for a given cubic system. From the synthesis equations (3.1-5) to (3.1-7), we define

$$\alpha_0 = \frac{a_3 \omega_0^3}{b_3} - a_0 \tag{3.7-1}$$

$$\alpha_1 = \frac{b_1 a_3 \omega_0^2}{b_3} - a_1 \tag{3.7-2}$$

$$\alpha_2 = \frac{b_2 a_3 \omega_0}{b_3} - a_2 \tag{3.7-3}$$

where in each case the first term on the right represents the desired value of the polynomial coefficient and the second term is the value of the coefficient as calculated from the analysis of the circuit or system. Next, we define a set of free variables, x_0 , x_1 , and $x_2 = x_j$ representing the dominant elements. For Design B, they would be G_F , C_F , and G_2 , for example. The three α_i will be functions of the x_j since we assume the x_j to be dominant in determining the coefficients:

$$\alpha_i = f_i(x_0, x_1, x_2) = f_i(x_j) \tag{3.7-4}$$

For each of the three α_i , we take partial derivatives with respect to each of the x_j , giving an array of nine partial derivatives known as the *Jacobian matrix*, or *Jacobian*, written compactly as

$$\frac{d(\alpha_0, \alpha_1, \alpha_2)}{d(x_0, x_1, x_2)} = \frac{\partial \alpha_i}{\partial x_j} \tag{3.7-5}$$

The total differential of α_0 may be written

$$d\alpha_0 = \frac{\partial \alpha_0}{\partial x_0} dx_0 + \frac{\partial \alpha_0}{\partial x_1} dx_1 + \frac{\partial \alpha_0}{\partial x_2} dx_2 \tag{3.7-6}$$

The total differential change of the error in the dc coefficient is made up of changes that arise from each of the three x_j . The change in all three α_i can be written as

$$d\alpha_i = \frac{\partial \alpha_i}{\partial x_j} dx_j \tag{3.7-7}$$

For Design B, this expression can be written as

$$\begin{bmatrix} d\alpha_0 \\ d\alpha_1 \\ d\alpha_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial \alpha_0}{\partial G_F} & \frac{\partial \alpha_0}{\partial C_F} & \frac{\partial \alpha_0}{\partial G_2} \\ \frac{\partial \alpha_1}{\partial G_F} & \frac{\partial \alpha_1}{\partial C_F} & \frac{\partial \alpha_1}{\partial G_2} \\ \frac{\partial \alpha_2}{\partial G_F} & \frac{\partial \alpha_2}{\partial C_F} & \frac{\partial \alpha_2}{\partial G_2} \end{bmatrix} \begin{bmatrix} dG_F \\ dC_F \\ dG_2 \end{bmatrix} \quad (3.7-8)$$

This equation tells us the *direction* each of the α_i will go in response to changes in G_F , C_F , and G_2 . It is easy to evaluate from the analysis equations, (3.3-3) to (3.3-6), and the equations defining the α_i , (3.7-1) to (3.7-3). Taking α_0 , for example, we use (3.7-1) to obtain

$$\frac{\partial \alpha_0}{\partial G_F} = \frac{\omega_0^3}{b_3} \frac{\partial a_3}{\partial G_F} - \frac{\partial a_0}{\partial G_F} = R_G \quad (3.7-9)$$

[We use (3.3-6) to obtain $\partial a_3/\partial G_F=0$ and (3.3-3) to obtain $\partial a_0/\partial G_F=-R_G$.] With the equations programmed on a calculator or a computer, we perturb the value of G_F by a small amount (say, 1%) and find the change in α_0 , α_1 , and α_2 . Dividing these changes by ΔG_F , we obtain the first column of the Jacobian. We then repeat the process for C_F and G_2 to obtain the whole matrix. For the set of circuit values used in Section 3.3 for Design B, for example, the Jacobian is

$$\begin{bmatrix} d\alpha_0 \\ d\alpha_1 \\ d\alpha_2 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0.001 \\ 0 & -1 & -0.090 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} dG_F \\ dC_F \\ dG_2 \end{bmatrix} \quad (3.7-10)$$

In this matrix the dominant relationships are along the principle diagonal; the off-diagonal elements are less important, arising from the small dependence of a_3 on G_2 . For Design A, this dependence is absent, and the Jacobian is a diagonal matrix (with no off-diagonal elements). The negative signs of the elements along the principal diagonal are merely an expression of the phase reversal in the amplifier; all polynomial coefficients are negative.

What is needed for design is the direction in which to change the elements G_F , C_F , and G_2 to bring the α_i to zero. We obtain the desired result by premultiplying both sides of (3.7-8) by the inverse of the Jacobian; in compact form, this is

$$dx_j = \left(\frac{\partial \alpha_i}{\partial x_j} \right)^{-1} d\alpha_i \quad (3.7-11)$$

The inverse exists by virtue of our assumption of a dominant element for each

α_i ; each element gives a nonzero entry along the principal diagonal. Equation (3.7-11) gives us just what we need to form an iterative design method to find the values of the dominant elements, the x_j . If we replace the differential of α_i by the total desired change, which is α_i itself, we obtain from the equation the change in each x_j required for the next iteration, that is, $x_{j(n+1)} - x_{j(n)} = \Delta x_j$. By analogy with Newton's method and eq. (2.3-1), we write

$$x_{j(n+1)} = x_{j(n)} - \left(\frac{\partial \alpha_i}{\partial x_j} \right)^{-1} \alpha_i \tag{3.7-12}$$

which is merely an extension of Newton's method to three dimensions.

The inverse of the Jacobian of (3.7-10) for Design B is found to be

$$\begin{bmatrix} G_{F(n+1)} \\ C_{F(n+1)} \\ G_{2(n+1)} \end{bmatrix} = \begin{bmatrix} G_{F(n)} \\ C_{F(n)} \\ G_{2(n)} \end{bmatrix} - \begin{bmatrix} -1 & 0 & -0.001 \\ 0 & -1 & 0.090 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} \tag{3.7-13}$$

Program "SJ", which implements this equation, is given in Appendix B. It finds the Jacobian as outlined previously, inverts it, and determines the values of the three x_j by iteration. The circuit analysis program is separate from the synthesis program, so that the synthesis can be applied to any circuit having three dominant elements and yielding cubic polynomial response. General purpose computer analysis programs such as SPICE can be used, but the polynomial coefficients must be obtained from loss and phase information usually supplied by such programs; program "RCU" can be used for this purpose.

An iterative process is itself a feedback process—the result of a computation is compared with a reference, and the difference (or error) is used in a new computation to tend to correct the error. Nonconvergence is the equivalent of instability. We can illustrate convergence in Design B. Where the approximate analysis equations of Section 3.3 are used, convergence is not only assured, but is completed in one step because the a_i are *linear* in the x_j for eqs. (3.3-3) to (3.3-6). Because of this linear relationship, all partial derivatives of the Jacobian are constants, not a function of the x_j . Hence the direction of movement of the x_j from the arbitrary starting point to the final value lies along a straight line; convergence is immediate.

If the program is modified to include the denominator, $1 - r_3 G_2$, left out of the approximate equations, convergence becomes dependent on the starting point. With $r_3 = 0.01 \text{ k}\Omega$, a value of G_2 in excess of 100 mS (an outlandish value) causes the sign of the denominator to reverse, so that the phase reversal of the third stage is lost. Hence the feedback of the outer path becomes regenerative, and the design converges on an incorrect set of values. More reasonable starting points will assure convergence on the correct values.

The device model shown in Fig. 3.1 is too crude to provide us with a viable design. It has been used to illustrate design principles and to develop sensitivity relationships in a definite way without burdening the development with more detail than is necessary. We develop accurate models later and show how they are to be incorporated into the design process. Many of the changes required in the model, such as adding collector capacitance (or drain to gate capacitance) and dc leakage at the input ($1/h_{fe}$ for bipolar transistors), do not change the general outlines of the design but do change the values of the feedback immittances obtained. One aspect left out here is central to feedback system design: signal delay may profoundly affect the design process. We study the effects of delay in Chapter 5, where we modify the analysis to include it. The design procedure given in this section remains, but delay is found to alter the Jacobian in such a way that for sufficiently large delays, the design becomes unrealizable. It is usually necessary, therefore, to consider delays in even an approximate model of feedback systems.

PROBLEMS

- 1 Compare the dc loss obtainable for an amplifier with the circuit in Fig. 3.2 that is to have response of ± 0.1 dB up to $\omega_0 = 0.1$ Grad/s using (a) Butterworth response and (b) Chebyshev response. Note that the ripple width for the Chebyshev case and the error tolerance for the Butterworth case are both 0.2 dB. Assume cubic responses. For equal dc losses, compare the amplifier bandwidths obtainable for the two responses.
- 2 An amplifier that has the configuration and device parameters of Fig. 3.2 is measured and found to have the following values of loss magnitude and phase at frequencies of

ω	Loss, dB	Phase
0.05	-28.60	70.53°
0.2	-10.67	-160.6°

Find the values of G_F , C_F , and G_3 if $G_L = 10$ mS and $R_G = 1.0$ k Ω .

- 3 Assume that $r_3 = kT/qI_{C3} = 0.026/I_{C3}$. Find the sensitivities of the four coefficients, a_0 to a_3 to r_3 for a redesigned version of Design B in which $I_{C3} = 10$ mA. The redesign is again to give Butterworth cubic response, and all other device parameters remain the same as in the text. Repeat the problem for Design A.
- 4 Using the transistor model in Fig. 3.1 with $r_1 = r_2 = r_3 = 0.010$ k Ω and $C_1 = C_2 = C_3 = 10$ pF, design a 200 MHz amplifier to operate from a 0.100 k Ω source into a 0.100 k Ω load. Let the response be Butterworth cubic. How much dc loss is there? Find the sensitivity of loss to r_3 at the

cutoff frequency. Repeat this problem, using Chebyshev response with 0.5 dB ripple.

REFERENCES

- 1 D. Hilberman, "An Approach to the Sensitivity and Statistical Variability of Biquadratic Filters," *IEEE Trans. CT-20* (4), 1973.
- 2 G. S. Moschytz, "*Linear Integrated Networks: Design*" Van Nostrand Reinhold, New York 1975, Chapter 1.

Chapter 4

Signal Flow Graphs of Polynomials, Rational Functions, and Circuits

The design process is in part a uniting of response requirements as polynomials or rational functions and the circuits that are to provide them. Thus far we have used the polynomial coefficients as a common ground for expressing the requirements on and the characteristics of a feedback system, obtaining the system coefficients from the circuit immittances by reasonably straightforward circuit analysis. In the design process, another step is required at the outset—to establish the circuit or system topology. This step is the one least amenable to codification into a cut-and-dried procedure. The existence of tens of thousands of circuit patents attests to the creative nature of this step.

It is often helpful to express system requirements in topological form; if this can be done, the requirements may themselves suggest possible system realizations. *Signal flow graphs* provide a way of expressing both circuit characteristics and system requirements topologically and provide a common topological language to talk about both. We begin the study of this valuable design tool by briefly stating the rules for forming signal flow graphs.

4.1 INTRODUCTION TO SIGNAL FLOW GRAPHS* AND SEQUENTIAL MATRICES

Signal flow graphs¹ provide us with a clear way of ordering our thoughts about the relationships among the constants and variables of a mathematical expression. More important, they can be used to relate systems and circuits to the mathematical expressions that describe them. They are a branch of the study of topology from a mathematical point of view and also bear a close relationship to the study of circuit configurations or circuit topology. In this section we develop rules to allow us to draw signal flow graphs directly from mathematical expressions, and we begin the process of drawing signal flow graphs from circuit diagrams.

Definitions

A signal flow graph is a *graph* containing *nodes* (or vertices) and *branches* (or edges) connecting them. The branches are *directed*, as indicated by the arrow in the simple two-node, single branch signal flow graph in Fig. 4.1a. This graph is a representation of the equation $y = bx$, where x is the *transmitting node*, y is the *receiving node*, and b is the value of the branch connecting them. Node x is *unaffected by the branch leaving it*. In Fig. 4.1b, for example, a second branch whose tail is connected to x also has signal x transmitted along it, to be multiplied by *its* branch value. In the figure, a second branch has its nose connected to y contributing signal z to y ; this graph depicts the equation $y = bx + z$. Figures 4.1c and 4.1d establish the addition and multiplication properties of signal flow graphs. The equivalence expressed in Fig. 4.1c follows directly from Fig. 4.1b. A node may act as both transmitting and receiving nodes, as in node z in Fig. 4.1d, where $z = ax$; in this graph we also have $y = bz$, so that the equivalence of Fig. 4.1d is established directly.

The operations of factoring and its opposite, distribution, are shown in Fig. 4.2 and follow directly from the preceding.

The terms “independent” and “dependent” nodes are relative terms relating to individual branches. An independent node transmits signals down a branch whose tail is connected to it, so it might equally well be called a *transmitting node*. Similarly, a dependent node is a *receiving node* that has at least one branch whose nose is connected to it.

A node that has no incoming branches is a *source node*; it is where the analysis of the system begins. Similarly, a *sink node* has no outgoing branches; it is where the analysis of the system terminates. There may be more than one sink or source node in a system.

Closed Loops

Where a closed loop exists, as in Fig. 4.3, the analysis may be carried out by solution of simultaneous equations. For clarity in system analysis, it is well to define source and sink nodes for the system in question. Where no such nodes

*Such a graph is sometimes called a *directed graph*, or *digraph*.

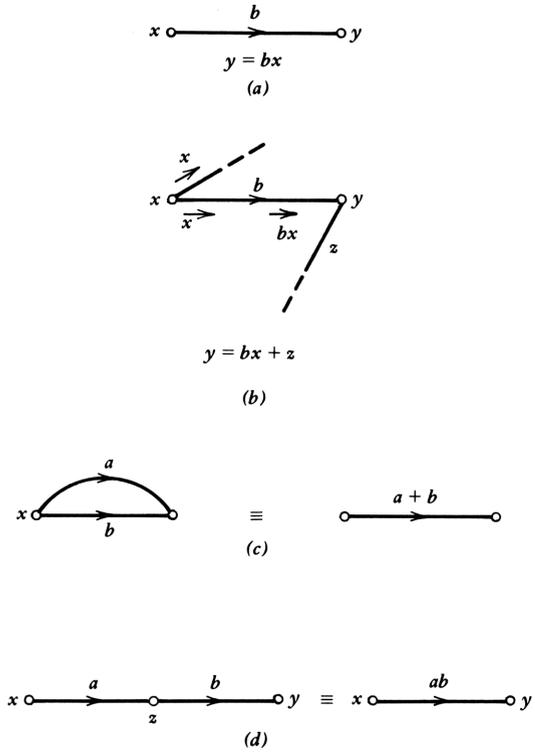


Figure 4.1. Basic signal flow graphs.

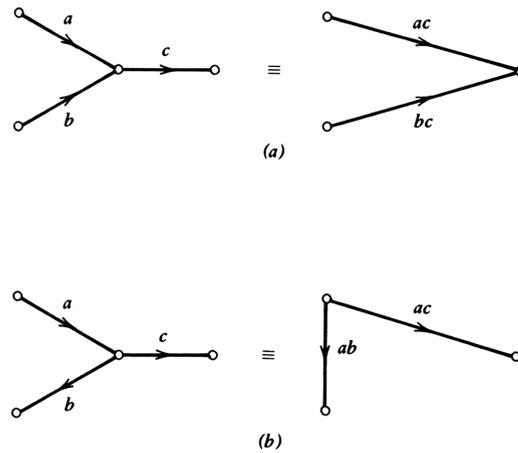


Figure 4.2. Factoring and distribution.

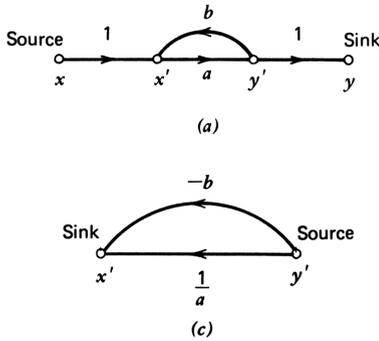


Figure 4.3. A signal flow graph containing a closed loop. Sequential matrices for this graph are shown in parts b and c.

are evident, they may be added by adding a branch with branch value unity between it and the existing graph node(s). Thus in Fig. 4.3, node x was added as a source node and y as a sink node, with unity gain branches connecting x to x' and y' to y . Thus, by inspection of the graph, we set

$$\begin{aligned}
 y &= y' \\
 y' &= ax' \\
 x' &= x + by'
 \end{aligned}
 \tag{4.1-1}$$

from which we obtain

$$y = \frac{ax}{1-ab}
 \tag{4.1-2}$$

Alternatively, we can sum all path products from input to output as the signal traverses from x to x' and y' to y . The first such path has path product a ; to this we add the path x to x' and y' to y through b to x' again through a to y' and to y ; a third path flows around the feedback path twice, and so forth for an infinite number of traverses, giving the sum

$$y = \{a[1 + ab + (ab)^2 + (ab)^3 + \dots]\}x
 \tag{4.1-3}$$

summing the series, with $ab < 1$,

$$y = \frac{ax}{1-ab}
 \tag{4.1-4}$$

as before. Where $|ab| \ll 1$, the first few terms of (4.1-3) may be sufficiently accurate to avoid having to deal with the denominator of (4.1-4).

Sequential Matrices

The signal flow graph shown in Fig. 4.3 is the same as the canonical feedback diagram in Fig. 1.1 as given by Black. The graph in Fig. 4.3a may be represented as a matrix, shown in Fig. 4.3b. Each *transmitting* node in the graph appears as a *column* label, with the source node at the left. Each *receiving* node appears as a *row* label with the sink node at the bottom. The value of a branch connecting two nodes appears in the matrix where the row and column intersect. For example, node y' receives signal transmitted from node x' through branch a , so that a appears in row y' and column x' .

In examining the topological structure of mathematical relationships, a similar matrix, called an *adjacency matrix* or *incidence matrix*, in which the branch values are replaced by ones merely to indicate their existence, is also used. For an interesting discussion of such matrices and some of the problems that they deal with, see Martin Gardner's column in *Scientific American*, March 1980.²

The matrix in Fig. 4.3b includes not only the branch values, but *orders* the nodes in sequence from source to sink to correspond with our concept of the flow of signals through the system. To emphasize this distinguishing feature—one that includes our concept of how the system works—we term this a *sequential matrix*. In Fig. 4.3b the principal diagonal describes the forward path causally, and we call the corresponding matrix a *causal sequential matrix*. Conversely, where the forward path is described under the reciprocal formulation, the associated matrix is termed an *anticausal sequential matrix*.

In this way, the forward path appears along the principal diagonal of the matrix, and the feedback path elements appear *below* it. We see in the following that in an *anticausal sequential matrix*, on the other hand, the feedback path elements appear *above* the principal diagonal.

Note that to write a sequential matrix from a signal flow graph, no nodes may exist off the path from source to sink; this path must contain all signal flow graph nodes. It is always possible to arrange a signal flow graph in this way (if we allow branches with zero branch value) since branches that span more than one node (in either direction) are allowed.

In Fig. 4.3c the signal flow graph is shown for the same structure, except that the forward path a is defined anticausally—the loss is $1/a$. The sign of b changes as a result of the revision of functional dependencies, as described in Chapter 1. Note that node x' is now a sink node, so that the unity gain branch between x and x' can be dropped; similarly, y' is now a source node, so that the other unity gain branch can also be dropped, leading to the two-node graph in Fig. 4.3c. The anticausal sequential matrix degenerates to a *single element* here.

A signal flow graph between source and sink nodes containing no closed loops is termed a *cascade graph*. It is particularly easy to evaluate because one merely adds all path products between the source and sink nodes. No denominators appear in the total. Conversely, loops appearing in a signal flow graph are equivalent to denominators that are not unity.

4.2 SIGNAL FLOW GRAPHS OF CIRCUITS

Signal flow graphs offer us a particularly clear way of deciding how we wish to analyze a circuit; it makes specific the ordering of functional dependencies.

Active Low-Pass Filter

As an example of the use of signal flow graphs and sequential matrices, consider the low-pass RC active filter in Fig. 4.4.³ The circuit incorporates an operational amplifier that we take (for the moment) to be ideal, as it has no input current to either input lead and no differential voltage between the two input leads for any finite output signal.

Taking the output voltage as the independent signal variable, the voltage between either input lead and ground is nv_o , where $n=R_4/(R_3+R_4)$, as shown in the signal flow graph. The rest of the flow graph follows directly, as does the sequential matrix. Thus v_i implies that current $v_i C_2 s$ flows through C_2 and R_2 , giving voltage v_n as $nv_o(1+R_2 C_2 s)$. Given this voltage and the output voltage, we can determine the current through C_1 as $[n(1+R_2 C_2 s)-1]C_1 s v_o$; finally, we add this to i_2 to determine the input current, which is multiplied by R_1 to obtain the drop across R_1 . To this we add v_n to obtain e_G . The process is straightforward because there are no feedback loops in the signal flow graph, thus allowing the loss to be written by enumerating all paths from v_o to e_G :

$$\frac{e_G}{v_o} = n \left\{ 1 + \left[(R_1 + R_2) C_2 + R_1 C_1 \left(1 - \frac{1}{n} \right) \right] s + R_1 R_2 C_1 C_2 s^2 \right\} \quad (4.2-1)$$

Since $0 \leq n \leq 1$, the damping factor for this circuit can be made as small as desired (even negative) by a suitable choice of n .

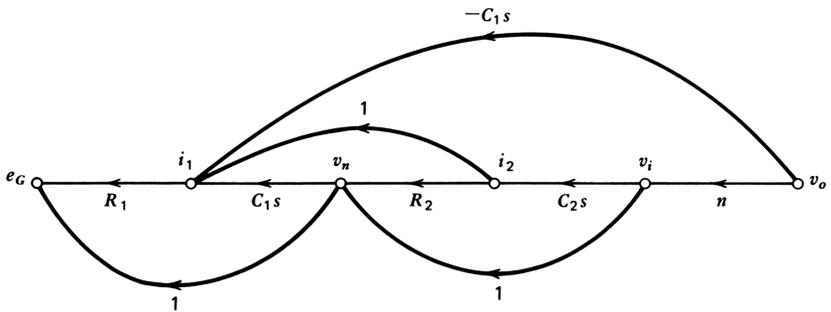
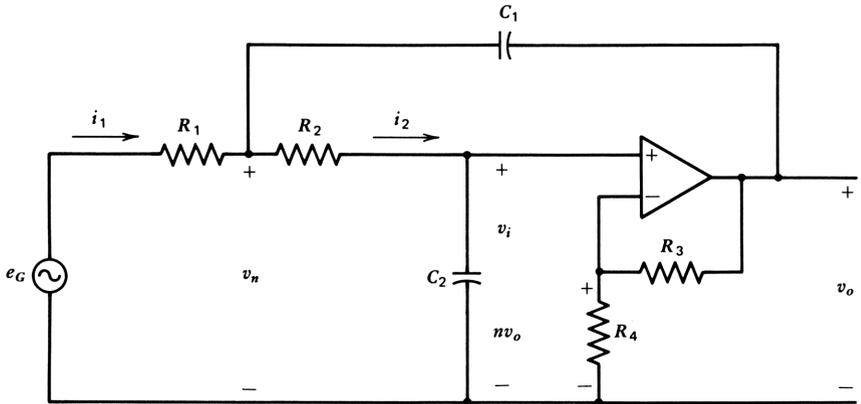
In this analysis the operational amplifier was assumed to be ideal. Actual operational amplifiers approach the ideal, except that the voltage between the two input leads increases with frequency; the voltage loss is not zero. Ordinary voice frequency operational amplifiers are arranged to have a differential input voltage v_d given approximately by

$$v_d = \tau_1 s v_o \quad (4.2-2)$$

where τ_1 is the *unity voltage loss time constant*.^{*} We can incorporate this nonideal behavior simply by replacing n in eq. (4.2-2) by $n + \tau_1 s$, which is equivalent to adding a branch $\tau_1 s$ between v_o and v_i in the signal flow graph. In the sequential matrix, $\tau_1 s$ is added to n wherever the latter appears.

The feedback network consisting of R_3 and R_4 adds an input *voltage* $[R_3/(R_3+R_4)]v_o$ to the operational amplifier input. This is quite different from the feedback networks of the case studies in Chapter 3, in which an input *current* proportional to the output voltage was added. We can develop an

^{*}For the type 741 operational amplifier, for example, $\tau_1 \approx 160$ ns, giving a unity gain (or loss) frequency of 1 MHz.



		Transmitting nodes				
		i_1	v_n	i_2	v_i	v_o
Receiving nodes	e_G	R_1	1	0	0	0
	i_1	0	$C_1 s$	1	0	$-C_1 s$
	v_n	0	0	R_2	1	0
	i_2	0	0	0	$C_2 s$	0
	v_i	0	0	0	0	n

Figure 4.4. A low-pass RC active filter (Sallen-Key), showing its anticausal signal flow graph and sequential matrix. No closed loops appear.

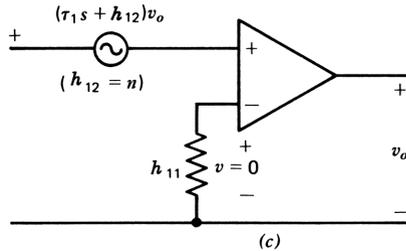
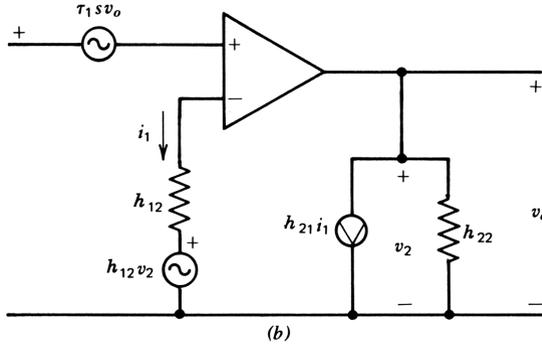
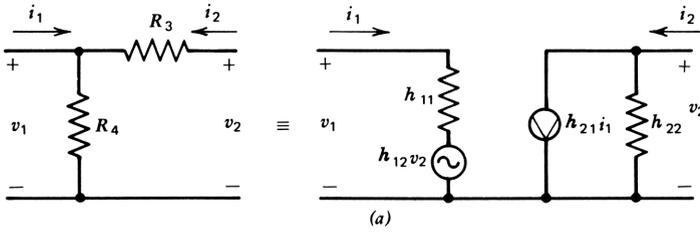


Figure 4.5. Development of the resistive feedback path as an equivalent ladder circuit.

equivalent ladder network for this case. In Fig. 4.5 the feedback network is replaced as before by a two-port equivalent circuit. In this case, however, we use the h parameters to describe the feedback network; the h parameters are defined by

$$\begin{aligned} v_1 &= h_{11}i_1 + h_{12}v_2 \\ i_2 &= h_{21}i_1 + h_{22}v_2 \end{aligned} \tag{4.2-3}$$

The h parameters are given in terms of the resistances R_3 and R_4 by

$$[h] = \begin{bmatrix} \frac{R_3 R_4}{R_3 + R_4} & \frac{R_4}{R_3 + R_4} \\ -\frac{R_4}{R_3 + R_4} & \frac{1}{R_3 + R_4} \end{bmatrix} \tag{4.2-4}$$

The equivalent network is connected to the operational amplifier as in Fig. 4.5*b*. Just as in Section 3.2, the four two-port parameters correspond to input loading (h_{11}), feedback (h_{12}), direct feedthrough (h_{21}), and output loading (h_{22}). Parameters h_{21} and h_{22} affect only the output *current* of the operational amplifier; by our assumption (a good one) that the input voltage and current are not functions of the output current, neither parameter affects circuit operation and does not appear in the expression for the loss. According to our assumption that the amplifier input current is zero, h_{11} is similarly absent since no voltage appears across h_{11} , so that the only effect of the feedback is to add voltage generator $h_{12}v_o$ in series with the negative input lead. This generator can equally well be placed in series with the positive input lead, as shown in Fig. 4.5*c*, where it is in series with the forward path amplifier loss generator $\tau_1 v_o$. The circuit in Fig. 4.5*c* can be used to draw an equivalent ladder circuit for the active filter. Such an analysis serves the function of making specific our assumptions about the circuit (e.g., that the amplifier input is unaffected by the output loading or direct feedthrough).

When the output is connected directly to the negative input in the circuit in Fig. 4.4, $n=1$, and the resulting circuit is called a *unity gain follower*. In this case (4.2-4) gives $h_{11}=h_{22}=0$ and $h_{12}=-h_{21}=1$. Since both loading terms are zero, this is an example of *lossless feedback*. Also, since the feedback network includes no storage elements (inductances or capacitances) the feedback network is termed *nonenergetic*. These topics will be considered further in Part 2.

Design B Output Stage

As a second example of the use of signal flow graphs and sequential matrices, we reconsider the output stage of Design B, previously analyzed in Section 3.2. In that stage the loss of the forward path was not sufficiently small to completely disable the effects of feedback network loading and direct feedthrough. We now see how these subsidiary effects of feedback enter the signal flow graph and sequential matrix.

In Fig. 4.6*a* we have reproduced the equivalent ladder circuit for the stage from Fig. 3.4 and immediately below it, the signal flow graph with nodes corresponding to the signals at circuit nodes in the equivalent ladder. Signal flow graph node i_{c3} , for examples, comprises three components: $G_L v_o$, $G_2 v_o$, and $-G_2 v_{b3}$; all nodes are established in this way. The *anticausal* sequential matrix is shown in Fig. 4.6*b*. Feedback conductance G_2 appears four times in the matrix, corresponding to the four roles of G_2 in the circuit: input loading y_{11} , input augmentation or feedback $-y_{12}$, direct feedthrough $-y_{21}$, and output loading y_{22} . Note that the matrix can be written as a sum of the active path and feedback circuit matrices:

$$\begin{bmatrix} i_{in3} \\ v_{b3} \\ i_{c3} \end{bmatrix} = \left[\begin{bmatrix} y_{11} & 0 & y_{12} \\ 0 & 0 & 0 \\ y_{21} & 0 & y_{22} \end{bmatrix} + \begin{bmatrix} C_3 s & 0 & 0 \\ 0 & -r_3 & 0 \\ 0 & 0 & G_L \end{bmatrix} \right] \begin{bmatrix} v_{b3} \\ i_{c3} \\ v_o \end{bmatrix} \quad (4.2-5)$$

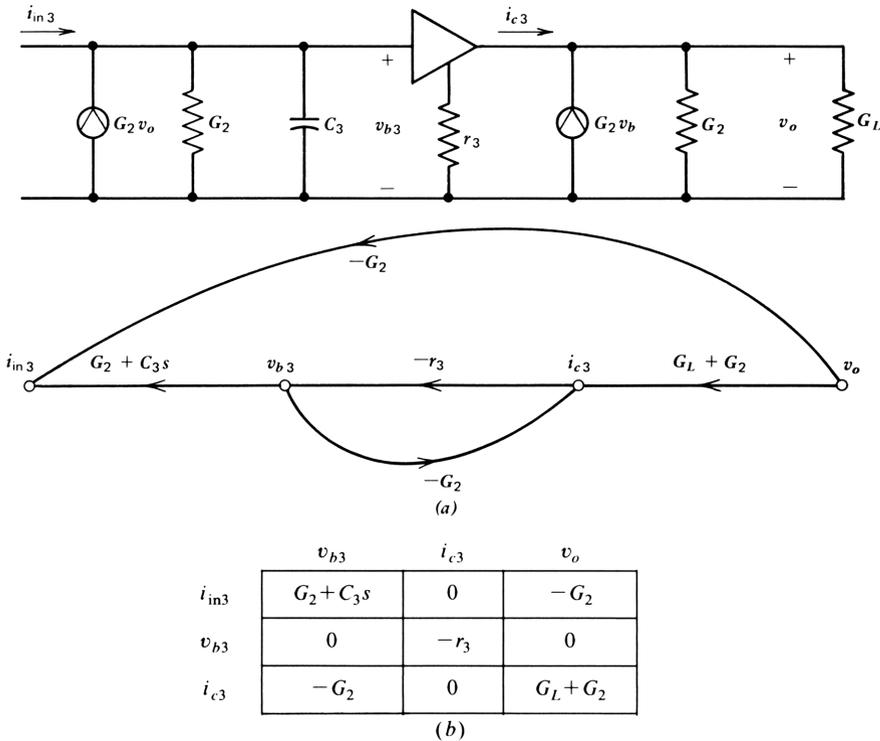


Figure 4.6. Signal flow graph and sequential matrix for the third stage of Design B for the anticausal formulation.

In the total, only the direct feedthrough element $y_{21} = -G_2$ appears below the principal diagonal. Even though $y_{21} = y_{12}$, y_{21} is usually unimportant because the input voltage is much smaller than the output voltage (we see counterexamples in Section 4.4). Hence the resulting matrix is essentially an *upper triangular matrix*, a matrix in which all elements below the principal diagonal are zero. The signal flow graph corresponding to an upper triangular matrix is a cascade graph. Note that the sequential matrix in Fig. 4.3 is not upper triangular but that the matrix in Fig. 4.4 is. The matrix on the left in eq. (4.2-5) is the two-port y matrix discussed in Section 3.2, built out to include signal node i_{c3} ; the latter is needed to describe the matrix on the right, a diagonal matrix that gives the loss of the active path.

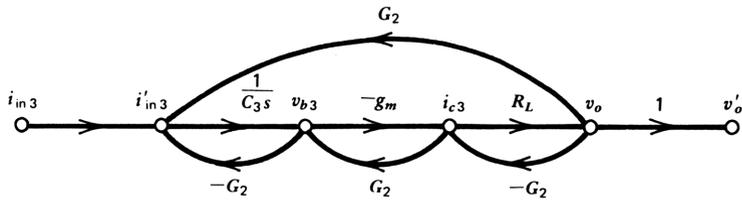
The graph contains one loop; it can be removed by replacing the branch from i_{c3} to v_{b3} by $-r_3/(1 - G_2r_3)$. In Design B this amounted to a 1% correction or r_3 . When this is done, the loss i_{in3}/v_o is found by adding the paths:

$$i_{in3}/v_o = -G_2 - \frac{(C_3s + G_2)r_3(G_L + G_2)}{1 - r_3G_2} \tag{4.2-6}$$

as before.

We can now compare the causal and anticausal formulations of the feedback problem for this simple (but practical) case. We begin as before with the active path, but we invert each of the branches, as shown in Fig. 4.7, to give the output as a function of the input. We then add the four effects of the feedback to it, as shown in Fig. 4.7. When this is done, we find that we no longer have source and sink nodes, so that we add new ones, giving the complete signal flow graph for the circuit. The (causal) sequential matrix is given in Fig. 4.7*d*; it has increased to a 5×5 matrix to accommodate the feedback network elements. The same y -parameter description of the feedback conductance has now moved below the principal diagonal, where it forces us to solve simultaneous equations to account for the effects of y_{11} , y_{12} , and y_{22} . The usually unimportant y_{21} is on the principal diagonal, and is incorporated by simple addition. This matrix can be split into two portions, as in the anticausal case:

$$\begin{bmatrix} v'_o \\ v_o \\ i_{c3} \\ v_{b3} \\ i'_{in} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ y_{22} & 0 & y_{21} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ y_{12} & 0 & y_{11} & 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & R_L & 0 & 0 & 0 \\ 0 & 0 & -g_m & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{C_3 s} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{bmatrix} \begin{bmatrix} v_o \\ i_{c3} \\ v_{b3} \\ i_{in} \\ i'_{in} \end{bmatrix} \tag{4.2-7}$$



	v_o	i_{c3}	v_{b3}	i'_{in3}	i_{in3}
v'_o	1	0	0	0	0
v_o	0	R_L	0	0	0
i_{c3}	$-G_2$	0	$-q_3 + G_2$	0	0
v_{b3}	0	0	0	$1/C_3 s$	0
i'_{in3}	G_2	0	$-G_2$	0	1

Figure 4.7. Causal formulation of the signal flow graph and sequential matrix for Design B.

The causal description is at its essence more complex than the anticausal description, not because the matrix has grown larger, necessarily, but because the feedback elements make their major contributions below the principal diagonal of the sequential matrix. Evaluation of the gain from (4.2-7) entails considerably more work than evaluation of the loss from eq. (4.2-5), although the results of the two computations must be reciprocals of one another. It might be argued that with the computer available, the work is easily done in either case. But we are interested in more than analysis; we wish to retain control over a *design*. For this purpose, and for the purpose of understanding, the simpler description is the more appropriate one.

The anticausal formulation is simpler because the effect of the feedforward of y_{21} is much smaller than that of the feedback of y_{12} . With $y_{12}=y_{21}$, the reason for this disparity resides in the active path itself. The signal level at the output is much larger than that at the input; hence the signal flow from output to input is much larger than that from input to output. Furthermore, the direct feedthrough signal at the output is compared with a large output signal, whereas the feedback signal at the input is compared with the very small signal at the active path input. Thus even a modest amount of gain allows us to ignore the direct feedthrough.

4.3 FEEDBACK DESCRIPTION OF AN ACTIVE RESONATOR

A circuit that combines the two types of feedback shown in Figs. 4.4 and 4.6 is the bandpass resonator shown in Fig. 4.8.⁴ Negative feedback of the y -parameter type is connected around the operational amplifier through R_A and C_B . Positive feedback of the h parameter type is applied to the positive input lead of the amplifier through a resistive divider, represented as a dependent voltage generator, as in Fig. 4.4. The circuit is called a *single-amplifier biquad*, so called because with the use of additional resistors (which may or may not be connected for different applications), it is capable of realizing any rational function consisting of a ratio of quadratic polynomials⁵:

$$L(s) = \frac{a_0 + a_1s + a_2s^2}{d_0 + d_1s + d_2s^2} \quad (4.3-1)$$

In the form shown in Fig. 4.8, the circuit is a simple resonator with d_0 and d_2 equal to zero and $d_1=1$. Control of the damping of the numerator is effected by the positive feedback network consisting of R_3 and R_4 , which applies a portion n of the output voltage to the *positive* input of the operational amplifier (opposite to that of the Sallen-Key circuit of Fig. 4.4). Our purpose is to understand how (1) the circuit functions and (2) to adjust the amount of positive feedback to minimize variations in the resonant frequency or the Q value of the resonator in the presence of variations in the passive circuit elements and the amplifier.⁶

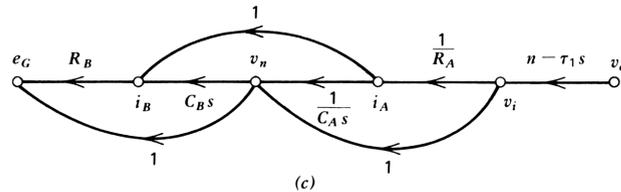
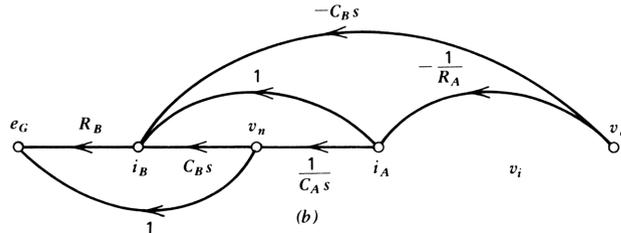
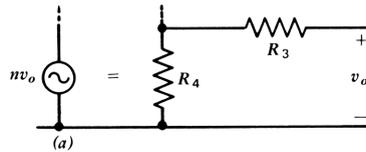
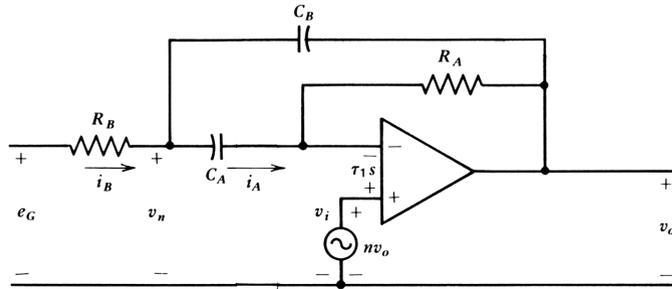


Figure 4.8. (a) Single-amplifier Biquad active resonator circuit; (b) signal flow graph for y -parameter feedback portion of circuit giving L_y ; (c) signal flow graph for h -parameter feedback and amplifier portion giving L_n and L_a .

Analysis by Separation

Analysis of the circuit is easily done by treating the y - and h -parameter feedback networks separately and superimposing their effects. A signal flow graph for the y feedback is shown in Fig. 4.8b; in this graph we assume the positive feedback to be zero and the amplifier to be ideal ($\tau_1=0$), so that the signal at node v_i is zero. The graph transmission is the portion of the circuit loss attributable to the y feedback and can be written by inspection:

$$L_y(s) = - \left[\frac{1}{R_A C_A s} + \frac{R_B}{R_A} \left(1 + \frac{C_B}{C_A} \right) + R_B C_B s \right] \tag{4.3-2}$$

A signal flow graph for evaluating the h -feedback loss is shown in Fig. 4.8c and is drawn assuming the y feedback to be absent (e.g., by activating the nv_o node, but assuming the signal at v_o itself to be zero). In this graph both the positive feedback and the amplifier loss voltages appear in series. Since we are interested in the separate contributions of the amplifier loss and the positive feedback, we write equations for their losses separately, again by inspection:

$$L_h(s) = n \left[\frac{1}{R_A C_A s} + \frac{R_B}{R_A} \left(1 + \frac{C_B}{C_A} \right) + 1 + R_B C_B s \right] \quad (4.3-3)$$

and for the amplifier

$$L_a(s) = -\tau_1 \left(\frac{1}{R_A C_A} + \right) \frac{R_B}{R_A} \left(1 + \frac{C_B}{C_A} \right) + 1 (s + R_B C_B s^2) \quad (4.3-4)$$

The loss of the circuit including the two types of feedback loss and the amplifier loss is given by

$$L(s) = L_y(s) + L_h(s) + L_a(s) \quad (4.3-5)$$

As usually designed, $L_y(s)$ provides a resonance with low Q , perhaps from 1 to 4 (a damping of 0.5 to 0.125). At resonance L_y and L_h are both real since the first and third terms of (4.3-2) cancel, as do those of (4.3-3). The two losses are of opposite sign, however, so the damping reduced, or the Q value is increased. The combined effect at resonance ($s = j\omega_0$) of $L_y(s)$ and $L_h(s)$ is to *reduce* the loss compared with that of each of them separately. Furthermore, $|L_y(j\omega_0)| > |L_h(j\omega_0)|$ for the circuit to be stable; the negative feedback is of larger magnitude than the positive feedback. Note that at resonance $L_a(j\omega_0)$ is purely imaginary because of the assumed form of the amplifier loss $-j\tau_1\omega$. In addition, the loss of the combined feedback networks should be much *larger* than the amplifier loss if the amplifier variations are not to affect circuit operation appreciably.

Resonator Description

These considerations are all expressed in the Bode diagram in Fig. 4.9 for a “good” resonator design. The three component losses are shown separately, as well as the combined feedback loss, and are drawn with straight-line (asymptotic) approximations except near resonance. Four frequencies are defined in the Bode diagram in Fig. 4.9 with respect to the unity loss line: from eq. (4.3-2) we obtain

$$\omega_A = \frac{1}{R_A C_A} \quad (4.3-6)$$

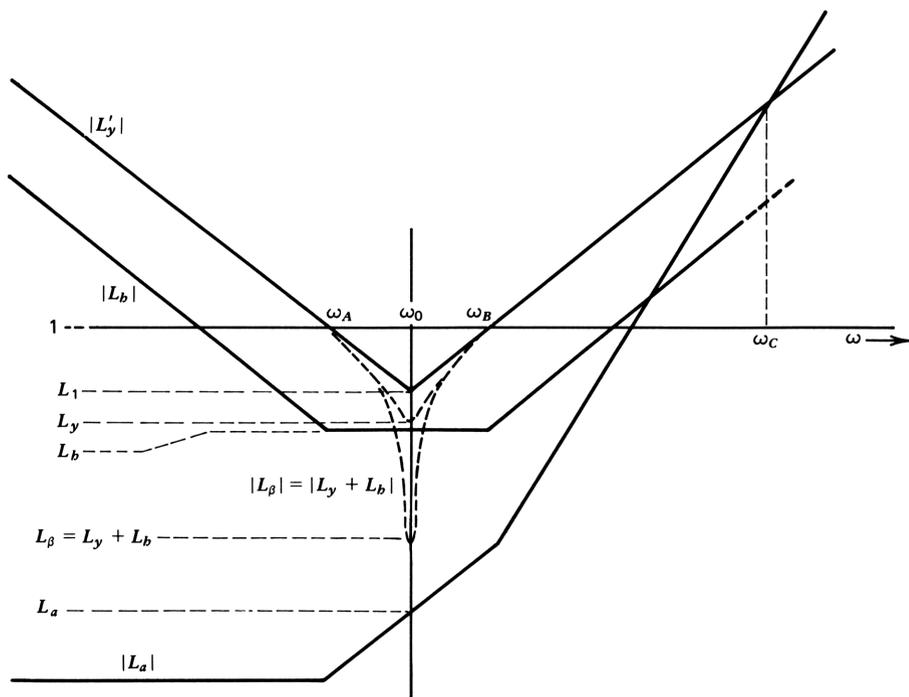


Figure 4.9. Bode diagram showing separate loss components of the active resonator.

and

$$\omega_B = \frac{1}{R_B C_B} \quad (4.3-7)$$

The resonant frequency is the geometric mean:

$$\omega_0 = \frac{1}{\sqrt{R_A C_A R_B C_B}} \quad (4.3-8)$$

Finally, the amplifier loss becomes dominant above the *crossover frequency*:

$$\omega_C = \frac{1-n}{\tau_1} \approx \frac{1}{\tau_1} \quad (4.3-9)$$

so that the crossover frequency is about at the unity loss frequency of the amplifier. In a good design, as we see later, $n \ll 1$ for this circuit.

At this point we introduce the mathematical simplification that $C_A = C_B$. It is easily shown that from eq. (4.3-2) the damping is minimized for L_y when this is the case, but it is not as easily shown that the sensitivities are thereby minimized. Since the development is clarified by the assumption we adopt it.

At resonance, five loss parameters are defined in the Bode diagram in Fig. 4.9, all of which are used presently. They are, in order of decreasing magnitude, L_1 , L_y , L_n , L_β , and L_a ; L_1 is the loss at which the asymptotes of L_y intersect. This is the point from which we measure the circuit Q values; from purely geometrical considerations we obtain

$$L_1 = -\sqrt{\frac{R_B}{R_A}} \quad (4.3-10)^*$$

From (4.3-2) the negative feedback loss of the y network at resonance is

$$L_y = -2\frac{R_B}{R_A} \quad (4.3-11)$$

From (4.3-3) the positive feedback of the h network at resonance is

$$L_h = n\left(1 + 2\frac{R_B}{R_A}\right) \simeq n \quad (4.3-12)$$

Loss L_β is the total feedback loss, the sum of L_y and L_h :

$$L_\beta = L_y + L_h = -2\frac{R_B}{R_A}(1 - n) + n \quad (4.3-14)$$

Note that L_y is negative and L_h is positive, so that L_β is smaller than either, as shown in Fig. 4.9. This is the mechanism by which Q is increased through the positive feedback. Finally, from (4.3-4) the amplifier loss at resonance is

$$L_a = -j\tau_1\omega_0\left(1 + 2\frac{R_B}{R_A}\right) \simeq -j\tau_1\omega_0 \quad (4.3-13)$$

For the case where the loss is feedback controlled (so that the amplifier does not affect circuit operation greatly), we also have the approximation that the loss L of the circuit at resonance is close to L_β .

We can now define Q (the nominal Q of the whole circuit) and Q_0 (the Q associated with L_y) in terms of the loss parameters. Thus from (4.3-10) and (4.3-11) we obtain

$$\frac{1}{Q_0} \triangleq \frac{L_y}{L_1} = 2\sqrt{\frac{R_B}{R_A}} \quad (4.3-15)$$

*The sign of L_1 is taken as negative to agree with L_y . For stable operation, the complete circuit includes a phase reversal:

and with (4.3-12),

$$\frac{1}{Q} = \frac{L_y + L_h}{(1-n)L_1} \simeq \frac{L_y + L_h}{L_1} = \frac{L_\beta}{L_1} \quad (4.3-16)$$

Also, the change in Q is given by

$$\frac{1}{Q} - \frac{1}{Q_0} = \frac{L_h}{L_1} \quad (4.3-17)$$

For a strictly mathematical development, it is unnecessary to introduce these expressions for Q , but it is intuitively useful to do so.

Sensitivities

We can now direct the sensitivities of loss to each of the three components of loss by direct application of the sum rule. The sensitivity of loss to L_y from (4.3-15) and (4.3-16) is

$$S_{L_y}^L = \frac{L_y}{L_y + L_h + L_a} \simeq \frac{L_y/L_1}{L_\beta/L_1} = \frac{Q}{Q_0} \quad (4.3-18)$$

and that to L_h from (4.3-17) is

$$S_{L_h}^L = \frac{L_h}{L_y + L_h + L_a} \simeq \frac{L_h/L_1}{L_\beta/L_1} = -\frac{1/Q - 1/Q_0}{1/Q}$$

$$S_{L_h}^L = 1 - \frac{Q}{Q_0} \simeq -\frac{Q}{Q_0} \quad (4.3-19)$$

The magnitudes of the sensitivities of loss to each of the feedback components at resonance is thus given by the vertical distance between the indicated loss parameters in the Bode diagram. The sensitivity to the amplifier loss is given by

$$S_{L_a}^L = \frac{L_a}{L_y + L_h + L_a} \simeq \frac{L_a}{L_\beta} = \frac{j\tau_1\omega_0 Q}{\sqrt{R_B/R_A}} \quad (4.3-20)$$

From (4.3-15) this becomes

$$S_{L_a}^L = 2jQQ_0\tau_1\omega_0$$

and is found by adding (and subtracting) the appropriate vertical intervals along the ordinate at ω_0 .

Effect of Amplifier Variations

The deviation of the loss due to variations in τ_1 is given by

$$\sigma_{L_a} = \sigma_A S_{L_a}^L S_{\tau_1}^{L_a} \quad (4.3-21)$$

in which σ_A is the standard deviation of τ_1 ; the magnitude of the sensitivity of L_a to τ_1 is unity, so we can write

$$\sigma_{L_a} = 2\sigma_A Q Q_0 \tau_1 \omega_0 \quad (4.3-22)$$

for the standard deviation of loss due to amplifier variations.

The sensitivity expressions developed in the preceding paragraphs also can be used to determine the variation of loss that results from variation in the passive elements, as shown in Chapter 3. To shorten the story, we ignore variations in the capacitors; it can be shown that if their variation is of the same order as that of the resistors, their contribution to the loss variation is small; the details can be found in Fleischer⁶. At resonance, it should be noted that variation in the capacitors varies the resonant frequency as well as the real part of the loss. Here, as in Chapter 3, we are interested in the *magnitude* of the loss variation, $|\Delta L/L|$; its real part represents the change in magnitude of L and its imaginary part, the phase. The change in resonant frequency can be found from the latter at resonance by the standard relationship between the phase and frequency of a tuned circuit at resonance:

$$\frac{\Delta\omega}{\omega} = \frac{\theta}{2Q} \quad (4.3-23)$$

where θ is the phase shift in radians corresponding to the imaginary part of $\Delta L/L$. By finding the magnitude of $\Delta L/L$, we are simply finding an upper bound on the change in the real and imaginary parts of the loss variation. Where the Q change or the resonant frequency change is individually of special interest, a two-dimensional analysis should be used.

Effect of Passive Component Variations

To find the sensitivities of L to the passive components, we use relation 9 from Table 1.1, which can be written (e.g., taking sensitivity of loss to R_B)

$$S_{R_B}^L = S_{L_v}^L S_{R_B}^{L_v} + S_{L_h}^L S_{R_B}^{L_h} + S_{L_a}^L S_{R_B}^{L_a} \quad (4.3-24)$$

At resonance, from (4.3-18) and (4.3-19),

$$S_{R_B}^L = \frac{Q}{Q_0} (S_{R_B}^{L_v} - S_{R_B}^{L_h})$$

The third term in (4.3-24) is negligible according to our assumption that the

network is feedback dominated. From (4.3-2), we obtain

$$S_{R_B}^{L_y} = 1 \quad (4.3-25)$$

and from (4.3-3) or (4.3-12), and applying the sum rule,

$$S_{R_B}^{L_h} = \frac{2R_B/R_A}{1 + 2R_B/R_A} \quad (4.3-26)$$

If we use (4.3-3), we must remember that at resonance the dc and quadratic terms cancel. From (4.3-10) and (4.3-15), this expression can be written

$$S_{R_B}^{L_h} = \frac{1}{2Q_0^2 + 1} \quad (4.3-27)$$

where L_h is set by the ratio of R_1 and R_2 and is insensitive to the values of R_A and R_B , so that the sensitivity of L to R_B is set by the L_y contribution. We ignore this term relative to unity, giving

$$S_{R_B}^L \simeq \frac{Q}{Q_0} \quad (4.3-28)$$

Since R_B and R_A always appear as a ratio, the sensitivity of L to R_A is $-Q/Q_0$. A similar argument gives the magnitude of sensitivity of loss to R_3 and R_4 also as about Q/Q_0 .

Physically, the chief determining element of L_y at resonance is the ratio R_B/R_A , and that of L_h is n ; the positive feedback increases the sensitivities of loss to each of the four circuit resistors by the factor Q/Q_0 by the sum rule in which the sum is the small *difference* between relatively large quantities. Hence the sensitivities to the passive elements are increased by the positive feedback by the factor Q/Q_0 .

The optimization procedure can be interpreted in another way. On the log plot in Fig. 4.9 vertical linear distances represent ratios of losses. The distance between the L_y curve and the L_β curve represents an *increase* in sensitivity of loss to the components that make up L_y (e.g., R_A and R_B). The vertical distance between L_β and L_a , on the other hand, represents the *decrease* in sensitivity of the circuit loss L_β to the amplifier time constant. The effect of amplifier time constant variations, which are large, is reduced by the factor L_a/L_β , whereas the effect of R_A and R_B variations, which are small, is increased by the factor L_y/L_β . An optimum is found where the two contributions to the variation are equal. (It is true that the amplifier variations tend to change the resonant *frequency* and the component changes tend to change the Q , but we have chosen to take the *magnitude* of the variation as our error measure.)

The total variation of loss from the resistors is found by adding the individual variations on a root sum of squares (rss) basis so that if the standard deviation of each of the resistors is σ_R , the total variation is $\sqrt{4}\sigma_R = 2\sigma_R$. We take $2\sigma_R$ as the composite standard deviation of the passive elements; for a more detailed summary of the variations from the passive elements, see Fleischer⁶. The deviation from passive element variations can thus be written approximately as

$$\sigma_{L\beta} = 2\sigma_R \frac{Q}{Q_0} \quad (4.3-29)$$

Optimization of the Resonator

The total deviation of the closed-circuit loss is found by adding (4.3-22) and (4.3-29) on an rss basis:

$$\sigma_L^2 = \sigma_{L\alpha}^2 + \sigma_{L\beta}^2$$

so that

$$\sigma_L^2 = (2\sigma_A Q Q_0 \tau_1 \omega_0)^2 + \left(2\sigma_R \frac{Q}{Q_0}\right)^2 \quad (4.3-30)$$

To minimize this expression, we choose Q_0 to make the two contributors equal, so that

$$\left(\frac{1}{Q_0}\right)^2 = \frac{\sigma_A \tau_1 \omega_0}{\sigma_R}$$

Then, from (4.3-15), the optimum ratio of resistors is

$$\frac{R_B}{R_A} = \frac{\sigma_A \tau_1 \omega_0}{4\sigma_R} \quad (4.3-31)$$

Note that the optimum ratio of R_B/R_A does not depend on the required circuit Q ; the effect of the amplifier deviations is reduced and the effect of the passive elements is increased by the same factor as the positive feedback is increased. The *variation* of the loss, however, is directly proportional to Q . For the optimum ratio of R_B/R_A , the closed circuit variation is given by (4.3-30) as

$$\sigma_L = 2Q\sqrt{2\sigma_R\sigma_A\tau_1\omega_0} \quad (4.3-32)$$

The loss variation includes both variation of flat loss and Q variation, and in equal amounts if the variations due to passive elements and the amplifier add

in phase (in general, they do not, but we are content here to ignore the phase information, obtaining an upper bound on the variation of either the loss or the resonant frequency variation.) The factor of 2 can be seen from the Bode diagram, in which both L_1 and $1/Q_0$ are proportional to $\sqrt{R_B/R_A}$. Loss L_1 is proportional to the flat loss variation and $1/Q_0$ to the Q variation, so that an upper bound on the Q variation is given by

$$\sigma_Q \leq Q \sqrt{2\sigma_R\sigma_A\tau_1\omega_0} \quad (4.3-33)$$

An upper bound on the variation of the resonant frequency is given through (4.3-23) by

$$\sigma_{\omega_0} \leq \sqrt{2\sigma_R\sigma_A\tau_1\omega_0} \quad (4.3-34)$$

which is invariant with circuit Q .

Program "OPTRES" in Appendix B calculates the circuit values and deviations for this circuit for given values of the standard deviations of the amplifier time constant, the impedance level, the amplifier time constant, the required Q , and the resonant frequency. For a 741-type amplifier with $\tau_1 = 0.158 \times 10^{-3}$ ms, typical values of $\sigma_A = 0.15$, $\sigma_R = 0.0012$, $Q = 20$, and $\omega_0 = 2\pi$ krad/s (2kHz), the optimum design is that plotted in the Bode diagram in Fig. 4.9. The standard deviation of loss is 3.4%, the Q variation 1.7%, and the percent variation of the resonant frequency, 0.085%. The latter two figures are an upper bound, assuming in each case that all the variation is concentrated in it.

Frequency Compensation

Performance variations of the active resonator are proportional to $\sqrt{\tau_1\omega_0}$, the magnitude of the amplifier loss at the resonant frequency. This loss can be decreased by the use of T frequency compensation, as shown in Fig. 4.10a.

As we see in Chapter 5 (Fig. 5.3), amplifier loss in voice frequency operational amplifiers is usually controlled by an internal feedback capacitance. This controls the amplifier response to give a loss magnitude proportional to frequency. By substituting the T network in Fig. 4.10a for the capacitor, we obtain a double integration over a limited frequency range. In the Bode diagram for the amplifier loss in Fig. 4.10b the loss $\tau_1 s$ has been replaced by

$$L_a(s) = \frac{\tau_1\tau_2 s^2}{1 + \tau_2 s} \quad (4.3-35)$$

in effect providing a double integration below a frequency of $1/\tau_2$, as shown. This frequency is chosen to be much lower than ω_c (to affect the phase of $L_a(j\omega_c)$ as little as possible at crossover) but higher than the resonant frequency of the resonator. At resonance the denominator magnitude of (4.3-35) is roughly unity, and the numerator is reduced by the factor $\tau_2\omega_0$, in effect giving more distance on the Bode diagram between $L_\beta(j\omega_0)$ and

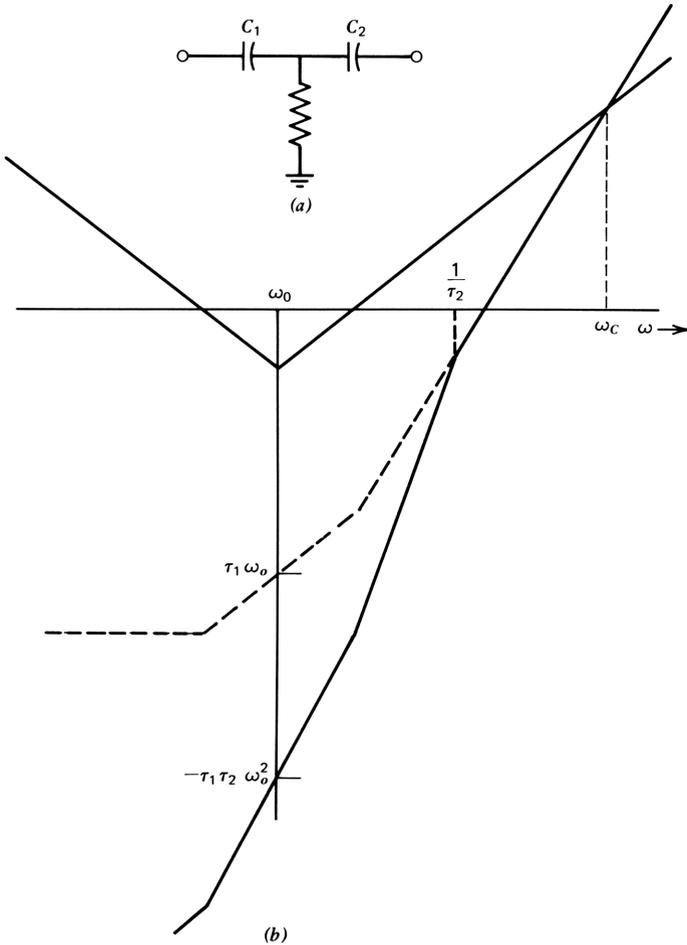


Figure 4.10. Effect of T frequency compensation in the operational amplifier.

$L_a(j\omega_0)$. Such an arrangement considerably enhances the Q that can be achieved at high audio frequencies for a given maximum deviation. In practice, the advantage of frequency compensation is somewhat less than indicated because the amplifier variation increases. As we see in Chapter 5, the loss $j\tau_1\omega$ of the simply compensated operational amplifier is achieved by internal feedback; when this is removed, as by eq. (4.3-35), internal amplifier variations are swamped out less.

The main theme of this section has been that we can separate a circuit such as the resonator into loss components corresponding to the different feedback paths, analyze them separately with individual signal flow graphs and loss equations, and analyze the result simply by adding the components. The sensitivities all have a direct physical interpretation, made explicit on the Bode diagram of the three components of loss.

4.4 SIGNAL FLOW GRAPH REPRESENTATION OF RATIONAL FUNCTIONS

In one sense the design of a feedback system may be regarded as a means for bringing the coefficients of the circuit polynomial into 1:1 correspondence with a performance polynomial, such as one of those described in Section 2.6. Since we have the means to express the circuit loss as a signal flow graph (or its associated sequential matrix), it is intuitively helpful to put the performance polynomial into this form as well. This topological interpretation of polynomials will move us closer to a solution of the problem of selecting a circuit configuration for a proposed feedback system. There are several forms in which polynomials may be represented as signal flow graphs; some useful ones for a cubic polynomial are given in this section. The source node for the graph is shown on the right (to correspond with the system output) and is set equal to unity; the input is then the loss polynomial, represented by the “signal” at the sink node of the graph.

The polynomial as a sum of powers of s (Fig. 4.11a) is represented as a graph with only a source and sink node; one branch is drawn for each term. In this form each branch represents a path product for an individual polynomial coefficient. By dividing an individual path product by the total $L(s)$, we obtain the sensitivity of the loss to the coefficient of that path as we saw in Section 2.6. Each of these paths may be broken down further into sums of products of component values, as shown in Fig. 4.11b for the cubic path, from which we may obtain the component sensitivities. Where the path consists of a single product, the sensitivities of the coefficient to the components of the path are all unity. Where there exist parallel paths, the sensitivities are obtained by the sum rule. The graph in Fig. 4.11a may be considered to contain six branches with values proportional to s , counting an s^i branch as having i s -branches in tandem. The six branches are redundant: only three s branches are required for a cubic polynomial. Two of several possible graphs using the minimum number are given in Fig. 4.12.

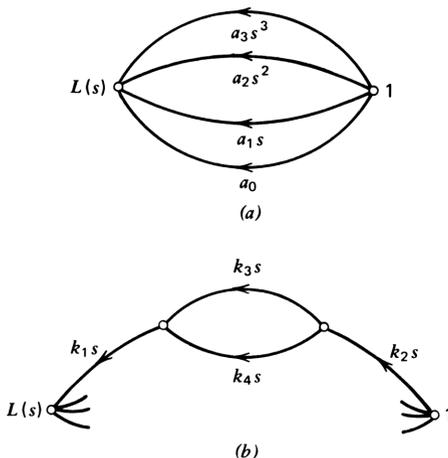
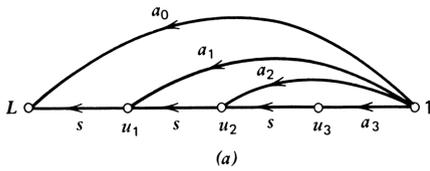
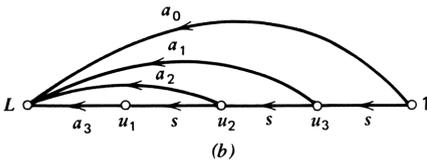


Figure 4.11. Signal flow graph for a polynomial expressed as a sum of powers of s . An example of factors of the cubic term is shown in *b*. Component sensitivities are: $S_{K_1}^{a_3} = S_{K_2}^{a_3} = 1$; $S_{K_3}^{a_3} = K_3/(K_3 + K_4)$; $S_{K_4}^{a_3} = K_4/(K_3 + K_4)$.



	u_1	u_2	u_3	1
L	s	0	0	a_0
u_1	0	s	0	a_1
u_2	0	0	s	a_2
u_3	0	0	0	a_3

(a)



	u_1	u_2	u_3	1
i	a_3	a_2	a_1	a_0
u_1	0	s	0	0
u_2	0	0	s	0
u_3	0	0	0	s

(b)

Figure 4.12. Signal flow graphs and sequential matrices of polynomials in nested form

The graph in Fig. 4.12a corresponds to writing the polynomial in nested form:

$$L(s) = a_0 + s(a_1 + s(a_2 + sa_3)) \tag{4.4-1}$$

The sequential matrix corresponding to the graph is upper triangular. An alternative representation is given in Fig. 4.12b. In each case the a_0 branch represents “overall feedback”—feedback over the entire graph from source to sink, and the a_2 branch represents “local feedback” over a single s branch. Other feedback branches represent intermediate cases; in general, the distance of an element in the matrix from the principal diagonal is a measure of the comprehensiveness of the feedback it supplies.

The polynomial can also be represented in factored form, as shown in Fig. 4.13 for the cubic. Where the polynomial is of odd degree, at least one root must be real, giving the parallel branches s and α in Fig. 4.13. Complex roots occur in pairs and can be represented as shown. The resulting matrix includes elements only along the principal diagonal and along the diagonal immediately above it. Since any polynomial can be uniquely represented by its factors, it

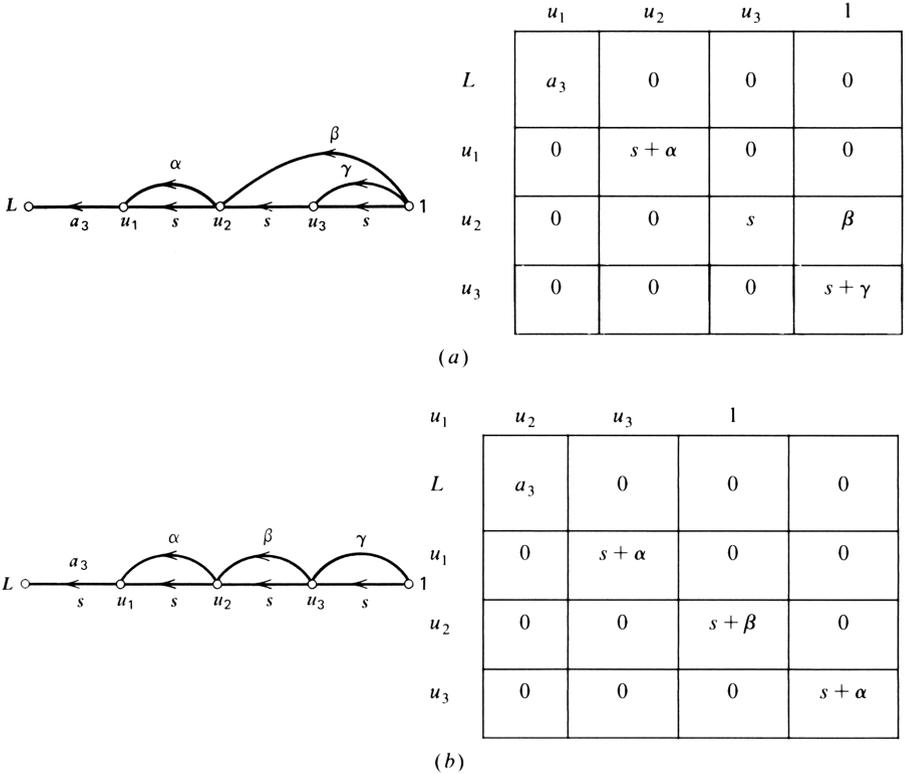


Figure 4.13. Signal flow graphs and sequential matrices of polynomials in factored form.

may be represented in this form, with all matrix elements real. Where all roots are real, the matrix consists only of elements along the principal diagonal. This special case is depicted in Fig. 4.13b.

What prevents us from realizing the polynomials of Fig. 4.12 directly, using a single active device for each place where s is called for, and a feedback path (e.g., passive conductance) wherever a coefficient is called for? The answer is that the feedback elements give components of loss that are of the wrong sign. Using a feedback conductance in the case study designs of Chapter 3, for example, created an input current of $-G_F v_o$, not $G_F v_o$. Feedback over a single stage works because the active element also includes a phase reversal. This is not a coincidence—it is a natural outcome since the single-stage characteristics also arose from passive feedback (e.g., the common lead r in the device in Fig. 2.1). It is also workable for feedback over three stages, or any odd number of stages, but when we come to feedback over two stages, the paradigm breaks down, and we have positive feedback. That is why we needed a capacitor C_F in the amplifier designs of Chapter 3.

In Chapter 2 we also represented polynomials in s by their magnitude and phase for $s=j\omega$ to plot Bode and Nyquist diagrams. Signal flow graphs for the

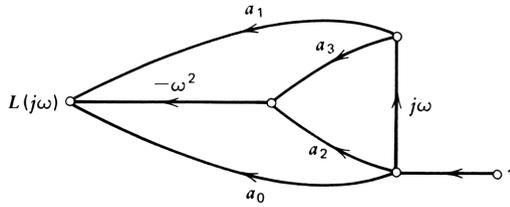
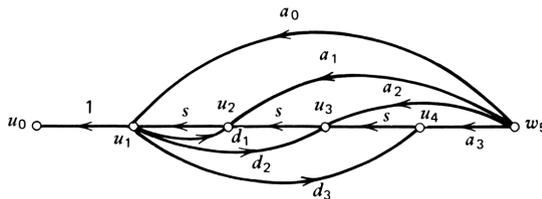


Figure 4.14. Signal flow graph for the evaluation of a polynomial for $s = j\omega$ (in nested form).

cubic in this form are shown in Fig. 4.14. The point here is that a signal flow graph can be drawn for any of the several ways of expressing polynomials that were discussed in Chapter 2. The signal flow graph is a way of representing algebraic relationships graphically. Whether equations, matrices, or flow graphs are used is a matter of personal preference; it is useful to be familiar with more than one method to give us the means for checking our results in difficult or (temporarily) confusing situations.

Signal Flow Graphs of Ratios of Polynomials

Rational functions or ratios of polynomials can also be represented by signal flow graphs or matrices. Where a denominator polynomial is present, the associated signal flow graph will contain loops—that is, will not be a cascade graph. The signal flow graph and sequential matrix for a ratio of two cubic polynomials are shown in Fig. 4.15. The denominator is found as the sum of all



u_0	1	0	0	0	0
u_1	0	s	0	0	a_0
u_2	d_1	0	s	0	a_1
u_3	d_2	0	0	s	a_2
u_4	d_3	0	0	0	a_3

Figure 4.15. Signal flow graph and sequential matrix for a rational function.

loop gains subtracted from unity for this case in which all feedback loops “touch”—that is, have at least one node in common. The graph gain is given by

$$L(s) = \frac{a_0 + a_1s + a_2s^2 + a_3s^3}{1 - d_1s - d_2s^2 - d_3s^3} \quad (4.4-2)$$

As before, loops in the graph give entries in the sequential matrix below the principal diagonal. For an anticausal sequential matrix, such entries do not affect the characteristic equation of the system, so that stability of the system, for example, is not affected. It should not be inferred that denominators never affect stability, however. Where the direct feedthrough loop does not touch a feedback branch as in the signal flow graph in Fig. 4.6*a*, for instance, the feedback branch stands alone when added to the forward-path loss, with the latter including a denominator. Therefore, the feedback branch must be multiplied by the denominator to put the whole expression over a common denominator, as in eq. (4.2-6).

In low-pass systems containing no denominator of loss, or where the departure of the denominator from unity is incidental, we have seen that we can eliminate loop gains from the system description by suitable choice of independent signal variables in the circuit. Can this also be done for systems that have nonunity denominators? As we have seen, such systems are characterized by having a feedforward path, or a direct feedthrough path. If we select some internal node as the independent variable, loop gains can indeed be eliminated in the general case, as we shall show. Let the system loss be defined by the ratio of two polynomials:

$$\frac{e_G(s)}{v_o(s)} = \frac{N(s)}{D(s)} \quad (4.4-3)$$

We can then write

$$\frac{e_G(s)}{N(s)} = \frac{v_o(s)}{D(s)} \triangleq v_n(s) \quad (4.4-4)$$

where $v_n(s)$ represents the internal signal node we seek, at least within a constant factor [since we do not have information on the separate constant factors that might multiply $N(s)$ and $D(s)$ independently]. Whereas v_n is depicted here as a signal voltage, it might be a current or other signal variable in mechanical systems, for example. We can then write

$$e_G = N(s)v_n \quad (4.4-5)$$

and

$$v_o = D(s)v_n \quad (4.4-6)$$

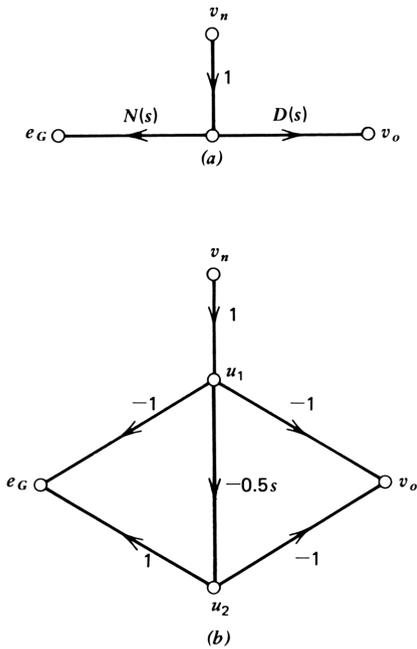


Figure 4.16. Cascade signal flow graphs for an all-pass rational function.

By this step we have divided the system into two: the first equation is the anticausal representation of the feedback portion of the system, and the second is the causal representation of the feedforward portion. Since both portions are defined in terms of polynomials, loop gains have been eliminated from both; each of the two equations is representable by a cascade signal flow graph or an upper triangular sequential matrix. The numerator and the denominator can then be considered separately, although in a design, the two portions will have common components, such as integrators. A signal flow graph for the two equations is given in Fig. 4.16a.

All-Pass Network

As an example of the use of this technique in the design of a circuit, suppose that we wish to realize a simple all-pass network (so called because the magnitude of the loss is uniform over the entire frequency range):

$$L(s) = \frac{1 + 0.5s}{1 - 0.5s} = \frac{N(s)}{D(s)} \tag{4.4-7}$$

Then we have

$$e_G = (1 + 0.5s)v_n \tag{4.4-8}$$

and

$$v_o = (1 - 0.5s)v_n \tag{4.4-9}$$

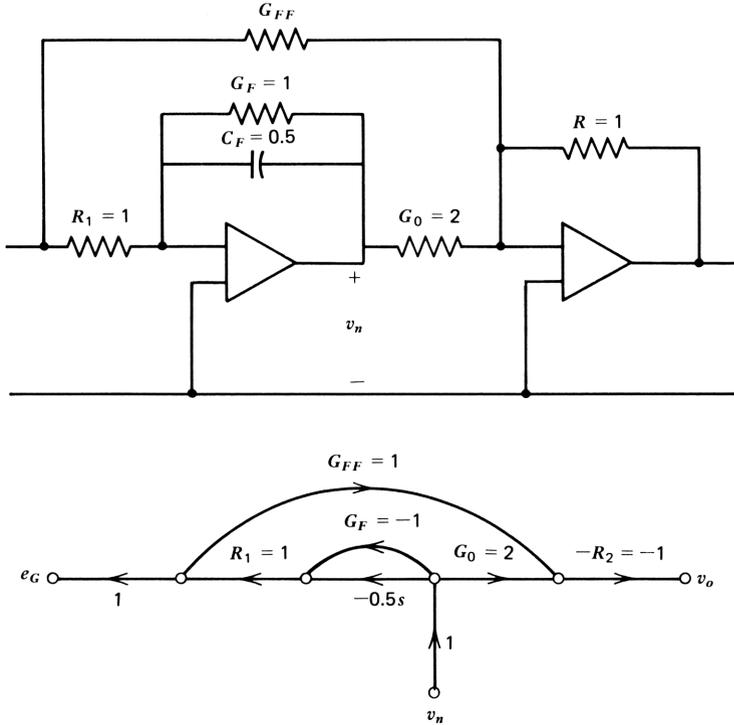


Figure 4.17. A circuit realization.

A signal flow graph for these two equations is given in Fig. 4.16b, in which we have chosen to give the $0.5s$ branch a negative sign for ease of realization since simple integrators usually include a phase reversal. A circuit that realizes the desired function is shown in Fig. 4.17a and its signal flow graph, in Fig. 4.17b. The latter differs from the signal flow graph in Fig. 4.16b because the summing node current of the integrator is not directly *observable*; we must pass it through the input resistor to convert it to a voltage. we then pass the entire input voltage to the feedforward summing node Σ , including the constant term. By making $G_0 = 2$, the input constant term is canceled and one of opposite polarity substituted, giving

$$e_G = -(1 + 0.5s)v_n \tag{4.4-10}$$

and

$$v_o = 1 + 0.5s - 2 = -(1 - 0.5s)v_n \tag{4.4-11}$$

as required. We have identified the internal node v_n within the constant factor of -1 .

Loudspeaker Equalizer

As a second example of extending the synthesis procedure to include poles of loss, we consider an equalizer to extend the bass response of a closed-box loudspeaker. It is well known that a loudspeaker mounted in a sealed enclosure exhibits a quadratic cutoff, such as shown in Fig. 4.18*a*. The cutoff is characterized by its mechanical resonant frequency established by the mass of the cone and the net compliance of the cone suspension and the air in the enclosure. The Q of the resonance—or the damping ($1/2Q$)—is affected by the degree of coupling to the source and by the source impedance. (The damping rises roughly with the size of the magnet and the reciprocal of the source impedance.) The resonant frequency varies with the reciprocal of the volume of the sealed box. The resonant frequency can be lowered by increasing the mass of the cone. Since the low-frequency response falls off at 12 dB per octave, however, each octave of response obtained in this way reduces the efficiency and increases the power required for a given sound level by about 16 times. Thus electronic equalization is attractive. The equalizer cancels the mechanical resonance of the loudspeaker and substitutes an electrical resonance at a lower frequency. The required equalizer loss is shown in Fig. 4.18*b*.

Taking ω_e and ω_m as the respective electrical and mechanical resonant frequencies and Q_e and Q_m the respective values of Q , we can write the loss of the equalizer as

$$L_e(s) = \frac{1 + (1/Q_e)(s/\omega_e) + (s/\omega_e)^2}{(\omega_m/\omega_e)^2 + (1/Q_m)(\omega_m/\omega_e)^2 s/\omega_m + (s/\omega_e)^2} \quad (4.4-12)$$

where we have scaled the denominator to be equal to the numerator at high frequencies. If we take the speaker resonance to be 0.1 kHz and the equalizer resonance to be 0.040 kHz, for example, and assume that the Q of both resonances are unity, the equalizer required is

$$L_e(s) = \frac{1 + 4s + 16s^2}{6.25 + 10s + 16s^2} \quad (4.4-13)$$

with s expressed in kilorads per second. (With resistances in kilohms, capacitances will be in microfarads.) The numerator of this expression is recognized as the loss of an active filter discussed in Section 4.2, the Sallen-Key circuit with $n=1$. From eq. (4.2-1), there are four circuit elements to realize a_1 and a_2 ; they are R_1 , R_2 , C_1 , and C_2 . Choosing $R_1 = R_2 = 100 \text{ k}\Omega$, we can find C_1 and C_2 from (4.2-1) with $n=1$:

$$\frac{e_G}{v_n} = 1 + (R_1 + R_2)C_2s + R_1R_2C_1C_2s^2 \quad (4.4-14)$$

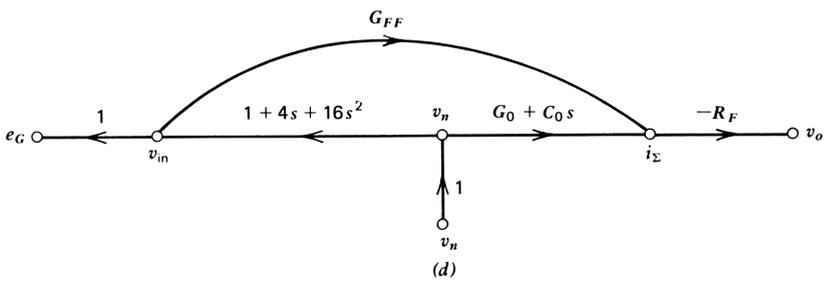
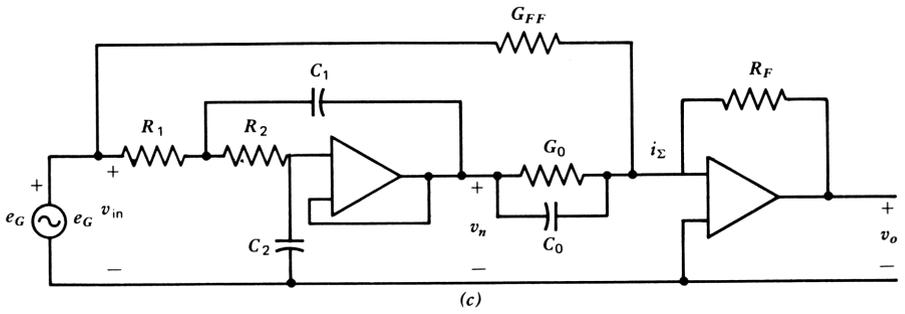
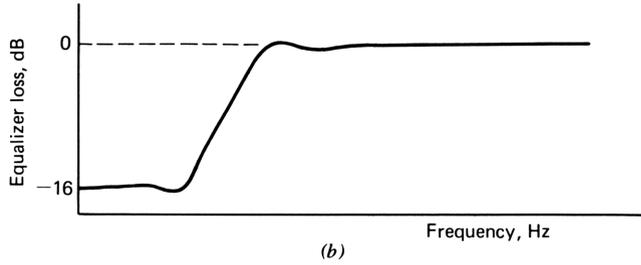
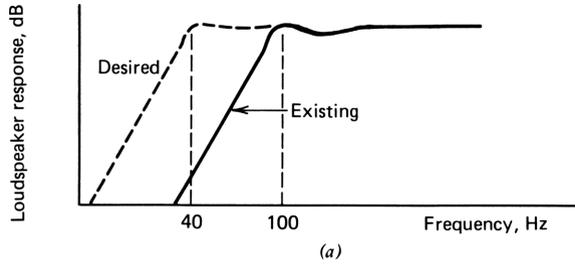


Figure 4.18. Development of an equalizer for extending the bass response of a closed-box loudspeaker.

Thus from the middle term we have $C_2 = 4/200 = 0.02 \mu\text{F}$ and from the quadratic term, $C_1 = 16/[(0.02)(100)^2] = 0.08 \mu\text{F}$. This realizes the numerator of (4.4-13):

$$e_G = (1 + 4s + 16s^2)v_n \quad (4.4-15)$$

with v_n the output voltage of the operational amplifier.

To realize the denominator, we use feedforward from the input, summing it with the signal from the output of the filter in a second operational amplifier, as shown in Fig. 4.18c. A signal flow graph is shown in Fig. 4.18d. With this arrangement, we can write for v_o :

$$\begin{aligned} v_o &= -[(1 + 4s + 16s^2)G_{FF}R_F + G_0R_F + C_0R_FS]v_n \\ &= \{(1 + G_0R_F) + (4 + C_0R_F)s + 16G_{FF}R_FS^2\}v_n \end{aligned} \quad (4.4-16)$$

But this is to be equal to the denominator in (4.3-13), so we can set the terms in powers of s equal individually. Setting the quadratic terms equal, we obtain $G_{FF}R_F = 1$ and setting the dc terms equal, $G_0R_F = 5.25$. If we let $R_F = 100 \text{ k}\Omega$, then $G_{FF} = 0.01 \text{ mS}$ and $G_0 = 0.0525 \text{ mS}$ ($19.0 \text{ k}\Omega$). Finally, we adjust the damping to match that of the loudspeaker by making $4 + C_0R_F = 10$, or $C_0 = 0.06 \mu\text{F}$, thus completing the design.

The key idea involved here is the change from feedback to feedforward. The value G_{FF} is a feedforward conductance only because the voltage at its right side is virtually zero and the voltage at its left side is the nonzero e_G . What little signal current fed back to the input through G_{FF} is short-circuited by the source. This same source, however, does feed signal current to the input summing node of the second amplifier, where it affects the output. Signal levels, then, determine the role—feedback or feedforward—that a conductance will play when it is connected from one point on a cascaded network to another.

Here we can only touch on the fascinating subject of active filter design. Fortunately, there are several sources available on this subject. Moshytz⁷ has characterized and typed the various possible active filter circuits, including detailed analysis. The main point is that cascade signal flow graphs for a proposed system can help us to formulate a system or circuit configuration to realize it.

In certain important cases it is impossible to secure the favorable ratio of signal levels that we have enjoyed here. Where feedforward is to be applied directly to the output of a power amplifier, such as for purposes of distortion reduction,⁸ the output signal is unavoidably high, leading to a significant amount of incidental feedback unless steps are taken to eliminate it. In such cases directional couplers—unidirectional coupling devices—are employed to eliminate or reduce the feedback.⁹

4.5 WHAT IS FEEDBACK? WHAT IS LOOP GAIN?

The arbitrary nature of models of feedback systems has been recognized for decades. Mason and Zimmermann put it most succinctly in 1960¹⁰:

A feedback loop is a closed chain of dependency, a closed path of signal flow in a system *diagram* [italics added]. Since the relationships among the signals in a given *physical* system may be represented as any one of the number of different system diagrams, some containing feedback loops and some not, it can be argued that the presence or absence of “feedback” in the “system” is more a matter of viewpoint than of physical reality. Nevertheless, many physical systems are, by the very philosophy of their design, strongly associated with diagrams containing loops, and in such cases, we find it comfortable to speak of the physical arrangement itself as a feedback system.

The key words here are “physical reality,” “very philosophy of design,” and “comfortable.” Feedback begins to emerge as a teleological concept—having to do with the purposes of the designer, and not with the physical system itself. It attaches itself to the philosophy of design for the comfort of the designer. Note, however, the association of “feedback” and “loop” in the quote. In this section we intend to break this association.

In these pages methods for analyzing and designing feedback systems have been developed that eliminate loop gains from the system description. Feedback itself has not been eliminated from the system: feedback is defined¹¹ as “*the returning of a portion of the output (of a system or a portion thereof) to the input.*” That aspect of feedback is retained in this book. What has been shown, however, is that there is a systematic way to reduce or eliminate the closed chain of dependencies that make thinking about feedback systems difficult. *One key connection* is broken, namely, the previously assumed equivalence of “feedback” and “loop gain.” By so doing, the analysis and design of feedback systems is made more transparent and direct. By use of anticausal analysis, feedback is taken as a returning of a portion of the output to the input, where it *augments* or *swamps out* the forward path input signal that was required to produce that output. Once that key connection—between feedback and loop gain—is broken, the “comfort” of the designer is enhanced by the relative ease of the mathematics of *addition* as opposed to *division*.

We value models for their simplicity as well as their accuracy. Ptolemy’s heliocentric model of the solar system, for example, gave results as accurate (at the time) as the Copernican geocentric one, but the latter was simpler and eventually prevailed. The changeover involved a conceptual leap from the geocentric to the heliocentric view, a leap that involved redefining “earth” as something that might actually move. As Thomas Kuhn relates¹²:

Consider... the men who called Copernicus mad because he proclaimed that the earth moved. They were not just wrong or quite wrong. Part of what they mean by “earth” was fixed position. Their earth, at least, could not be moved. Correspondingly, Copernicus’ innovation was not simply to

move the earth. Rather, it was a whole new way of regarding the problems of physics and astronomy, one that necessarily changed the meaning of both “earth” and “motion.” Without those changes the concept of a moving earth was mad.

Kuhn calls such changes of the terms in which a problem is seen as a “paradigm shift” and discusses the difficulty of communication across the divide between the two competing paradigms. The parallel to our subject is direct. If what one means by “feedback” is loop gain, to say that feedback exists without loop gain is mad. To get a clear picture of what is meant by “anticausal analysis,” it is essential to distinguish between these two concepts of feedback and loop gain. Loop gain is defined precisely for any given *description* of a physical situation. Feedback, on the other hand, is a higher-level concept that is difficult to define precisely in any given situation but that is useful nonetheless.

The conceptual leap needed in anticausal analysis is involved with our notions of causality and cannot be dismissed lightly. No departure from the principle of causality is involved; the *analysis* of the active path, in proceeding from output to input, is done in a direction from effect to cause and hence is termed *anticausal*. The active path itself is, of course, a causal structure. As we have seen, however, many of the effects in active devices such as emitter resistance in transistors can be viewed as feedback; what does constitute a cause-and-effect relationship? We take a closer look at this question in Chapter 7, but for present purposes, we can say that there is a *transport phenomenon*, complete with delay and dispersion, involved in every active device. In the transistor it does not involve the emitter resistance, which is a feedback element in the dictionary sense of returning a portion of the output (current) to the input (as a voltage). The transport of charge carriers defines cause and effect in the transistor.

To sharpen the distinction between feedback and loop gain, consider the two descriptions of a source connected to a load in Fig. 4.19. A Thevenin source e_G, R_G is connected to a load conductance G_L . There is no feedback here—no returning of the output to the input—except under a strained definition. There is loop gain, however, as seen in the flow graph, equal to $R_G G_L$. With a high-resistance source (relative to $1/G_L$), the loop gain is large. Figure 4.19 also shows a Norton source i_G, G_G connected to a load resistance. Here, the loop gain is $G_G R_L$. If the two descriptions are of the same circuit, the loop gains are reciprocals of one another—if one is 10, the other is 0.1. The anticausal representation of either source, on the other hand, gives no loop gain, as seen in the signal flow graphs.

Suppose that the source resistance is high. Then we know that the Norton representation is in some sense more *appropriate* than the Thevenin representation—we can pretty much ignore the source conductance in finding the load voltage or current. This forms a basis for rating models on a criterion of appropriateness: one model of a system is more *appropriate* than another if it involves *less loop gain*. Thus two representations of the same circuit—with no

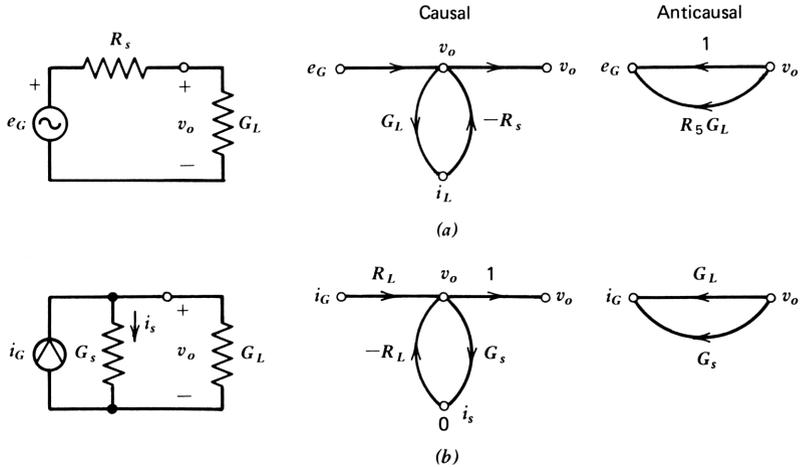


Figure 4.19. Three signal flow graphs and models of a source connected to a load with varying amounts of loop gains from 10 to zero.

feedback—yield two different values of loop gain in the causal description; the one with the smaller loop gain is the more appropriate. But in the anticausal description, neither representation has any loop gain; both are equally appropriate.

The break between “feedback” and “loop gain” is clear under the anticausal formulation of the feedback problem. As we have seen, loop gains (usually small or incidental) arise in anticausal analysis, but they are not connected with the feedback aspects of the problem; rather, they involve feedforward. The problem of defining the feedback network and the forward path network is not a central concern if the concept of loop gain is not needed.

Consider the transistor stage with feedback shown in Fig. 4.20; here, the feedback network is simply the conductance G_F connected from collector to base. But the transistor includes (internally) a collector junction capacitance that can be shown separately from the transistor, as in Fig. 4.20*b*. Are we justified in assigning this capacitance to the forward path, when its circuit function and topological position are similar to those of the feedback network? Tradition, starting with Bode,¹³ has it that the collector capacitance belongs with the feedback path. On the other hand, the capacitance is nonlinear with voltage; thus if feedback is applied to reduce distortion, it should be considered part of the forward path. The question is moot for the anticausal formulation because the loss is simply the sum of the three contributions (the active path without the capacitance, the capacitance, and the feedback conductance). Sensitivity analysis replaces considerations of loop transmission and return difference, and the latter concepts are no longer needed.

Now that we have demonstrated the existence of loop gain without feedback, and feedback without loop gain, the break between these concepts is on a firm logical foundation.

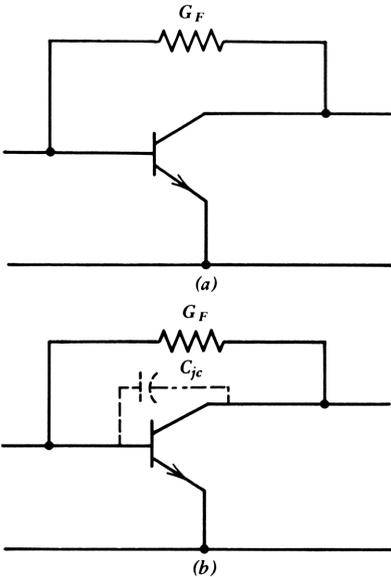


Figure 4.20. Variable interpretation of feedback and forward paths in a simple structure.

PROBLEMS

- 1 Write an expression for the sensitivity of the loss of the low-pass filter in Fig. 4.4 to the unity loss time constant of the operational amplifier used. (Use the sum rule.)
- 2 The single-amplifier biquad (SAB) is a type of active filter that realizes transfer functions that consist of a ratio of quadratic functions. The loss of an SAB is given by

$$\frac{e_G}{v_o} = \frac{a_0 + a_1s + a_2s^2}{1 + d_1s + d_2s^2}$$

In Fig. 4.21 two such SABs are shown; in Fig. 4.21a the loss is low pass, with $d_1 = d_2 = 0$, and in Fig. 4.21b the loss is bandpass, with $d_0 = d_2 = 0$. Draw signal flow graphs and write the sequential matrices for each of these circuits. (Note that the voltage divider is returned to the positive input of the amplifier, unlike the Sallen-Key circuit in Fig. 4.4.)

- 3 Show that the polynomial whose sequential matrix is

$$\begin{bmatrix} s - k_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & s - k_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & s & b_3 & 0 & 0 \\ 0 & 0 & 0 & s - k_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & s & b_1 \\ 0 & 0 & 0 & 0 & 0 & s - k_4 \end{bmatrix}$$

is stable if and only if all k_i are negative.

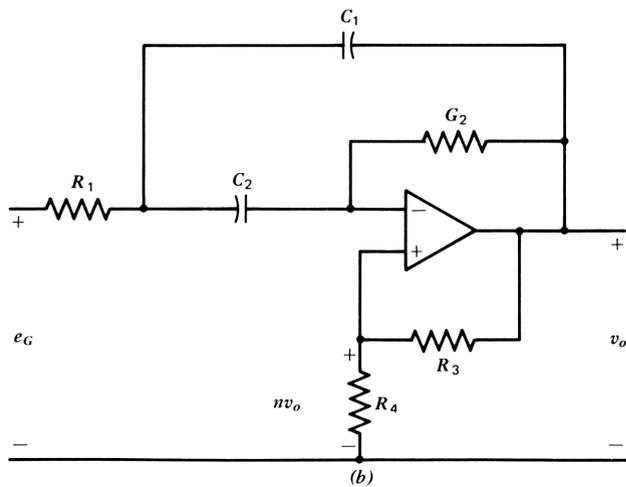
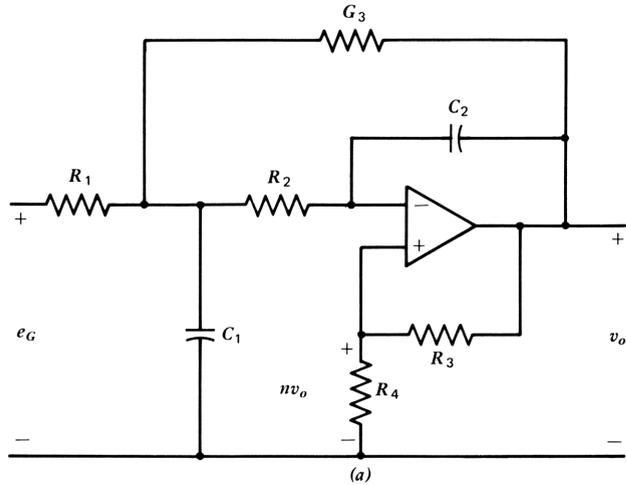


Figure 4.21.

- 4 Write the causal and anticausal sequential matrices for the feedforward and feedback portions of the circuit in Fig. 4.17c. Compare these with the causal and anticausal sequential matrices of the circuit taken as a whole.
- 5 Design an active filter section that has the transfer function

$$L(s) = \frac{1 + s + s^2}{1 + 0.5s^2}$$

REFERENCES

- 1 S. J. Mason, "Feedback Theory—Some Properties of Signal Flow Graphs," *Proc. IRE* **41** (9), 1144–1157 (September 1953); S. J. Mason, "Feedback Theory—Some Further Properties of Signal Flow Graphs," *Proc. IRE* **44** (7), 920–926 (July 1956).
- 2 M. Gardner, "Mathematical Games," *Sci. Am.* (March 1980).
- 3 R. P. Sallen and E. L. Key, "A Practical Method of Designing RC Active Filters," *IRE Trans. CT-2*, 74–85 (March 1955).
- 4 T. Deliyannis, "High Q Factor Circuit with Reduced Sensitivity," *Electron. Lett.* **4**, 577 (December 1968).
- 5 J. J. Friend, "A Single Operational Amplifier Biquadratic Filter Section," *IEEE Internat. Symp. Circuit Theory Digest*, IEEE Cat. No. 70 C 61-CT, December 1970, p. 179.
- 6 P. E. Fleischer, "Sensitivity Minimization in a Single Amplifier Biquad Circuit," *IEEE Trans. CAS-23* (1) Ann No. 601CA006 (January 1976).
- 7 G. S. Moshytz, *Linear Integrated Networks; Design*, Van Nostrand Reinhold, New York, 1975.
- 8 H. S. Black, "Translating System," U.S. Patent No. 1,686,792, issued October 9, 1928.
- 9 H. Seidel, "A Microwave Feed-Forward Experiment," *BSTJ* **50** (9), 2879–2916 (November 1971).
- 10 S. J. Mason and H. J. Zimmermann, *Electronic Circuits, Signals, and Systems*, The MIT Press, Cambridge, MA, 1970 (Wiley, New York, 1960) pp. 565 ff.
- 11 F. Jay, Ed., *IEEE Standard Dictionary of Electrical and Electronic Terms*, 2nd ed., IEEE, 1977.
- 12 T. S. Kuhn, *The Structure of Scientific Revolutions*, 2nd ed., University of Chicago Press, 1972.
- 13 H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1947, p. 46.

Chapter 5

Signal Delay in Feedback Systems

Delay plays a central role in feedback systems. When the gain is turned up too far in a public address system, some of the loudspeaker output finds its way to the microphone, and the system oscillates at one or more of the natural frequencies of the system, including the modes of vibration of the room, the loudspeaker, and the microphone element. Acoustical delay between the loudspeaker and the microphone make control of the phase shift impossible, and the phase shift itself may be many thousands of degrees. The solution, of course, is to increase the loss of the feedback path by directing the loudspeaker energy away from the microphone and finally, to turn down the gain. In this example feedback is an undesirable side effect in the effort to make an orator audible. Similar problems attend the design of systems and circuits that include feedback to improve performance.

Feedback and feedforward both involve the addition of signals at a circuit node. We have treated the addition by use of dependent generators in an equivalent ladder circuit. Wherever addition of signals occurs, the possibility of signal cancellation—somewhere on the finite s plane—arises. Where feedback is involved, such cancellation causes the *loss* to go to zero at points on the s plane at which the cancellation occurs, thus leading to the possibility of

instability. With feedforward, on the other hand, such cancellation causes the *gain* to fall to zero: no instability is thereby created. Thus, in feedback systems, a fundamental limit on the amount of delay is imposed by stability considerations. No such fundamental limit exists for feedforward systems, where large delays can be used, for example, to bring a set of signals into synchronism after time-consuming signal processing has produced results at different times with an unwanted time variation.

5.1 FEEDBACK AROUND A PURE DELAY

In the time domain a signal $f(t)$ may be delayed as it travels along a path, so that its delayed version is $f(t-\tau_d)$, where τ_d is the amount of delay. This delay may be represented in the frequency domain by the shifting theorem; the Laplace transform of the delayed signal is

$$\mathcal{L}\{f(t-\tau_d)\} = e^{-\tau_d s} F(s) \quad (5.1-1)$$

where $F(s)$ is the Laplace transform of $f(t)$. If we set $s=j\omega$, we see that the effect of the delay is to add an amount of phase $e^{-j\tau_d\omega}$ to the phase of $F(j\omega)$, an amount of phase that increases linearly with frequency.

If we place feedback around an active forward path consisting of a pure delay as shown in Fig. 5.1a, we can represent its loss as

$$L(s) = b + ae^{\tau_d s} \quad (5.1-2)$$

Roots of the characteristic equation $L(s)=0$ are found by taking the logarithm:

$$s = \frac{1}{\tau_d} \ln\left(-\frac{b}{a}\right) \pm j \frac{2n\pi}{\tau_d}, \quad \frac{b}{a} < 0, \quad n=0, 1, 2, \dots$$

$$s = \frac{1}{\tau_d} \ln\left(\frac{b}{a}\right) \pm j \frac{(2n-1)\pi}{\tau_d}, \quad \frac{b}{a} > 0, \quad n=1, 2, 3, \dots \quad (5.1-3)$$

In either case the roots lie in the right half plane for $|b/a| > 1$, so that for stability, the loss of the active path must exceed the loss of the feedback path.

The Nyquist diagram in Fig. 5.1b shows the separate terms of the sum of eq. (5.1-2); b is a fixed phasor, whereas the exponential multiplying a rotates its phase in a counterclockwise direction. Since this is a map of the $j\omega$ axis, the area to the left of the curve, the inside area of the circle, is a map of the left half plane. Unless $a < b$, the point $L(j\omega)=0$ will lie in the right half plane.

The Bode diagram for this function exhibits a *comb filter* response typical of delay networks (Fig. 5.1c).

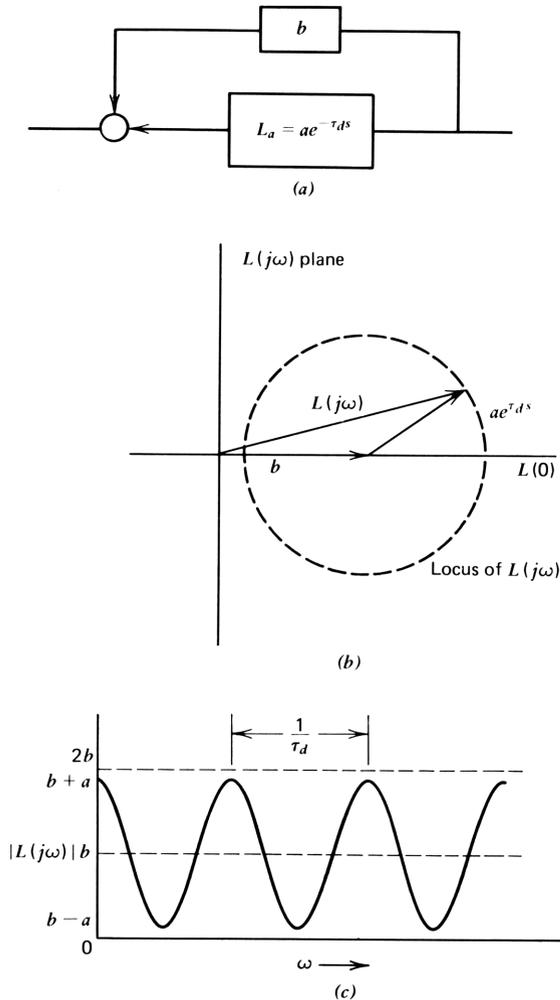


Figure 5.1. Feedback around an active path consisting of pure delay giving comb-filter response.

5.2 A CLASSIC FEEDBACK EQUATION

In Section 5.1 we saw that where the bandwidth is unlimited in a feedback system, the loss of the active path must, for stability, exceed that of the feedback path, and that when the limit is approached, the loss will oscillate between zero and twice the forward-path loss, with a period of oscillation (in the frequency domain) equal to $1/\tau_d$, where τ_d is the delay. Thus benefits of feedback in improving sensitivity performance, which rely on the feedback path loss substantially exceeding the forward-path loss, are not available for this case. To obtain these benefits, the bandwidth must be restricted by an

amount that is related to the total delay of the feedback and forward paths, the loop delay.

The simplest way to restrict the bandwidth is to introduce an integration in the forward path, leading to the equation for the loss:

$$L'(s) = b + \tau_1' s e^{\tau_d s} \quad (5.2-1)$$

Since this equation expresses in its simplest form the limitations in the application of feedback, we term it the *classic feedback equation*. When it is normalized to a dc value of unity, it becomes the loss equation for a *unity gain follower*: setting $\tau_1 = \tau_1'/b$ and $L(s) = L'(s)/b$,

$$L(s) = 1 + \tau_1 s e^{\tau_d s} \quad (5.2-2)$$

In addition, we may normalize the frequency variable as we did in Chapter 2 for polynomial loss ratios. By changing the frequency variable to $\tau_1 s$, we can write (5.2-2) in a normalized form:

$$L(\tau_1 s) = 1 + (\tau_1 s) \exp \left[\frac{\tau_d}{\tau_1} (\tau_1 s) \right] \quad (5.2-3)$$

The character of the equation is expressed, within a translation in magnitude or in frequency, by only one parameter, τ_d/τ_1 . Because of its simplicity in expressing the basic limitation imposed by delay in feedback systems, the equation is worth studying in some detail. It is also of practical significance and turns up surprisingly often.

Application of the Classic Feedback Equation

Before studying the equation, we take a brief look at one such example, the 741-type operational amplifier. We look into operational amplifiers later in some detail; here, we are interested in its frequency response, the source of the delay, and how τ_1 is controlled in the circuit.

An approximation to the circuit of the amplifier is shown in Fig. 5.2a. The circuit consists of four parts. Transistors Q_1 and Q_3 form an input *differential pair*; Q_2 and Q_4 provide *dc level shifting* and are *lateral pnp* transistors (described below); these transistors feed a *current mirror*, which functions to add i_2 to i_4 as shown; and finally, there is a high-gain amplifier whose response is dominated by feedback capacitor C_F . For our purposes, we can simplify the diagram further by analyzing the half-circuit in Fig. 5.2b, since the input stage is assumed to be balanced. A signal flow graph for the circuit (Fig. 5.2c) has four tandem branches corresponding to the four parts just described. The input voltage is $v_d/2$, half the differential input voltage, since we are considering the half-circuit. This voltage consists of the drop across the emitter resistances of Q_1 and Q_2 , a total of $2i_1 r_e$.

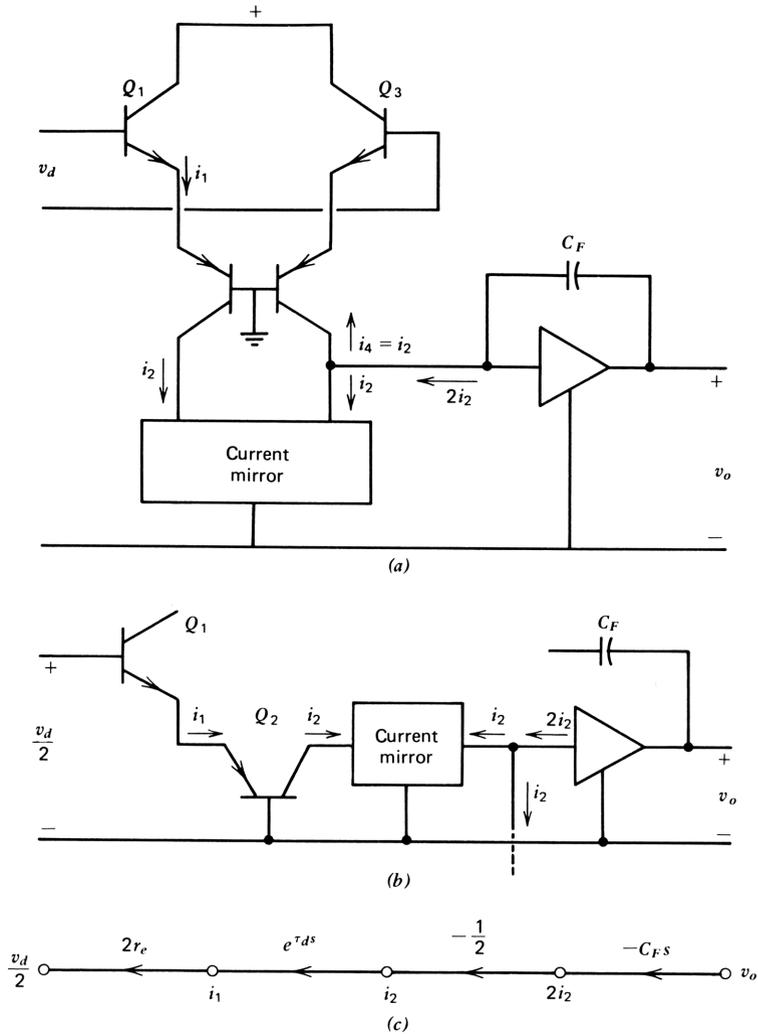


Figure 5.2. A representation of the 741-type operational amplifier.

The delay in the circuit is accounted for by Q_2 , the lateral *pn*p transistor, connected in the *common base* configuration. All other transistors concerned with signal processing (as opposed to dc bias) are *npn* transistors with dimensions controlled by impurity diffusion vertically into a slice of silicon, giving exceedingly small dimensions along which the signal must travel between emitter and collector. To make transistors of the *pn*p type without increasing the complexity of transistor processing, these transistors are made with emitter, base, and collector placed side by side. Their critical dimensions are defined by the horizontal geometry (photolithographic masks) and are

between 10 and 100 times those of the *npn* transistors. As a result, signals traversing Q_2 (and Q_4) encounter significant delay τ_d ; thus $i_1 = i_2 e^{\tau_d s}$.

The current mirror, using *npn* transistors, gives a broadband loss of $\frac{1}{2}$. Finally, the amplifier with feedback capacitor C_F around it is of extremely high gain and is virtually totally controlled by the capacitance. Multiplying the four branches of Fig. 5.2c, we obtain

$$L(s) = \frac{v_d}{v_o} = 2r_e C_F s e^{\tau_d s} \quad (5.2-4)$$

Comparing this with (5.2-2), we can identify the time constant $2r_e C_F$ with τ_1 .

The 741 operational amplifier was designed as a general-purpose amplifier for which a stringent application from a stability standpoint is the unity gain follower shown in Fig. 5.3. The circuit is shown in Fig. 5.3a, and its equivalent circuit is given in Fig. 5.3b. We find the loss of this circuit by adding unity to (5.2-4), giving the classic feedback equation of (5.2-2). The 741 is a particularly pure case of a feedback design controlled almost entirely by delay.

Nyquist Diagrams

We can plot a Nyquist diagram for $L(j\omega)$ in eq. (5.2-2) by calculating the magnitude and phase of the active path:

$$\begin{aligned} [L_a(j\omega)] &= \tau_1 \omega \\ \arg L_a(j\omega) &= \frac{\pi}{2} + \tau_d \omega \end{aligned} \quad (5.2-5)$$

We then convert these to rectangular form and add unity to the real part, giving the rectangular coordinates of $L(j\omega)$. Figure 5.3c shows Nyquist diagrams for three values of τ_d/τ_1 . They are similar in shape to those of Fig. 2.15 for linear, quadratic, and cubic polynomials. This is hardly surprising since the power series expansion for $e^{\tau_d s}$ is

$$e^{\tau_d s} = 1 + \tau_d s + \frac{(\tau_d s)^2}{2!} + \frac{(\tau_d s)^3}{3!} + \dots \quad (5.2-6)$$

For sufficiently small values of τ_d , we are justified in truncating the series; as τ_d rises, we need more and more terms to make the approximation accurate. The curve for the largest value of τ_d/τ_1 in Fig. 5.3c is drawn for an unstable case; the point $L(j\omega) = 0$ lies in the map of the right half plane. There are two such roots in the right half plane. As the frequency is increased, the Nyquist diagrams (for positive frequency) are all logarithmic spirals with an infinite number of encirclings of the origin. There are an infinite number of left half plane roots for all three curves. The one for maximum delay puts a single pair of roots in the right half plane. The value $\tau_d/\tau_1 = \pi/2$ is the boundary value between stable and unstable performance.

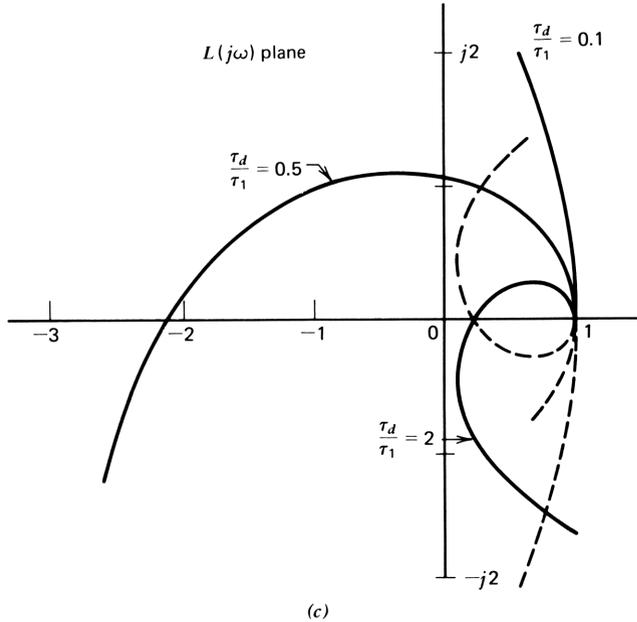
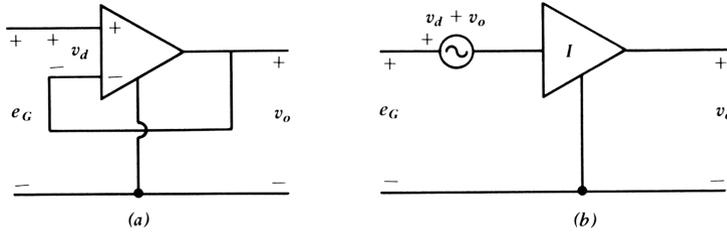


Figure 5.3. Equivalent representations of the amplifier connected as a unity gain follower. Nyquist diagrams for the classic feedback equation with three ratios of delay to control time constants are shown in part c.

Root Locus Diagram for the Classic Feedback Equation: Analysis

The Nyquist diagrams give insight into the behavior of the classic feedback equation. Further insight is afforded by finding the root locations themselves on the s plane as τ_d/τ_1 is varied. To find the roots, we rewrite the characteristic equation $L(s)=0$ from (5.2-2) as

$$\begin{aligned} \tau_1 s &= -e^{-\tau_d s} \\ &= e^{j\pi} \exp[-\tau_d(\sigma + j\omega)] \end{aligned} \tag{5.2-7}$$

in which we have substituted $s = \sigma + j\omega$, and $e^{j\pi} = -1$. We can also express the

frequency variable s in polar form: $s = \rho e^{j\theta}$; when this is substituted on the left side of the equation, we have

$$\tau_1 \rho e^{j\theta} = \exp[-\tau_d \sigma + j(\pi - \tau_d \omega)] \quad (5.2-8)$$

Taking logarithms of the two sides and equating the real and imaginary parts separately, we have

$$\ln \tau_1 \rho = -\tau_d \sigma \quad (5.2-9)$$

and

$$\theta = \pi - \tau_d \omega \quad (5.2-10)$$

These equations relate the magnitude and phase of the roots of the equation to the real and imaginary parts of the roots. The previous sentence provides the key to solving (implicitly) for the roots for any value of delay. For zero delay, the equation has one real root at $\tau_1 \sigma = -1$. As the delay is increased, we see that a second real root moves into the finite part of the $\tau_1 s$ plane from $-\infty$ and the first root moves to the left from -1 . As the delay is increased further, the two roots coalesce and then become complex. We show that the point of coalescence occurs at $\tau_1 \sigma = -e$, and it occurs for a value of $\tau_d/\tau_1 = 1/e$, as shown in Fig. 5.4a. For mathematical convenience, we divide the process of finding the locus of roots in two: first we find the locus for real roots and then for complex roots.

For real roots, $\omega = 0$, so that $\tau_d \omega = 0$, or $\theta = \pi$, as seen in eq. (5.2-10). This equation gives no information about the root locations, but for this case, $\sigma = \rho$, so that (5.2-9) can be written

$$\ln(-\tau_1 \sigma) = -\tau_d \sigma \quad (5.2-11)$$

To plot the root locus for real roots, we assign values to $\tau_1 \sigma$ and find the resulting values of τ_d and σ :

$$\frac{\tau_d}{\tau_1} = -\frac{\ln(-\tau_1 \sigma)}{\tau_1 \sigma} \quad (5.2-12)$$

If we assign $\tau_1 \sigma = -2$, for example, this equation gives $\tau_d/\tau_1 = 0.347$. A value of -4 for $\tau_1 \sigma$ also gives $\tau_d/\tau_1 = 0.347$; therefore, the two roots corresponding to $\tau_d/\tau_1 = 0.347$ are at $\tau_1 \sigma = -2$ and -4 . The maximum value that can be found for τ_d/τ_1 using this equation occurs for $\tau_1 \sigma = -e$, and τ_d/τ_1 for this value is (obviously) $1/e$. This is the point of coalescence for the two real roots. Larger values of delay give a pair of complex roots.

To calculate the locus for *complex roots*, we assign values of the *angle* of the root. Thus we take $\tau_d \omega$ as a dummy variable. Since $\omega/\sigma = \tan \theta = -\tan \tau_d \omega$ by

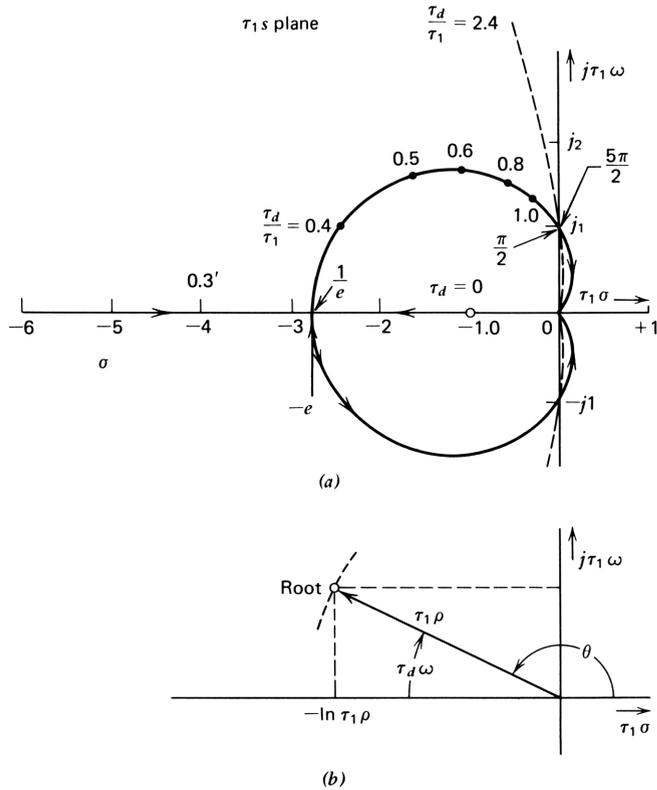


Figure 5.4. Root locus diagram for varying τ_d/τ_1 for the classic feedback equation. Higher-degree roots are shown in part *b*.

definition, we can write (5.2-9) as

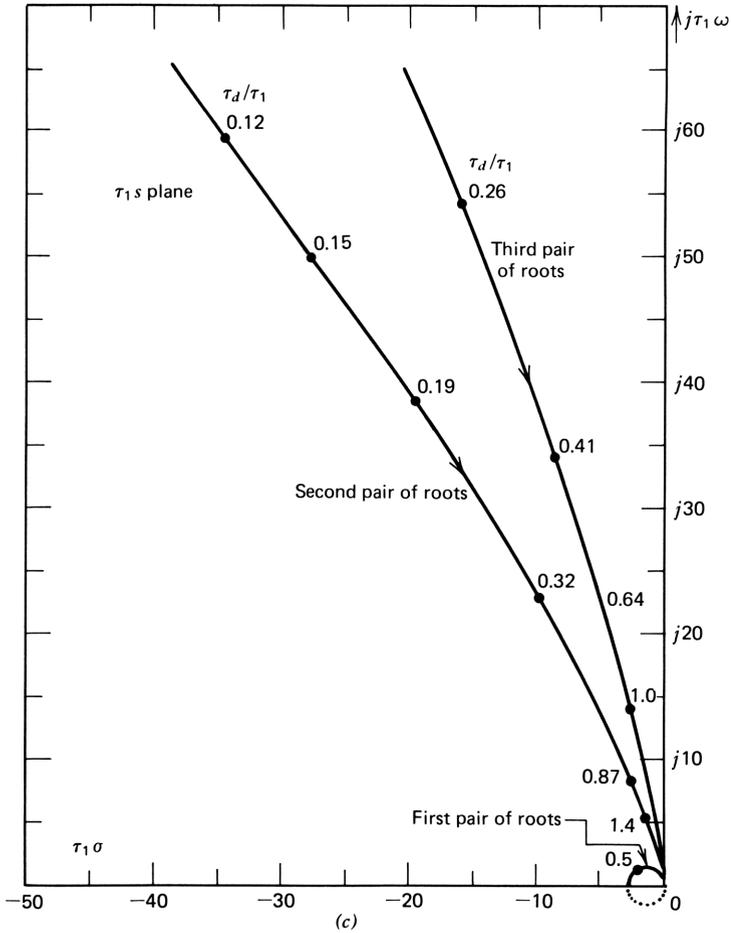
$$\frac{\tau_d \omega}{\tau_1 \sigma} = \frac{\tau_d \omega}{\ln \tau_1 \rho} = \tan \tau_d \omega$$

The geometric relationships are shown in Fig. 5.4*b*; thus

$$\tau_1 \rho = \exp \frac{\tau_d \omega}{\tan \tau_d \omega} \tag{5.2-13}$$

Since we can obtain θ from (5.2-10) using the given dummy variable, we have found the position of the roots of the characteristic equation. Note that ρ is ω_0 for a quadratic pair of roots as defined in Chapter 2 and that θ is the angle whose cosine is the negative of the damping factor ζ ; thus

$$\zeta = -\cos \theta = \cos \tau_d \omega$$



Next, we must find the value of τ_d/τ_1 associated with the root position that we have just determined.

Expressing the root positions in rectangular coordinates, we have

$$\tau_1 \sigma = -\exp\left(\frac{\tau_d \omega}{\tan \tau_d \omega}\right) \cos \tau_d \omega \tag{5.2-14}$$

and

$$\tau_1 \omega = \exp\left(\frac{\tau_d \omega}{\tan \tau_d \omega}\right) \sin \tau_d \omega \tag{5.2-15}$$

From the latter equation, we find τ_d/τ_1 as

$$\frac{\tau_d}{\tau_1} = \frac{\tau_d \omega}{\tau_1 \omega} \quad (5.2-16)$$

These equations were programmed on the calculator to obtain the root locus diagrams in Fig. 5.4; the program, "CFE", is given in Appendix B. It is an example of a program written to clarify relationships, since a fair amount of mental juggling is involved. Once the root locus is found, there is no need for the program, since all root loci for the classic feedback equation are similar.

In Fig. 5.4a, as noted previously, the locus begins at $\tau_1 s = -1 + j0$ for $\tau_d = 0$. As the delay is increased, the root moves along the real axis away from the imaginary axis, and a second root moves in from $-\infty$. The two roots coalesce at $s = -e + j0$ at a value of τ_d/τ_1 of $1/e$, whereupon the roots become complex. For values of τ_d/τ_1 above $1/e$, there are no real roots in the finite s plane. The phasor diagram in Fig. 5.4b shows the relationships among the quantities of eqs. (5.2-13) to (5.2-16). At a value of $\tau_d \omega$ of $\pi/4$ (a damping value of 0.707), τ_d/τ_1 is about 0.5, and for $\tau_d \omega$ of $\pi/2$, the damping value is zero and the roots enter the right half plane for $\omega = 1$. As noted earlier, there are an infinite number of roots; to find them, τ_d may be increased without bound. The dashed line shows the second pair of roots to appear. This set of roots crosses over into the right half plane for $\tau_d \omega$ equal to $5\pi/2$, or one complete revolution more than the first set.

Significance of the Root Locus

The important thing to notice about this root locus is that for values of delay that are sufficiently small to prevent instability (a value of τ_d/τ_1 less than $\pi/2$, or 1.57) *only the first pair of roots are important in determining the dynamics of the system*. All other roots are remote. Therefore, the response of the classic feedback equation is essentially *quadratic* even for values of τ_d large enough to reach instability.

Thus, even for τ_d/τ_1 equal to $\pi/2$, the nearest higher-degree roots are over four times more remote from the origin than the unstable roots. For smaller values of delay yielding more useful loss functions, these higher-degree roots are much more remote. In Fig. 5.4c we show the root locus over a wider range of root locations on the s plane. The second and third pair of roots that appear have larger values of delay. Such roots are of consequence only after the first pair of roots have passed into the right half plane, except that they do cause delay of the overall closed-circuit transfer function.

Such delays are normally inconsequential in a feedback system since the dynamics of the system are unaffected except for the translation in time. However, if the system is, in turn, part of a larger feedback system, a *metasystem*, such external delays must be taken into account in the design of *that* system. We see an example of this later in this chapter, where we add delay to Design B of the Chapter 3 case study.

Thus the loss of the classic feedback equation has only one pair of essential roots and is a quadratic. This greatly simplifies the study of delay in feedback structures. We now obtain the quadratic constants that approximate the delay function with the use of a technique that will be useful in the design of feedback systems incorporating delay.

Linear and Quadratic Approximations to the Classic Feedback Equation

The function $e^{\tau_d s}$ can be represented by its power series, given by eq. (5.2-6). When this series is truncated, both magnitude and phase errors attend the approximation. Figure 5.5 gives these errors as a function of the phase angle being approximated for the linear approximation

$$e^{\tau_d s} = 1 + \tau_d s \quad (5.2-17)$$

and the quadratic approximation

$$e^{\tau_d s} = 1 + \tau_d s + \frac{(\tau_d s)^2}{2} \quad (5.2-18)$$

Thus, for example, the linear approximation is in error by 0.5 dB at a frequency at which the delay causes 20° of phase shift. Similarly, the quadratic approximation error is 0.5 dB where the delay introduces 48° of phase shift. Figure 5.5 also gives the error in phase, as well as the total magnitude of error in percent (without information as to whether the phase or the magnitude is in error).

If we try to apply this approximation to the classic feedback equation directly by substituting the linear approximation for $e^{\tau_d s}$ in (5.2-2), we indeed obtain a quadratic, but the error curves in Fig. 5.5 do not lead us to expect good accuracy for phase angles greater than 15 or 20° . Still, we know that a quadratic function can approximate the classic feedback equation. The question is how to find it. The key is to remove the phase of the remote roots of the equation before attempting to find the quadratic.

This can be done by writing the classic feedback equation as follows:

$$L(s) = e^{B\tau_d s} (e^{-B\tau_d s} + \tau_1 s e^{A\tau_d s}) \quad (5.2-19)$$

in which the delay has been divided between the forward and feedback paths as shown in the signal flow graph in Fig. 5.6. The delay term outside the brackets may be considered the effect of all roots except the two smallest roots, which are to be approximated by a quadratic that replaces the terms within the brackets. To force (5.2-19) into the desired form (quadratic plus delay), we represent the second delay term by the first *two* terms of the power series:

$$e^{A\tau_d s} = 1 + A\tau_d s \quad (5.2-20)$$

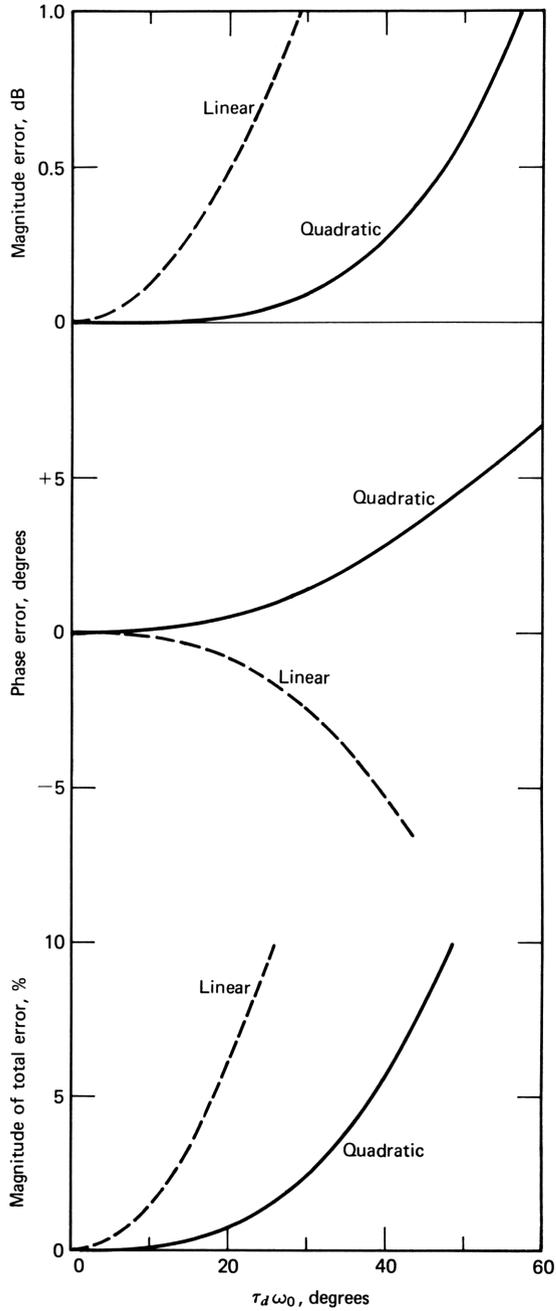


Figure 5.5. Error of linear and quadratic delay approximations as a function of the phase angle being approximated.

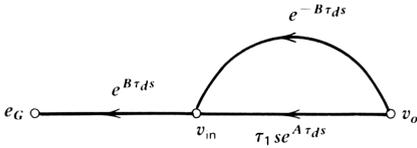


Figure 5.6. Signal flow graph showing an arbitrary division of delay between the causal (feedback) and anticausal (forward) paths.

and the first delay term by the first *three* terms of the power series:

$$e^{-B\tau_d s} = 1 - B\tau_d s + \frac{(B\tau_d s)^2}{2} \tag{5.2-21}$$

Multiplying (5.2-20) by $\tau_1 s$ and adding it to (5.2-21), we obtain the quadratic approximation to the classic feedback equation. The sum of A and B should add to unity and be proportioned to minimize the approximation error. Roughly speaking, the quadratic approximation of (5.2-21) is twice as good as the linear approximation of (5.2-20), so that $B = 2A = 0.67$. We can find values of A and B that match the quadratic roots exactly for any given delay; the values for $\tau_d/\tau_1 = 0.5$, for example, are $A = 0.29$ and $B = 0.70$, as calculated by equating the approximation to the exact value term by term. The approximation is

$$L'(s) = e^{B\tau_d s} \left\{ 1 + (\tau_1 - B\tau_d)s + \left[A\tau_1\tau_d + \frac{(B\tau_d)^2}{s} \right] s^2 \right\} \tag{5.2-22}$$

where B is obtained by setting the second term equal to $(2 \cos \theta)/\rho$, and A is obtained by setting the third term equal to $1/\rho^2$, with θ and ρ given by (5.2-10) and (5.2-13). The approximation is good: a root locus for the approximating quadratic as τ_d/τ_1 is varied in Fig. 5.7 for fixed values of A and B of 0.3 and

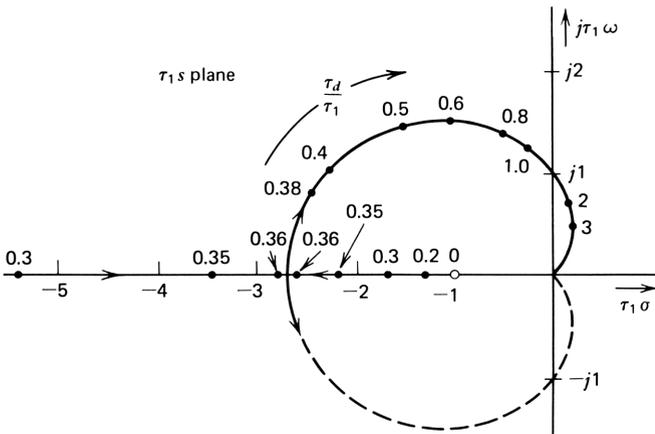


Figure 5.7. Root locus diagram for the quadratic approximation to the classic feedback equation for varying τ_d/τ_1 .

0.7. Bode plots of the exact function and the approximation are virtually indistinguishable for usable amounts of delay; at the point of instability, the delay for the approximation is about 9% low.

Where the delay is sufficiently small to cause a pair of well-separated real roots, a linear approximation will suffice for delay. For this approximation, we let $A=0$ in (5.2-20) and let $e^{-B\tau_d s} = 1 - B\tau_d s$ in (5.2-21), with $B=1$. The expression for loss is

$$L(s) = e^{\tau_d s} [1 + (\tau_1 - \tau_d)s] \quad (5.2-23)$$

Again, the external delay does not affect the dynamics of the system under study but must be taken into account if it is a part of a metasystem.

Design of Systems Governed by the Classic Feedback Equation

Design of a feedback system described by the classic feedback equation requires that we find only one parameter to be specified for a given low-frequency loss, namely, the damping factor. In the unity gain follower using the 741 amplifier, for example, we wish to know the value of C_F (the compensation capacitor) to give a desired quadratic polynomial loss.

For *design*, it is convenient to use a frequency normalization that is different from that given in (5.2-3). Here, we take $\tau_d s$ as our normalized frequency variable, so that the classic feedback equation becomes

$$L(\tau_d s) = 1 + \frac{\tau_1}{\tau_d} (\tau_d s) e^{\tau_d s} \quad (5.2-24)$$

The locus of roots of this equation for varying τ_1/τ_d is given in the root locus diagram in Fig. 5.8a, plotted for a fixed unit delay. The numbers along the curve are values of τ_1/τ_d needed to obtain the desired root positions. The curve is drawn for unity dc loss, as in the unity gain follower application. For other values of dc loss, the indicated value of τ_1 should be multiplied by the dc loss. For given values of τ_d and the desired damping factor, we obtain the bandwidth and control time constant by the previously developed equations; thus

$$\tau_d \omega = \cos^{-1} \zeta \quad (5.2-25)$$

From the geometry in Fig. 5.8b,

$$\rho = \frac{\tau_d \omega}{\tau_d \sin \tau_d \omega} \quad (5.2-26)$$

and from (5.2-13)

$$\tau_1 = \frac{1}{\rho} \exp\left(\frac{\tau_d \omega}{\tan \tau_d \omega}\right) \quad (5.2-27)$$

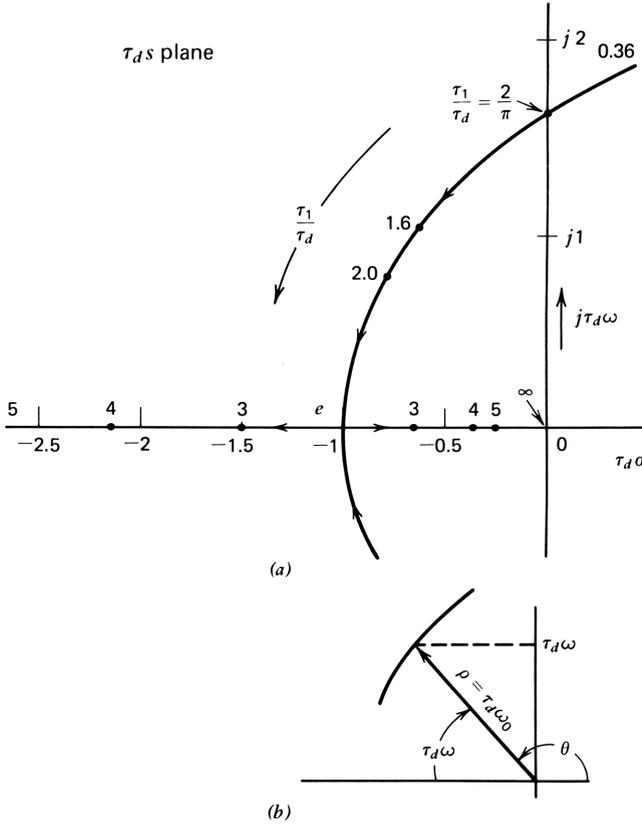


Figure 5.8. Root locus diagram for the classic feedback equation as τ_1/τ_d is varied. This is the root locus for design of a system.

For the 741 operational amplifier connected as a unity gain follower, the delay of the lateral *pnp* transistor is about $0.08 \mu\text{s}$. If we wish the damping factor to be 0.707 , we have

$$\tau_d \omega = \frac{\pi}{4} \text{ rad}$$

so that

$$2\pi f_0 = \frac{\pi/4}{0.08 \times \sin \pi/4} = 13.9 \text{ Mrad/s}$$

or $f_0 = 2.21 \text{ MHz}$. Finally, we need a control time constant of

$$\tau_1 = \frac{1}{13.9} \exp\left(\frac{\pi/4}{\tan \pi/4}\right) = 0.158 \mu\text{s}$$

In the 741 this time constant is $2r_e C_F$, as we saw in eq. (5.2-4). The value of r_e for the bipolar transistors is $kT/qI_C = 0.026/I_C$, with I_C in milliamperes. The collector currents for the input stage are about 0.01 mA, so that $2r_e = 5.2 \text{ k}\Omega$; then $C_F = 0.158/5.2 = 0.03 \text{ nF}$, or 30 pF.

Slew Limiting

Another type of delay encountered in feedback systems is nonlinear in character and is also clearly illustrated by the 741 operational amplifier. If a sufficiently large signal is applied to the input one side of the differential pair will be turned off (i.e., will conduct no current), whereas the current through the other side will double, from 0.01 to 0.02 mA. Under these conditions, the output will not follow the input since C_F can only charge at the rate set by the maximum current flowing into it, 0.02 mA. The rate of change of output voltage with time is constant under the slew-limited condition and is

$$\frac{dv_o}{dt} = \frac{I}{C_F} = \frac{0.02 \text{ mA}}{0.03 \text{ nF}} = 0.67 \text{ V}/\mu\text{s} \quad (5.2-28)$$

Thus for the output voltage to change from -12 to $+12 \text{ V}$, for example, takes $36 \mu\text{s}$, a much longer delay time than applies under linear conditions. This does not affect the stability of the unity gain follower since the forward-path loss is high under these conditions, but if the amplifier is itself a part of a metasystem, the external delay caused by the slew limitation may indeed cause instability.

If we attempt to reduce slew-limiting delay by increasing the dc collector current of the first stage to charge the capacitor faster, the value of r_e drops in proportion. To maintain the control time constant, C_F will have to rise; hence there is no net improvement in slew rate. Thus the lateral transistor delay indirectly controls the slew rate as well as the control time constant. One method to break this chain of dependencies is to add a resistor in series with r_e . Such a resistor will not, of course, be dependent on collector current, so that the control time constant can be set independently of the first-stage collector current.

We can conclude a key fact about the effect of delay of feedback systems from this section, as well as certain relationships about polynomial approximations to delay. Feedback is an additive process, where signals to be added have a common cause (the input) but arise in differing time frames, or signal *epochs*. We have seen in this section that the stability requirement puts severe limits on the amount of delay that can be accommodated. We have taken the classic feedback equation as our case study here because of its simplicity and practical importance.

This equation places 90° of phase (from $j\tau_1\omega$) in the forward path plus that induced by delay. The limitation on delay for this case can be simply stated: at the frequency at which the *magnitude* of the feedback signal equals that of the forward-path input signal (the *loss crossover frequency*) delay must cause less

than 90° of additional phase shift. With 90° of additional phase, the loss goes to zero on the $j\omega$ axis; with more than 90° , the zero of loss is in the right half plane. The amount by which the total phase is less than 180° is called the *phase margin*. The amount of phase that can come from the delay must be less than 90° by an amount equal to the phase margin. We do not emphasize the concept of phase margin here because it, like loop gain, is a subjective concept that depends on how the system is viewed. Here, the meaning of the feedback path and forward path are clear, so we can define the phase margin without ambiguity.

The most important result of this section is that the delay in the classic feedback equation can be accurately approximated by increasing the degree of the polynomial of loss by one—a quadratic approximation to the closed circuit loss is amply accurate for even small phase margins. For systems of higher degree, the loss usually includes additional phase from lumped elements giving more phase than the 90° of the classic equation. For these systems, the amount of delay must be less than 90° , making the approximation even better.

5.3 DEVICE TRANSIT-TIME DELAY AND KIRCHOFF'S LAW

The limitation on the bandwidth of the 741-type amplifier is directly attributable to the delay in a *common base stage* utilizing the slow lateral *pnp* transistor structure. A mere $0.08 \mu\text{s}$ of delay in this transistor requires correction (through the control time constant τ_1) that limits the performance of the amplifier to audio frequencies, as we saw in the active resonator circuit in Chapter 4.

All transistors include this same delay mechanism—the time it takes for charge carriers to traverse the physical distance from emitter to collector, or from source to drain. If all the carriers starting out at a given time from the emitter or source move in lockstep as they traverse this distance, the output signal current waveform will be a perfect replica of the emitter input signal current, except that it will be delayed in time. In physical devices, the transit time of carriers is a random variable, leading to *dispersion* of the signal, as shown in Fig. 5.9a.

Dispersion of the signal also arises from *RC* time constants involving emitter and base resistances and junction capacitances in the bipolar transistor and similar effects in other devices such as FETs. The delay is itself the essential limitation to performance, however, so that we can gain insight into the high-frequency performance of transistors by initially ignoring the dispersion and assuming that the carriers remain in phase in their trip from emitter or source to collector or drain.

The question addressed here is as follows. When the transistor is connected in the common emitter configuration, as in the amplifiers of the case study in Chapter 3, we assumed there that it behaves as a perfect integrator. In light of the preceding discussion, is this a valid assumption? Or is there a delay associated with the integration that must be taken into account?

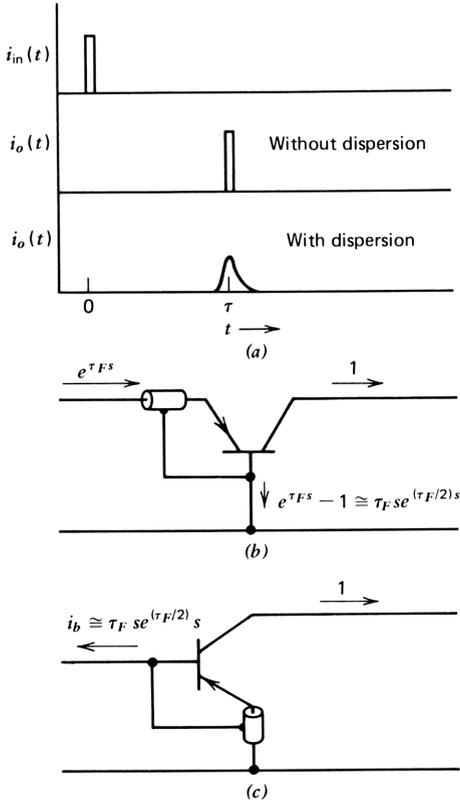


Figure 5.9. Transit time delay in a transistor connected in the common base configuration.

The answer is provided by applying Kirchoff's current law to the transistor. If we assume unit current flowing out of the collector, then the current flowing into the emitter is *advanced* in time by τ_F , the time for carriers leaving the emitter to reach the collector. We can represent the transit time delay in the common base transistor roughly as an ideal transistor (with no delay) that has a delay line connected in cascade with its emitter lead, as shown in Fig. 5.9b. The delay line, represented in the figure by a short piece of coaxial cable, has a time delay equal to τ_F . By Kirchoff's current law, the current flowing *out* of the base is the difference, so that in the frequency domain, we obtain

$$\frac{i_b}{i_c} = e^{\tau_F s} - 1 \tag{5.3-1}$$

This can be written

$$\frac{i_b}{i_c} = e^{\tau_F s/2} (e^{\tau_F s/2} - e^{-\tau_F s/2}) \tag{5.3-2}$$

If we substitute $s=j\omega$, we have

$$\begin{aligned}\frac{i_b}{i_c} &= e^{j(\tau_F\omega/2)}(e^{j(\tau_F\omega/2)} - e^{-j(\tau_F\omega/2)}) \\ &= e^{j(\tau_F\omega/2)}j2 \sin \frac{\tau_F\omega}{2}\end{aligned}\quad (5.3-3)$$

At low frequencies the small-angle approximation $\sin x = x$ applies, so that

$$\frac{i_b}{i_c} \simeq \tau_F\omega \exp j\left(\frac{\tau_F\omega}{2} + \frac{\pi}{2}\right) \quad (5.3-4)$$

which is the equation for the loss of an integration with delay.

Bode Plots for a Device with Transit Time Delay

In Fig. 5.10 the solid-line plots give the magnitude and phase calculated from eq. (5.3-3). The broken straight line on the magnitude curve is the low-frequency asymptotic loss, equal to the approximation of (5.3-4). The asymptote is a good approximation up to the unity loss frequency, where the magnitude is 0.5 dB greater than the function itself. The phase is represented exactly by (5.3-4). In terms of the complex frequency variable, (5.3-4) can be written

$$\frac{i_b}{i_c} = \tau_F s e^{\tau_F s/2} \quad (5.3-5)$$

as in Fig. 5.9c. Thus in the extreme case of ignoring dispersion, the transistor in the common emitter connection does behave as an ideal integrator with delay of half the transit time. Note that the *transit time itself* provides the integrating time constant in the common emitter connection.

The dashed curve in Fig. 5.10 is drawn for a less extreme case, in which τ_F is made up of two components, one the transit time delay without dispersion and the other an RC time constant arising from the combination of the emitter resistance and emitter junction capacitance. Thus, letting τ_{ec} represent the total delay time from emitter to collector, we obtain

$$\tau_{ec} = r_e C_{je} + \tau_F \quad (5.3-6)$$

The case shown is for these two time constants equal to half the previous transit time. The resulting delay is three-fourths that of the extreme curve, and the magnitude is equal to the low-frequency asymptote to well beyond the unity loss frequency.

In practical circuits, particularly integrated circuits, there will be additional sources of phase shift from parasitic capacitances in combination with lossy elements. A chief example is the phase shift induced by the base resistance of

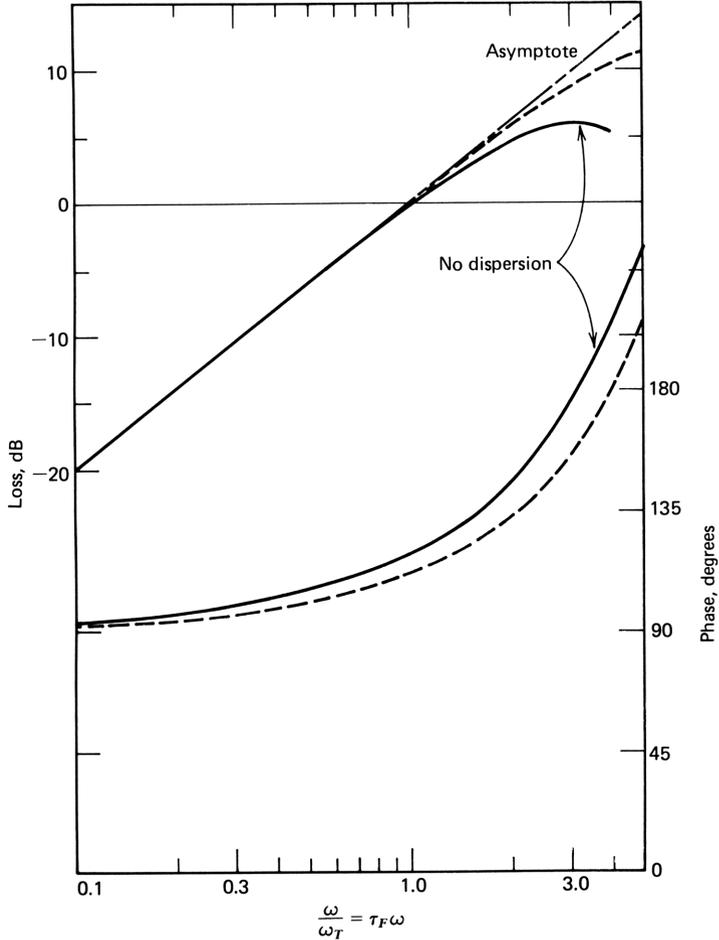


Figure 5.10. Current loss in the common emitter configuration. This solid curve shows the effect of pure transit time delay; the dashed curves represent the same transit time delay with one half from pure delay and the other half from $r_e C_{je}$.

the transistor in combination with interstage shunt parasitic capacitance arising from, for example, the collector-to-substrate capacitance of the previous stage.

The circuit and the signal flow graph are shown in Fig. 5.11, from which we obtain the current loss:

$$\begin{aligned}
 L = \frac{i_{in}}{i_o} &= - \left[r_e (C_1 + C_p) s + r_e r_b C_1 C_p s^2 \right] e^{\tau_d s} \\
 &= - \left[r_e (C_1 + C_p) s \left(1 + r_b \frac{C_1 C_p}{C_1 + C_p} s \right) \right] e^{\tau_d s} \quad (5.3-7)
 \end{aligned}$$

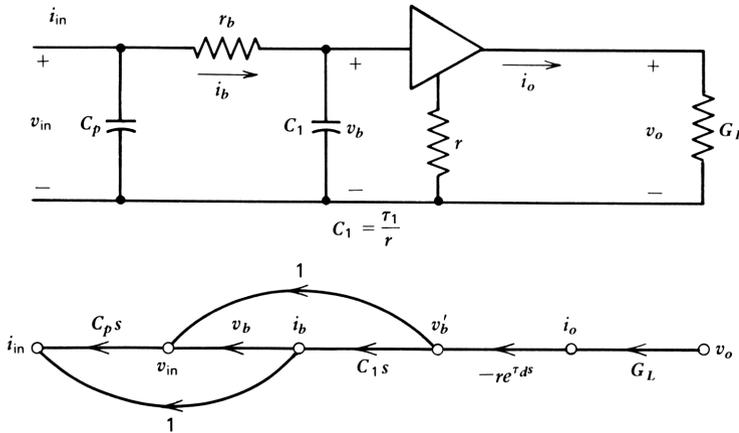


Figure 5.11. Parasitic delay from shunt capacitance and series base resistance in an integrated circuit.

The second factor in the brackets gives the additional phase shift referred to. Where the phase shift is small, we can replace the factor by a delay term $e^{\tau_B s}$, where

$$\tau_B = r_b \frac{C_1 C_p}{C_1 + C_p} \tag{5.3-8}$$

The advantage of treating this factor as a delay is that it may then be lumped with (added to) τ_d and the delay of other stages.

The conclusion here is that a common emitter stage behaves as an integrator with an amount of delay that is of the same order as the unity loss time constant. Base resistance contributes to this delay when the effects of parasitic capacitances commonly encountered in integrated circuits are included.

5.4 INCORPORATING DELAY INTO DESIGN B

The equations for Design B introduced in Section 3.3 used a crude model for the devices, leaving out not only delay, but several other effects such as current defect ratio, $\delta = 1/h_{fe}$, of the transistors. When all such effects are incorporated (we develop an accurate transistor model later), the equations become impossibly tedious. We now take an important step in maintaining control and understanding of the design process, namely, consolidating or “chunking” the equations into more than one hierarchical level.* For our purposes in this book, three such levels are sufficient: (1) the system level, (2) the circuit level, and (3) the device level. For Design B, the system level corresponds to the

*The inelegant term “chunking” is taken from D. R. Hofstadter’s elegant book, *Godel, Escher, and Bach*,¹ which suggested ways of organizing this section.

overall loss equations, and the circuit level corresponds to the analysis of the individual stages. We postpone a detailed discussion of the third level until later, being content to get along with an informally derived device model for the time being.

At the system level, Design B can be represented by the signal flow graph in Fig. 5.12a, in which $A_F(s)$ represents the outer feedback path and Z_1 , D_2 , and Y_3 , suitably delayed, represent the individual stage losses. In Fig. 5.12b the individual stage delays have been moved out of the forward path and placed to give equivalent performance, with the feedback path now incorporating the delay. The external delay is unimportant to the analysis of the system, as noted in Section 5.2.

The reason for the transition from Fig. 5.12a to 5.12b is to maintain a loss polynomial of minimum degree when the delay function is approximated by a polynomial. In Section 5.2 we showed that in a stable feedback system delay may be approximated by a quadratic polynomial. The feedback path loss function is of first degree; when multiplied by the quadratic delay approximation, the feedback and forward paths are each cubic functions, so that their sum is also cubic.

The equation for the voltage loss ratio of the amplifier is the only equation written at the system level:

$$L(s) e^{-\tau_D s} = A_F(s) e^{-\tau_D s} + Z_1(s) D_2(s) Y_3(s) \tag{5.4-1}$$

in which external delay has been incorporated on the left side and τ_D is the *sum* of the stage delays. To program this equation on the calculator or computer, the stage loss functions must be multiplied together to obtain the forward-path loss polynomial; $A_F(s)$ is then multiplied by the quadratic delay approximation and the resulting polynomial added term by term to the forward-path loss polynomial.

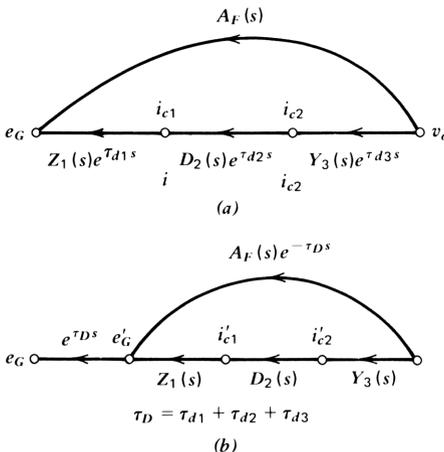


Figure 5.12. Signal flow graph of the Design B amplifier at the system level.

Feedback-Path Analysis

The feedback path consists of G_F and C_F in parallel; these supply a current that flows through R_G , so that with the quadratic approximation, we have

$$A_F(s) e^{-\tau_D s} \doteq -R_G \left[G_F + (C_F - G_F \tau_D) s + \left(\frac{\tau_D G_F}{2} - C_F \right) \tau_D s^2 + \frac{\tau_D^2 C_F}{2} s^3 \right] \tag{5.4-2}$$

Analysis of the First and Second Stages

The circuits and signal flow graphs for the first two stages are given in Fig. 5.13; the stage losses (with delay removed) can be written by inspection. The term $Z_1(s)$ is the generator voltage divided by the first-stage output current

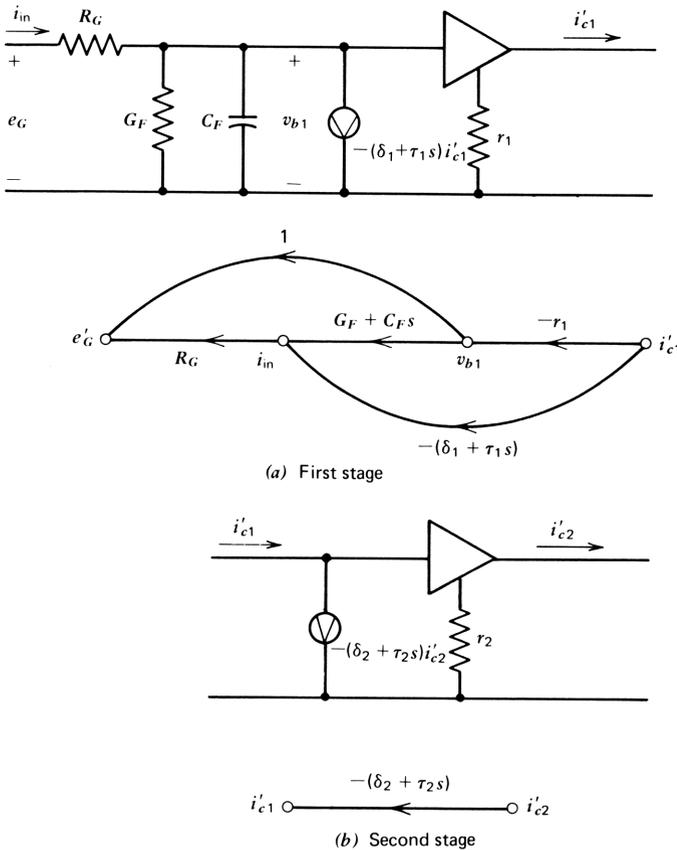


Figure 5.13. Circuit and signal flow graph for (a) first stage and (b) second stage.

and has units of impedance:

$$Z_1(s) = -[r_1 + (\delta_1 + G_F r_1)R_G + (\tau_1 + C_F r_1)R_G s] \quad (5.4-3)$$

The G_F and C_F terms account for the input loading of the outer feedback path, and δ_1 is the current defect ratio of the first stage, ignored in Chapter 3. The second-stage current loss ratio is simply

$$D_2(s) = -(\delta_2 + \tau_2 s) \quad (5.4-4)$$

Third-Stage Analysis

The third stage, with its local feedback, is also a feedback system in the sense of this section. It can be analyzed in a two-step hierarchy. This has already been done in deriving eq. (3.2-9), rewritten here with some notational changes and adding delay τ_{d3} to the forward path:

$$e^{\tau_{d3}s} Y_3(s) = - \left[G_2 + \frac{G'_L (\delta_3 + G_2 r_3 + \tau_3 s) e^{\tau_{d3}s}}{1 - r_3 G_2} \right] \quad (5.4-5)$$

in which $\tau_3 = r_3 C_3$ of the earlier equation; we have also added the current defect ratio term. The delay term on the left makes this equation correspond to the signal flow graph in Fig. 5.12a; delay has not yet been removed. When the delay is removed, this equation can be written

$$Y_3(s) = -[G_2 e^{-\tau_{d3}s} + G'_L (\Delta_3 + \tau'_3 s)] \quad (5.4-6)$$

where

$$G'_L = G_L + G_2 + G_F \quad (5.4-7)$$

$$\Delta_3 = \frac{\delta_3 + G_2 r_3}{1 - G_2 r_3} \quad (5.4-8)$$

and

$$\tau'_3 = \frac{\tau_3}{1 - G_2 r_3} \quad (5.4-9)$$

The output loading of both feedback paths is incorporated into G'_L , and the leakage caused by the input loading of G_2 is incorporated in Δ_3 ; the direct feedthrough denominator is incorporated in the latter two equations.

For small values of delay, we can approximate the delay by a linear term, so that

$$Y_3(s) = -[G_2 + \Delta_3 G'_L + (\tau'_3 G'_L - G_2 \tau_{d3})s] \quad (5.4-10)$$

a binomial in the frequency variable.

Analysis of the Complete Circuit

To analyze the circuit on the calculator, eq. (5.4-10) is programmed and the coefficients stored. Then eq. (5.4-4) gives the coefficients of the second stage, and these are multiplied by those of the third stage; the process is repeated for the first stage. Finally, eq. (5.4-2) is evaluated and added to the result, giving the loss of the entire circuit, less the external delay. This is done in program “AN2” in Appendix B.

In the design of the amplifier we established the values of three dominant elements to realize the desired loss polynomial. In the earlier designs we ignored delay; we might suspect that delay affects the design of the amplifier in an important way and that if delay becomes sufficient, the design will become unrealizable, at least for a prescribed bandwidth. For lesser amounts of delay, the dominant elements G_F , C_F , and G_2 will change as the delay is increased.

Synthesis of Designs for Various Values of Delay

To see how delay affects the design of the amplifier, program “AN2” was used with program “SJ”, the synthesis program described in Chapter 3. In this way, several designs were made for various assumed amounts of individual stage delays from zero to 3.0 ns (all three delays were assumed equal). Naturally, only one value of delay is correct for a given device; we vary the delay for tutorial purposes to show the effect on the design.

As shown in Section 5.3, for normal transistors the delay will be roughly between one-half and one times the transit time, or 0.5–1 ns for the devices assumed here. The required values of G_2 , C_F , and G_F (the dominant elements) were found by synthesizing the amplifier assuming various values of delay. Each such synthesis resulted in a set of values for the dominant elements; these are plotted as a function of delay in Fig. 5.14. The value of dc loss obtained by the synthesis is also plotted.

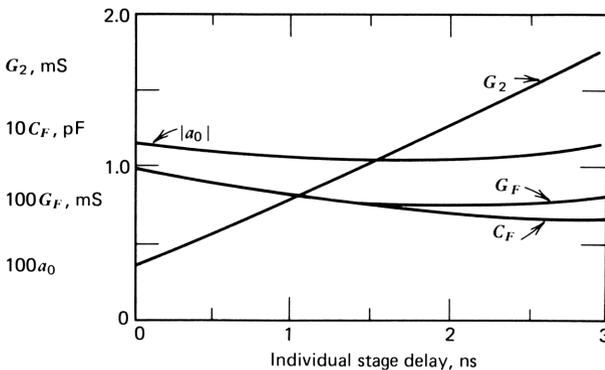


Figure 5.14. Dominant element values as a function of delay for Design B.

As the delay is increased G_2 must be increased correspondingly. The reason is clear qualitatively from eq. (5.4-2), in which the delay decreases the quadratic coefficient by $R_G C_F \tau_D$; G_2 must rise to replace the lost portion of the quadratic coefficient. [Note that in these curves the individual stage delay is shown on the abscissa, whereas τ_D in eq. (5.4-2) is the total delay of the three stages.]

In the approximate analysis in Chapter 3 the defect current ratio of the transistors was ignored. When it is included, it increases the quadratic loss term so that G_2 is not required to be as high as before. Figure 5.15 shows the effect on G_2 of varying (all three) δ values from 0 to 0.02. The design is not adversely affected by increasing current defect ratio, provided that this ratio is well controlled. The sensitivity of the loss to δ can be found by using program "SLX" with program "AN2", both in Appendix B. Clearly, the sensitivities of loss to any or all components and device parameters of the amplifier can be found in this way. If a statistical analysis of the effect of component variations is required, program "STAT" can be invoked. In all, a complete picture of the performance of an amplifier realized as shown can be derived from the simple programs described in Appendix B, or from similar programs written for the computer. The calculator programs are compatible in using memory register locations in common.

In this design the greatest complications arose in the analysis of the third stage, rather than in either the first two stages, the feedback path, or their combination. The reason is that where the gain of an amplifier stage is low, we cannot justify simplifying assumptions that are appropriate in the high-gain

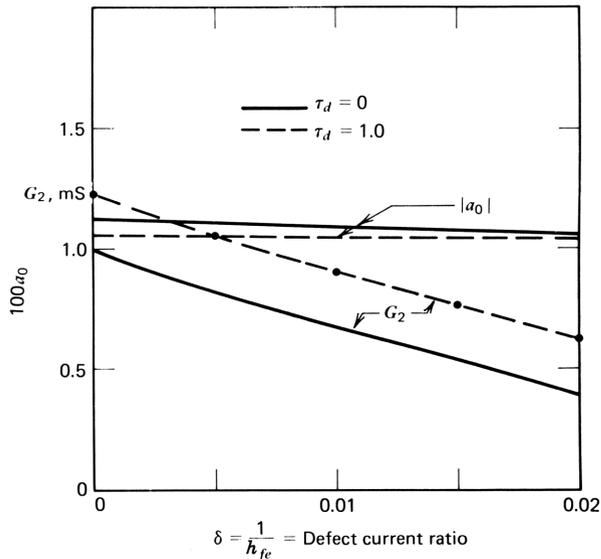


Figure 5.15. Resulting design as a function of the defect current ratio $\delta = 1/h_{fe}$ (equal for the three stages).

case (e.g., ignoring direct feedthrough is an example); thus a more complete analysis is required. The third stage is a good case in point; the local feedback must be analyzed fully to maintain accuracy. In Section 5.5 we extend the accuracy of the third-stage analysis by using a quadratic delay approximation.

5.5 QUARTIC SYNTHESIS: ADDITION OF LOAD CAPACITANCE

If we use a quadratic approximation for the delay of the output stage, the loss polynomial of the amplifier becomes a quartic. The extension to a quartic is worthwhile; it becomes unavoidable when a sufficiently large load capacitance is added to the amplifier. The first question to be settled concerns what polynomial we should use as a specification for the design. In the cubic design we used a Butterworth polynomial, an arbitrary but reasonable choice. Can we substitute a quartic polynomial that gives essentially the same performance, and if so, what is it?

Finding a Suitable Quartic Polynomial

By “same performance” we adopt the following notion. When a higher-degree polynomial is substituted, both the in-band sensitivities and the out-of-band stability margin are potentially affected. The in-band sensitivities of loss to the polynomial coefficients (for the polynomial normalized to unity loss) are given by $b_i(\omega/\omega_0)^i$ by the sum rule since the in-band loss is unity. Hence, to maintain the same in-band sensitivities, the coefficients that are important in band (the ones of lower degree) should be the same for both the quartic and the cubic. The out-of-band stability margin can be measured by the damping of the roots nearest the $j\omega$ axis: if this damping is kept the same, the margin against instability will be essentially unaffected. These two rules tell us what to look for in a quartic polynomial. For the cubic Butterworth, $b_0=1$ and $b_1=2$. The damping of the roots nearest the $j\omega$ axis is 0.5 for the cubic (the angle of the roots with the $j\omega$ axis is 30°). These are the numbers to be emulated by the required quartic.

We could try a quartic Butterworth polynomial, but as the discussion of Section 2.5 shows, the angle of the roots nearest the $j\omega$ axis is only 22.5° for this polynomial, giving an inadequate stability margin. The Bessel polynomials have better stability margins, so that we should go in this direction. Consider, for example, the transitional polynomial with $m=0.5$ given in Table 2.1 or by program “POLYTBL”:

$$L(p) = 1 + 2.89p + 3.87p^2 + 2.86p^3 + p^4 \quad (5.5-1)$$

Next, we normalize or scale this polynomial by frequency transformation (using Program “N”) to obtain $b_1=2$, so that the in-band sensitivity to b_1 is the same as the cubic: letting $q=(2.89/2.0)p$, we have

$$L(q) = 1 + 2q + 1.85q^2 + 0.94q^3 + 0.23q^4 \quad (5.5-2)$$

The roots of this polynomial (from program "ROOTS") are at $s_{1,2} = 1.40/\pm 158.7^\circ$ and $s_{3,4} = 1.49/\pm 119.8^\circ$. The pair of roots nearer the $j\omega$ axis are at an angle of 29.8° , so that the out-of-band stability margin remains the same as the cubic. (Had we not been so lucky on the first try, we would have repeated the procedure for different values of m in the transitional polynomial until the requirements were met.)

Equation (5.5-2) gives us a normalized quartic polynomial that has essentially cubic Butterworth behavior. The frequency response also remains the same in band, as a computation using program "BODE" would show. The quartic term in (5.5-2) represents an upper limit dictated by stability requirements. If the amplifier has a value of b_4 that is less than the value shown, the stability margin becomes greater. The in-band response scarcely changes at all since the sensitivity to the quartic coefficient in band is $b_4(\omega/\omega_0)^4$, negligible for ω less than ω_0 . Therefore, we can *monitor* the value of b_4 in the analysis of the amplifier, making sure that it remains within bounds. We *do not* have to *synthesize* this coefficient—just make sure that it is smaller than or equal to the requirement.

Control of the Cubic Coefficient

When a sufficiently large load capacitance is added to the output of the amplifier, the third-stage loss becomes a quadratic function even in the absence of delay. Delay tends to reduce the damping of this quadratic. A fourth control element is thus needed in addition to G_F , C_F , and G_2 , one to control the cubic coefficient of the entire amplifier, or the damping of the output-stage quadratic. A capacitor C_B connected in parallel with G_2 , shown in Fig. 5.16, will serve this function; it is analogous to the feedback capacitance in the 741 amplifier. It provides the control time constant for the output stage. The equation for the output stage may be derived from (5.4-6) by replacing G'_L and G_2 by Y'_L and Y_2 , where

$$\begin{aligned} Y'_L &= G'_L + C'_L s \\ &= G_L + G_2 + G_F + (C_L + C_B + C_F)s \end{aligned} \quad (5.5-3)$$

and

$$Y_2 = G_2 + C_B s \quad (5.5-4)$$

With this replacement, the third stage is characterized by the equation

$$Y_3(s) = -(G_2 + C_B s)e^{-\tau_{D3}s} + (\Delta' + \tau'_3 s)(G'_L + C'_L s) \quad (5.5-5)$$

where

$$\Delta' = \frac{\delta_3 + G_2 r_3}{1 - G_2 r_3} \quad (5.5-6)$$

$$\tau'_3 = \frac{\tau_3 + C_B r_3}{1 - G_2 r_3} \quad (5.5-7)$$

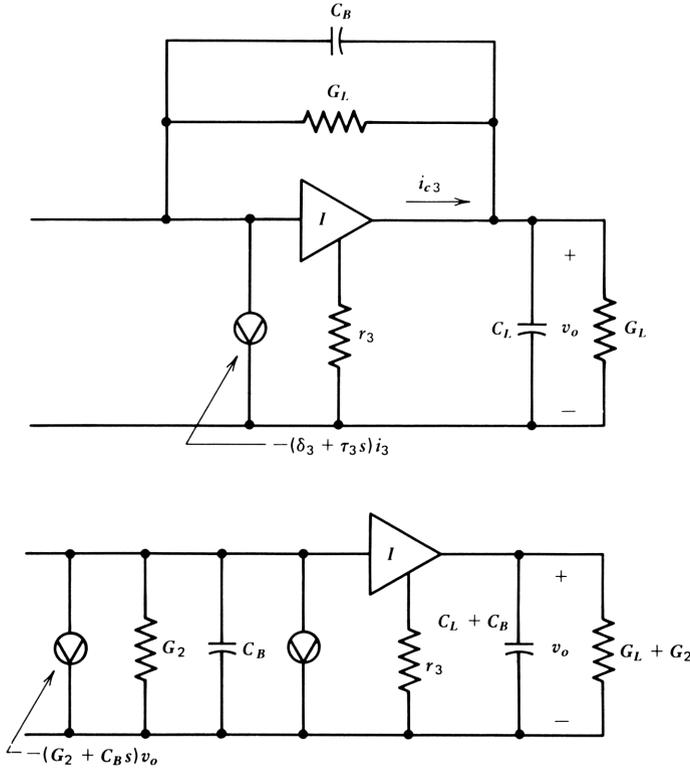


Figure 5.16. Circuit diagram of the third stage showing addition of control capacitance C_B .

and

$$\tau_{D3} = \tau_{d3} + \frac{r_3(C_B + G_2\tau_d)}{1 - G_2r_3} \tag{5.5-8}$$

The denominators in these equations represent the increase in dc loss due to direct feedthrough, which is usually small. Equation (5.5-8) includes a small increase in third-stage delay arising from the direct feedthrough. This direct-feedthrough delay arises because the base input voltage of the third stage causes a current to flow directly to the output through C_B . This adds the term $-r_3(C_B + G_2\tau_d)$ to the denominator and gives a right half plane pole of loss. Since it is a small effect, we can replace the pole by an equivalent delay that gives the same phase response, as expressed by the second term in (5.5-8).

We now represent the third stage delay by the quadratic approximation, so that

$$Y_3(s) = -G_2 + \Delta'G'_L + (C_B - G_2\tau_{D3} + \tau'_3G'_L + \Delta'C'_L)s + \left[\left(\frac{G_2\tau_{D3}}{2} - C_B \right) \tau_{D3} + \tau'_3C'_L \right] s^2 \tag{5.5-9}$$

This equation, and eqs. (5.4-2) to (5.4-4) (for the feedback path and the first two stages), constitute a complete analysis of the amplifier with capacitive load and control (by C_B). The quadratic delay approximation assures accuracy in all practical cases. An analysis program for Design B incorporating this equation is given as program "AN3" in Appendix B.

The effect of load capacitance on the design is shown in Fig. 5.17, showing the polynomial coefficients in Fig. 5.17a and the dominant circuit elements in

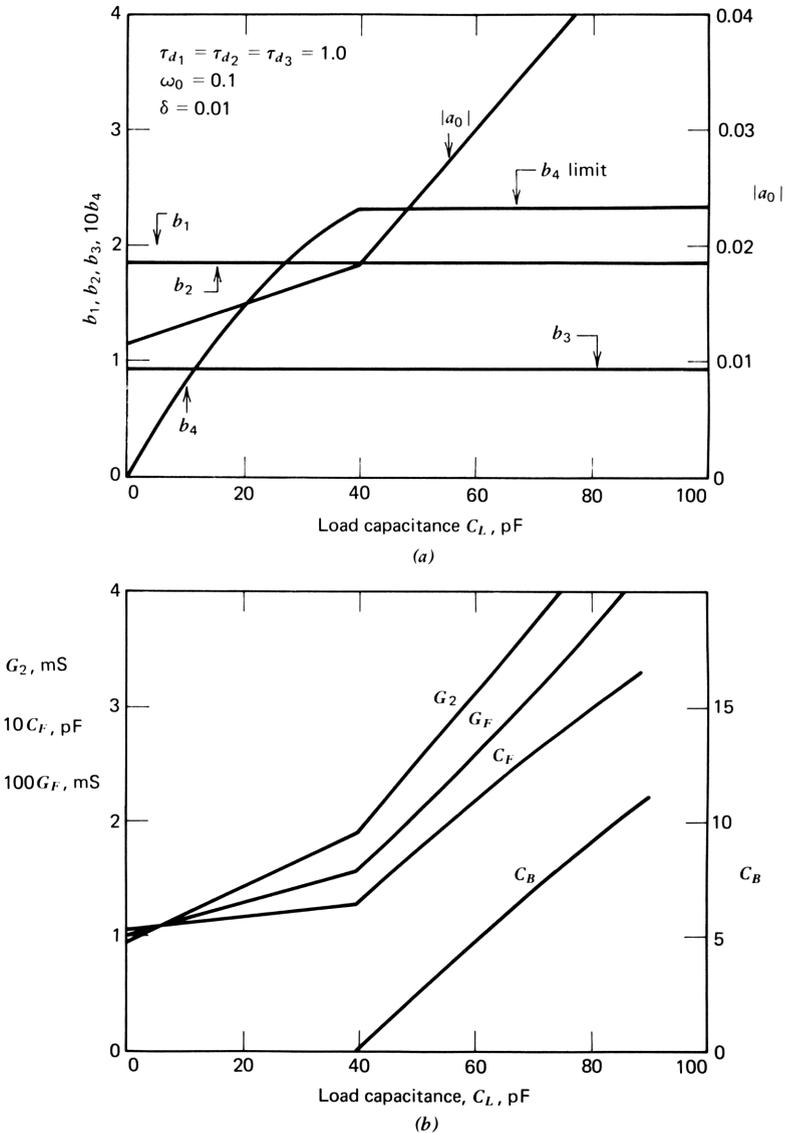


Figure 5.17. Resulting design as a function of load capacitance for a quartic design.

Fig. 5.17b as functions of the load capacitance. The plots were obtained using program “AN3” with “SJ”, with C_B chosen as zero until b_4 exceeds the limit set by eq. (5.5-2). For larger values of C_L , C_B is chosen (by trial and error, or “manual iteration”) to give $b_4=0.23$.

Values of load capacitance greater than 40 pF require correction by C_B . In this region the increase in the cubic coefficient caused by C_B means that the dc loss increases rapidly, and G_2 also increases rapidly. For sufficiently high values of load capacitance, the design becomes impractical at the given cutoff frequency. In such cases an emitter follower can be added at the output, as shown in Fig. 5.18.

In this case the loss polynomial of the amplifier becomes a quintic. A suitable quintic may be found as shown previously for the quartic, in which the in-band sensitivities and the out-of-band stability margin are preserved. To realize such a quintic as a circuit, means must be provided to synthesize the quartic term; a straightforward method is to add a capacitance from the input of the emitter follower to ground, as shown in Fig. 5.18. In this case the cubic term would be provided largely by C_B since the equivalent load conductance at the output of the third stage is small. Use of an emitter follower is then a way of reducing the sensitivity of loss to the load conductance and capacitance.

As we have noted, the transistor model used here is a crude one; better ones are developed later. It is important to emphasize here that changes in the model *will not change the design in any fundamental way*. It will model more accurately the values of the time constants and delay and the low-frequency loss. But the physical device is adequately represented qualitatively by the time constant, the delay, and the low-frequency loss. Even parasitics encountered in

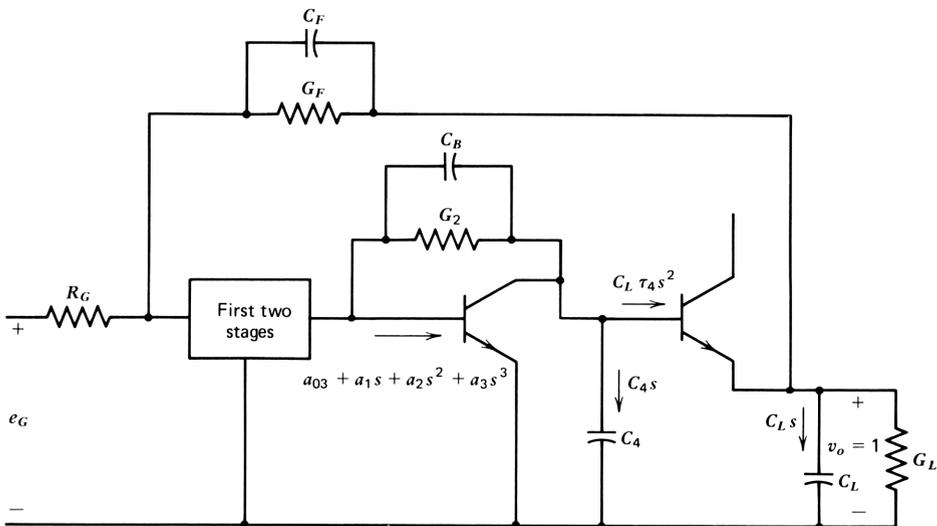


Figure 5.18. Use of an emitter follower to increase the load capacitance that can be driven by the amplifier; C_4 controls the quartic coefficient.

integrated circuits will be accommodated within the present framework. It is because of this fact that the separation into hierarchies works well: there is not a large interaction among the hierarchies, in this case between the device level and the circuit level, thus allowing us to solve subproblems of quite manageable size.

5.6 A PRACTICAL 300 MHz AMPLIFIER

Now that delays can be accounted for, we can use the synthesis procedure to design a practical version of the case study amplifier with 300 MHz bandwidth, useful in repeaters for high-speed digital transmission. In lieu of a complete transistor model, we estimate the time constants and delays on an informal basis, anticipating results to be derived later.

The amplifier, shown in Fig. 5.19, operates between 75Ω impedances, with matching resistors included in the amplifier to minimize signal reflections in the cables connected to the amplifier. AC coupling is used at both input and output. The amplifier is assumed to be realized as a silicon integrated circuit in which the transistors are isolated by reverse-biased semiconductor junctions. The circuit is similar to Design B, except that diodes have been connected in series with the base leads of the second and third stages to provide dc voltage bias for the first and second transistor collectors. Shunt conductances connected between base and ground draw 1.0 mA through these diodes, giving them a dynamic resistance of about $0.030 \text{ k}\Omega$. The dc drop across the diodes is roughly 0.75 V, as is the base to ground voltage of the transistors, so that the dc collector voltage of the first and second stages is 1.5 V. The dynamic resistance of the diodes adds to the effective base resistance of the second and third stages.

To obtain the values of the stage time constants and delays as well as the dc current losses, we must descend to the third level of the hierarchy of equations — to the device level. For the present, we take the following equations to approximate the stage characteristics. First, the individual stage dc current losses are

$$\Delta_i = \delta_i + g_i r_{Ei} \quad (5.6-1)$$

where r_{Ei} is the total emitter resistance of the i th stage, including the dynamic diode resistance and the *emitter contact resistance*, r'_{ei} :

$$r_{Ei} = \frac{kT}{qI_{C_i}} + r'_{ei} \quad (5.6-2)$$

The voltage drop across this resistance, proportional to output current, drives a current through g_i , the conductance used to draw current through the biasing diodes, increasing the dc loss of the stage.

This same voltage drop drives a frequency-dependent current through any shunt capacitance at the base, thereby increasing the stage time constant,

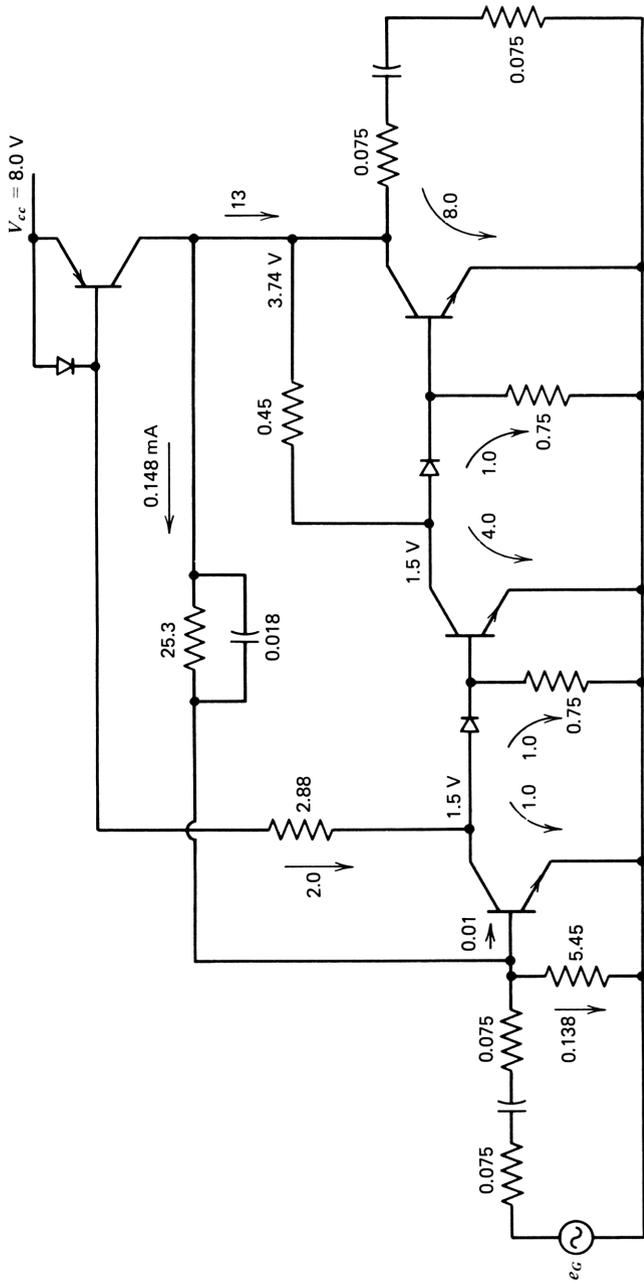


Figure 5.19. Circuit design of a 300 MHz amplifier showing dc biasing.

which is given by

$$\tau_i = \tau_{Fi} + r_{Ei}(C_{jei} + C_{PTi}) + r_{b(i+1)}C_{jci} \quad (5.6-3)$$

in which τ_{Fi} is the forward transit time of the transistor, C_{jei} is the emitter junction capacitance, and C_{PTi} is the total parasitic shunt capacitance at the base node of the transistor. Where applicable, this parasitic capacitance includes the collector-to-substrate capacitance of the previous transistor (obviously absent for the first stage), the collector junction capacitance of both the previous stage and the i th stage (output loading of the previous stage and input loading of the i th stage), and any wiring or other shunt capacitance of the interstage. The series biasing diodes add shunt parasitic capacitance. Thus for the three stages, we obtain

$$C_{PT1} = C_{jc1} + C_{p1} \quad (5.6-4)$$

$$C_{PT2} = C_{cs1} + C_{jc1} + C_{jc2} + C_{p2} \quad (5.6-5)$$

$$C_{PT3} = C_{cs2} + C_{jc2} + C_{jc3} + C_{p3} \quad (5.6-6)$$

The last term on the right of eq. (5.6-3) adds to the stage time constant in the following way. The input current to the following ($i+1$)st stage almost all flows through the total base resistance of that stage, creating a voltage at the collector of the i th stage. This voltage causes a current to flow through the collector junction capacitance that adds to the input current of the i th stage and augments the time constants of the stage. The base resistance r_B includes the series 0.03 k Ω dynamic resistance of the biasing diodes.

Finally, the stage delays are estimated by the equation

$$\tau_{di} = \frac{\tau_{Fi}}{2} + r_{ei}C_{jci} + R_{Bi} \frac{\tau_i C_{PTi}}{\tau_i + r_{Ei}C_{PTi}} \quad (5.6-7)$$

The first term on the right is the transit time delay, discussed in Section 5.3, assuming no dispersion. The second term is the equivalent delay introduced by the direct-feedthrough term. The third term is the equivalent delay introduced by the base resistance and parasitic capacitance, discussed in Section 5.3.

These device-level equations are programmed on the calculator in program "DEV" in Appendix B, which stores the results in memory locations required by program "AN2" or "AN3". Program "SJ", in conjunction with the analysis program, then does the amplifier synthesis. Thus program "DEV", working at the device level, provides the data for program "AN3", working at the circuit level. This program allows program "SJ", working at the system level, to synthesize the feedback amplifier, giving the circuit values in Fig. 5.19. Device parameters and details of the computation are given in Appendix B.

DC Bias Design

DC voltages for the first and second stages have been provided by the series diodes; the desired collector currents of 1.0, 4.0, and 8.0 mA and the output collector voltage must now be supplied. The latter voltage should be 3.0–4.0 V to handle the output signal voltage linearly. The output voltage is the sum of the collector voltage of the second stage and the drop across G_2 , found by the synthesis to be 2.23 mS; hence

$$V_{CE3} = 1.5 + \frac{4 + 1}{2.23} = 3.74 \text{ V}$$

a satisfactory value.

The dc current flowing through G_F is $(V_{CE3} - V_{BE1})G_F$, or

$$I_{GF} = (3.74 - 0.75)(0.0395) = 0.148 \text{ mA}$$

The first-stage base current is $I_{C1} = 0.01$ mA; the remainder of I_{GF} must be drawn off by a conductance between base and ground of the first stage. Its resistance is

$$R_{BE1} = \frac{0.75}{0.148 - 0.01} = 5.45 \text{ k}\Omega$$

Direct currents for the output and second stages are provided by a *pnp* transistor acting as a current source. Such circuits are described in detail later; the *pnp* transistor must provide the sum of I_{C3} , I_{G2} , and I_{GF} , or 13.15 mA. The sum of the first-stage collector current and the diode bias current (2 mA) is provided through a resistor that activates the diode of the current source. Its value is $(V_{CC} - V_{CE1} - V_{dCS})/2.0$, where V_{dCS} is the diode drop of the current source (~ 0.75 V). Hence the first-stage load resistance is $(8 - 1.5 - 0.75)/2 = 2.88$ k Ω . The current source is designed to give the required ratio of diode to transistor current (2/13.15), by making the areas of the diode and transistor in this ratio. The only effect of the current source on the synthesis is to add a shunt capacitance to the output collector of about 2.0 pF. This added to the collector-to-substrate capacitance of the output transistor and wiring capacitance brings the total parasitic load capacitance to about 2.5 pF, a value incorporated into the design. The collector junction capacitance of the output stage is incorporated into the design as C_B in the analysis program.

Variation of Loss

A complete sensitivity study of the amplifier may be made by using program “SCX” to find the coefficient-to-component sensitivities. This program operates in conjunction with “AN2” or “AN3”. The sensitivities thus calculated can be used to find the sensitivity of loss to the components at any given frequency by use of program “SLX”. This program is a “stand alone”; each set

of coefficient-to-component sensitivities must be entered manually. Finally, the standard deviation of loss and phase can be calculated by using program “STAT”. This is illustrated with numerical results for this design in Appendix B.

5.7 QUANTIZED FEEDBACK

A feedback technique that illustrates the handling of delay in the time domain as opposed to the frequency domain analysis used in Section 5.17 is quantized feedback. It is used in digital transmission (PCM) systems to eliminate the necessity for transmission of low frequencies.* In this brief treatment, we do not discuss the details of digital transmission, but rather focus on the quantized feedback arrangement itself. A practical quantized feedback arrangement and design are given in this section. The design is programmed in Appendix B if further study is desired.

In Fig. 5.20 the *decision circuit* of a digital regenerator is shown, with a noisy digital signal at its input. The function of the decision circuit is to observe the signal at a time near the center of the pulse or time slot and to generate a new pulse for transmission to the following regenerator some distance away. The decision circuit may be a clocked flip-flop, for example, and must have a minimum of a one-half time-slot delay. At high digital rates, a master-slave flip-flop is often used, causing a full time slot of delay.

If the signal is ac coupled, the pulse waveform will exhibit sag, as shown in Fig. 5.20*b*. Since the flip-flop threshold is set at the center of the extremes of voltage, this sag reduces the margin against errors that might be induced by a large noise spike, for example. The worst case of such sag would occur with a long string of pulses of like polarity.†

The output of the decision circuit is a (delayed) replica of the input, so that the possibility exists of removing the sag by adding a filtered signal from the output to the input of the decision circuit. This is quantized feedback—the output signal that is filtered and fed back to the input is quantized to a discrete set of signaling levels, two in the example discussed here.

Simple Quantized Feedback System with Single Forward-Path Cutoff

A simple quantized feedback system is shown in Fig. 5.21*a*, in which a simple low-frequency cutoff is placed in the forward path and a complementary filter is placed in the feedback path. The time responses of the forward-path signal and the feedback signal and their sum are shown for a long string of positive pulses. The forward-path signal is a simple exponential decay, given by the equation

$$u(t) * h(t) = k_1 e^{-t/\tau_1} \quad (5.7-1)$$

*Other methods of suppression of low-frequency transmission in digital systems include line coding (e.g., *bipolar* or *AMI*—alternate mark inversion—in which successive pulses are inverted in polarity) and dc restoration, but these are substantially less efficient than quantized feedback.

†In this discussion we assume 100% duty factor pulses.

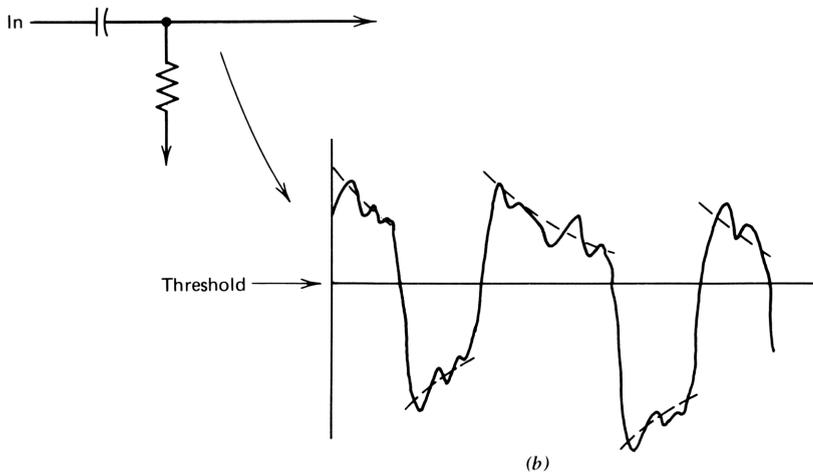
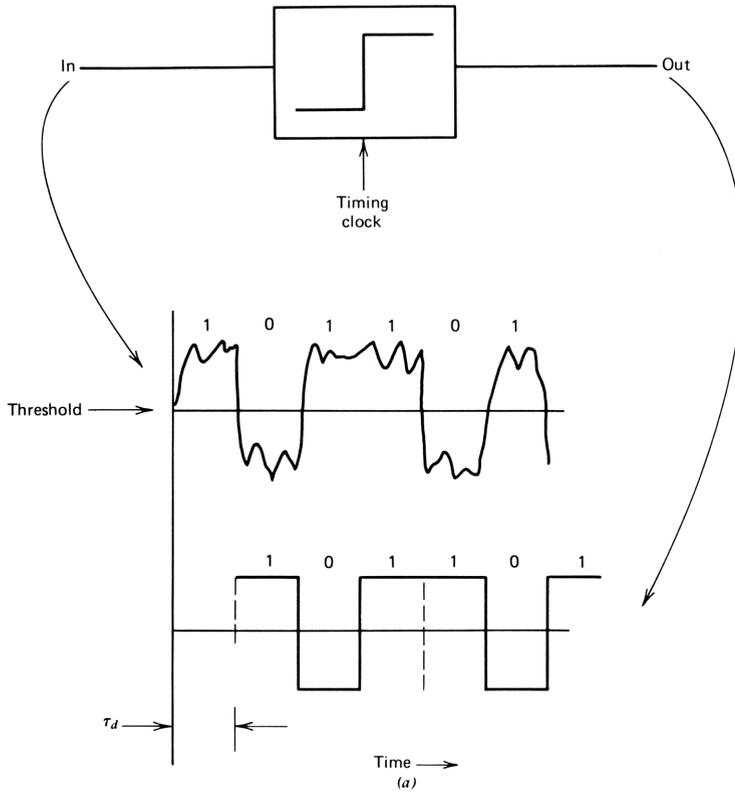


Figure 5.20. Decision circuit of a digital regenerator showing regeneration of a string of pulses and sampling delay. The effect of ac coupling on the incoming pulse stream is shown in part *b*.

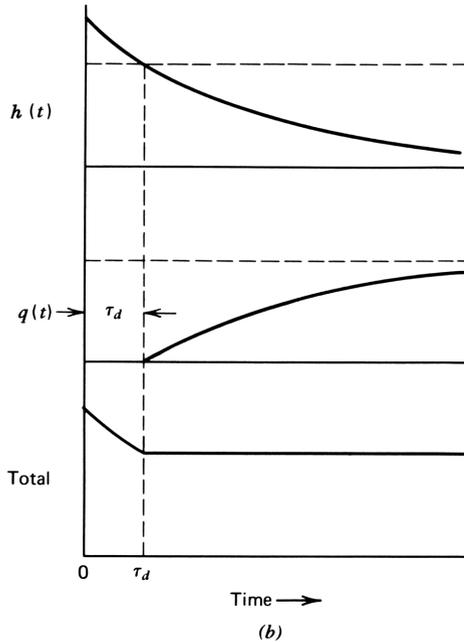
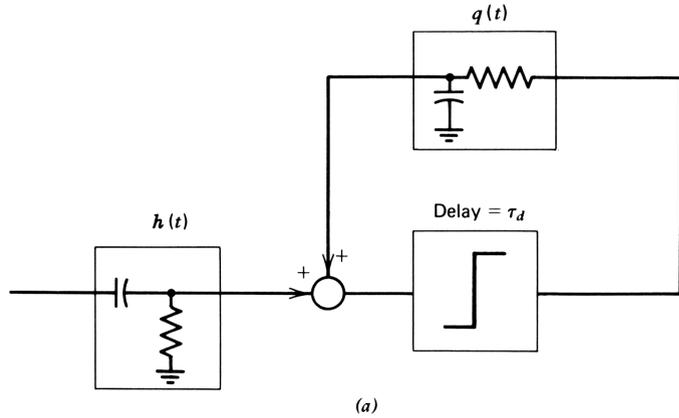


Figure 5.21. Simple quantized feedback arrangement showing waveforms for a long string of positive pulses beginning at $t = 0$. Feedback restores pulse baseline starting after $t = \tau_d$.

in which $u(t)$ is the unit step function used to represent the long string of pulses and is convolved with $h(t)$, the impulse response of the forward-path cutoff; k_1 is the residue in the root of the ac cutoff; and τ_1 is the cutoff time constant. After time $t = \tau_d$, the decision circuit generates an output signal of fixed amplitude, and the quantized feedback signal becomes available. It must be complementary to the forward-path signal:

$$u(t - \tau_d) * q(t - \tau_d) = 1 - k_2 e^{-(t - \tau_d)/\tau_1} \tag{5.7-2}$$

in which $u(t-\tau_d)$ is the unit step delayed by the decision circuit, convolved with $q(t-\tau_d)$, the impulse response of the feedback filter; and k_2 is the residue of the feedback filter root. For the two signals to cancel after $t=\tau_d$, the two time constants must be equal, so we can write

$$k_1 e^{-t/\tau_1} = k_2 e^{-(t-\tau_d)/\tau_1} \quad (5.7-3)$$

or

$$k_1 = k_2 e^{\tau_d/\tau_1} \quad (5.7-4)$$

Thus if the signal amplitude of the input signal to the forward path equals that of the decision circuit output, perfect cancellation after the delay time requires a gain factor in the forward path of e^{τ_d/τ_1} . The time responses are shown in Fig. 5.21*b*.

For delay times greater than one-half time slot, the error at the sampling instant in the first time slot, as $t=\tau_B/2$ (where τ_B is the width of a time slot) is not zero; it is given by

$$E_1 = \exp \frac{(\tau_d - \tau_B)/2}{\tau_1} - 1$$

where E_1 is the error, normalized to unity amplitude. If the delay is equal to one time slot, for example, and the low-frequency cutoff is $0.02 f_B$ ($\tau_1 = 2\pi\tau_B/0.02$), the error will be 6.5%.

The resulting filters are given by translation back to the frequency domain:

$$\frac{1}{s} H(s) = \frac{k_1 \tau_1}{1 + \tau_1 s} \quad (5.7-5)$$

or

$$H(s) = k_1 \frac{\tau_1 s}{1 + \tau_1 s} \quad (5.7-6)$$

a simple low-frequency cutoff with a gain factor. Similarly,

$$\frac{1}{s} Q(s) = \frac{k_0}{s} - \frac{k_2 \tau_1}{1 + \tau_1 s} \quad (5.7-7)$$

Since at $t=\tau_d$, $q(t)=0$, and by our normalization to unit amplitude, $k_0=1$, we have $k_2=1$, so that

$$Q(s) = \frac{1}{1 + \tau_1 s} \quad (5.7-8)$$

Here, we have found the filter functions by working in the time domain, using the residues as derived by observing the constraints on the time domain

performance. Essentially the same result can be obtained in the frequency domain by observing that

$$H(s) + Q(s)e^{-\tau_d s} = 1 \quad (5.7-9)$$

If we take the quantized feedback filter as that given above, we obtain $H(s)$ as

$$H(s) = \frac{1 + \tau_1 s - e^{-\tau_d s}}{1 + \tau_1 s} \quad (5.7-10)$$

Replacing the delay term by $1 - \tau_d s$, we have

$$H(s) \doteq \frac{(\tau_1 + \tau_d)s}{1 + \tau_1 s} \quad (5.7-11)$$

which again is a low-frequency cutoff with a gain factor of $(\tau_1 + \tau_d)/\tau_1 \approx e^{\tau_d/\tau_1}$, as before. This involves a straight-line approximation to the time response during the first time slot, which is accurate for usable values of delay.

The simple quantized feedback system described previously is seldom used because the error in the first (and all succeeding) time slots can be removed for all practical purposes by use of a *quadratic* cutoff in the quantized feedback filter instead of the linear one used previously, with negligible extra cost.

Quantized Feedback System with a Quadratic Cutoff

To see why addition of a single-filter section enhances the operation of quantized feedback greatly, consider the quantized feedback filter shown in Fig. 5.22. It uses three equal resistors and two equal capacitors. The transfer function is easily written by inspection of the signal flow graph (which takes the quantized feedback current as the independent variable and finds the decision circuit output voltage):

$$\begin{aligned} \frac{v_3}{i_{\text{in}}} &= \frac{1}{Q(s)} = 3R + 4R^2Cs + R^3C^2s^2 \\ &= 3R(1 + RCs) \left(1 + \frac{RC}{3}s \right) \end{aligned} \quad (5.7-12)$$

A second root, at three times the frequency of the first, is supplied by the extra section. The step response of this filter is given by the inverse Laplace transform of

$$\frac{1}{s} Q(s) = \frac{1}{s(1 + \tau_1 s)(1 + \tau_2 s)} \quad (5.7-13)$$

where $\tau_1 = RC$ and $\tau_2 = RC/3$; thus

$$u(t) * q(t) = 1 - k_3 e^{-t/\tau_1} - k_4 e^{-t/\tau_2} \quad (5.7-14)$$

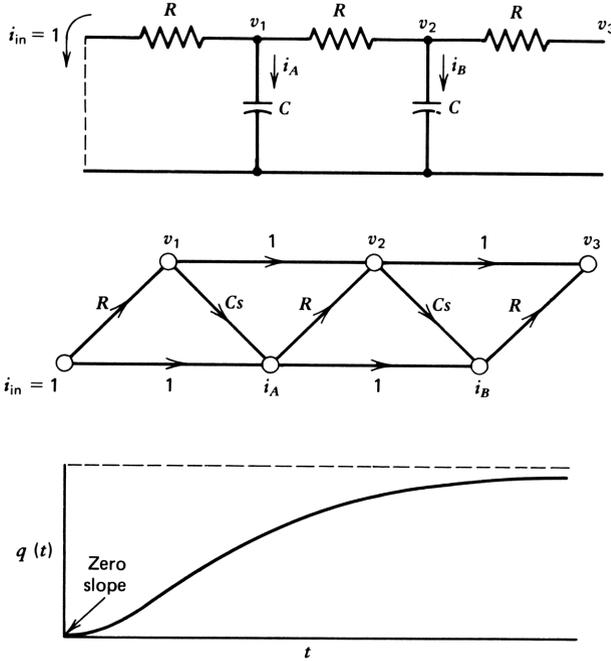


Figure 5.22. Quadratic feedback filter with equal R and C values and signal flow graph for analysis. Step response has zero slope at the beginning of the step.

The key advantage of the extra section is that the initial slope of the feedback time response is zero, as shown in Fig. 5.22b, instead of the large slope for the single section filter. When this response is delayed and added to a complementary $u(t) * h(t)$, the slope of $h(t)$ must also be zero at $t = \tau_d$. Thus, during the first time slot the slope is low and the error is consequently small at the first sampling instant.

To calculate the complementary forward path filter, we can work in either the frequency or time domains. In the frequency domain, using eq. (5.7-9), we have

$$H(s) = \frac{(1 + \tau_1 s)(1 + \tau_2 s) - e^{-\tau_d s}}{(1 + \tau_1 s)(1 + \tau_2 s)} \tag{5.7-15}$$

Using a quadratic approximation to $e^{-\tau_d s}$, we obtain

$$H(s) = \frac{(\tau_1 + \tau_2 + \tau_d)s + [\tau_1 \tau_2 - (\tau_d^2/2)]s^2}{(1 + \tau_1 s)(1 + \tau_2 s)} \tag{5.7-16}$$

which can be split into an ac cutoff, a gain factor, and a doublet (pole-zero

pair):

$$H(s) = \frac{\tau_1 s}{1 + \tau_1 s} \cdot K \cdot \frac{1 + \tau_{NF} s}{1 + \tau_2 s} \quad (5.7-17)$$

with

$$K = 1 + \frac{\tau_2 + \tau_d}{\tau_1} \quad (5.7-18)$$

and

$$\tau_{NF} = \frac{\tau_1 \tau_2 - (\tau_d^2 / 2)}{\tau_1 + \tau_2 + \tau_d} \quad (5.7-19)$$

Figure 5.23 shows the two-section filter, which altogether includes two capacitors and four resistors. The values of the elements were determined for a digital transmission system with a signaling rate of 50 Mbaud (50 MHz), a regenerator delay of $0.03 \mu\text{s}$ (1.5 time slots), and a quantized feedback cutoff frequency ($1/2\pi\tau_1$) of 1 MHz. The feedback filter resistors were (arbitrarily) chosen as $3.0 \text{ k}\Omega$, and the impedance level of the doublet section was chosen to give a ratio of decision circuit output voltage to the input voltage to the doublet section of 10. An asymptotic Bode plot of the feedback and forward path filters is shown in Fig. 5.23*b*.

Although the frequency domain design is direct, it does not tell us the error in the time response. Of course, we can find this error easily enough by use of the inverse Laplace transform, but since the time domain performance is the essential point here, we rederive the preceding result directly in the time domain.

For the quadratic quantized feedback filter, the time response of the forward path is

$$u(t) * h(t) = k_1 e^{-t/\tau_1} + k_2 e^{-t/\tau_2} \quad (5.7-20)$$

When $t = \tau_d$, we know that this step response must be exactly unity, giving us one equation relating the two unknowns, the residues. But we also know that the *slope* of the time response is zero. We take the derivative of the equation

$$\frac{d}{dt} [u(t) * h(t)] = -\frac{k_1}{\tau_1} e^{-t/\tau_1} - \frac{k_2}{\tau_2} e^{-t/\tau_2} \quad (5.7-21)$$

giving us a second equation relating the residues. Setting $t = \tau_d$ in these two equations, we have

$$k_1 e^{-\tau_d/\tau_1} + k_2 e^{-\tau_d/\tau_2} = 1 \quad (5.7-22)$$

$$\frac{k_1}{\tau_1} e^{-\tau_d/\tau_1} + \frac{k_2}{\tau_2} e^{-\tau_d/\tau_2} = 0 \quad (5.7-23)$$

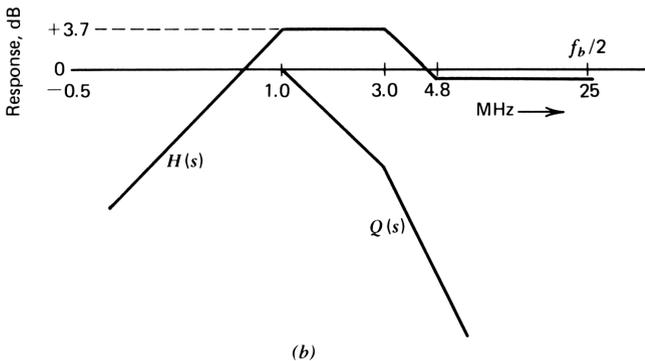
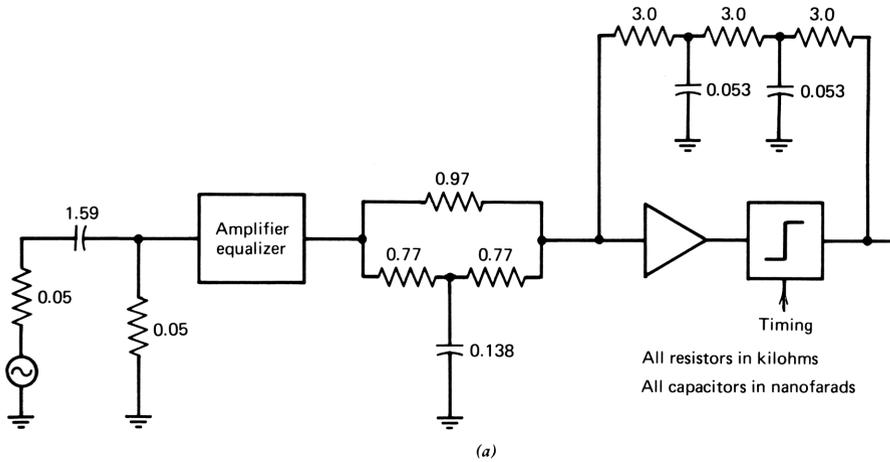


Figure 5.23. (a) Complete quadratic quantized feedback arrangement for a 50 Mbaud digital regenerator; (b) Asymptotic Bode plots for the feedback and forward path filters.

from which we obtain

$$k_1 = \frac{\tau_1}{\tau_1 - \tau_2} e^{\tau_d/\tau_1} \tag{5.7-24}$$

$$k_2 = - \frac{\tau_2}{\tau_1 - \tau_2} e^{\tau_d/\tau_2} \tag{5.7-25}$$

With these residues we can obtain the filter parameters directly, but we also obtain the time response from (5.7-20) and can thus find the time domain performance.

For the values of system parameters used to obtain the design of Fig. 5.23, the time response is shown in Fig. 5.24, expanded to show the error. As expected, the error is zero at $t = \tau_d$ and is maximum at $t = 0$. At the first sampling instant, the error is 3%.

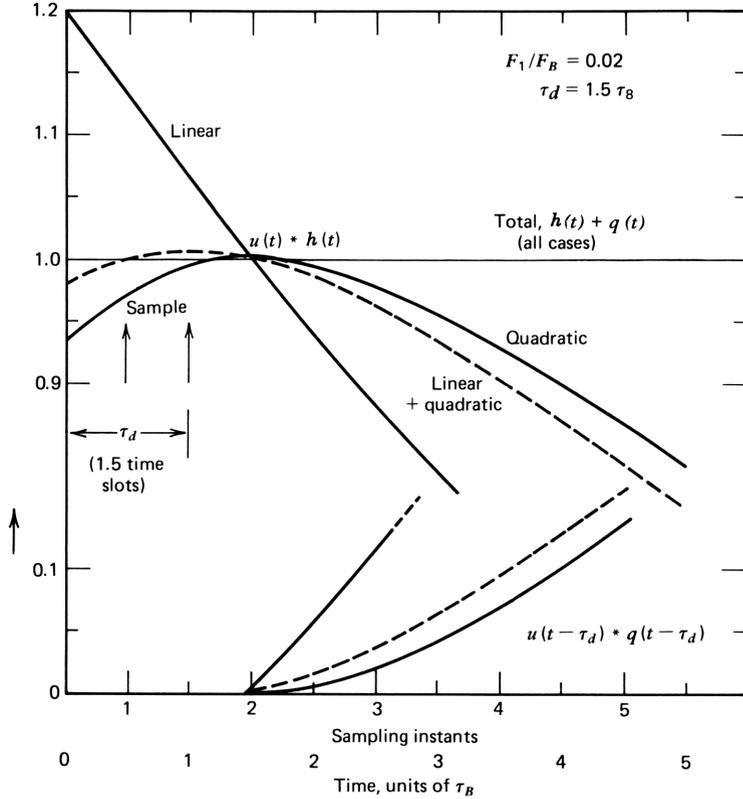


Figure 5.24. Time responses for linear, quadratic, and combination quantized feedback configurations.

Combination of Quadratic and Linear Filter Cutoff

This approach—of setting the derivative of the time response to zero at the delay time—is far better than the simple linear quantized feedback. We can do somewhat better by removing the restriction on the derivative and forcing the response to be correct at the first sampling instant. For this case, the error at *all* sampling instants is zero for decision circuit delays up to 1.5 time slots. One consequence of this is that the quantized feedback filter must provide a signal with finite slope at $t = \tau_d$, which means that it will also include a numerator. Since this is a minor correction, and will cost only one resistor, it is worthwhile. The procedure for the design of this *linear plus quadratic* quantized feedback configuration is given in Appendix B, with calculator program “PCM” that calculates all circuit values and prints out the time domain performance.

A comparison of the time domain performance for the linear, quadratic, and linear-plus-quadratic quantized feedback configurations is given in Fig. 5.24. In all three cases the ratio of the quantized feedback cutoff frequency to the

signaling rate is 0.02, and the delay is 1.5 time slots. Only in the linear plus quadratic configuration discussed in Appendix B is the correction perfect at all sampling instants, but the error in the pure quadratic case is small. On the other hand, the extra cost of the quadratic plus linear configuration is only one resistor, and Appendix B gives a program for finding it and the remaining components; thus it must be regarded as the appropriate choice for systems that have decision circuits with up to 1.5 time slots of delay.

For greater amounts of delay, higher-degree systems are appropriate. Such large amounts of delay are found only in unusual cases, one of which is in the T4M digital transmission system, in commercial service since 1975. In this system, operating at 274 Mbaud/s, the decision circuit is realized with discrete components, with considerable analog delay in addition to a full time slot of sampling delay. Total delay is nearly three time slots, and a special filter maintains the response of the forward path for the delay time until the quantized feedback signal becomes available. Integrated circuit systems have considerably less delay; lower-speed systems will usually have delays approaching the sampling delay of the decision circuit.

The pulse amplitude of the decision circuit output and that of the incoming pulse must be equal if zero error is to be attained. Automatic amplitude adjustment of the incoming pulse to a reference voltage [e.g., automatic gain control (AGC) or an automatic equalization adjustment] is virtually always used in regenerators for pulse transmission, so that close tracking of the AGC reference and the power supply that controls the output pulse amplitude of the decision circuit is advisable.

The sensitivity of the quantized feedback system to its components is not difficult to ascertain. If we consider the sensitivity in the frequency domain, we see from Fig. 5.23*b* that the overall channel is made up of two additive components whose sum is roughly unity. According to the sum rule, any contributor that is larger than unity will have a sensitivity larger than unity. For the case of $H(s)$, the asymptotic response in midband is +3.7 dB, but the actual response never reaches the asymptote and has a maximum of about 1 dB, giving a maximum sensitivity of 1.1. This is sufficiently close to unity to consider quantized feedback components as having roughly the same sensitivity as other equalizer components.

We have by now established a framework for considering feedback system design, illustrated at some length by reference to the case study amplifier. We have explained the design method in detail for a single amplifier to enable us to thoroughly understand what is involved. In the following chapters we show how this method can be extended to any circuit or feedback configuration. We are able to draw an equivalent ladder network and write the circuit loss equations by inspection as we have learned to do it here for the case study amplifier. To do this, we must introduce another level into our hierarchy of circuit equations: to avoid getting bogged down in detail, we need a method of chunking voltage and current into a single variable. This, in turn, requires a deeper understanding of two-port networks, which is the subject of Chapter 6.

PROBLEMS

- 1 Show that a stable quadratic that has complex roots

$$L(s) = a_0 + a_1s + a_2s^2$$

can be represented accurately by

$$L(s) \doteq a_0 e^{-0.7\tau_d s} (1 + \tau_1 s e^{\tau_d s})$$

where

$$\tau_d = \frac{\theta}{\rho \sin \theta}$$

$$\tau_1 = \frac{1}{\rho} \exp \frac{\theta}{\tan \theta}$$

and

$$\rho = \left(\frac{a_0}{a_2} \right)^{1/2}$$

$$\theta = \cos^{-1} \frac{a_1 \rho}{2a_0}$$

REFERENCE

- 1 D. R. Hofstadter, *Gödel, Escher, and Bach, An Eternal Golden Braid*, Basic Books, New York, 1979.

Part 2

Circuits and Devices

Chapter 6

Two-Port Analysis of Circuits and Devices

The representations of the loss of circuits in the previous chapters were scalar equations, usually giving the ratio of input generator voltage to output voltage as a single scalar variable. Since both input and output current must also be taken into account, the source and load immittances became a part of the equations for the loss. Since any network exists as a separate entity, apart from particular terminating immittances, characterization of such a network requires that we recognize both voltage *and* current explicitly at its input and output. This is the function of two-port analysis of networks; it is applicable at each of the three hierarchical levels studied in this book: the system level, the circuit level, and the device level.

The purpose of this chapter is to introduce two-port network descriptions of circuits and devices with particular focus on their transmission properties. The two-port network description most appropriate for this purpose is the *ABCD* or transmission matrix, as we show. This matrix corresponds to the reciprocal formulation of the feedback problem and allows us to extend the type of analysis already presented to more general systems.

The two-port network description of a circuit or device allows us to express its properties independent of its source and load. We can then build up more

elaborate structures from combinations of two-port networks. If such a structure has a single input and a single output port, it can be described by a single two-port matrix; the equation for this matrix can be written as a function of the matrices of the devices and circuits of which it is constructed.

6.1 ALTERNATIVE DESCRIPTIONS OF TWO-PORT NETWORKS

A two-port network, shown in Fig. 6.1a is characterized completely by constraints among the signal variables at its two ports. The constraint that distinguishes a two-port network from a network that has four access leads is that the current entering one lead of a given port must be equal to the current leaving the other lead of the same port. The signal variables may be taken in pairs: we might, for example, take the port voltages as one pair and the port currents as the other. If we wish to find the port currents in terms of the port voltages, for example, we may write

$$i_1 = y_{11}v_1 + y_{12}v_2 \quad (6.1-1)$$

$$i_2 = y_{21}v_1 + y_{22}v_2 \quad (6.1-2)$$

The y_{ij} are all admittances and are defined implicitly by these equations. Their definitions are

$$y_{11} = \left. \frac{\partial i_1}{\partial v_1} \right|_{v_2=0}, \quad \text{short-circuit input admittance} \quad (6.1-3)$$

$$y_{12} = \left. \frac{\partial i_1}{\partial v_2} \right|_{v_1=0}, \quad \text{reverse transadmittance} \quad (6.1-4)$$

$$y_{21} = \left. \frac{\partial i_2}{\partial v_1} \right|_{v_2=0}, \quad \text{forward transadmittance} \quad (6.1-5)$$

$$y_{22} = \left. \frac{\partial i_2}{\partial v_2} \right|_{v_1=0}, \quad \text{short-circuit output admittance} \quad (6.1-6)$$

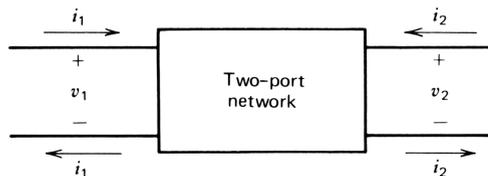


Figure 6.1. A two-port network signal variables defined.

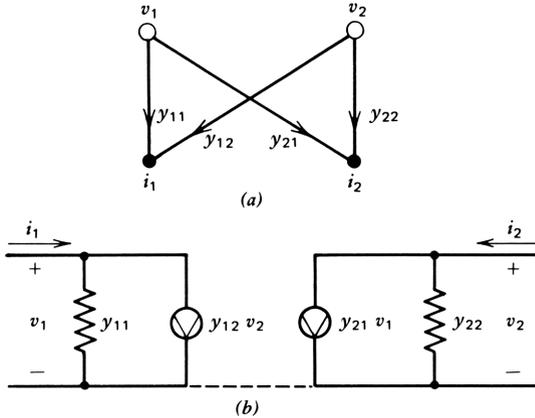


Figure 6.2. Graphical representations of the y -parameter equations: (a) signal flow graph; (b) equivalent circuit.

In matrix form eqs. (6.1-1) and (6.1-2) may be written

$$di_j = y_{jk} dv_k \quad (j, k = 1, 2) \tag{6.1-7}$$

where di_j is the signal current vector, dv_k is the signal voltage vector, and the y_{jk} are the admittance or y parameters. For the y -parameter description of a two-port network, di_j is the *dependent signal vector*, and dv_k is the *independent signal vector*.

Graphic interpretations of the preceding relationships are given in Fig. 6.2. In Fig. 6.2a eqs. (6.1-1) and (6.1-2) are represented by a signal flow graph in which the two nodes representing the independent variables are shown as hollow circles. An equivalent circuit is given in Fig. 6.2b, in which admittances represent y_{11} and y_{22} and dependent generators represent the transadmittances. The graphical representations are completely equivalent to the equations.

The character of a two-port description depends on our choice of the pair of signal variables that is to be the independent signal vector. (The other pair, of course, becomes the dependent signal vector.) There are six possible permutations in this choice, leading to six sets of two-port parameters. For each of these choices, we may write a set of equations analogous to eqs. (6.1-1) to (6.1-7).

ABCD or General Circuit Parameters

The two-port matrix of particular interest in this chapter is the transmission, or *ABCD* matrix.* For this matrix, the output voltage and current are the independent variables, and the input voltage and current are the dependent

*This matrix has also been called the *chain matrix*, and its elements have been termed the *chain, or general circuit parameters*.

variables, so that

$$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} v_2 \\ -i_2 \end{bmatrix} \quad (6.1-8)$$

The output current for which these parameters are defined is $-i_2$, to make the direction of current flow at the output agree with that of the input of a following cascaded network. This two-port description gives the input signal vector, the input *excitation*, as a function of the output signal vector, or the output *response*. Therefore, it corresponds to the reciprocal formulation described in Part 1.

The definitions of the parameters are implicit in eq. (6.1-8):

$$\begin{aligned} A &= \left. \frac{\partial v_1}{\partial v_2} \right|_{i_2=0}, && \text{reciprocal of} \\ &&& \text{open-circuit voltage gain} \\ B &= \left. \frac{\partial v_1}{\partial(-i_2)} \right|_{v_2=0}, && \text{negative reciprocal of} \\ &&& \text{forward transadmittance} \\ C &= \left. \frac{\partial i_1}{\partial v_2} \right|_{i_2=0}, && \text{reciprocal of} \\ &&& \text{forward transimpedance} \\ D &= \left. \frac{\partial i_1}{\partial(-i_2)} \right|_{v_2=0}, && \text{negative reciprocal of} \\ &&& \text{short-circuit current gain} \end{aligned} \quad (6.1-9)$$

Note that the $ABCD$ parameters are all reciprocals or negative reciprocals of the four forward transfer or gain parameters. For example, B is $-1/y_{21}$ of the y parameters described previously. The other forward transfer parameters are described in the following paragraphs.

A signal flow graph and equivalent circuit for a two-port described by its $ABCD$ parameters is shown in Fig. 6.3. The tail of each branch of the signal flow graph originates at a node representing an independent circuit variable, and the nose points toward a dependent variable node. The branch value multiplies the value of the independent variable and adds the result to the dependent variable.

The equivalent circuit in Fig. 6.3*b* shows the four parameters as dependent voltage and current generators connected at the input of an *ideal amplifier*, defined as an amplifier whose input voltage and current are zero for any finite output voltage and current.¹ For an ideal amplifier, each element of the matrix in eq. (6.1-8) is zero, and the matrix is the *null matrix*.

The equivalent circuit in Fig. 6.3*b* expresses the essence of anticausal analysis as applied to two ports. Since the input voltage and current of the ideal amplifier are identically zero, the actual input voltage and current of the network are attributed entirely to the dependent voltage and current generators at the input.

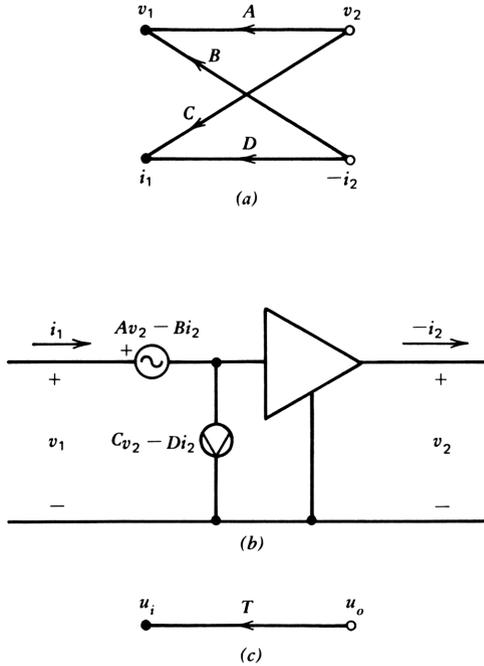


Figure 6.3. Signal flow graph and equivalent circuit for the $ABCD$ parameters.

Since these generators are dependent on the *output* current and voltage, they provide *feedback* around the ideal amplifier by the definition of feedback given in Section 4.5. Furthermore, the feedback is of four types corresponding to the four parameters.

A Feedback, also termed *series input parallel output feedback*, augments the input voltage in proportion to the output voltage. B Feedback, or series input series output feedback, augments the input voltage in proportion to output current. Similarly, C and D feedback augment the input *current* in proportion to the output voltage and current, respectively.

The matrix equation (6.1-8) may be written in compact form:

$$u_1 = Tu_2 \tag{6.1-10}$$

where u_1 and u_2 are the input and output signal vectors, respectively

$$u_1 = \begin{bmatrix} v_1 \\ i_1 \end{bmatrix}, \quad u_2 = \begin{bmatrix} v_2 \\ i_2 \end{bmatrix}$$

and

$$T = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

In Fig. 6.3c this equation is represented by a simpler graph, a *transmission matrix signal flow graph* (TMSFG), which connects u_2 to u_1 through T .

In this graph the matrix T premultiplies the output signal vector and adds the result to the input signal vector. The transmission matrix signal flow graph is similar to an ordinary (scalar) signal flow graph, except that each graph node now represents a signal vector—a voltage *and* a current at a particular point in a circuit. Such a point is termed a *circuit vector node*, defined as a node of the circuit that has only two connections to it. This restriction allows the current at the node to be defined uniquely. The voltage is also defined uniquely, so that the vector represented by the TMSFG node is unequivocally defined. We can thus refer to the *vector* in our equations rather than to the currents and voltages themselves.

The simplification of the equations thereby effected is important to one of our main goals here and in Chapter 7: this enables us to write the circuit equations of most feedback circuits by inspection. By chunking voltage and current together, the equations become simpler; the computer can be called on to do the detailed work of sorting out currents and voltages.

So far, we have defined two of the six possible sets of two-port parameters obtainable by permuting the signal voltage and current variables. All six are summarized in Fig. 6.4, which shows the signal flow graph and equivalent circuit for each.² As in Fig. 6.2a, the independent variable nodes of the signal flow graph are shown as open circles. In the first four parameter sets, one independent variable is taken from the input and the other from the output. Therefore, one branch of each must proceed from an input node to an output node as, for example, the branch y_{21} of the y parameters. Each such branch carries a 21 subscript. These are the four forward transfer parameters whose reciprocals comprise the $ABCD$ parameters; thus

$$\begin{aligned} A &= \frac{1}{g_{21}} & B &= -\frac{1}{y_{21}} \\ C &= \frac{1}{z_{21}} & D &= -\frac{1}{h_{21}} \end{aligned} \quad (6.1-11)$$

In last entry in Fig. 6.4 the inverted $ABCD$ parameters are obtained by taking the inverse of the $ABCD$ matrix; Δ_A is the determinant of the $ABCD$ matrix.

Parameter Conversions

The six descriptions of Fig. 6.4 are six views of the same network, so that it must be possible to convert from one set of parameters to another by performing the necessary algebra. Since all six are of use in the text that follows, Appendix C gives a general procedure and a calculator program “CNV” for effecting the conversion.

By use of this procedure (perhaps in two steps), any set of the six parameter sets may be expressed in terms of any other set. In Table 6.1 the conversion of

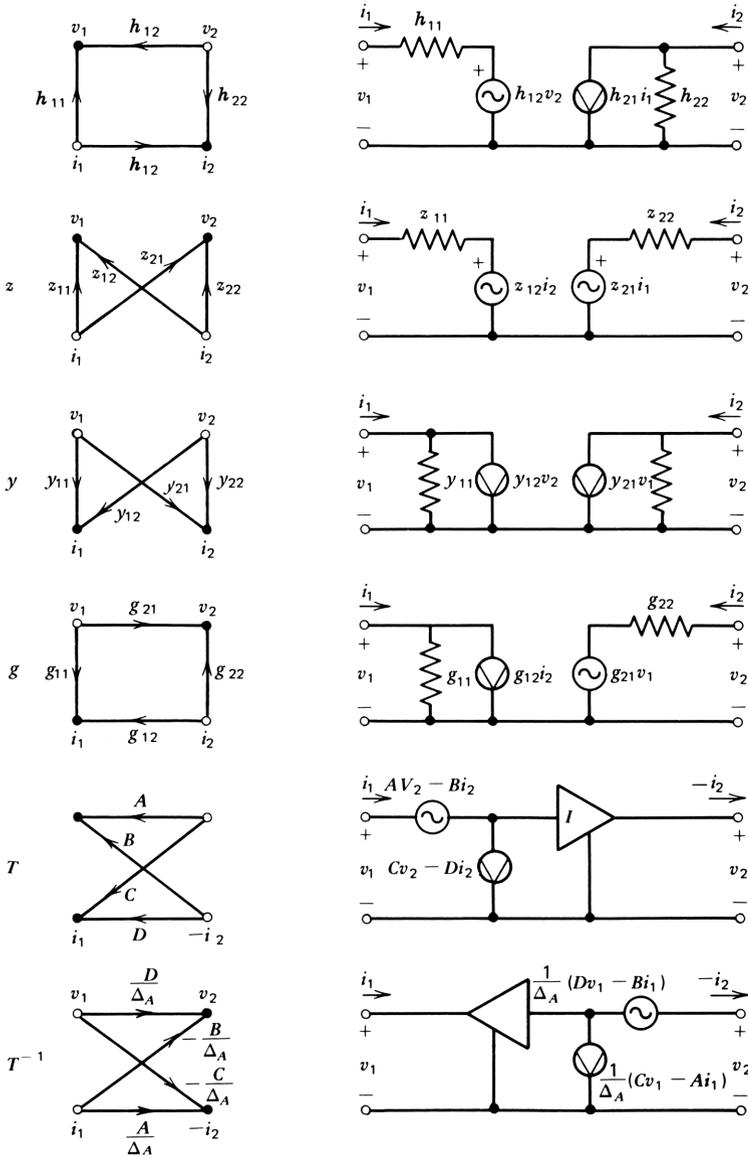


Figure 6.4. Signal flow graph and equivalent circuits for six possible sets of two-port parameters.

Table 6.1 Translations Among Sets of Two-Port Parameters

	From $ABCD$		To $ABCD$	
h	$h_{11} = \frac{B}{D}$ $h_{21} = \frac{-1}{D}$	$h_{12} = \frac{\Delta_A}{D}$ $h_{22} = \frac{C}{D}$	$A = \frac{-\Delta_h}{h_{21}}$ $C = \frac{-h_{22}}{h_{21}}$	$B = \frac{-h_{11}}{h_{21}}$ $D = \frac{-1}{h_{21}}$
z	$z_{11} = \frac{A}{C}$ $z_{21} = \frac{1}{C}$	$z_{12} = \frac{\Delta_A}{C}$ $z_{22} = \frac{D}{C}$	$A = \frac{z_{11}}{z_{21}}$ $C = \frac{1}{z_{21}}$	$B = \frac{\Delta_z}{z_{21}}$ $D = \frac{z_{22}}{z_{21}}$
y	$y_{11} = \frac{D}{B}$ $y_{21} = \frac{-1}{B}$	$y_{12} = \frac{-\Delta_A}{B}$ $y_{22} = \frac{A}{B}$	$A = \frac{-y_{22}}{y_{21}}$ $C = \frac{-\Delta_y}{y_{21}}$	$B = \frac{-1}{y_{21}}$ $D = \frac{-y_{11}}{y_{21}}$
g	$g_{11} = \frac{C}{A}$ $g_{21} = \frac{1}{A}$	$g_{12} = \frac{-\Delta_A}{A}$ $g_{22} = \frac{B}{A}$	$A = \frac{1}{g_{21}}$ $C = \frac{g_{11}}{g_{21}}$	$B = \frac{g_{22}}{g_{21}}$ $D = \frac{\Delta_g}{g_{21}}$
S	$S_{11} = \frac{A + \hat{B} - \hat{C} - D}{A + \hat{B} + \hat{C} + D}$ $S_{21} = \frac{2}{A + \hat{B} + \hat{C} + D}$	$S_{12} = \frac{2\Delta_A}{A + \hat{B} + \hat{C} + D}$ $S_{22} = \frac{-A + \hat{B} - \hat{C} + D}{A + \hat{B} + \hat{C} + D}$	$A = \frac{-\Delta_S + S_{11} - S_{22} + 1}{2S_{21}}$ $C = \frac{\Delta_S - S_{11} - S_{22} + 1}{2S_{21}}$	$B = \frac{\Delta_S + S_{11} + S_{22} + 1}{2S_{21}}$ $D = \frac{-\Delta_S - S_{11} + S_{22} + 1}{2S_{21}}$
				$\Delta_A = -\frac{h_{12}}{h_{21}}$ $\Delta_A = \frac{z_{12}}{z_{21}}$ $\Delta_A = \frac{y_{12}}{y_{21}}$ $\Delta_A = -\frac{g_{12}}{g_{21}}$ $\Delta_A = \frac{S_{12}}{S_{21}}$

the z , y , h , and g parameters from and to the $ABCD$ parameters is given. The table also includes the conversion back and forth between the $ABCD$ parameters and a set of parameters to be discussed in the following paragraphs, the *scattering parameters*.

S Parameters

Two-port measurements of components and devices are often expressed by their scattering or S parameters.

The S parameters are not obtained by simply permuting the signal variables, but rather by forming linear combinations of voltage and current at each port. These parameters were originally introduced to characterize waveguide and microwave components, where a *wave* rather than a voltage and current description of the signal is more appropriate.^{3,4} Although we describe the S parameters with respect to the port voltages and currents of a two-port network, we briefly introduce them in their original wave formulation.

In Fig. 6.5 we show a network that has *incident waves* a_1 and a_2 *impinging* on its input and output ports and *reflected waves* b_1 and b_2 *emanating* from these same two ports. The S parameters are defined under the wave formulation in terms of these four wave signals:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{6.1-12}$$

These equations give the values of the waves emanating from the two ports as functions of the waves incident on the ports.

The definitions of the S parameters, implicit in eq. (6.1-12), are

$$\begin{aligned} S_{12} &= \left. \frac{b_1}{a_1} \right|_{a_2=0}, && \text{input reflection coefficient} \\ S_{12} &= \left. \frac{b_1}{a_2} \right|_{a_1=0}, && \text{reverse transfer coefficient} \\ S_{21} &= \left. \frac{b_2}{a_1} \right|_{a_2=0}, && \text{forward transfer coefficient} \\ S_{22} &= \left. \frac{b_2}{a_2} \right|_{a_1=0}, && \text{output reflection coefficient} \end{aligned} \tag{6.1-13}$$

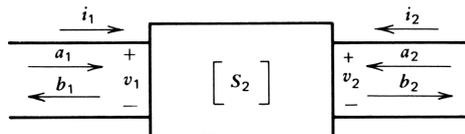


Figure 6.5. Two-port network with signal variables defined as incident and reflected waves at input and output.

where S_{11} and S_{22} simply give the ratio of the reflected wave to the incident wave at the input and output ports, respectively, with the incident wave at the opposite port equal to zero. The transfer coefficients, on the other hand, give the ratio of the wave *emanating* from one port due solely to the incident wave at the other port. For example, S_{21} gives the ratio of b_2 , the wave traveling to the right at the output in Fig. 6.5, to a_1 , the wave impinging on the input from the left. This coefficient is also called the *insertion gain*, a quantity we have occasion to use later.

To relate the S parameters to the six parameter sets discussed previously, the incident and reflected waves must be translated into port voltages and currents. We define the S parameters for the useful special case in which the parameters are related to a resistive characteristic impedance $R_0 = 1/G_0$. (For the extension to the more general cases including differing, complex characteristic impedances at input and output, and for n ports, see Weinberg.⁴) The incident and reflected waves can be written in terms of the port voltages and currents as

$$a_k = \frac{1}{\sqrt{R_0}} (v_k + i_k R_0) \quad (6.1-14)$$

$$b_k = \frac{1}{\sqrt{R_0}} (v_k - i_k R_0) \quad (6.1-15)$$

The port signal variables for the S parameters are not simply permutations of the voltages and currents of the parameter sets studied earlier but are *linear combinations* of them. Thus eq. (6.1-12) can be rewritten

$$\begin{bmatrix} v_1 - i_1 R_0 \\ v_2 - i_2 R_0 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} v_1 + i_1 R_0 \\ v_2 + i_2 R_0 \end{bmatrix} \quad (6.1-16)$$

in which the factor $1/\sqrt{R_0}$, appearing on both sides, has been canceled.

The S parameters are appropriately used with matched or nearly matched circuits, where the reflections are small. The impedance may be 50 or 75 Ω , for example. The $ABCD$ parameters may also be normalized to such an impedance (rather than being normalized to the impedance unit in the system of units being employed, such as 1 k Ω or 1 Ω). Thus $\hat{B} = B/R_0$, and $\hat{C} = CR_0$. Table 6.1 gives the conversion between $ABCD$ and S parameters for the case in which both parameter sets are normalized to the same impedance.

F Parameters

Still another set of two-port parameters, the *transfer*, or F parameters, is also based on the wave formulation.^{4,5} The set is analogous to the $ABCD$ parameters in that the two input signal variables constitute the pair of dependent variables and the two output signal variables constitute the independent

variables. The signal pairs are the incident and reflected waves at a given port:

$$\begin{bmatrix} v_1 - i_1 R_0 \\ v_1 + i_1 R_0 \end{bmatrix} = \begin{bmatrix} F_{ri} & F_{rr} \\ F_{ii} & F_{ir} \end{bmatrix} \begin{bmatrix} v_2 + i_2 R_0 \\ v_2 - i_2 R_0 \end{bmatrix} \quad (6.1-17)$$

These parameters are useful in the cascading of waveguide and microwave components. Note that the output signal vector is written such that $v_2 - i_2 R_0$, the wave emanating from the output, is the incident wave for a two-port connected in cascade at the output. This is analogous to the use of $-i_2$ as the output signal current variable for the $ABCD$ matrix.

The theory developed here is based on the $ABCD$ parameters. It might also have been developed by using F parameters, since both parameter sets are at their essence feedback or anticausal in nature. The $ABCD$ parameters are more appropriate for lumped parameter circuits and the F parameters for the wave or distributed formulation.

As technology advances, the frequencies dealt with increase; this tends to favor a distributed, or F parameter, formulation. A countervailing tendency is for the size of circuits to be scaled down, favoring the simpler lumped parameter approach. In the recent past, size reduction has been realized to an extent that is greater than bandwidth expansion (with some important exceptions). Hence the lumped parameter $ABCD$ description appears more appropriate to present-day technology.

Classification of Two-Port Networks

To facilitate the classification of two-port parameter sets, we use Fig. 6.6 to introduce the concept of *signal epochs*. In the figure a generator is connected through a transmission line to a two-port network, the output of which is then transmitted to a load network through another transmission line. Suppose that the generator emits a short pulse. As we trace the signal transmission from generator to load, we can identify four “time frames” or *epochs* corresponding to the arrival of the pulse at (1) the generator output, (2) the two-port input, (3) the two-port output, and (4) the load input. These epochs will differ from each other by the delay of the intervening networks.

The networks may cause the pulse to be dispersed in time—spread out—making the definition of delay somewhat fuzzy, but in a general way, we can say that the voltage and current at a given port occur simultaneously, whereas the voltage and currents at different ports are separated into individual epochs. The distinction is one that is useful in cases (as in Chapter 5) where a transfer function is expressed as a rational function with a delay. The delay defines the difference in epochs at input and output.

For the classification of two-port parameters, we can conveniently divide them into *mixed-epoch* and *separate-epoch* parameter sets. A mixed epoch set is one whose associated signal vectors include components from more than one epoch, as, for example, the y parameters, in which the independent signal vector consists of v_1 from the input epoch and v_2 from the output epoch. Each

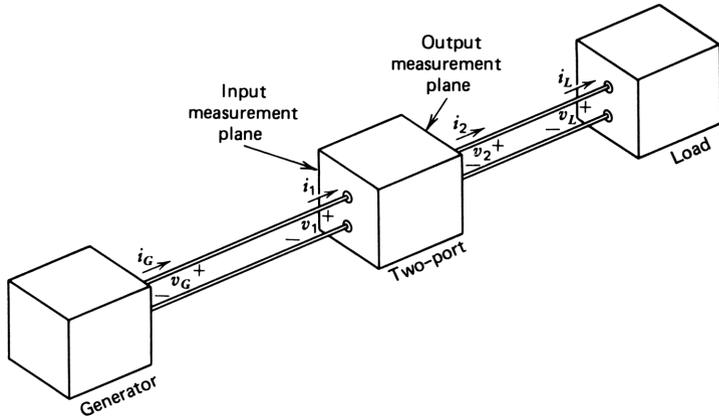


Figure 6.6. Introduction of the concept of signal epochs.

Table 6.2 Classification of Two-Port Networks

	Separate-Epoch Sets	Mixed-Epoch Sets
Voltage and current	$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} v_2 \\ -i_2 \end{bmatrix}$ $\begin{bmatrix} v_2 \\ i_2 \end{bmatrix} = \frac{1}{\Delta_A} \begin{bmatrix} D & B \\ C & A \end{bmatrix} \begin{bmatrix} v_1 \\ -i_1 \end{bmatrix}$	$\begin{bmatrix} v_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} i_1 \\ v_2 \end{bmatrix} \quad \left \quad \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix}$ $\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad \left \quad \begin{bmatrix} i_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ i_2 \end{bmatrix}$
Wave	$\begin{bmatrix} v_1 - i_1 Z_0 \\ v_1 + i_1 Z_0 \end{bmatrix} = \begin{bmatrix} F_{ri} & F_{rr} \\ F_{ri} & F_{rr} \end{bmatrix} \begin{bmatrix} v_2 + i_2 Z_0 \\ v_2 - i_2 Z_0 \end{bmatrix}$	$\begin{bmatrix} v_1 - i_1 Z_0 \\ v_2 - i_2 Z_0 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} v_1 + i_1 Z_0 \\ v_2 + i_2 Z_0 \end{bmatrix}$

signal vector of a separate-epoch parameter set, on the other hand, consists of components exclusively from a single epoch, as for the *ABCD* parameters, where the independent signal vector consists of v_2 and i_2 .

Table 6.2 summarizes the two-port descriptions of this section. The parameter sets have been divided into separate- and mixed-epoch sets; they are further divided into parameters that employ voltages and currents separately as signal variables and those employing waves as signal variables.*

*The subscripts employed for the *F* parameters, *i* and *r*, refer to “incident” and “reflected” waves, respectively. The order of the subscripts agrees with the mixed-epoch parameter convention, where y_{21} , for example, refers to the ratio v_2/i_1 . A similar convention could be employed for the *ABCD* parameters, called the *K* parameters by Belevitch.⁵ Thus K_{ve} would refer to the input voltage: output voltage ratio, equal to *A*. Similarly, $B = K_{vi}$, $C = K_{ic}$, and $D = K_{ii}$.

6.2 APPROPRIATE APPLICATIONS OF THE VARIOUS TWO-PORT DESCRIPTIONS

Why do we introduce so many sets of two-port parameters? In principle, any one of them can be used for the analysis of any two-port; hence the introduction of the eight sets of parameters in Table 6.2 may seem redundant, or superfluous. Having access to these various descriptions may be likened to having an adequate vocabulary with which to utter a good sentence. Each parameter set introduced in Section 6.1 is uniquely appropriate under certain circumstances and so conveys a particularly direct understanding of the particular circumstance for which it is appropriate. In this section we begin the task of finding where each parameter set previously introduced fits into circuit analysis and design.

The term “appropriateness” is related in some sense to whether a certain circuit description “feels” right. A less fuzzy measure of appropriateness is proposed here: one circuit description is considered more *appropriate* than another if its use leads to less loop gain in the circuit description. Since loop gains form an “endless chain of dependencies,” analysis is clearer without them.

Suppose, for example, that we apply known voltages to each port of a two-port network and that we wish to find the port currents. The appropriate description of this two-port is the y -parameter description since we need to merely write the matrix equation $i_j = y_{jk}v_k$, obtaining the currents directly without solution of simultaneous equations and without the introduction of any feedback loops, or denominators. However, if we wish to apply a known voltage to the output and a known current to the input, the h parameters would be the appropriate choice (see Fig. 6.4). Similar considerations attend the use of z and g parameters. We use all four descriptions later.

Causal Analysis of an Amplifier

Suppose now that we have an amplifier operated between a Thevenin source and a load conductance as shown in Fig. 6.7. The amplifier is a two-port network described by its y parameters. For this circuit, we can write the following five equations:

$$i_1 = y_{11}v_1 + y_{12}v_2 \quad (6.2-1)$$

$$i_2 = y_{21}v_1 + y_{22}v_2 \quad (6.2-2)$$

$$v_1 = e_G - i_1 R_G \quad (6.2-3)$$

$$v_2 = v_o \quad (6.2-4)$$

and

$$v_o = -R_L i_2 \quad (6.2-5)$$

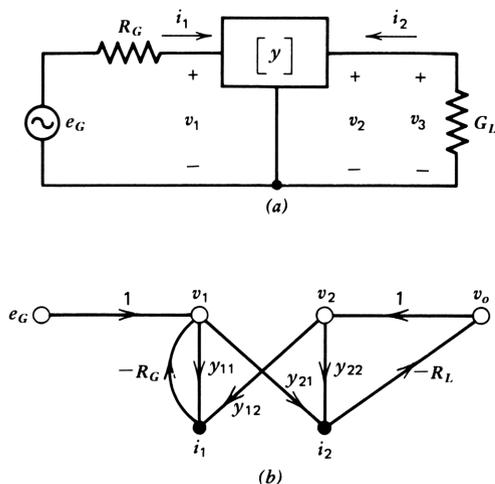


Figure 6.7. Analysis of a two-port amplifier described by its y parameters. $v_o/e_G = -y_{21}R_L/[1 + R_G y_{11} + R_L y_{22} + R_G R_L(y_{11}y_{22} - y_{12}y_{21})]$.

These five equations are diagrammed in the signal flow graph in Fig. 6.7*b*, which is drawn to evaluate the *gain* of the circuit, proceeding from e_G to v_o . Note that three loops appear in the signal flow graph. At the input the loop gain is $-y_{11}R_G$ and at the output, $-y_{22}R_L$. The third loop includes both input and output and has the gain $y_{12}y_{21}R_LR_G$. They introduce corresponding loop gain terms in the denominator of the gain expression, as shown. By either signal flow graph reduction or simultaneous solution of eqs. (6.2-1) to (6.2-5), the gain is evaluated as shown by the gain expression in Fig. 6.7. For sufficiently low values of R_L and R_G , the loop gains can be small, and the circuit description can be an appropriate one. More often, however, these feedback loops constitute a block to a direct, intuitive understanding of circuit operation.

Anticausal Analysis

The preceding analysis corresponds to a causal approach—finding the output for a given input. Consider what happens when the analysis proceeds in the opposite direction—what input is required to give a preassigned output? Of the five equations given above, we first change two: eq. (6.2-3) must now define e_G :

$$e_G = v_1 + i_1 R_G \quad (6.2-6)$$

and since v_o is now a source node, we must now invert eq. (6.2-5) to give

$$i_2 = -v_o G_L \quad (6.2-7)$$

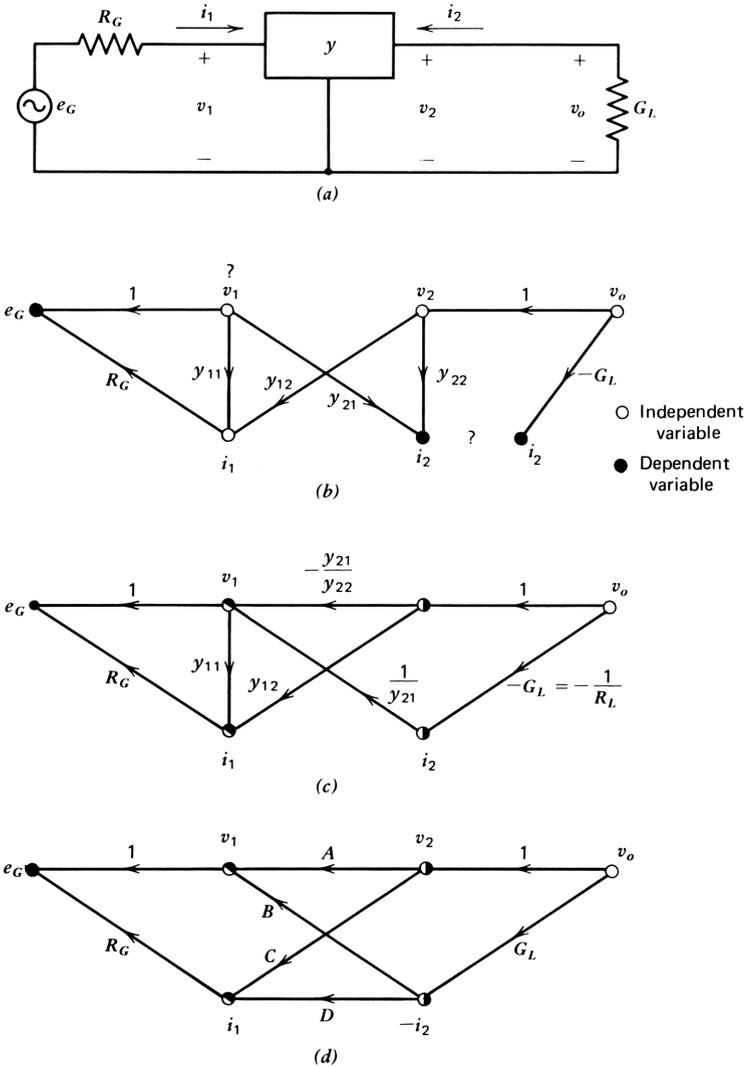


Figure 6.8. Anticausal analysis of an amplifier described by its y parameters.

where $G_L = 1/R_L$. These two equation modifications serve to interchange the roles of the source and sink nodes of the signal flow graph in Fig. 6.7; the independent variable is now v_o , at the output. In Fig. 6.8 the revised five equations are diagrammed in a signal flow graph that exhibits two peculiarities: (1) the node voltage v_1 is no longer defined by any signal variables since no branches lead to it; and (2) i_2 is defined twice, as $v_o G_L$ and as $y_{21}v_1 + y_{22}v_2$. Clearly, some functional dependencies must be reversed to make a usable signal flow graph.

We note parenthetically that the graph in Fig. 6.8*b* has been termed a *disconnected graph* because it has two dependent nodes for the same variable.⁶ It is also peculiar in that it has two source nodes; one, v_o , is intended, and the other, v_1 , is not.

The resolution of these improprieties in the graph in Fig. 6.8*b* is found by reversing the signal dependencies of certain branches of the graph. To maintain the anticausal direction of analysis, we cannot reverse the signal dependencies of the terminating networks at either the input or the output, so *we are forced to alter the signal dependencies in the two-port network itself*. Therefore, we solve for v_1 in (6.2-2):

$$v_1 = \frac{1}{y_{21}} i_2 - \frac{y_{22}}{y_{21}} v_2 \quad (6.2-8)$$

Equations (6.2-1), (6.2-4), (6.2-6), (6.2-7), and (6.2-8) constitute a consistent set of equations for the anticausal direction of analysis. These five equations are diagrammed in Fig. 6.8*c*, in which the output voltage is the source node of the graph and e_G is the sink node.

Take a careful look at Fig. 6.8*c*. It is a cascade graph, one without feedback loops. By our criterion of appropriateness introduced at the beginning of this section, this description is appropriate. By reversing the signal dependencies in eq. (6.2-8), we made v_1 a dependent variable for the two-port network. But i_1 was already a dependent variable in the y -parameter description, so that the reversal of the $v_1 - i_2$ signal dependency has altered the two-port description from one using y parameters to one using $ABCD$ parameters, as shown in Fig. 6.8*d*. The equivalence of Figs. 6.8*c* and 6.8*d* can be seen by separate evaluation of the $ABCD$ parameters in Fig. 6.8*c* and comparing them with those given in Table 6.1:

$$\begin{aligned} A &= -\frac{y_{22}}{y_{21}} & B &= -\frac{1}{y_{21}} \\ C &= y_{12} - \frac{y_{11}y_{22}}{y_{21}} & D &= -\frac{y_{11}}{y_{21}} \end{aligned} \quad (6.2-9)$$

Note that the minus signs on B and D originate from the $ABCD$ parameter sign convention on the output current.

By use of the anticausal direction of analysis, we were forced to an $ABCD$ parameter description of the two-port network. Although we originated the discussion with y parameters in this example, the result would be the same if we were to use h , z , or g parameters. The same doubling of one of the output variables and the same lack of definition of one of the input variables would occur. By solving for the undefined input variable, we would find that we have again expressed the two-port in terms of its $ABCD$ parameters.

The mixed-epoch parameters in Table 6.2 each have a set of input and output *terminations* that remove feedback or loop gain from a circuit in which the two-port is connected between a Thevenin source and a load. These are the

Table 6.3 Terminating Impedances for Which Loop Gains are Zero for Various Two-Port Parameter Descriptions

	Parameter Set	Z_G	Z_L	Direction of Analysis
Mixed-epoch sets	h	Infinite	0	Causal
	z	Infinite	Infinite	Causal
	y	0	0	Causal
	g	0	Infinite	Causal
	S	R_0	R_0	Causal
Separate-epoch sets	$ABCD$	All	All	Anticausal
	F	All	All	Anticausal
	$DBCA$	None ^a	None ^a	Causal

^aYields disconnected signal flow graph (see text).

terminations for which these parameter sets are uniquely appropriate; they are shown in Table 6.3. Under any other terminations, the circuit description contains loop gains that tend to obscure circuit operation, making the description a less appropriate one.

Table 6.3 also includes the three separate-epoch parameter sets discussed previously and shows that no feedback loop gains arise under any terminating conditions for the $ABCD$ and transfer parameter sets, making them uniquely appropriate for describing the circuit. Both sets of parameters operate in the anticausal direction of analysis. The one parameter set that operates in the causal direction, the $DBCA$, or inverted $ABCD$ parameters, does not yield a connected signal flow graph. When path inversion is used to correct this difficulty, the description is converted to one of the other descriptions listed in Table 6.3; thus the inverted $ABCD$ parameters are not appropriate for describing the circuit for any terminations. We conclude that the $ABCD$ parameters are uniquely appropriate to describe the simple circuit shown using a voltage-current (rather than wave) description of the signal variables. Similarly, the transfer parameters are uniquely appropriate where the wave description is used.

Figure 6.9 illustrates the calculation of the loss (reciprocal gain) of an amplifier circuit whose amplifying element is described at the outset by its $ABCD$ parameters. Note that the output current i_o is defined as flowing outward from the output port, so that $i_o = -i_2$ in Fig. 6.8d. The signal flow graph in Fig. 6.9b and the TMSFG in Fig. 6.9c contain no feedback loops, so that we may obtain the voltage loss ratio e_G/v_o as the sum of all path products in Fig. 6.9b or, alternatively, as the matrix product in Fig. 6.9c:

$$L = \begin{bmatrix} 1 & R_G \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} 1 \\ G_L \end{bmatrix} \tag{6.2-10}$$

$$= A + BG_L + R_G C + R_G DG_L \tag{6.2-11}$$

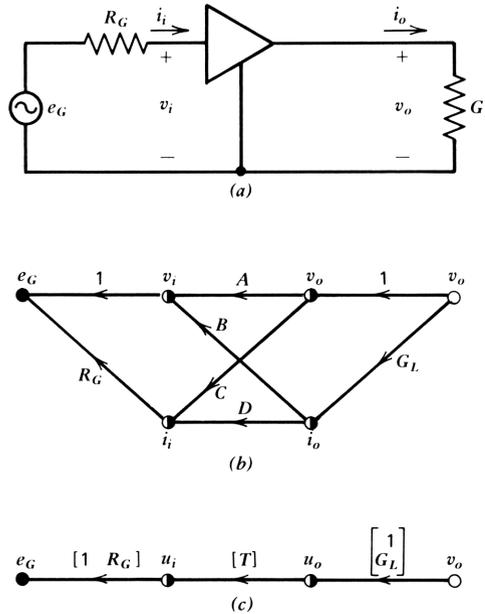


Figure 6.9. Direct analysis of the loss of an amplifier by use of the $ABCD$ matrix.

Several properties of the $ABCD$ matrix are illustrated in Fig. 6.10. In Fig. 6.10a the cascade property is shown, where $i_{2A} = i_{1B}$ and $v_{2A} = v_{1B}$. The $ABCD$ matrix of several cascaded networks is the matrix product of the $ABCD$ matrices of the constituent networks, taken in order. The $ABCD$ matrices of a series impedance and a shunt admittance are shown in Fig. 6.10b and that of their cascade connection, in Fig. 6.10c. Note from Table 6.1 that for any network

$$AD - BC = \Delta_A = -\frac{h_{12}}{h_{21}} = \frac{z_{12}}{z_{21}} = \frac{y_{12}}{y_{21}} = -\frac{g_{12}}{g_{21}} \quad (6.2-12)$$

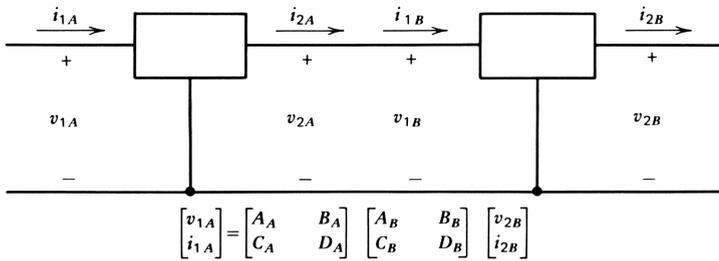
For a *passive* network, $AD - BC = 1$; for a *unilateral* network, $AD - BC = 0$. For the latter case, excitation at the output of the network produces no response at the input.

It is worthwhile at this point to restate the definition of an ideal amplifier, also termed a *nullor*⁷:

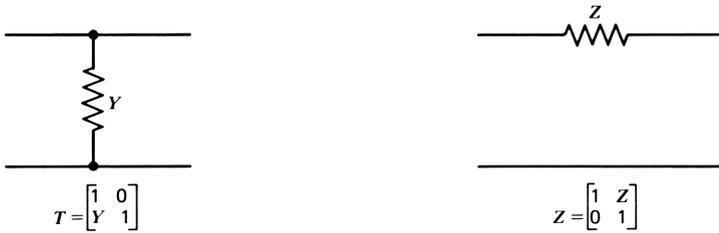
Definition: An ideal amplifier is an amplifier whose $ABCD$ matrix is the null matrix.

From eq. (6.2-11) we can write for the ideal amplifier

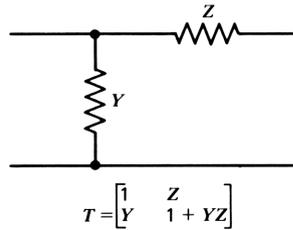
$$L = A + BG_L + CR_G + R_G DG_L = 0 \quad (6.2-13)$$



(a)



(b)



(c)

Figure 6.10. The ABCD matrices of simple circuits.

since $e_G/v_o=0$ by definition. Since the terminations are arbitrary and are assumed to be nonzero, A , B , C , and D must be zero individually.

An Intuitive Approach to the ABCD Parameters

An important benefit of the ABCD parameters is that they facilitate the understanding of circuits. To realize this benefit, one has to develop a direct “feel” for what each parameter represents. We have already begun this process in this chapter; here, we mention some relationships that will be familiar to many in their causal form but perhaps less so in anticausal form.

Parameter A is the voltage loss with open-circuited output. In Fig. 6.11a we have a circuit whose voltage gain is the ratio R_L/R_E . Parameter A is the reciprocal of this, G_L/R_E . In either case the guiding principle is that the current through R_L and R_E is the same, so that the voltage drops are in proportion to

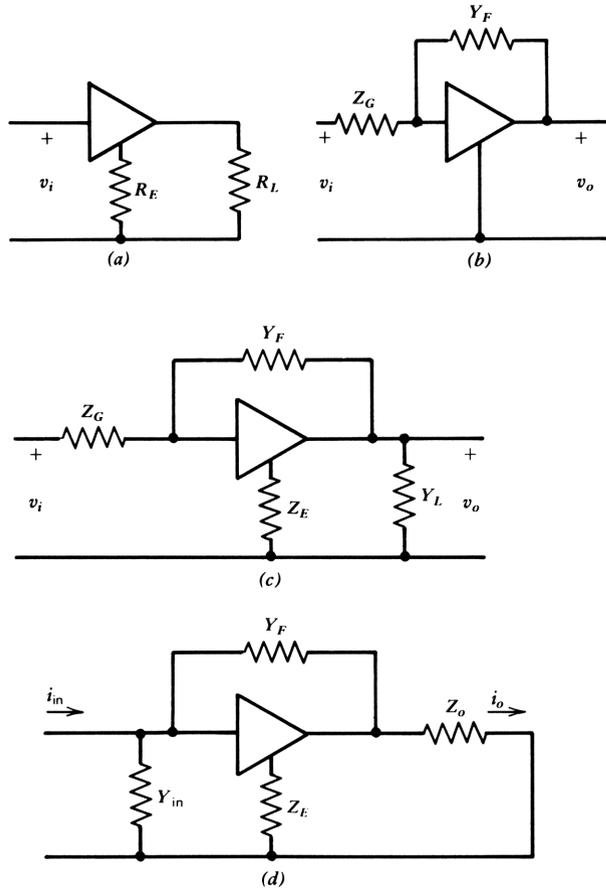


Figure 6.11. Circuits illustrating combining relationships in $ABCD$ matrices of a single gain stage.

the resistances. The A parameter is made up of terms of this sort—the product of an output shunt admittance and a series impedance in the common lead. There is another set of products that go to make up A , shown in Fig. 6.11*b*, and terms of this type are also familiar; they are the product of a series input impedance and an output-to-input feedback conductance. Where both effects exist in the same circuit as in Fig. 6.11*c*, A is approximately the sum of the two effects:

$$A = Y_L Z_E + Y_F Z_G \quad (6.2-14)$$

The relationship is approximate because direct feedthrough causes a denominator to appear, equal to $1 - Y_F Z_E$. The approximation is a good one where the feedback immittances are not too large and provides a convenient way of thinking about such circuits. These two effects can be added together,

at least approximately, because we deal here with loss rather than gain. *Losses* add directly, as do series impedances, whereas *gains* combine in a way similar to parallel impedances.

Similarly, D is the current loss with short-circuit output. It can also be regarded as being made up of two types of term, as illustrated in Fig. 6.11*d*, which includes a shunt input admittance and a series output impedance as well as Z_E and Y_F . For this circuit

$$D \simeq Y_{in} Z_E + Y_F Z_o \quad (6.2-15)$$

Parameters B and C are simpler and may be considered central to single-stage analysis. Roughly,

$$B = Z_E \quad (6.2-16)$$

and

$$C = Y_F$$

An exact analysis of the circuit in Fig. 6.11*d* is simple to write:

$$T_d = \frac{-1}{1 - Z_E Y_F} \begin{bmatrix} Z_E Y_F & Z_E \\ Y_F & Z_E Y_F \end{bmatrix} \quad (6.2-17)$$

Where $Z_E Y_F$ is much smaller than unity, we obtain

$$T_d \simeq - \begin{bmatrix} 0 & Z_E \\ Y_F & 0 \end{bmatrix} \quad (6.2-18)$$

in which the approximation merely involves dropping the effects of input and output loading and direct feedthrough.

If we replace the ideal amplifier by a transistor, several of the transistor effects combine directly with the circuit effects; the transistor includes emitter resistance ($1/g_m$), which can be combined with Z_E ; it also includes collector-to-base capacitance, which can be combined with Y_F . In the following chapter we develop a more detailed transistor equivalent circuit, but the simple combining relationships developed here will facilitate our understanding of transistor operation.

REFERENCES

- 1 F. D. Waldhauer, "Anticausal Analysis of Feedback Amplifiers," *BSTJ* **56** (10), 1337-1386 (October 1977). The central theory of Part 2 (of the present work) was developed in this article.
- 2 S. J. Mason and H. J. Zimmermann, *Electronic Circuits, Systems, and Signals*, Wiley, New York, 1960, Chapter 5.

- 3 C. G. Montgomery, R. H. Dicke, and E. M. Purcell, *Principles of Microwave Circuits*, McGraw-Hill, New York, 1948. Reprinted by Dover Publications, New York, 1964 (MIT Radiation Laboratory Series, Vol. 8).
- 4 L. Weinberg, "Scattering Matrix and Transfer Scattering Matrix," in R. F. Shea, Ed., *Amplifiers*, McGraw-Hill, New York, 1966.
- 5 V. Belevitch, "Four-Dimensional Transformations of 4-Pole Matrices with Applications to the Synthesis of Reactance 4-Poles," *IRE Trans. CT-3* (2), 105–111 (June 1956).
- 6 See Fig. 4.25 of Reference 2 and relevant discussion.
- 7 H. J. Carlin, "Singular Network Elements," *IEEE Trans. Circuit Theory CT-11*, 67–72 (1964).

Chapter 7

Feedback Analysis of the Bipolar Transistor

Every form of feedback is exhibited by the bipolar transistor. Therefore, it serves well as an introduction to two-port feedback as it relates to a physical device. To understand the feedback aspects of transistor operation, we must describe the physical operation of the transistor in some detail. By so doing, physical interactions are clarified and the reasons for adopting a feedback description (or $ABCD$ matrix description) are demonstrated.

At its core, the $ABCD$ matrix of a bipolar transistor is given by

$$\begin{bmatrix} v_{be} \\ i_b \end{bmatrix} = - \begin{bmatrix} 0 & r_e \\ C_{jc}s & \delta + \tau_T s \end{bmatrix} \begin{bmatrix} v_{ce} \\ -i_c \end{bmatrix} \quad (7.0-1)$$

in which we have expressed the base-emitter voltage and base current in terms of the collector-emitter voltage and the collector current. Four “core” parameters describe the transistor itself: r_e is the emitter resistance kT/qI_C , C_{jc} is the collector junction capacitance, δ is the defect current ratio, equal to $1/h_{fe}$ or $1/\beta$ in conventional terms, and τ_T is the unity current loss (or gain) time constant, equal to $1/2\pi f_T$, where f_T is the unity current loss frequency.

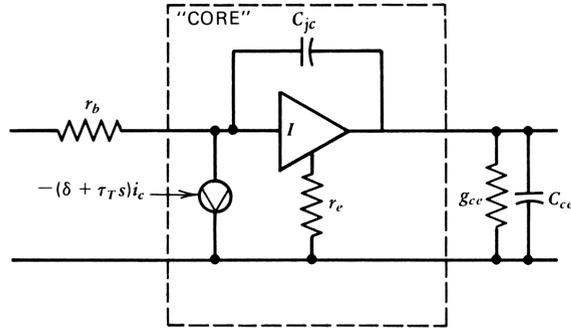


Figure 7.1. A “core” equivalent circuit for the bipolar transistor.

An equivalent circuit that expresses this core matrix is given in Fig. 7.1. The main effects not included in the core equivalent circuit are the base resistance r_b at the input and a shunt admittance at the output. In addition, transit time delay must be taken into account at frequencies above about one-tenth the unity gain cutoff frequency. In the sections to follow, we show how this core matrix and equivalent circuit are related to the physical properties of the transistor. We begin with dc analysis of the transistor and show how the equivalent circuit relates to standard equivalent circuits now used. Dynamic effects are then *added*. Finally, a set of equations is developed that allows us to derive the $ABCD$ parameters of the complete transistor from the equivalent circuit at any voltage and current bias and at any frequency, within wide limits. Another set of equations is developed that allows us to find the equivalent circuit elements from $ABCD$ parameters, enabling us to use two-port measurements directly to find the equivalent circuit. Calculator programs for both equation sets are given in Appendix C.

7.1 PHYSICAL DESCRIPTION

Transistors are available in many sizes tailored to various applications from minuscule microwave transistors to low-frequency power transistors. The principles of operation are remarkably similar over many orders of magnitude of collector current. To establish a clear point of reference, we focus on a particular transistor design, an integrated circuit transistor whose *horizontal geometry* is shown in the plan view of Fig. 7.2a. A sketch of the transistor showing its vertical geometry is given in Fig. 7.2b, and a filamentary slice of the active portion of the transistor is shown in Fig. 7.2c. The processing steps by which this structure is manufactured is beyond our scope here but is amply described in the literature.¹⁻⁵

In Fig. 7.2a the emitter contact and the emitter itself is a stripe of length Z_E and width Y_E . (The coordinate directions agree with convention—the x direction is reserved for the vertical geometry, with x increasing toward the

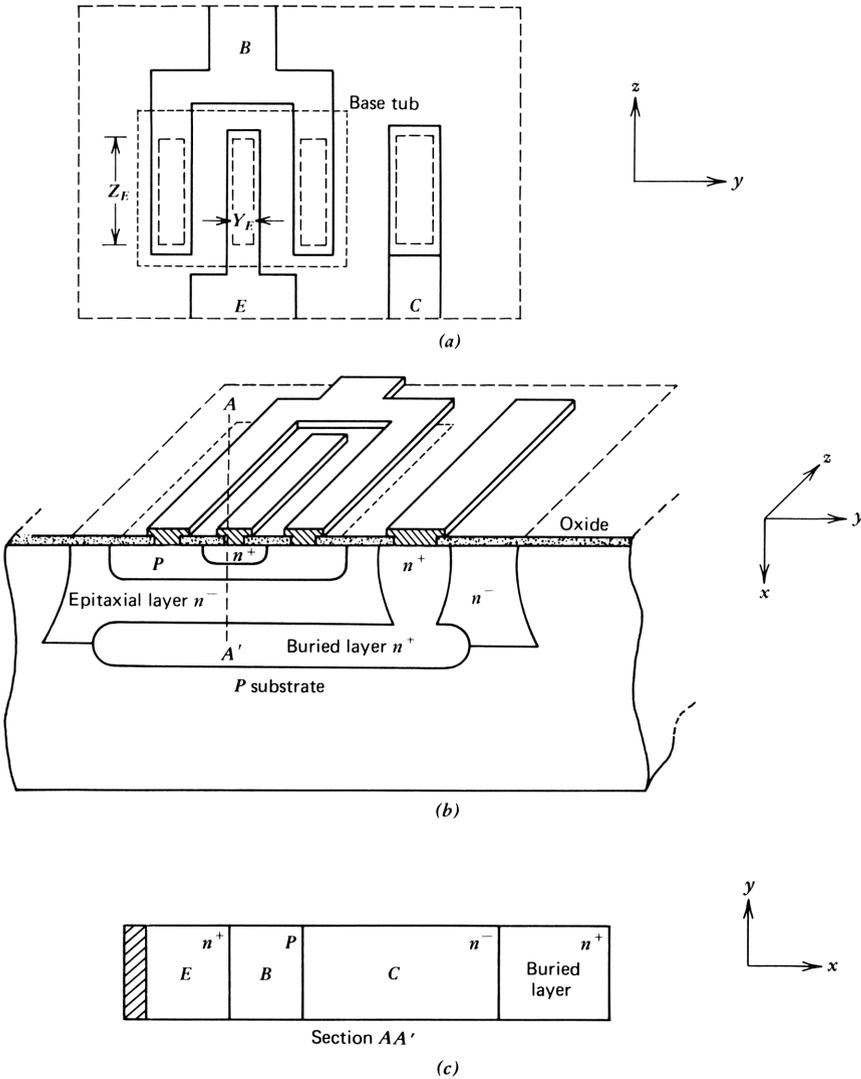


Figure 7.2. A bipolar transistor in a monolithic junction-isolated integrated circuit.

collector.) To maximize the bandwidth, the width of the emitter Y_E is as small as possible, consistent with economical manufacture.⁶ Many considerations of the horizontal geometry are established by the accuracy with which areas can be delineated—the minimum feature size that can be reproduced with good yields at low cost. If we take the minimum feature size as Y_m , then $Y_E = Y_m$. Different features will have differing values of Y_m in general, but this is beyond the level of detail to be covered here. Where Y_m is $5 \mu\text{m}$, for example, the processing rules by which the integrated circuit are made will be characterized

as having a $5\ \mu\text{m}$ geometry. The length of the emitter Z_E will be established by the maximum collector current to be conducted, with a minimum value, of course, given by Y_m .

The base contact is made by stripes on either side of the emitter, also Z_E long. Where Z_E is longer than about $10Y_E$, several interdigitated stripes can be used. For low-current transistors where Z_E would be very short, a wraparound base structure can be used, and a circular emitter is sometimes employed. The object here is to make the ohmic base resistance small. Where this is not important, simple rectangles can be used for both base and emitter.

The collector contact is spaced near the base contact but may be farther than Y_m away to avoid voltage breakdown laterally between the collector and the base. The collector contact is made to a highly doped n^+ region that serves as a low-resistance path to the *epitaxial layer*, which serves as the collector region of the transistor. The epitaxial layer is a thin skin of lightly doped silicon, grown by vapor deposition. As we shall see later, the light doping gives the collector desirable characteristics.

In an integrated circuit the transistor is isolated from the rest of the semiconductor chip by an isolation diffusion that interposes a reverse-biased junction completely around the transistor, extending down to the substrate on which the integrated circuit is made.

Vertical Geometry

The vertical geometry is that under the emitter contact in Fig. 7.2*b* along the line AA' . This line passes through the active area of the transistor. From top to bottom it encounters the emitter contact metal, the n^+ emitter region (n^+ designates an n region of heavy doping), the p -type base region, the (lightly doped) n^- collector epitaxial region, and the n^+ buried layer, which provides a low-resistance conductive path to the collector contact. Below the buried layer is the p -type substrate, the starting material or wafer.

The processing steps by which this vertical geometry is realized are succinctly described, for example, in Gray and Meyer.⁷ We are interested here in the function of each layer.

Figure 7.2*c* shows a filamentary slice of the transistor along the line AA' . The transistor section is divided into emitter, base, collector (epitaxial), and buried layers. The emitter depth may be as small as $0.2\ \mu\text{m}$ for a microwave transistor, or upward of $5\ \mu\text{m}$ for a power transistor. Base widths are of the same order as the emitter. The collector epitaxial layer width varies over a wide range: for microwave transistors, it may be less than $1\ \mu\text{m}$, whereas a power transistor may have a width of $100\ \mu\text{m}$ or more. Metallurgical *junctions* separate base and emitter, base and collector, and buried layer and substrate. Note that there is no junction between the epitaxial layer and the buried layer since the semiconductor does not change conductivity type.

Derivation of Electrical Properties

Intrinsic silicon, silicon that has no impurities, conducts electricity by means of movable charge carriers, both holes and electrons. The number of such charge

carriers is small: there are only 1.45×10^{-2} such carriers of each type per cubic micrometer at room temperature. When impurities are added, the conductivity rises. Donor-type impurity atoms that have a valence of 5 donate an electron to the conduction process since only four of the outer electrons are needed to complete the lattice. Similarly, acceptor atoms with a valence of 3 remove an electron from the valence band, leaving a hole that becomes available for conduction. In equilibrium, it can be shown that

$$np = n_i^2 \quad (7.1-1)$$

where n and p are the densities of electrons and holes in a given semiconductor and n_i is the density of both holes and electrons in intrinsic material. If 10^4 donor atoms per cubic micrometer were added to intrinsic silicon, for example, n would increase by $10^4 \mu\text{m}^{-3}$ and p would decrease to $(1.45 \times 10^{-2})^2 / 10^4 = 2.1 \times 10^{-8} \mu\text{m}^{-3}$. For this material, electrons are the *majority carriers* and holes, the *minority carriers*.

When equilibrium is disturbed, the charge carriers move under two influences; (1) if the density of carriers is not uniform, a *density gradient* will cause the carriers to move in the direction of decreasing density, like perfume molecules diffusing through a room (this motion of carriers is *diffusion*) and (2) when there is a field present, the carriers are accelerated by it; this is motion by *drift*. The velocity of carriers in a drift field does not continue to accelerate as it would in free space; energy is removed from the carriers by collisions with the lattice. The process can be likened to the travel of steel balls bumping down an inclined plane studded with nails. At sufficiently high voltages, the velocity of carriers saturates at a value termed the *scattering limited velocity*.

When oppositely doped semiconductors are in intimate contact, they form a junction. In the acceptor-doped p region, holes (*the majority carriers*) tend to diffuse across the junction under the influence of the density gradient since the number of holes in the n region is extremely small. Similarly, electrons in the donor-doped n region tend to diffuse into the p region. This process cannot continue for long, because as the carriers move, they leave behind the fixed charges on the nuclei of the doping atoms. An electric field is thereby created that halts the movement of charge. This field, integrated across the junction, produces a *built-in voltage* that creates a *barrier* to further movement of majority carrier charge.

On the other hand, any *minority carriers* that find themselves near the junction will be swept across it, accelerated by the built-in voltage. Therefore, density of minority carriers near the junction is virtually zero.

The fundamental basis of transistor action is to introduce controlled numbers of charge carriers (electrons for an *npn* transistor) in the immediate vicinity of the collector junction (on the base side). Once there, they will be given energy by being swept across the reverse-biased collector junction. This is the basis of gain in the bipolar transistor. At low frequencies the action of the collector junction itself is virtually ideal, with all charge carriers swept across the junction.

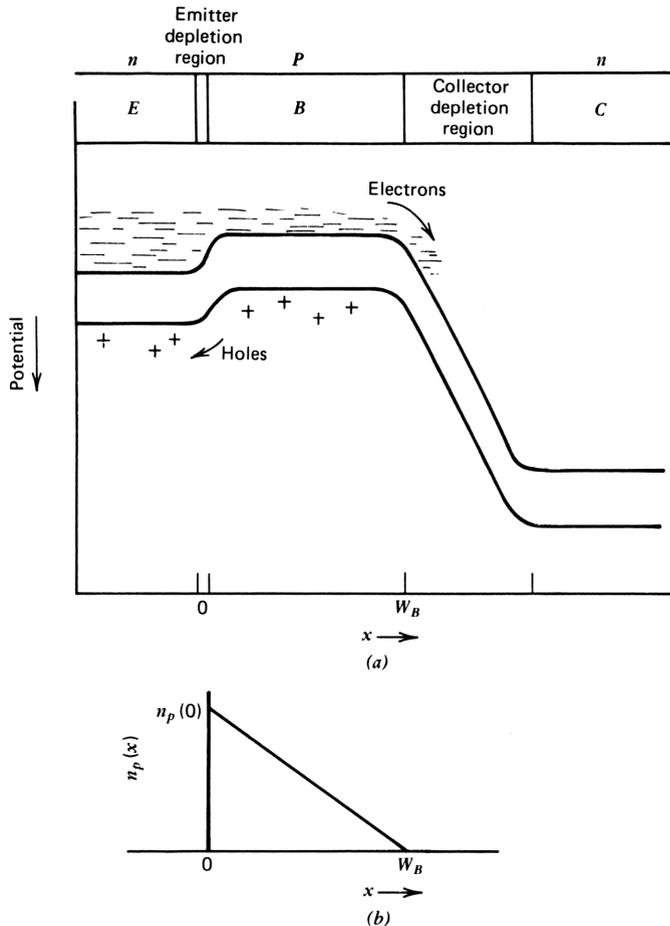


Figure 7.3. Physical processes in the biopolar transistor: (a) potential diagram; (b) minority carrier density in the base region.

The charge carriers are introduced near the collector junction by injection across the emitter junction into the base region under the influence of an applied forward bias potential. Once in the base, the charges diffuse toward the collector under the influence of a density gradient. Two separate physical processes are evident here; (1) the injection of minority carrier electrons into the base region and (2) transport across the base region by diffusion (see Figs. 7.3a and 7.3b)

The potential diagram in Fig. 7.3a shows a large reservoir of electrons in the emitter region, held back from flowing into the base by the potential barrier created by the fixed charge mentioned previously. When the barrier is lowered by applying forward bias to the emitter junction, some electrons have enough thermal energy to surmount the barrier and increase the concentration of

minority carrier electrons on the base side of the emitter junction. This process is governed by Maxwell-Boltzmann statistics at ordinary temperatures; the electron concentration is given by

$$n_p(0) = n_{po} \exp \frac{q}{kT} V_{be} \quad (7.1-2)$$

which states that the charge density in the base region at the edge of the emitter junction $n_p(0)$ is exponentially related to the applied voltage across the emitter junction V_{be} . The equilibrium charge density n_{po} is the charge density with zero applied bias and is given by eq. (7.1-1) and the following discussion as

$$n_{po} = \frac{n_i^2}{N_{AB}} \quad (7.1-3)$$

where N_{AB} is the acceptor doping density in the base.

Charges thus injected into the base region diffuse toward the collector junction under the influence of a density gradient; the density of minority carrier charge at the collector edge of the base region is virtually zero. The equation governing this motion is the diffusion equation:

$$I_n = -qA_e D_n \frac{dn_p(x)}{dx} \quad (7.1-4)$$

in which q is the magnitude of the electronic charge, A_e is the area of the emitter, D_n is a constant of proportionality (in $\mu\text{m}^2/\text{ns}$) called the *diffusion constant*, and $dn_p(x)/dx$ is the density gradient. As seen in Fig. 7.3b, the density gradient is assumed to be constant (departure from the assumption does not greatly alter the basic notions involved). Under these conditions we can write

$$\frac{dn_p(x)}{dx} = -\frac{n_p(0)}{W_B} \quad (7.1-5)$$

in which W_B is the width of the base region. The minority carrier current can then be written

$$I_n = \frac{qA_e D_n n_p(0)}{W_B} \quad (7.1-6)$$

Note that q is the *magnitude* of the electronic charge of either an electron or a hole. The sign is carried by the assumed direction of I_n . This equation holds for the condition of zero bias as well as forward bias. Hence we can write an expression for I_s , the saturation current, defined as the minority carrier

(electron) current that flows under conditions of zero emitter bias voltage:

$$I_S = \frac{qA_e D_n n_{po}}{W_B} \quad (7.1-7)$$

Substituting (7.1-2) for n_{po} , we have

$$I_S = \frac{qA_e D_n n_i^2}{W_B N_{AB}} = \frac{qA_e D_n n_i^2}{N_G} \quad (7.1-8)$$

in which we have substituted the *Gummel number* N_G for $W_B N_{AB}$. The Gummel number is the number of dopant atoms in the base per unit area of the emitter and is typically 3×10^4 atoms per square micrometer of emitter area. This saturation current is small, indeed. For every square micrometer of emitter area, the saturation current is only some 10 electrons per second.

Dividing (7.1-6) by (7.1-7), we obtain

$$\frac{I_n}{I_S} = \frac{n_p(0)}{n_{po}} \quad (7.1-9)$$

Substituting this equation into (7.1-2), we obtain the base-emitter voltage as a function of the electron current;

$$V_{be} = \frac{kT}{q} \ln \frac{I_n}{I_S} \quad (7.1-10)$$

Circuit Characteristics Related to the Physical Description

When the minority carriers reach the collector, they fall over the potential drop and appear at the collector lead as collector current, so that

$$I_c = I_n \quad (7.1-11)$$

This seemingly trivial equation expresses the gain mechanism of the transistor in idealized form and should not be dismissed lightly. This equation is not exactly true because some of the minority carrier electrons recombine with holes in the base region and are lost before they reach the collector. In a water-over-the-dam analogy, this is equivalent to evaporation of the water before it reaches the edge of the dam. The effect is negligible in most modern transistors. Another, more important, departure to be considered later is that I_c is delayed relative to I_n by the transit time in the device, taking I_n as the current at the base side of the emitter junction. For now, we use (7.1-11) in its idealized form, so that from (7.1-10) we have the base emitter voltage as a function of collector current;

$$V_{be} = \frac{kT}{q} \ln \frac{I_c}{I_S} \quad (7.1-12)$$

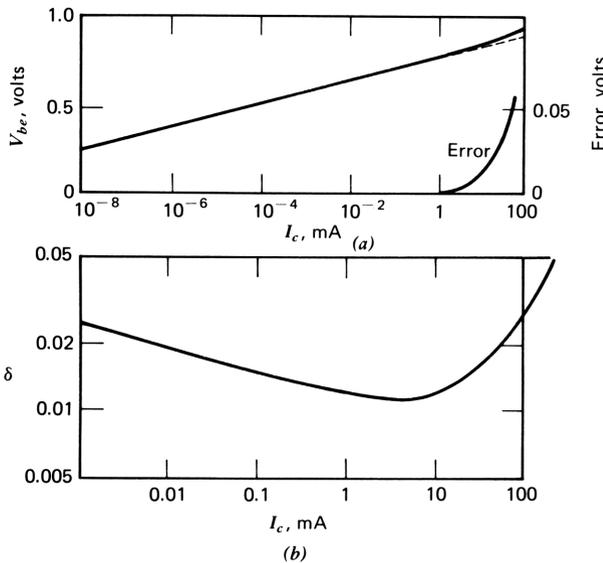


Figure 7.4. Variation of V_{be} and defect current ratio δ with collector current.

The logarithmic relationship between V_{be} and I_c of (7.1-12) holds over a remarkable range of currents of a billion to one. The accuracy is limited at high currents by emitter contact resistance, giving a linear increase in voltage with currents in addition to the logarithmic one. Also, at high currents high-level injection effects increase the slope of the V_{be}/I_c curve by a factor of roughly 2. The plot in Fig. 7.4a shows the relationship. The error curve gives a typical departure from the logarithmic relationship.

Base current arises in modern transistors primarily by injection of holes from the base into the emitter. This is also shown in the potential diagram in Fig. 7.3. Holes are majority carriers in the base, and they encounter the same voltage barrier (in the reverse x direction and with opposite polarities) as for electrons. Therefore, we might expect an analogous hole injection into the emitter, governed by an equation like that of (7.1-2), except that the densities are hole densities rather than electron densities. These holes must come from the base contact since there are essentially no holes in the collector. Thus the injected hole current becomes the base current, or the “defect” current. In good transistors, this current is held to a minimum by doping the base more lightly than the emitter. In this way, the density of holes (majority carriers) in the base is much smaller than the density of electrons (majority carriers) in the emitter. Thus, we can write

$$I_b = K \frac{N_{AB}}{N_{DE}} I_c = \delta I_c \tag{7.1-13}$$

where N_{AB} is the impurity (acceptor) density in the base and N_{DE} is the impurity (donor) density in the emitter. The term K is a constant, given

roughly by

$$K = \frac{W_B D_p}{L_e D_n} \quad (7.1-14)$$

in which D_p is the diffusion constant for holes in the n -type emitter, D_n is the diffusion constant for electrons in the p -type base, W_B is the width of the base, and L_e is an analogous quantity for the emitter. It is the diffusion length for holes in the emitter; since the emitter is heavily doped and the population of electrons is high, holes reaching the emitter region tend to *recombine* with electrons before they reach the emitter contact.

The defect current ratio is not constant with collector current; a typical variation is given in Fig. 7.4*b*, which shows that δ increases at extremes of low and high currents. At low currents there is an increase in recombination in the emitter depletion region, primarily near the surface of the semiconductor.⁸

When the density of minority carriers injected into the base becomes sufficient to be comparable with the density of acceptor atoms in the base, performance begins to deteriorate. A good way to grasp this intuitively is to consider that the minority carrier charge passing through the base region acts effectively to *increase* the effective acceptor doping density in the base region (remember that the acceptor atoms have a negative fixed charge, which is effectively increased by the minority carriers). It does not matter whether the minority carriers are flowing—they still change the character of the base region in the same way as do the stationary acceptor atoms. That is to say, the number of holes in the base is increased (by charge neutrality). This causes increased hole injection into the emitter (see Fig. 7.3*b*) with its consequent increase of the defect current. Hence δ increases under high-level injection conditions. This was first noted by Webster⁹ and is termed *base region conductivity modulation*. (As the doping density rises, so does the conductivity.)

The three key concepts of the bipolar transistor are expressed by eqs. (7.1-11), (7.1-12), and (7.1-13). They are summarized in an historical account of the development of the transistor by one of its inventors.¹⁰ In his words, they are the following:

1. Minority carrier injection into the base layer which increases exponentially with forward emitter bias,
2. Application of reverse voltage at the collector junction,
3. Favorable geometry and doping levels so as to obtain good emitter to collector efficiency.

7.2 STATIC EQUIVALENT CIRCUITS

The three key concepts are represented in Fig. 7.5; in Fig. 7.5*a* current flow in the silicon semiconductor is shown, and in Fig. 7.5*b* each of the three effects is represented by an equivalent circuit element in the static case. The reverse-biased collector junction is represented by the ideal amplifier; when minority

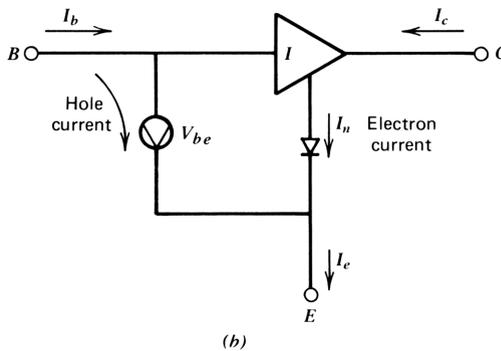
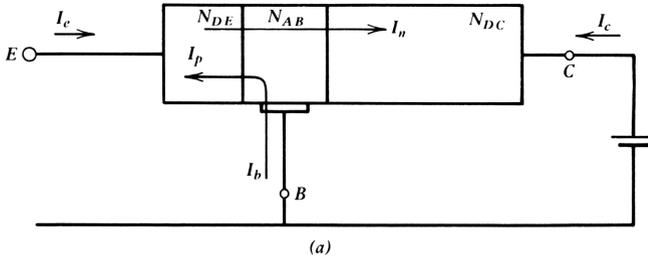


Figure 7.5. Representation of the three key concepts of the transistor: (a) current flowing in silicon; (b) an equivalent circuit representation.

carriers reach the base side of the collector junction (represented by the common lead of the ideal amplifier), they disappear, leaving no trace in the base circuit. In the act of “falling off the edge,” they produce neither any base current nor any base-emitter voltage.*

Thus the ideal amplifier is a circuit representation of the collector junction and its action is described by eq. (7.1-11). The diode in series with the common lead represents the emitter junction and produces the base-emitter voltage, given by eq. (7.1-12). The incremental resistance of this diode is r_e , the emitter resistance; this resistance is a feedback resistance because it returns a signal (voltage) to the input in proportion to the output signal (current). The dependent current generator connected in shunt with the input represents the hole injection into the emitter, given by eq. (7.1-13). It is proportional to the collector current (through the defect current ratio) and is likewise a feedback generator, returning a portion of the output signal (current) to the input (as a signal current).

*This is an approximation, but a good one. We explore departures later when we consider the effects of collector voltage.

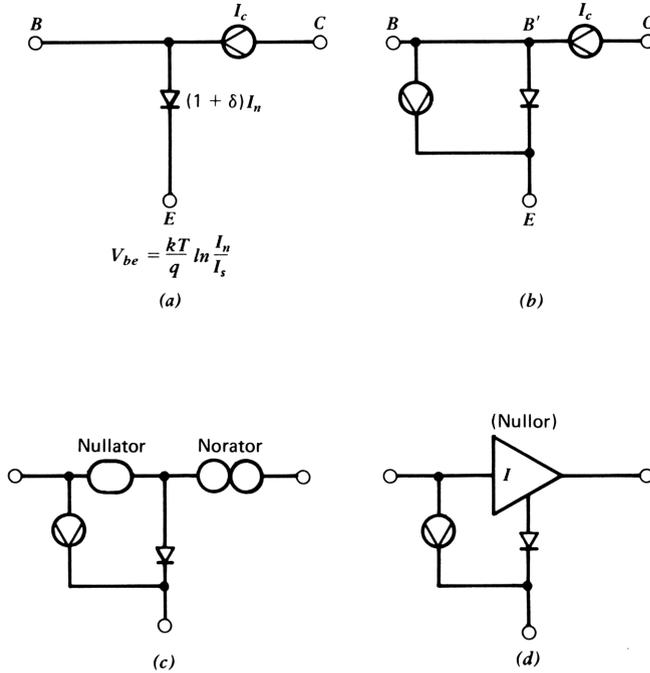


Figure 7.6. Development of the relationship between the Ebers-Moll transport model and the circuit in Fig. 7.5b (forward-active operation).

The transistor may be considered to be a feedback structure, in which the ideal amplifier representing the collector junction is the forward path and both the emitter diode and the defect generator are feedback elements. Thus the circuit in Fig. 7.5b represents the three key effects in the transistor with 1:1 correspondence between the equivalent circuit elements and the physical phenomena, with transistor action specifically represented by the ideal amplifier. This brings us to a clearer view of anticausal analysis; the feedback elements entirely determine the characteristics of the transistor; without them, the $ABCD$ matrix is null. By analyzing the *feedback* paths *causally*, that is, from *output* to *input*, we find the values of the two-port parameters. Since we think of cause and effect flowing from input to output, the analysis is called *anticausal*, but it is worthwhile noting that *causal* evaluation and summation of feedback path signals define the method.

Relationship to the Ebers-Moll Model

The representation in Fig. 7.5b is fully equivalent to the Ebers-Moll transport* model of the transistor biased for *forward-active* operation, that is, with

*The term “transport model” is used here to distinguish it from the *injection model*; the latter is the model originally proposed by Ebers and Moll. The distinction between the two is described in Reference 11, and the advantages of the transport model are pointed out.

reverse-biased collector and forward-biased emitter. This circuit is shown in Fig. 7.6a. “In this latter circuit, which has formed the basis for much transistor modeling, a collector current is represented by a current generator, and the current flowing through the emitter diode includes both the hole and electron currents; the voltage across the emitter diode is assumed to be created by electron current acting alone. It is an awkward circuit because too much is left for interpretation that is not specifically given in the circuit itself. Since the circuit has been widely used, it is worth relating the new circuit to it.

The first step is to remove the hole current from the diode, simply by placing a dependent defect-current generator around the diode as shown in Fig. 7.6b. When this is done, the current through the short circuit between nodes *B* and *B'* is identically zero. This short circuit shares the qualities of an open circuit; the short circuit cannot be removed, however, because the voltage between *B* and *B'* must be identically zero. The element required to replace the short circuit is the *nullator*, introduced by Carlin.* It is a pathological element that simultaneously constrains the voltage across it and the current through it to be zero.

The current generator I_c representing the collector junction is also inappropriate. By labeling the generator I_c , we make its current arbitrary—an independent variable. The voltage across it (or across any current generator) is also arbitrary. This element precisely fits the definition of a *norator* also introduced by Carlin. He defined it as a two-terminal singular (pathological) element whose voltage and current are both arbitrary. Replacing the short circuit by a nullator and the current generator by a norator, we obtain the equivalent circuit in Fig. 7.6c. The use of the nullator and the norator to represent the transistor was first recognized by Mitra¹² and has been discussed by Moschytz.¹³

Where a nullator and a norator are combined as in Fig. 7.6c, they form a *nullor*, or an ideal amplifier. As pointed out by Carlin, whereas the nullator and the norator are pathological elements, the nullor or the ideal amplifier is not. Its use in the transistor equivalent circuit makes the addition of various circuit and parasitic effects in the transistor particularly easy to accommodate. By putting the transport model on a firmer circuit theoretical basis, we have transformed it to the equivalent circuit in Fig. 7.5b.

A corresponding small-signal equivalent circuit is given in Fig. 7.7a, drawn by replacing the diode by its dynamic resistance and the generator by its small-signal equivalent. The difference between the large- and small-signal defect current generators is not great; in most designs the difference can be neglected.

The two-port *ABCD* matrix for this simple model is given directly as

$$\begin{bmatrix} v_{be} \\ i_b \end{bmatrix} = \begin{bmatrix} 0 & r_e \\ 0 & \delta \end{bmatrix} \begin{bmatrix} v_{ce} \\ i_c \end{bmatrix} \tag{7.2-1}$$

*See Reference 7 of Chapter 6.

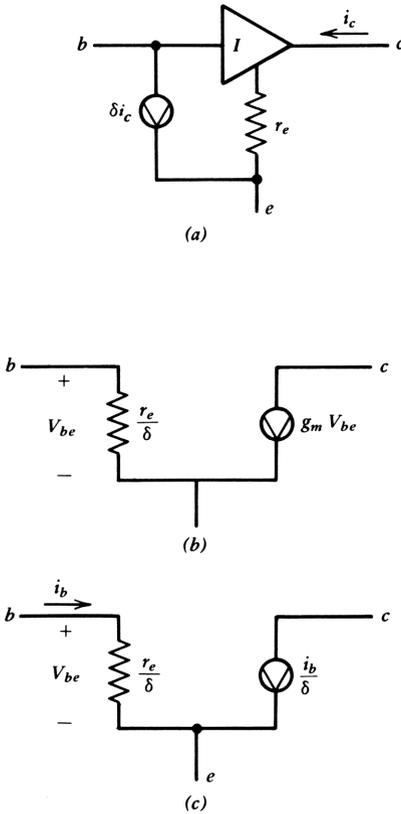


Figure 7.7. Small signal equivalent circuits corresponding to the Ebers-Moll model.

As noted in Chapter 1, these equations can be integrated to obtain the nonlinear behavior of the transistor by replacing r_e and δ by their respective functions of collector current. Only the boundary conditions have been lost in deriving the small-signal model.

Two small-signal equivalent circuits often used for conventional analysis are shown in Figs. 7.7b and 7.7c. In Fig. 7.7b the functions of the ideal amplifier (collector junction) and the emitter diode resistance are combined to form a dependent g_m generator, ($g_m = 1/r_e$). Separate representation of the collector junction is inconvenient or impossible in the conventional formulation because of the infinite gain involved. Defect current is represented by a resistance connected from base to the common lead and has the value $1/\delta g_m$ or $h_{fe} r_e$. Similarly, the circuit in Fig. 7.7c combines the function of the (ideal) collector junction with the feedback of the defect generator feedback in a current-controlled current source i_b/δ , or $h_{fe} i_b$. Base-emitter input voltage arises by connecting a resistor in series with the base lead.

Although all three circuits in Fig. 7.8 give the same answers to network problems, the conventional circuits are thermodynamically incorrect in predicting too much noise. There is shot noise associated with both the collector and

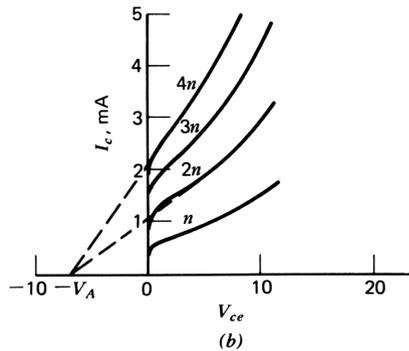
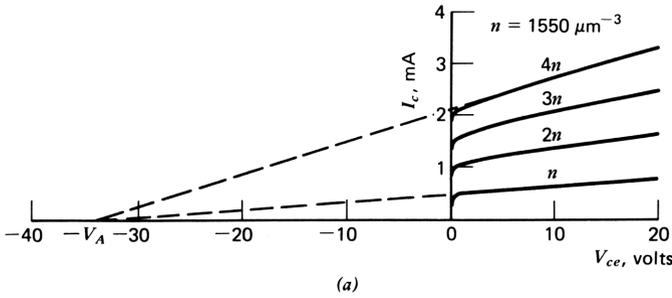


Figure 7.8. Static collector characteristics of two transistors: (a) light collector doping [$N_{DC} = 10^4 \mu\text{m}^{-3}$, $N_{AB} = 10^5 \mu\text{m}^{-3}$, $W_B(5V) = 0.3 \mu\text{m}$]; (b) heavy collector doping [$N_{DC} = 10^6 \mu\text{m}^{-3}$, $N_{AB} = 10^5 \mu\text{m}^{-3}$, $W_B(5V) = 0.3 \mu\text{m}$].

base currents, but no thermal noise as would be expected in the circuits in Figs. 7.8b and 7.8c. We discuss the subject of noise in devices in Chapter 11.

The circuit in Fig. 7.8a is appropriate for anticausal analysis and can be used to define a reference condition for the transistor. The reference condition is obtained by setting feedback effects elements r_e and δ to zero, giving an $ABCD$ matrix that is null. This is an example of an extremely useful type of equivalent circuit and associated matrix; we term it the *null reference equivalent circuit*.

7.3 EFFECTS OF COLLECTOR VOLTAGE

The equivalent circuit and two-port parameters derived in Section 7.2 are completely insensitive to collector voltage and must be considered approximations. There are three effects of collector voltage on the static characteristics of transistors:

- 1 Early effect, in which the width of the base region decreases as the collector voltage increases (the width of the collector depletion layer increases with voltage).

- 2 Saturation, where the reverse bias on the collector voltage is insufficient to prevent injection of minority carrier electrons from the collector into the base region.
- 3 Avalanche multiplication, which causes breakdown of the collector junction at sufficiently high voltages.

A fourth effect, encountered where the collector is highly doped, is punch-through, in which the base narrows so much due to Early effect that the collector reaches the emitter. We concentrate first on Early effect since this affects the characteristics of the transistor over the entire useful operating range of collector voltages. Saturation and avalanche effects tend to set limits on signal voltage excursion at the collector; these are discussed in the paragraphs that follow.

Early Effect

As described in Section 7.1 the field across the collector junction is provided by fixed charge of the dopant atoms caused by diffusion of the movable charge across the junction barrier. Movement of this charge stops when enough fixed charge is “exposed” to create this field. As the collector voltage is increased, increasing amounts of fixed charge must be so exposed, so that the depletion region—the region devoid of movable charge—must widen to support the increased voltage. Where the doping is light, more movement of the depletion region is required to expose a given amount of charge. In most transistors, the collector is more lightly doped than the base, so that the depletion region widens primarily into the collector. Nevertheless, movement into the base does take place, so that the base width becomes narrower as the collector voltage is increased.

The chief effect of this narrowing is to affect the diffusion of carriers across the base. According to eq. (7.1-6), as W_B becomes smaller, the minority carrier current I_n increases for a given value of $n_p(0)$. Since $n_p(0)$ is set by V_{be} , the minority carrier current (or the collector current) must increase for a fixed value of V_{be} . Hole injection into the emitter is unaffected by the base width, so that the defect current—the base current—is likewise unaffected. This can be seen graphically in the static characteristics of two transistors plotted in Fig. 7.8. In Fig. 7.8a the transistor is of the normal variety in which the collector doping level is lower than that of the base (by a factor of 10 here). The transistor in Fig. 7.8b is of a different type* in which the collector doping is 10 times *higher* than the base doping. In each case the tangent to the static characteristics (drawn for constant values of base current) intersects the current axis at a voltage of $-V_A$, called the *Early voltage*. Although this voltage varies with collector voltage and current, it is roughly constant for a given transistor. In Fig. 7.9a V_A is 35 V and in Fig. 7.9b, 7 V.

From the geometry of these characteristics we see that we can model Early effect by adding a shunt conductance between collector and emitter, as shown

*This is *isoplanar* transistor, a proprietary process of Fairchild Semiconductor, Inc.

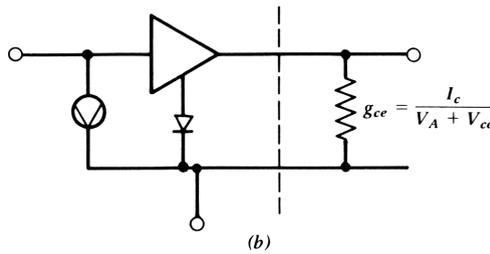
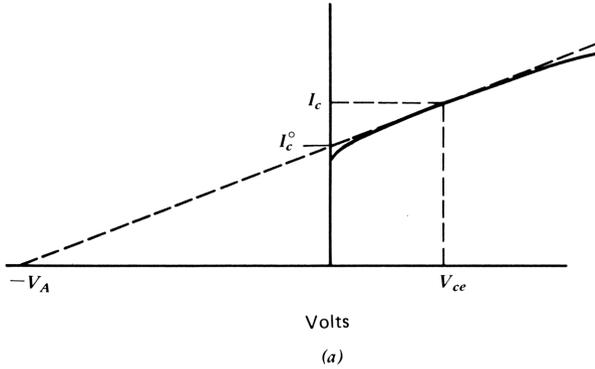


Figure 7.9. Derivation of collector conductance from geometrical considerations. The effect on the equivalent circuit is shown in part *b*.

in Fig. 7.9*b*. The value of the conductance is given by the geometry in Fig. 7.9*a* as

$$g_{ce} = \frac{I_c}{V_A + V_{ce}} \tag{7.3-1}$$

The increased minority carrier current (or collector current) is seen to pass through the external conductance rather than through the transistor model itself. In this way the base-emitter voltage relationship is preserved while the collector (and emitter) currents are increased by the amount observed.

In terms of the transistor geometry and doping levels, it can be shown that the Early voltage can be expressed by the equation

$$V_A + V_{ce} = \frac{q}{\epsilon} N_G W_{TC} \tag{7.3-2}$$

where N_G is the Gummel number in the base and W_{TC} is the width of the collector depletion region; q is the magnitude of the electronic charge, 1.6×10^{-7} pC, and ϵ is the dielectric permittivity in silicon, 1×10^{-4} pF/ μ m. The

width of the depletion region is given by

$$W_{TC} = \left(\frac{2\epsilon V_{ce}}{qN_{PC}} \right)^{1/2} \quad (7.3-3)$$

in which N_{PC} is a function of the base and collector doping, given by the sum of their reciprocals

$$\frac{1}{N_{PC}} = \frac{1}{N_{AB}} + \frac{1}{N_{DC}} \quad (7.3-4)$$

where N_{DC} is the donor concentration in the collector region. The curves shown in Fig. 7.8 were drawn by applying these equations but are typical of measured static characteristics. Equation (7.3-3) assumes uniform doping levels in the base and collector. Although this is not a good approximation to the actual doping profile, the results given here are not greatly affected. For example, with $V_{ce}=3$ V, $N_{AB}=10^5 \mu\text{m}^{-3}$, and $N_{DC}=10^4 \mu\text{m}^{-3}$, the depletion width is $0.64 \mu\text{m}$ from eq. (7.3-3). Then, from (7.3-2), with $N_G=3 \times 10^4$, $V_A + V_{ce}=31$ V, and $V_A=28$ V.

The effect on the $ABCD$ matrix of the Ebers Moll transport model is found simply by post multiplying its $ABCD$ matrix by that for the shunt conductance:

$$\begin{bmatrix} v_{be} \\ i_b \end{bmatrix} = - \begin{bmatrix} 0 & r_e \\ 0 & \delta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ g_{ce} & 1 \end{bmatrix} \begin{bmatrix} v_{ce} \\ i_c \end{bmatrix} \quad (7.3-5)$$

$$\begin{bmatrix} v_{be} \\ i_e \end{bmatrix} = - \begin{bmatrix} r_e g_{ce} & r_e \\ \delta g_{ce} & \delta \end{bmatrix} \begin{bmatrix} v_{ce} \\ i_c \end{bmatrix} \quad (7.3-6)$$

If we replace r_e and g_{ce} by their functions of current and voltage, we obtain

$$\begin{bmatrix} v_{be} \\ i_b \end{bmatrix} = - \begin{bmatrix} \frac{kT}{q(V_A + V_{ce})} & \frac{kT}{qI_C} \\ \frac{\delta I_C}{V_A + V_{ce}} & \delta \end{bmatrix} \begin{bmatrix} v_{ce} \\ i_c \end{bmatrix} \quad (7.3-7)$$

Suppose that a transistor operating at $V_{CE}=3.0$ V and $I_C=1.0$ mA has $\delta=0.01$ and $V_A=20$ V. It should thus have an $ABCD$ matrix at room temperature, from (7.3-7), of

$$T_a(0) = - \begin{bmatrix} 0.00113 & 0.026 \text{ k}\Omega \\ 0.00043 \text{ mS} & 0.01 \end{bmatrix} \quad (7.3-8)$$

Is this an accurate picture of the low-frequency two-port parameters of a transistor? Two-port measurements were made on a microwave transistor with

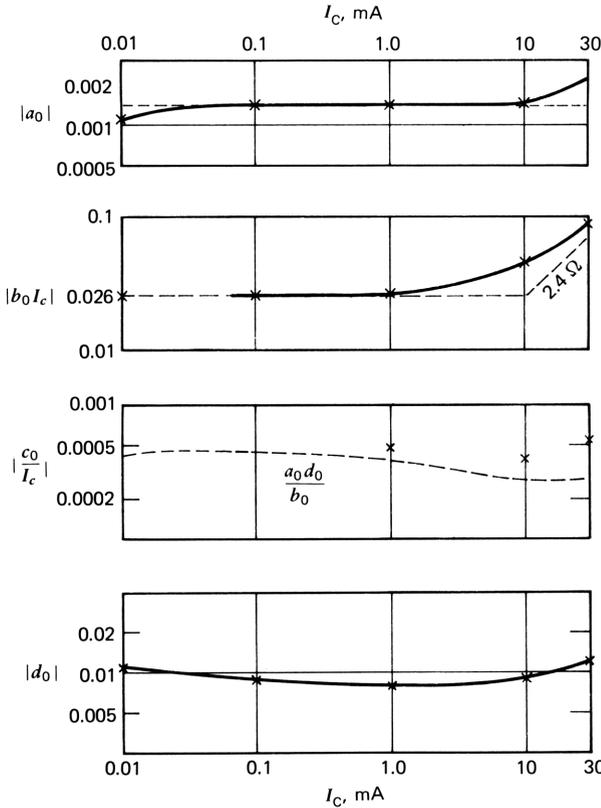


Figure 7.10. Comparison of measured and calculated $ABCD$ parameters of a microwave transistor for currents in the range 0.01–30 mA at dc.

$2.5 \mu\text{m}$ geometry at low frequencies with the results shown in Fig. 7.10; agreement is close to the calculated values even without accounting for the variation of V_A with collector current. At high currents B departs from the calculated values because of emitter contact resistance r'_e , as does A .

Saturation

In a topological sense the transistor is symmetrical—a base region sandwiched between two junctions. By reversing the voltages on both junctions, we convert the emitter junction to a collector junction and vice versa. In this *reverse-active* operation, the transistor is a poor one because it violates the third of the three key concepts listed in Section 7.1; the geometry and doping levels are distinctly unfavorable in most transistors. The collector junction is usually not doped more heavily than the base; when it is made into an emitter by application of forward bias, the emitter efficiency is small and the dc current loss is large. Hole injection across the forward biased collector junction is consequently high, with values of the defect current ratio δ_r typically greater than unity.

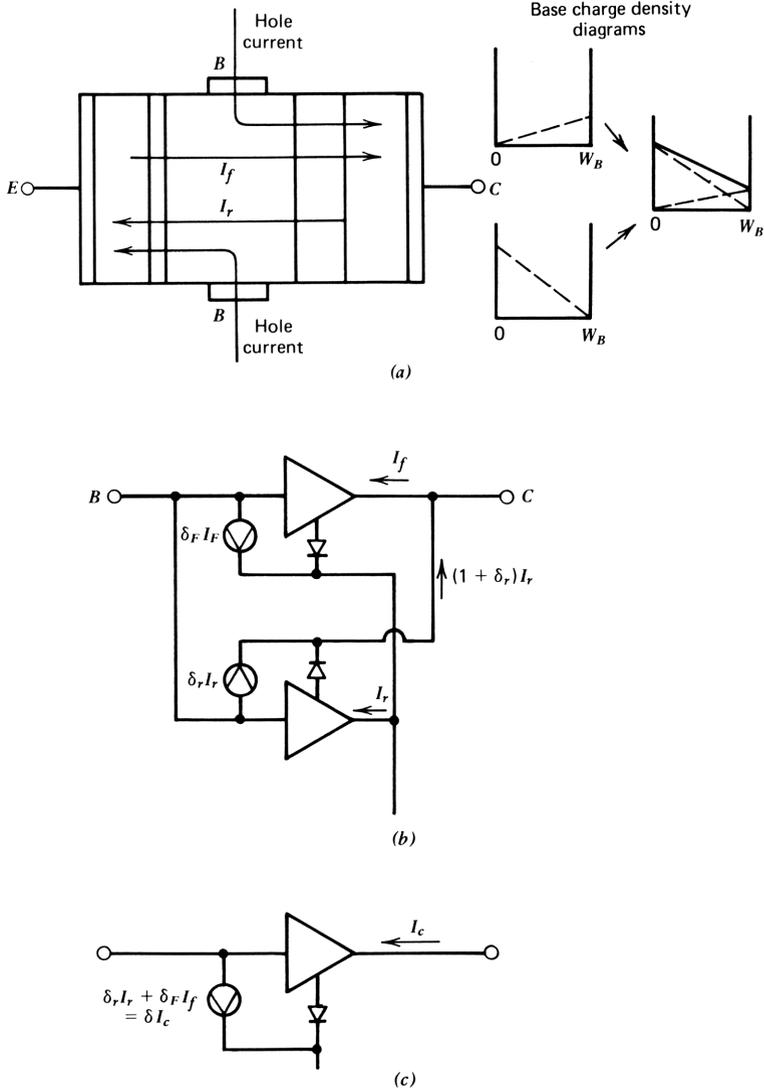


Figure 7.11. Operation in the saturated region: (a) hole and electron flow; (b) equivalent circuit as the superposition of two transistors; (c) equivalent circuit for operation at edge of the forward-active region.

In the saturated region forward bias is applied across both junctions simultaneously, so that electrons are emitted across both junctions as shown in Fig. 7.11a. The analysis of this more complicated case is eased by the concept of superposition. As in normal, forward-active operation, electrons crossing the emitter junction are collected by the collector junction, even though the collector bias is not in the reverse direction. Collector action takes place

because of the built-in voltage at the collector junction: as long as the forward bias is smaller than the built-in voltage, a small potential hill exists, allowing the electrons to be collected. (Such a small potential hill exists at the emitter junction in Fig. 7.3, for example.)

The same situation exists for electrons emitted across the collector junction, now forward biased. For these electrons, the small potential hill at the emitter allows their collection. In effect, we have two transistors in one, with electrons being emitted across both junctions and being collected by their opposite counterparts. The fact that their effects can be superimposed was first recognized by Ebers and Moll, who then developed the equivalent circuit bearing their name. The circuit is shown in Fig. 7.11*b*, modified from the original by the development given in Fig. 7.6. In Fig. 7.11*a*, electron and hole currents for the two transistors are shown schematically, with the minority carrier concentration diagrams shown separately for the two, along with their superposition. The equivalent circuit is shown in Fig. 7.11*b*. When the collector is reverse biased, the lower transistor becomes inoperative and disappears from the circuit. Typically, δ_r is greater than unity, and the currents supplied by the lower transistor to the base and collector of the upper transistor are of the same order of magnitude.

From the standpoint of the junction voltages, the regions of operation are shown in Fig. 7.12, which plots V_{be} against V_{cb} for various values of collector current. Normal operation of the transistor occurs in the relatively narrow shaded area in the first quadrant; inverse operation is shown in the similar shaded area in the third quadrant. In either case normal operation extends a little way into the second quadrant, but this area is dominantly controlled by saturation. Below and to the right of the shaded regions is the cutoff range. The remainder of the $V_{be}-V_{cb}$ plane is the burn-out region: if a transistor is measured to be operating in this region, you may be sure it is burned out since a healthy transistor cannot sustain forward voltages of more than 0.8–1.0 V. The facts represented by Fig. 7.3 are a useful aid to troubleshooting. By dc measurement of V_{be} and V_{cb} , the region of operation of a given transistor in a circuit can be established and an estimate of its health made.

We are concerned mainly with the effect of saturation as it affects the limits of operation in the forward-active region. In this region, when the collector voltage becomes sufficiently low, the lower transistor in Fig. 7.11*b* begins to conduct. Since the contributions of this transistor are roughly the same at both base and collector of the upper transistor and the base current is two orders of magnitude less than the collector (by the factor δ_f), the only significant effect is on the base current. We can represent the effect of saturation on operation at the edge of the forward-active region by the circuit in Fig. 7.11*c*, in which saturation is represented by an added current generator $\delta_r I_r$, connected from base to emitter. To complete the equivalent circuit, we must now find I_r in terms of the collector current and voltage. Although the $\delta_r I_r$ generator is actually connected between the base and the collector, which can be represented by two identical current generators, one from base to emitter and a

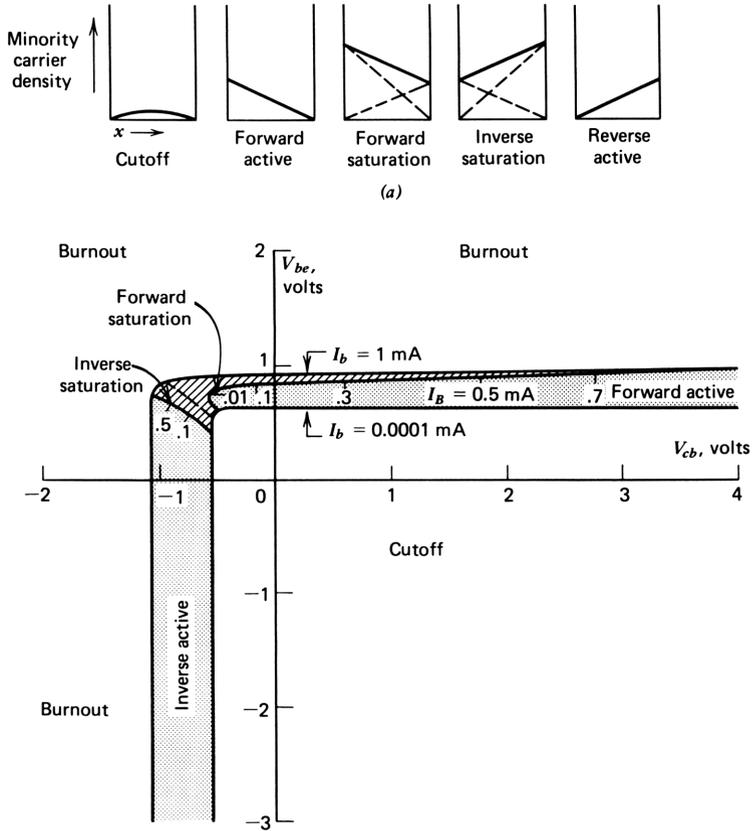


Figure 7.12. Operating regions of the bipolar transistor.

second from emitter to collector, the second generator can be ignored, since it is small compared with the collector current.

From eq. (7.1-10) we can write for the forward transport current

$$I_f = I_S \exp \frac{qV_{be}}{kT} \tag{7.3-9}$$

Similarly, for the reverse transport current, we have

$$I_r = I_S \exp - \frac{qV_{cb}}{kT} \tag{7.3-10}$$

Note that I_S is the same for the two equations; I_S is a property of the base region, in common for the two junctions. Since $V_{cb} = V_{ce} - V_{be}$, the latter

equation can be written

$$\begin{aligned} I_r &= I_S \exp \frac{qV_{be}}{kT} \exp - \frac{qV_{ce}}{kT} \\ &= I_f \exp - \frac{qV_{ce}}{kT} \end{aligned} \quad (7.3-11)$$

$$\doteq I_c \exp - \frac{qV_{ce}}{kT} \quad (7.3-12)$$

which gives the reverse transport current in terms of the collector-emitter voltage of the intrinsic transistor—the voltage not including voltage drops in bulk resistances, notably the collector series resistance.

The effect of saturation on the *ABCD* parameters is found immediately from (7.3-12). Since the input generator is

$$I_b = \delta I_c = \delta_f I_f + \delta_r I_r \quad (7.3-13)$$

we have

$$-D = \delta_f + \delta_r \exp - \frac{qV_{ce}}{kT} \quad (7.3-14)$$

Hence the effect on *D* is to add the second term on the right (reversed in sign to agree with the output current sign convention). Although δ_r is large, the exponential term causes it to disappear for even a small positive V_{ce} . Saturation also affects *C* since the input current is a function of the collector voltage. We find *C* by taking the partial derivative of the base current with respect to V_{ce} :

$$C = - \frac{\delta_r q I_c}{kt} \exp - \frac{qV_{ce}}{kT}, \quad V_{ce} > 0 \quad (7.3-15)$$

Saturation is thus characterized by a sudden increase in the magnitude of both *C* and *D* and establishes a sharp limit on signal excursion at low collector voltages. The increase is due to a large amount of highly nonlinear feedback of both output voltage and current to the input current.

Avalanche Multiplication and Breakdown¹⁴

At sufficiently high electric fields in the junction depletion region, hole-electron pairs are generated by impact ionization. This occurs chiefly at the metallurgical junction, where the field is highest. When a hole-electric pair is generated, the field causes the hole to accelerate toward the base region. If it reaches the base, it contributes to the base current; similarly, the electron accelerates toward the collector, where it contributes to the collector current. As they

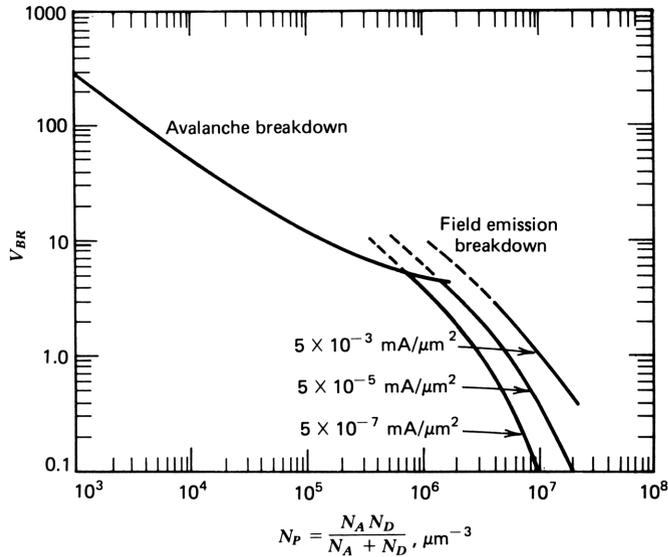


Figure 7.13. Avalanche breakdown voltage as a function of the doping densities on either side of the junction.¹⁴ For high doping levels, breakdown is by field emission or tunneling.¹⁹

move, they may collide with an atom in the lattice. If they have gained enough energy, they may loosen one of the bound electrons, creating another hole-electron pair, which also contributes to the flow of holes to the base and electrons to the collector. Avalanche breakdown occurs when the field becomes sufficiently high that the junction voltage no longer increases with the current through it. The breakdown voltage is a function of the doping level on either side of the junction; the relationship is shown in Fig. 7.13, which plots the breakdown voltage as a function of the composite doping level N_p , given by

$$N_p = \frac{N_A N_D}{N_A + N_D} \tag{7.3-16}$$

The breakdown voltage is given by the empirical equation

$$V_{BR} = \frac{\epsilon}{2qN_p} \left(\frac{40}{1 + \frac{1}{3} \log(10^4/N_p)} \right)^2 \tag{7.3-17}$$

At high doping concentrations, field emission or tunneling breakdown occurs before avalanche breakdown. This is discussed further in the paragraphs that follow.

At voltages much lower than breakdown, avalanche multiplication causes significant changes in transistor characteristics. Empirically, the process is described as a multiplication of the collector current by a factor M given by

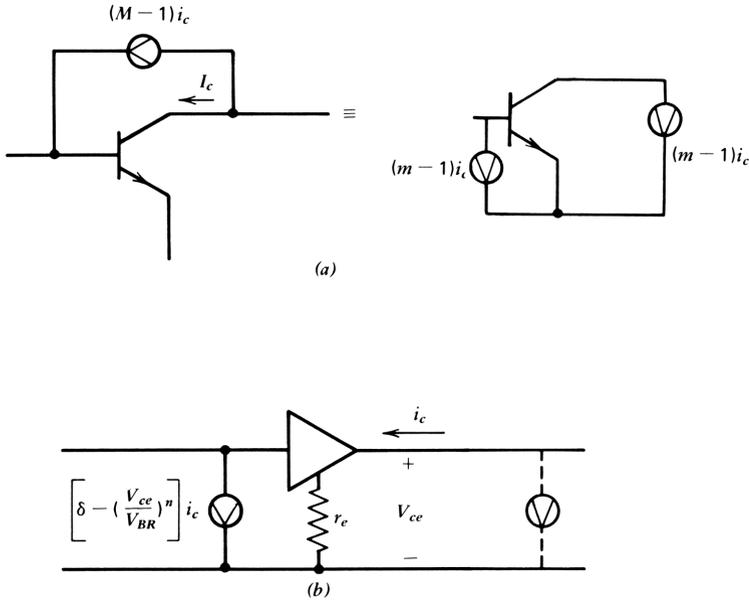


Figure 7.14. (a) Effect of avalanche multiplication represented as a current generator; (b) effect on equivalent circuit.

Miller’s formula¹⁵

$$M = \frac{1}{1 - (V_{ce}/V_{BR})^n} \quad 3 < n < 7 \quad (7.3-18)$$

Since the change in collector current is equal to the change in base current, the effect can be represented as shown in Fig. 7.14a, by a current generator between collector and base that has the value $(M-1)I_c$. The total current flowing into the collector is MI_c . The change in base current alters the operation of the transistor long before a significant change in collector current occurs, just as for saturation discussed previously. Therefore, the equivalent circuit in Fig. 6.4a is a good approximation; the change in base current due to avalanche multiplication is then given by

$$\Delta I_b = - \left(\frac{V_{ce}}{V_{BR}} \right)^n I_c \quad (7.3-19)$$

which is a function of both V_{ce} and I_c . As for saturation, both C and D are changed:

$$C = \frac{\partial I_b}{\partial V_{ce}} = - \frac{n I_c}{V_{BR}} \left(\frac{V_{ce}}{V_{BR}} \right)^{n-1} \quad (7.3-20)$$

and

$$D = -\frac{\partial I_b}{\partial I_c} = \delta_0 - \left(\frac{V_{ce}}{V_{BR}} \right)^n \quad (7.3-21)$$

where δ_0 is the current defect ratio in the absence of avalanche multiplication. Note that the base current vanishes when $\delta_0 = (V_{cc}/V_{BR})^n$; the voltage at which this occurs is called the *sustain voltage*, given by

$$V_{sus} = \delta_0^{1/n} V_{BR} \quad (7.3-22)$$

The effect of avalanche multiplication becomes important under high-voltage conditions. Its effect on circuit design is illustrated in the discussion of output stages in Chapter 8.

7.4 DYNAMIC PHYSICAL EFFECTS IN THE BIPOLAR TRANSISTOR

In this section we describe the physical effects in the bipolar transistor that affect accurate modeling of high-frequency performance. We are interested in the transit time and junction capacitances as well as the variation with collector current and voltage of these parameters. We can then develop an equivalent circuit suitable for a wide range of biases and frequencies, from dc to the vicinity of current gain cutoff.

Transit Time

Minority carriers traveling across the base region constitute a “block” of moving charge. The amount of charge involved can be calculated by integrating the charge density over the volume of the base region under the emitter. The charge density is given by the plot in Fig. 7.3c; the cross-sectional area of the emitter is A_e , and the base width is W_B . This block of charge must be equal to the product of the minority carrier current and the base transit time, so that we can write

$$\tau_B I_n = q A_e W_B \frac{n_p(0)}{2} \quad (7.4-1)$$

where $n_p(0)/2$ is the area under the plot in Fig. 7.3c and τ_B is the base transit time. Substituting the expression for I_n from (7.1-6) into this equation, we obtain the base transit time as

$$\tau_B = \frac{W_B^2}{2D_n} \quad (7.4-2)$$

This expression is correct only for the case of uniform base doping and low-level injection. In most transistors the carriers move under the influence of

an electric field as well as by diffusion. When this is taken into account, the minority carrier density departs from the straight-line relationship in Fig. 7.3c. This is usually accounted for by an additional factor η in the denominator, which has a value between 1 and 5:

$$\tau_B = \frac{W_B^2}{2\eta D_n} \quad (7.4-3)$$

It can be shown¹⁶ that η is given in terms of the acceptor concentrations at the emitter junction $N_{AB}(0)$ and at the collector junction $N_{AB}(W_B)$ by

$$\eta = \frac{\ln(k)^2}{2[(\ln k) - 1]} \quad (7.4-4)$$

where

$$k = \frac{N_{AB}(0)}{N_{AB}(W_B)} \quad (7.4-5)$$

Where $k=10$, for example, $\eta=2$, so that the transit time is halved by the presence of a field. With $N_{AB}(0)=10^5 \mu^{-3}$ and $N_{AB}(W_B)=10^4 \mu\text{m}^{-3}$ and an average diffusion constant of $2.0 \mu\text{m}^2/\text{ns}$, a base width of $0.3 \mu\text{m}$ would yield a base transit time of 0.011 ns .

In the collector depletion region the electrons move at the scattering limited velocity $v_x=100 \mu\text{m}/\text{ns}$, so that the transit time of carriers through this region is

$$\tau_x = \frac{W_{TC}}{v_x} \quad (7.4-6)$$

where W_{TC} is given by (7.3-3). The transit time is not the signal delay through this region. Unlike the base transit time where the signal delay and carrier delay are the same, the signal current at the collector is created by the *movement of the collector-side edge of the depletion region*. This movement begins at the instant that electrons enter the depletion region from the base. Electrons entering the collector depletion region induce a charge on the opposite side of the region. This takes place at the speed of light, virtually instantaneously, so that the delay through the collector depletion region is less than the delay of carriers. A step change in the minority carrier current entering at the base will cause a ramp of output collector current until the wave front of the step reaches the collector edge of the depletion region, as shown in Fig. 7.15. Therefore, the average delay of the carriers is only one-half the transit time. With a depletion width of $1.0 \mu\text{m}$, for example, the delay signal delay through the collector depletion region would be $1/(2 \times 100) = 0.005 \text{ ns}$. Adding this to the previously calculated 0.011 ns of base transit time gives

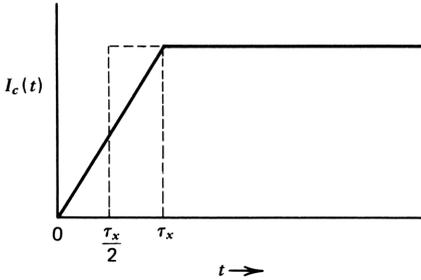


Figure 7.15. Collector current as a function of time for a step change in minority carriers from the base entering the collector depletion region.

a total transit time through the base region and collector depletion region of 0.016 ns.

Kirk Effect^{17, 18}

We have seen in the preceding paragraphs that the presence of minority carriers in the base acts to *increase* the effective base doping level (Webster effect). When these same minority carriers enter the collector depletion region, they become majority carriers and act to *reduce* the effective doping density there. When the minority carrier density entering the collector depletion region becomes greater than the collector doping level, this collector region is transformed into a part of the base region! In effect, the base region is *pushed out* toward the collector. Since this increases the base width, the base transit time is increased. Furthermore, as the effective doping density in the collector depletion region is reduced, the latter region widens. Transit time τ_x increases, at least until the depletion region widens into the buried layer. (The higher doping density in the buried layer effectively stops this movement.) Figure 7.16 illustrates both effects: a typical net doping profile is given for both zero collector current and for a collector current large enough to make the density of moving carriers in the collector depletion region equal to half the collector doping density.

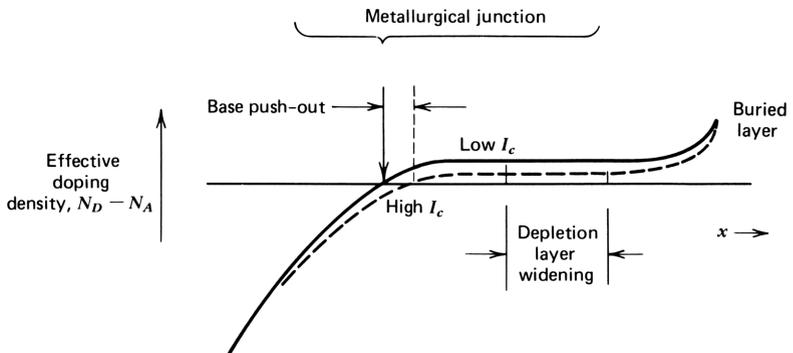


Figure 7.16. Change in effective doping density in the collector due to electron flow, illustrating Kirk effect.

The *Kirk current* is defined as the current that effectively cancels the collector doping. It can be estimated by assuming a carrier velocity equal to the scattering limited velocity in the collector depletion region, so that

$$I_{\text{Kirk}} = qA_e N_{DC} v_x \tag{7.4-7}$$

For $N_{DC} = 10^4 \mu\text{m}^{-3}$ and $A_e = 500 \mu\text{m}^2$, for example, $I_{\text{Kirk}} = 80 \text{ mA}$. The details of the increase of transit time with collector current depend on the shape of the doping density in the collector, but a good general rule is that τ_f increases according to the expression

$$\tau_f(I_c) = \tau_f(\text{low}) \left[1 + \left(\frac{I_c}{I_{\text{Kirk}}} \right)^2 \right] \tag{7.4-8}$$

for currents well below the Kirk current.

The transit time is also a function of collector voltage; as the collector voltage is increased; the base region becomes narrower and the depletion region widens. The net effect is to shorten the transit time. The effect can be modeled by the equation

$$\tau_f(V_{ce1}) = \tau_f(V_{ce2}) \left(\frac{V_{ce1}}{V_{ce2}} \right)^{\Gamma_\tau} \tag{7.4-9}$$

in which the constant Γ_τ is to be determined by measurements at different voltages. A typical variation of τ_f with voltage is shown in Fig. 7.17, for which Γ_τ is approximately 0.18.

Collector Junction Capacitance

The junction capacitances are given by the standard parallel plate capacitance formula, $C = \epsilon A/d$, where ϵ is the dielectric permittivity in silicon, $1 \times 10^{-4} \text{ pF}/\mu\text{m}$; A is the plate area; and d is the distance between the plates, in this

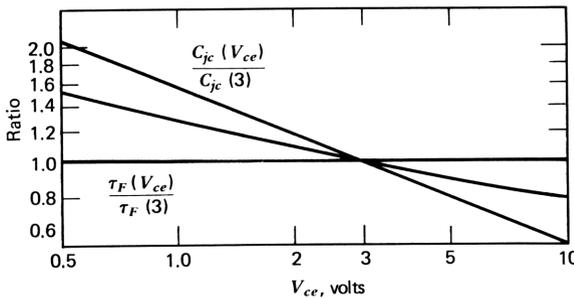


Figure 7.17. Typical variation of τ_f and collector capacitance with collector voltage.

case the width of the depletion region. Thus

$$C_{jc} = \frac{\epsilon A_e}{W_{TC}} \quad (7.4-10)$$

and

$$C_{je} = \frac{\epsilon A_e}{W_{TE}} \quad (7.4-11)$$

For the collector, the depletion region width is given by

$$W_{TC} = \left[\frac{2\epsilon(V_{cb} + \phi_{ic})}{qN_{PC}} \right]^{1/2} \quad (7.4-12)$$

$$\doteq \left(\frac{2\epsilon V_{ce}}{qN_{PC}} \right)^{1/2} \quad (7.4-13)$$

where N_{PC} is given by (7.3-16), and ϕ_{ic} is the collector “built-in” voltage. This is the equation for a step junction; a graded junction would have a variation with voltage to the one-third power. In practice, the variation of collector voltage is determined experimentally, and the collector capacitance variation with voltage is determined by

$$C_{jc}(V_{ce2}) = C_{jc}(V_{ce1}) \left(\frac{V_{ce1}}{V_{ce2}} \right)^{\Gamma_C} \quad (7.4-14)$$

where Γ_C is the collector grading coefficient, usually 0.33–0.5. Typical variation is shown in Fig. 7.17.

Emitter Junction Capacitance

The emitter junction is a step junction, for which the depletion width is

$$W_{TE} = \left[\frac{2\epsilon(\phi_{ie} - V_{be})}{qN_{AB}(0)} \right]^{1/2} \quad (7.4-15)$$

in which we have made the (very good) approximation that $N_{PE} = N_{AB}(0)$, where the latter is the base doping density at the emitter edge of the base region. The emitter is doped so heavily relative to the base that the approximation is valid. Combining (7.4-15) with (7.4-11), we have

$$C_{je} = A_e \left[\frac{\epsilon q N_{AB}(0)}{2(\phi_{ie} - V_{be})} \right]^{1/2} \quad (7.4-16)$$

Notice that C_{je} is insensitive to collector voltage but is a function of collector current through V_{be} .

Evaluation of (7.4-16) poses two problems: we must evaluate both $N_{AB}(0)$ and $\phi_{ie} - V_{be}$. We can estimate $N_{AB}(0)$ from the emitter breakdown voltage BV_{EB0} , shown in Fig. 7.13.¹⁹ At base doping densities above about $7 \times 10^5 \mu\text{m}^{-3}$, breakdown occurs by field emission. Transistors are usually designed to avoid this region since operation under field emission breakdown conditions tends to cause an irreversible increase in the defect current. In the avalanche breakdown region, the curve in Fig. 7.13 can be modeled by the equation

$$N_{AB}(0) = 3.2 \times 10^5 \left(\frac{7.0}{BV_{EB0}} \right)^{2.5} \quad (7.4-17)$$

The voltage $\phi_{ie} - V_{be}$ is obtained from the relation

$$\phi_{ie} - V_{be} = \frac{kT}{q} \ln \frac{N_{DE}}{n_p(0)} \quad (7.4-18)$$

where N_{DE} is the emitter doping density, about $2 \times 10^7 \mu\text{m}^{-3}$, and $n_p(0)$ is the minority carrier density at the emitter edge of the base region. We can find $n_p(0)$ from eq. (7.1-6), substituting I_c for I_n :

$$n_p(0) = \frac{I_c W_B}{A_e D_n} \quad (7.4-19)$$

for uniform base doping. Where the base doping is nonuniform, we can substitute $N_G/N_{AB}(0)$ for W_B . Thus

$$\phi_{ie} - V_{be} = \frac{kT}{q} \ln \frac{\Gamma_E A_e}{I_c} \quad (7.4-20)$$

where Γ_E is a function of the doping and geometry and is given by

$$\Gamma_E = \frac{q N_{DE} N_{AB}(0) D_n}{N_G} \quad (7.4-21)$$

With $D_n = 1.2 \mu\text{m}^2/\text{ns}$, $N_{DE} = 2 \times 10^7 \mu\text{m}^{-3}$, $N_{AB}(0) = 3 \times 10^5 \mu\text{m}^{-3}$, and $N_G = 3 \times 10^4 \mu\text{m}^{-2}$, $\Gamma_E = 40$.

Combining eqs. (7.4-16), (7.4-17), and (7.4-20), we obtain for C_{je}

$$C_{je} = 400 A_e \left[\frac{\epsilon q}{(kT/q) \ln(\Gamma_E A_e / I_c)} \right]^{1/2} \left(\frac{7.0}{BV_{EB0}} \right)^{1.25} \quad (7.4-22)$$

This equation expresses the dependence of C_{je} on the collector current (or more

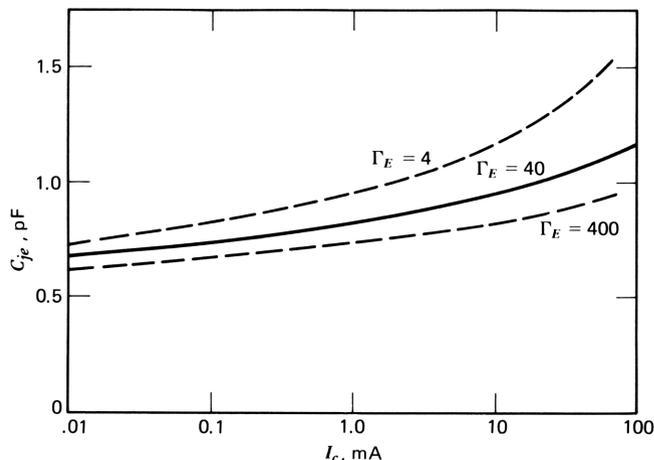


Figure 7.18. Emitter junction capacitance as a function of collector current calculated from (7.4-23).

accurately, on the current density) and can be written

$$C_{je} = 0.01 A_e \left(\ln \frac{\Gamma_E A_e}{I_c} \right)^{-1/2} \left(\frac{7.0}{BV_{EB0}} \right)^{1.25} \quad (7.4-23)$$

at room temperature. The variation of C_{je} with collector current for a transistor with $A_e = 250 \mu\text{m}^2$, $\Gamma_E = 40$ and $BV_{EB0} = 7.0 \text{ V}$, is shown in Fig. 7.18. The dashed curves are intended to illustrate the sensitivity of C_{je} to Γ_E and do not necessarily represent practical transistor designs. These curves show C_{je} when Γ_E is changed by an order of magnitude in either direction.

The term C_{je} is usually modeled as a function of V_{be} rather than I_c as used here. Although the V_{be} functional dependence is direct, it is inappropriate in the forward bias region. This is because the sensitivity of C_{je} to V_{be} is relatively high, giving rise to unnecessary inaccuracy in its estimation. It has often been noted that with enough forward bias, C_{je} becomes infinite. This occurs when $\phi_{ie} - V_{be}$ goes to zero; clearly, this can not occur until the injected minority carrier density is equal to the emitter doping density. At this point the collector current is far above the Kirk current (since $N_{DE} \gg N_{DC}$); thus the transistor is not operative. Attention to modeling this portion of the C_{je} versus V_{be} curve is therefore unnecessary. The difficulty is automatically avoided when C_{je} is expressed in terms of either charge densities or collector current rather than in terms of V_{be} .

Summary of Bias Dependencies

This completes the analysis of the elements that control high-frequency behavior in the transistor and their variation with bias. The variation of $r_e = kT/qI_c$ with current will also affect the performance at high frequencies and will dominate the variation of parameters with bias at low currents.

Table 7.1 Variation of Equivalent Circuit Parameters with Transistor Size and Collector BiasSize A_e

$$\frac{C_{je2}}{C_{je1}} = \frac{C_{jc2}}{C_{jc1}} = \frac{C_{ce2}}{C_{ce1}} = \frac{I_{Kirk2}}{I_{Kirk1}} = \frac{r_{b1}}{r_{b2}} = \frac{r'_{e1}}{r'_{e2}}$$

$$= \frac{A_{e2}}{A_{e1}}$$

Collector voltage V_{ce}

$$\frac{C_{ce2}}{C_{ce1}} = \frac{C_{jc2}}{C_{jc1}} = \left(\frac{V_{ce1}}{V_{ce2}} \right)^{\Gamma_C}$$

$$\frac{\tau_{F2}}{\tau_{F1}} = \left(\frac{V_{ce1}}{V_{ce2}} \right)^{\Gamma_\tau}$$

$$\frac{\delta_2}{\delta_1} = \frac{g_{ce2}}{g_{ce1}} = \frac{V_A + V_{ce1}}{V_A + V_{ce2}}$$

Collector current I_c

$$r_{e2} = \frac{kT}{qI_{c2}}$$

$$\frac{C_{je2}}{C_{je1}} = \sqrt{\frac{\ln(40A_{e2}/I_{c1})}{\ln(40A_{e1}/I_{c2})}} \quad (40 = \Gamma_E)$$

$$\frac{\tau_{F2}}{\tau_{F1}} = \frac{I_{c2}^2 + I_{Kirk2}^2}{I_{c1}^2 + I_{Kirk1}^2}$$

Unity loss time constant

$$\tau_{T2} = \tau_{F2} + r_{e2}C_{je2} + r_{E2}C_{jc2}$$

where

$$r_e = r_e + r_{e'}$$

The variations of the equivalent circuit element values with bias are summarized in Table 7.1. Also included in this table are the variations of these values as the horizontal geometry of the transistor is scaled; these variations are approximate but are useful in optimizing transistor size in an analog integrated circuit.

7.5 BROADBAND EQUIVALENT CIRCUIT

The purpose of this section is to obtain a set of equations for the $ABCD$ parameters in terms of the equivalent circuit parameters and also to derive the

equivalent circuit parameters in terms of the $ABCD$ parameters. Both descriptions are useful: the $ABCD$ description can be used directly in circuit design, but it must be determined under the bias conditions of the circuit. The collector voltage and current are directly related to the equivalent circuit parameters, as we saw in Section 7.4.

Rather than measure the transistor $ABCD$ parameters at each bias condition used in the circuit, we can evaluate the $ABCD$ parameters at one or two bias conditions and translate these $ABCD$ parameters to equivalent circuit parameters. We can then revise the equivalent circuit for the desired bias condition and calculate the $ABCD$ parameters under the new bias condition.

For analog integrated circuit design, the characteristics of the transistor must be accurately determined since mistakes are costly. This evaluation need only be done once, however, if an accurate equivalent circuit is developed along with an accurate evaluation of the bias dependencies defined in Section 7.4. The equations developed in this section for translating back and forth between the $ABCD$ and equivalent circuit parameters are programmed for the calculator in Appendix C.

A complete functional equivalent circuit that models the two-port parameters with good accuracy from dc to well beyond the unity loss frequency is given in Fig. 7.19a. Since we rarely need an equivalent circuit to such high frequencies, we derive from it the simpler circuit shown in Fig. 7.19b, good to about $\omega_T/3$. We use the more accurate circuit as a guide to developing the simpler circuit. Where analysis and design are limited to less than $\omega_T/10$, the delay terms in the generator of in Fig. 7.19b may be dropped, giving phase errors in the range of 2–6°.

To show transit time delay schematically, the circuit in Fig. 7.19a includes a delay line (represented as a small piece of coaxial cable) in the common lead of the ideal amplifier with an amount of delay, $\tau_B + \tau_x/2 = \tau'_F$. From the transit time argument in Section 5.3, this delay line causes a base current component to flow equal to $\tau'_F s i_c$ and also includes a delay equal to $\tau'_F/2$ if dispersion in the delay line is ignored. The emitter junction capacitance C_{je} is connected from base to emitter; the current through it is induced by the voltage across r_e , which is

$$v_{be} = r_e e^{\tau'_F s} i_c \quad (7.5-1)$$

The current through C_{je} adds to the base current.

Development of τ_i and τ_T

The total transit time from emitter to collector is increased by the time constants of two RC sections illustrated in the transistor common base equivalent circuit in Fig. 7.20. The first consists of the $r_e C_{je}$ section at the emitter, whose current loss is $1 + r_e C_{je} s$. Since the time constant is small, we can approximate it (as in Chapter 5) by the delay term $e^{r_e C_{je} s}$. Similarly, the bulk collector resistance between the depletion layer and the collector contact and the collector junction capacitance give a collector delay equal to $r_c C_{jc}$.

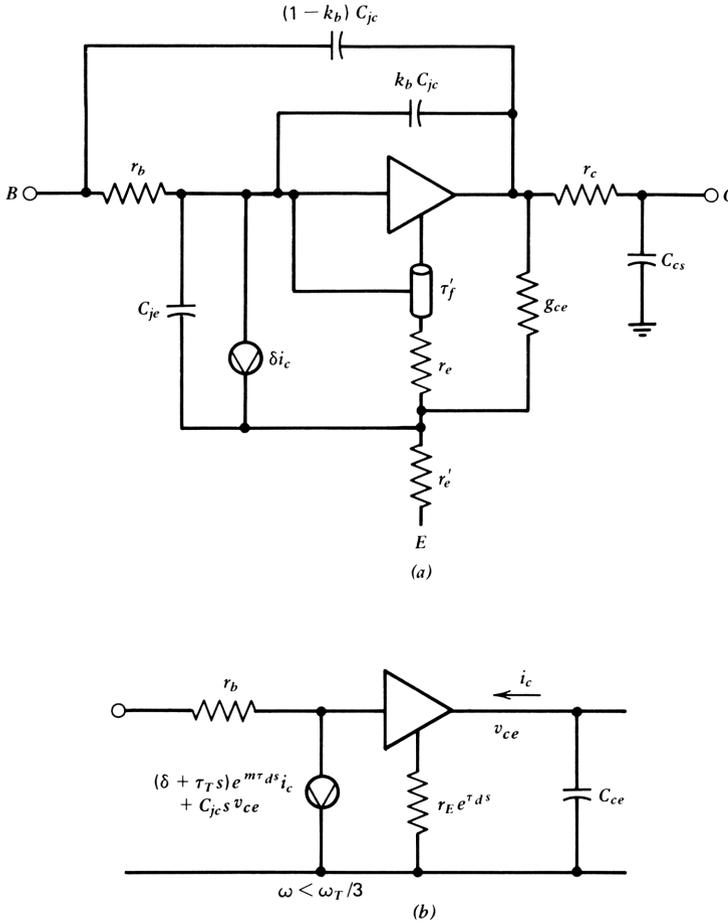


Figure 7.19. Broadband equivalent circuits for the bipolar transistor: (a) complete model, accurate to beyond f_T ; (b) simplified model, accurate to about $f_T/3$.

Transit time delay may be combined with the emitter time constant $r_e C_{je}$ and the collector time constant $r_c C_{jc}$ to give a total emitter-to-collector delay of

$$\tau_t = r_e C_{je} + \tau_B + \frac{\tau_x}{2} + r_c C_{jc} \tag{7.5-2}$$

where τ_F includes the collector time constant $r_c C_{jc}$.

When the transistor is connected in the common emitter configuration, the input loading term of C_{jc} is in parallel with C_{je} , giving an additional term proportional to frequency. When this is added to τ_t , we obtain the common emitter unity gain time constant τ_T , given by

$$\tau_T = \tau_t = r_E C_{jc} \tag{7.5-3}$$

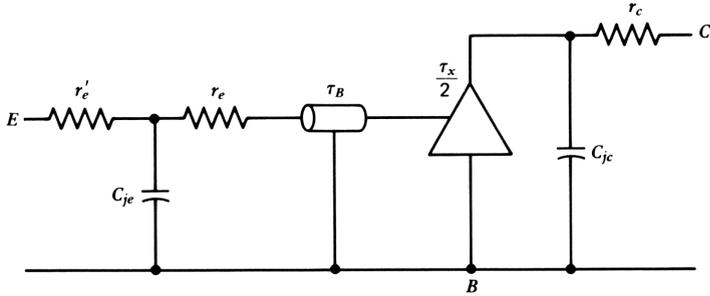


Figure 7.20. Development of τ_T as a cascade of delay terms.

The variation of τ_T with collector current is dominated at low collector currents by the emitter time constant and at high currents by Kirk effect. The variation of τ_T with collector current is shown in Fig. 7.21.

If we are interested in separating τ_F from τ_T , we can do so by measuring D or τ_T at two (or more) collector currents; we then form the products of τ_T and the current at which it is measured:

$$\begin{aligned} \tau_{T1} I_{C1} &= \tau_F I_{C1} + \frac{kT}{q} (C_{je} + C_{jc}) \\ \tau_{T2} I_{C2} &= \tau_F I_{C2} + \frac{kT}{q} (C_{je} + C_{jc}) \end{aligned} \tag{7.5-4}$$

Subtracting the two equations, we obtain

$$\tau_F = \frac{\tau_{T1} I_{C1} - \tau_{T2} I_{C2}}{I_{C1} - I_{C2}} \tag{7.5-5}$$

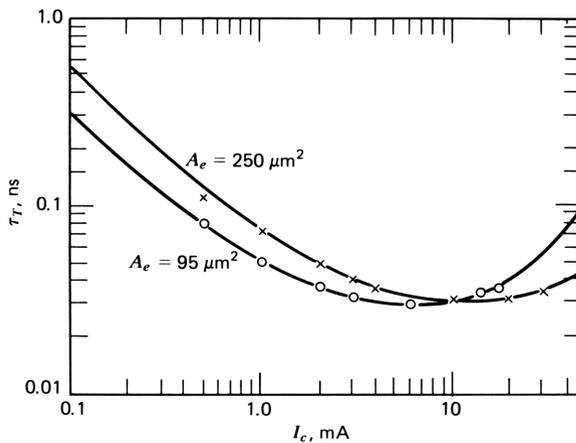


Figure 7.21. Variation of τ_T with collector current, showing the effects of emitter time constant at low currents and Kirk effect at high currents. The circles are measured points.

This equation omits the effects of high-level injection, so that the currents should not be too high. Alternatively, we can modify the equation to take this into account. It is not difficult to show that if I_{Kirk} is the Kirk current, then

$$\tau_F = \frac{\tau_{T1}I_{C1} - \tau_{T2}I_{C2}}{I_{C1}(1 + I_{C1}/I_{\text{Kirk}})^2 - I_{C2}(1 + I_{C2}/I_{\text{Kirk}})^2} \quad (7.5-6)$$

The reason for separating τ_F from τ_T is to be able to derive the *ABCD* parameters for use in circuit analysis at any set of bias conditions (except deep saturation).

Collector Capacitance as a Feedback Element

The collector capacitance C_{jc} is a feedback element topologically the same as the G_2 feedback analyzed in Section 3.2. The feedback gives a frequency-dependent input current that is dependent on the output voltage, drawn as a voltage-dependent current source connected across the input terminals. The direct feedthrough can be represented as a voltage-dependent current generator connected across the output; this generator puts a loop in a signal flow graph that might be drawn for the equivalent circuit.

By the same set of equations used in Section 3.2, replacing G_2 by $C_{jc}s$, direct feedthrough causes a denominator to appear:

$$\text{Denominator} = 1 - r_e C_{jc} s \quad (7.5-8)$$

This denominator has a magnitude close to unity to beyond the unity loss crossover. When it is retained, the equivalent circuit is valid to beyond f_T . At frequencies somewhat below this, it increases the phase linearly with frequency and can be approximated as a delay of all four two-port *ABCD* parameters, so that the denominator can be eliminated.

Bulk Resistances

The output loading places a capacitance of C_{jc} across the output. It can be lumped with other output parasitic capacitances, including C_{cs} , the collector-to-substrate capacitance, where the transistor is connected in the common emitter configuration.

The equivalent circuit includes bulk resistances in series with the three leads. In the collector this is the resistance between the collector depletion region and the collector contact (see Fig. 7.2) including the portion of the epitaxial region not swept out by the depletion region, the buried layer, and the connection from the buried layer to the collector contact. This resistance separates C_{jc} from C_{cs} so that they are not strictly in parallel; the effect is to add a delay $r_e C_{jc}$ to the transit time delay, as indicated earlier.

The resistance in series with the emitter lead is the emitter contact resistance r'_e , only of importance at high collector currents; it is about 1–2 Ω .

ABCD Parameters From the Equivalent Circuit

The complete *ABCD* parameters are found by premultiplying the matrix of the core parameters by the series r_b matrix and postmultiplying the result by the shunt $C'_{ce}s$ matrix. When this is done, we obtain the following expressions for the *ABCD* parameters, suitable for programming on the computer or calculator:

$$\begin{aligned}
 -A &= g_{ce}r_e + r_b(k_b C_{jc} + \delta C'_{ce})s + r_b \tau_T C'_{ce} s^2 \\
 -B &= (r_E + r_b \delta + r_i \tau_T s) e^{\tau_d s} \\
 -C &= \delta g_{ce} + g_{cb} + (C_{jc} + \delta C'_{ce})s + \tau_T C'_{ce} s^2 \\
 -D &= (\delta + \tau_T s) e^{m \tau_d s}
 \end{aligned} \tag{7.5-9}$$

Several terms (e.g., $g_{ce} \tau_T$, omitted from $-C$) have been omitted as negligible; the remaining terms give an accurate set of parameters at any bias and frequency up to $\omega_T/3$. Some of the terms of A and C should include delay, but this has been neglected. A new term, k_b , has been introduced in the linear term of A to account for the distributed nature of the base end of the collector capacitance. The collector capacitance appears across the entire length of r_b , so that it is incorrect to place it all inside the base resistance. This only affects A significantly in the frequency range of interest. The factor k_b , which is less than one, is the proportion of the total collector capacitance connected to the inner end of r_b in a two-lump model of C_{jc} . Factor k_b is necessary to avoid overestimating the value of A .

The equations for obtaining the *ABCD* parameters from the equivalent circuit, as well as the bias dependencies of the equivalent circuit elements, have been incorporated into program "E-A" in Appendix C. A program such as this, written for the calculator or computer, allows us to find the *ABCD* parameters under any bias conditions or frequency, suitable for calculating the properties of circuits incorporating transistors. The equivalent circuit element values form the starting point for finding the *ABCD* parameters in this program.

Equivalent Circuit Elements from Two-Port Parameters

If the *ABCD* parameters at a known set of biases and frequency are available, the equivalent circuit elements can be found by turning the equations around to derive the element values from the *ABCD* parameters. To separate τ_F from τ_T , D must be evaluated at two frequencies (or more); alternatively, τ_F can be estimated from known physical properties of the transistor. Similarly, if high current injection levels are to be used, the Kirk current for the transistor must be known, or measurements must be made at several high currents to model the Kirk current. If the collector doping density is known, the Kirk current can be estimated from eq. (7.4-7). Similarly, the voltage dependence factors must

be known or estimated. The sensitivities are low so that rough estimates are usually satisfactory.

Program “A-E” in Appendix C gives the equations used to obtain the equivalent circuit values from the $ABCD$ parameters. If a library of transistor characteristics as used in computer-aided design programs is available, the transistors may be analyzed by such programs to obtain the $ABCD$ parameters (from the defining equations for these parameters); this may be used as a starting point.

If such information is not available and a discrete transistor design is to be used, manufacturer’s specifications usually give typical values of at least the core parameters, $\tau_T=1/2\pi f_T$, δ , and C_{jc} ; r_e can be estimated from kT/qI_C . These are the sensitive parameters; errors in the other parameters will not greatly affect the design. The delay should be estimated in feedback applications at high frequencies. A suitable value of delay is $\tau_T/2$ at reasonably high currents; this increases as the current is decreased and is $r_e C_{jc}$ (from direct feedthrough) at low currents.

Measurements of the $ABCD$ parameters of a microwave integrated circuit transistor are shown in the plots in Fig. 7.22. The crosses on the figure are points calculated by using program “E-A” to find the $ABCD$ parameters from

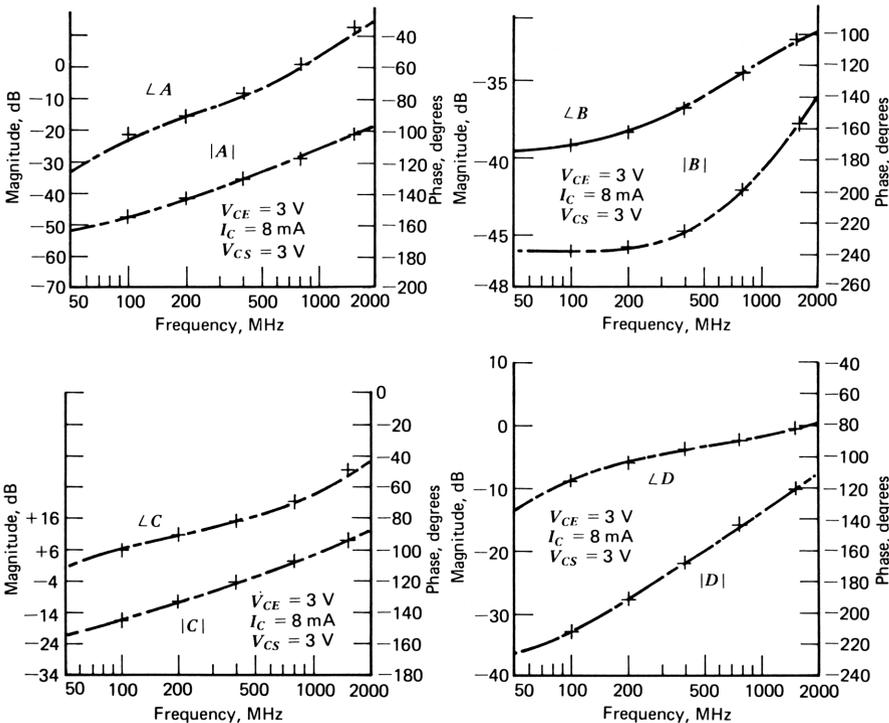


Figure 7.22. Measured $ABCD$ parameters on an integrated microwave transistor. Crosses represent values calculated from the equivalent circuit of Fig. 7.19b.

the equivalent circuit. The equivalent circuit parameters were obtained from the measurements by using program "A-E" and averaging the results of measurements at six frequencies. Appendix C gives the details.

The purpose of this chapter has been to show that the basic building block of active circuits can be described in terms of feedback and loading effects on a zero-loss forward path with delay. Each physical phenomenon taking place in the transistor, linear or nonlinear, can be described compactly and efficiently by its effect on one or more of the $ABCD$ parameters. The most important physical effects, those described by the core equivalent circuit, affect only one of the $ABCD$ parameters. All other effects—and there are many—can be evaluated and compared with the core parameters to assess their importance. With an $ABCD$ parameter description, the various effects are simply additive at the input of the device. The $ABCD$ parameters afford a way of cataloging the many effects in a readily understandable way.

REFERENCES

- 1 A. S. Grove, *Physics and Technology of Semiconductor Devices*, Wiley, New York, 1967.
- 2 R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, Wiley, New York, 1977.
- 3 H. R. Camenzind, *Electronic Integrated Systems Design*, Van Nostrand Reinhold, New York, 1972.
- 4 D. J. Hamilton and W. G. Howard, *Basic Integrated Circuit Engineering*, McGraw-Hill, New York, 1975.
- 5 A. B. Phillips, *Transistor Engineering*, McGraw-Hill, New York, 1962.
- 6 J. M. Early, "Structure-Determined Gain-Band Product of Junction Triode Transistors," *Proc. IRE* **46** (12), 1924, (December 1958).
- 7 P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Wiley, New York, 1977, pp. 73–78.
- 8 C. T. Sah, R. N. Noyce, and W. Shockley, "Carrier Generation and Recombination in P - N Junctions and P - N Junction Characteristics," *Proc. IRE* **45**, 1228 (1957).
- 9 W. M. Webster, "On the Variation of Current Gain with Emitter Current," *Proc. IRE* **42**, 914 (June 1954).
- 10 W. Shockley, "The Path to the Conception of the Junction Transistor," *IEEE Trans.* **ED-23** 597 (July 1976).
- 11 J. Logan, "Characterization and Modeling for Statistical Design," *BSTJ* **50** (4), 1105–1147 (April 1971).
- 12 S. J. Mitra, *Analysis and Synthesis of Linear Active Networks*, Wiley, New York, 1969.
- 13 G. S. Moschytz, *Linear Integrated Networks Fundamentals*, Van Nostrand Reinhold, New York, 1974.
- 14 S. M. Sze, *Physics of Semiconductor Devices*, Wiley-Interscience, New York, 1969.
- 15 S. L. Miller, "Avalanche Breakdown in Germanium," *Phys. Rev.* **99**, 1234 (1955).
- 16 J. G. Linvill and J. Gibbons, *Transistors and Active Circuits*, McGraw-Hill, New York, p. 120.
- 17 C. T. Kirk, "A Theory of Transistor Cutoff Frequency (f_T) Falloff at High Current Densities," *IRE Trans.* **ED-9** (2), 164, (March 1962).
- 18 R. J. Whittier and D. A. Tremer, "Current Gain and Cutoff Frequency Falloff at High Currents," *IEEE Trans.* **ED-16** (1), January 1969.
- 19 B. Hoeneisen and C. A. Mead, "Fundamental Limitations in Microelectronics—I. MOS Technology," *Solid State Electron.* **15**, 819 (Also see Appendix.) (1972).

Chapter 8

Two-Port Feedback Analysis

Selection of a circuit configuration capable of meeting a set of system requirements is perhaps the most subtle problem in circuit design. Once a satisfactory configuration is at hand, many methods exist for optimizing performance by use of the computer. Here and in Chapter 9 we develop methods by which the computer or calculator can aid in the choice of and in the appropriate modification of circuit configurations to meet these requirements.

There is a close relation between circuit structure and the polynomial coefficients of the $ABCD$ parameters, as we have seen for the bipolar transistor. The same is true of circuits that employ combinations of transistors. The first-degree coefficient of C in an inverting amplifier, for example, should bring to mind an equivalent capacitance connected from input to output, as for the collector capacitance of the transistor. When the characteristics of the entire network are expressed by its $ABCD$ parameters, we can find not only where in the circuit a given configuration is deficient, but what to do about it.

8.1 CLASSIFICATION OF FEEDBACK TYPES

Each element of the $ABCD$ matrix corresponds to one of four possible feedback types listed in Table 8.1. Each of these elements may be considered the feedback loss around an ideal amplifier. An example of this is the bipolar

Table 8.1 Classification of Feedback types

Feedback Type	Observe	Control	Controlling Parameter
<i>A</i> or SIPO	v_o	v_{in}	<i>A</i>
<i>B</i> or SISO	i_o	v_{in}	<i>B</i>
<i>C</i> or PIPO	v_o	i_{in}	<i>C</i>
<i>D</i> or PISO	i_o	i_{in}	<i>D</i>

transistor itself. For the core equivalent circuit in Fig. 7.1, the active path consists of the ideal collector junction with the null *ABCD* matrix representing it; the only nonzero elements are feedback elements. The equation $B = -r_e$ gives series input–series output (SISO) feedback, or more simply *B* feedback; $C = -C_{jcs}$ gives parallel input–parallel output (PIPO), or *C* feedback; $D = \delta + \tau_{Ts}$ gives parallel input series output (PISO), or *D* feedback. Series input–parallel output (SIPO), or *A* feedback, is virtually absent in the core equivalent circuit.

Operational amplifiers such as the 741 amplifier, on the other hand, are characterized almost solely by *A* feedback: As seen in Fig. 5.2, the feedback comes about as a combination of *B* feedback on the input stage and *C* feedback on the high-gain stage. In *ABCD* matrix notation, the (simplified) 741 amplifier can be represented as a cascade of a *B*-feedback amplifier and a *C*-feedback amplifier:

$$T_{741} = - \begin{bmatrix} 0 & 2r_E \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ C_{FS} & 0 \end{bmatrix} \quad (8.1-1)$$

$$= - \begin{bmatrix} 2r_E C_{FS} & 0 \\ 0 & 0 \end{bmatrix} \quad (8.1-2)$$

We have ignored the delay for simplicity. If we were to invert the order of the stages, with the differential pair at the output (a “gedanken experiment”), we would obtain a different result:

$$T_{147} = - \begin{bmatrix} 0 & 0 \\ C_{FS} & 0 \end{bmatrix} \begin{bmatrix} 0 & 2r_e \\ 0 & 0 \end{bmatrix} \quad (8.1-3)$$

$$= - \begin{bmatrix} 0 & 0 \\ 0 & 2r_e C_{FS} \end{bmatrix} \quad (8.1-4)$$

so that a *C*-feedback amplifier followed by a *B*-feedback amplifier yields a *D*-feedback amplifier.

Note that the tandem combination of two (ideal) *B*-feedback amplifiers (or two *C*-feedback amplifiers) yields the null matrix (an ideal amplifier).

(a)	(b)	(c)
<p>ABCD MATRIX OF TRANSISTOR</p> <p>FREQ., GHZ = 0.1000</p> <p>ABCD, MAG.+PH:</p> <p>A:</p> <p>R01= 0.0038 R02= -102.2474</p> <p>B:</p> <p>R03= 0.0052 R04= -170.4143</p> <p>C:</p> <p>R05= 0.1439 R06= -88.7339</p> <p>D:</p> <p>R07= 0.0225 R08= -116.3418</p>	<p>ABCD MATRIX OF TRANSISTOR</p> <p>FREQ., GHZ = 0.4000</p> <p>ABCD, MAG.+PH:</p> <p>A:</p> <p>R01= 0.0154 R02= -76.1781</p> <p>B:</p> <p>R03= 0.0061 R04= -145.3320</p> <p>C:</p> <p>R05= 0.5846 R06= -79.7403</p> <p>D:</p> <p>R07= 0.0011 R08= -94.9995</p>	<p>POLYS</p> <p>A-POLY:</p> <p>R00= -0.0011 R01= -0.0060 R02= -0.0008 R03= -8.1432-07</p> <p>B:</p> <p>R04= -0.0051 ◀ R05= -0.0014 R06= -2.0615-05 R07= -1.5796-07</p> <p>C</p> <p>R08= -0.0035 R09= -0.2290 ◀ R10= -0.0170 R11= -1.8214-05</p> <p>D</p> <p>R12= -0.0102 ◀◀ R13= -0.0322 ◀◀ R14= -0.0005 R15= -3.8380-06</p>

Figure 8.1. The *ABCD* matrices of a transistor at two frequencies and the corresponding *ABCD* polynomial coefficients.

A convenient classification system for amplifying networks is provided by the *ABCD* matrix. Whether the feedback is *unitary* (only one nonzero element as in the 741) or *hybrid* (more than one nonzero element as in the transistor), the network is characterized by the dominant element(s) in its *ABCD* matrix, as expressed by Table 8.1. Thus the 741 is a unitary feedback amplifier with *A* feedback, and the transistor is a hybrid feedback structure, containing (dominantly) *B*, *C*, and *D* feedback. In the latter case, which of these types of feedback are more important depends on the circuit in which the transistor is used. With a low admittance source and a low impedance load, for example, *D* will tend to dominate.

Several circuits are described here and in Chapter 9—circuits that may form parts of analog integrated circuits in common use. The properties of these circuits are perhaps most easily appreciated through a feedback description—that is, an *ABCD* description. To become familiar with numerical values, we illustrate circuit calculations using the transistor whose *ABCD* parameters were plotted in Fig. 7.21. Its *ABCD* parameters at $V_{ce} = 3V$, $I_c = 8$ mA are listed in

the calculator printout in Fig. 8.1; the listing gives the magnitude and phase of the $ABCD$ parameters at frequencies of 0.1 and 0.4 GHz.

Polynomial Coefficients of the $ABCD$ Parameters

These 16 numbers can be translated into four polynomial coefficients for each of the four $ABCD$ parameters by the method described in Section 2.5, implemented on the calculator by program "RCU" in Appendix A. Since we often have occasion to make this transformation for the four parameters together, in one operation, this program has been expanded to give this facility. Appendix C gives the modified program, "T>P3"; it is used as a subroutine in circuit analysis procedures to be described in this chapter.

The polynomial coefficients derived through use of the program are shown in Fig. 8.1c. As we expect, the core coefficients of the transistor are plainly in evidence. We might even recall the sizes of the numbers from Sections 7.4 and 7.5: referring to eq. (7.5-9), the dc coefficient of B , $b_0 = -0.0051$, is $-r_E$ in kilohms; the first-degree coefficient of C , $c_1 = -0.229$, is the negative of the collector capacitance in picofarads; similarly, $d_0 = -0.0102$ is $-\delta$ and $d_1 = -0.0322$ is $-\tau_T$. The negative signs give the phase reversal. Higher-degree coefficients represent the more complex effects described in Chapter 7, including excess phase effects. They are small since the transistor has a single high-frequency cutoff.

We have thus represented the transistor as a *hybrid feedback amplifier*, obtaining its $ABCD$ polynomial coefficients from the magnitude and phase of its $ABCD$ parameters at two frequencies. By combining the matrices of transistors in ways to be described, we can find the polynomial coefficients of circuits with many transistors. Before proceeding to this central concern of the chapter, we take a brief look at what we can do with these coefficients once we have them.

We restrict the examples to descriptions by cubic polynomials, although it should be recognized that by carrying along the matrix calculations at three frequencies, we can obtain quintic polynomial coefficients; seventh-degree coefficients would require magnitude and phase at four frequencies, and so forth. The extension of this method to finding the coefficients of the numerator and denominator of rational functions is given in Chapter 9.

All circuits described here are adequately represented by a cubic polynomial description. This serves to show the method without undue complexity in the calculations. When the method is realized on the computer (as opposed to the calculator used here), the additional burden imposed by going to higher-degree polynomials is not large.

8.2 NETWORK LOSS AND PORT IMMITTANCES

Given the $ABCD$ matrix of a network that has load conductance G_L connected to the output port, we can write

$$v_i = (A + BG_L)v_o \quad (8.2-1)$$

$$i_i = (C + DG_L)v_o \quad (8.2-2)$$

Since the Thevenin source voltage e_G is $v_i + R_G i_i$, we obtain the loss L_A as in eq. (6.2-11);

$$L_A = \frac{e_G}{v_o} = A + B G_L + R_G C + R_G D G_L \tag{8.2-3}$$

The *input admittance* Y_i is the ratio of (8.2-2) to (8.2-1):

$$Y_i = \frac{i_i}{v_i} = \frac{C + D G_L}{A + B G_L} \tag{8.2-4}$$

The *output impedance* is found by setting e_G to zero and exciting the network from the output. To do this, we must find the inverse of the $ABCD$ matrix:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \frac{1}{AD - BC} \begin{bmatrix} D & -B \\ -C & A \end{bmatrix} \tag{8.2-5}$$

This matrix relates the output voltage and current to the input signal variables with the signs of the currents shown in Fig. 8.2a; we take the positive direction of i_1 and i_2 into the network, so that the defining equation for the $ABCD$ matrix is written

$$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} v_2 \\ -i_2 \end{bmatrix} \tag{8.2-6}$$

When we take the inverse, we have

$$\begin{bmatrix} v_2 \\ -i_2 \end{bmatrix} = \frac{1}{AD - BC} \begin{bmatrix} D & -B \\ -C & A \end{bmatrix} \begin{bmatrix} v_1 \\ i_1 \end{bmatrix} \tag{8.2-7}$$

or

$$\begin{bmatrix} v_2 \\ i_2 \end{bmatrix} = \frac{1}{AD - BC} \begin{bmatrix} D & B \\ C & A \end{bmatrix} \begin{bmatrix} v_1 \\ -i_1 \end{bmatrix} \tag{8.2-8}$$

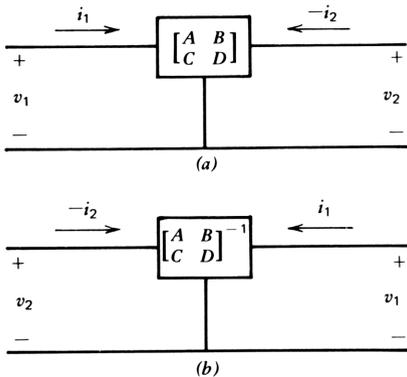


Figure 8.2. Sign convention for currents in a two-port for finding the inverse of the $ABCD$ matrix.

Which gives the proper sign relationships for the inverse $ABCD$ matrix in accordance with the $ABCD$ sign convention. We now can find the output impedance as the v_2/i_2 ratio; setting e_G to zero, and noting that $v_1 = -i_1 R_G$, we have

$$Z_o = \frac{B + DR_G}{A + CR_G} \quad (8.2-9)$$

Network Characteristics from the Polynomial Coefficients

We can use the polynomial coefficients in Fig. 8.1c to find rational expressions for the voltage loss, input admittance, and output impedance of the transistor for given source and load impedances. These operations are illustrated in Fig. 8.3a; eqs. (8.2-3), (8.2-4), and (8.2-9) are implemented in program "P3>LA" described in Appendix C. Figure 8.3a shows the polynomial coefficients of L_A and the numerator and denominator coefficients of Y_i and Z_o , using 0.075 K Ω source and load resistances. The program also finds the normalized polynomial coefficients as discussed in Section 2.2. For the single transistor, it is appropriate to scale the polynomial to its first-degree coefficient since the transistor has a single dominant cutoff. The cutoff frequency is calculated to be 0.173 GHz, and the dc loss is -22.0 dB, as shown in Fig. 8.3b. In the normalized polynomial (NP) listing the quadratic coefficient is virtually negligible, as is the cubic coefficient.

The polynomials give us the performance of the amplifier over the frequency range, as seen in Fig. 8.3c. These numbers are obtained by using program "FR", also described in Appendix C. The loss in Fig. 8.3c is that of a simple cutoff. The input admittance at dc from eq. (8.2-4) is effectively D/B since C and A are small at dc; hence $Y_i(0) = \delta/r_E = 0.0102/0.00511 = 2.0$ mS. At higher frequencies the imaginary part gives the input capacitance as 8.5 pF. The output impedance is more complicated; at dc it is $1/g_{ce}$ (the Early conductance); in parallel with this is the series combination of a capacitance of about 4.0 pF and a series resistance.

The analysis of the network is the beginning of the design process. Later in this chapter we see how the response and immittances can be tailored to arrive at desired values. To this end, we now find the relative contributions of A , BG_L , $R_G C$, and $R_G DG_L$ of (8.2-3) to the network loss L_A . In Fig. 8.4a these components are listed directly as calculated by program "NMR". At dc the loss is -0.0797, of which -0.0682 is contributed by B and -0.0102 is contributed by D , with the contributions of A and C virtually negligible as seen by comparing the dc coefficients.

At higher frequencies, inspection of the first-degree coefficients shows that D is the main contributor (through τ_T). In Fig. 8.4b the coefficients from Fig. 8.4a are divided by the coefficients of L_A of corresponding degree. This gives a list of *sensitivities* of each of the coefficients of L_A to its four contributors. This enables us to find the major contributors to loss and also those that may be

```

R-P3>LA
RG, RL:
R44= 0.0750
R45= 0.0750
OK?
RUN
LA POLY:
R20= -0.0797
R21= -0.0736
R22= -0.0028
R23= -8.1246-06
YI:
N
R24= -0.1395
R25= -0.6578
R26= -0.0237
R27= -0.0001
D
R28= -0.0693
R29= -0.0243
R30= -0.0010
R31= -2.9205-06
Z0:
N
R32= -0.0059
R33= -0.0038
R34= -0.0001
R35= -4.4581-07
D
R36= -0.0014
R37= -0.0231
R38= -0.0020
R39= -2.1804-06
NRM
bM=1,M=
R46= 1.0000
F0= 0.1725 ◀
a0
R20= -0.0797
-21.9664dB ◀
NP:
R40= 1.0000
R41= 1.0000
R42= 0.0414
R43= 0.0001
GTO M?
R-FR
2PI FMIN, MAX: INC.
R47= 0.0628
R48= 6.2832
R49= 2.1544
P3>S, SF06
OK?
(b)
LA
F L, dB PH
0.010 -21.95 -176.68
0.022 -21.90 -172.87
0.046 -21.69 -164.89
0.100 -20.80 -149.54
0.215 -18.10 -126.83
0.464 -13.09 -104.60
1.000 -6.72 -86.13
2.154 0.58 -65.97
YIN
F RE IM
0.010 2.025 0.552
0.022 2.065 1.188
0.046 2.248 2.543
0.100 3.069 5.322
0.215 6.356 10.115
0.464 15.225 13.901
1.000 25.263 9.973
2.154 28.138 2.526
Z0
F RE IM
0.010 2.104 -2.057
0.022 0.817 -1.526
0.046 0.309 -0.815
0.100 0.179 -0.396
0.215 0.149 -0.195
0.464 0.137 -0.112
1.000 0.115 -0.088
2.154 0.074 -0.074
(c)

```

Figure 8.3. Calculated network values for the transistor of Fig. 8.1 with 75Ω source and load: (a) polynomial coefficients of loss and rational functions giving the input admittance and output impedance; (b) scaled polynomial loss values; (c) loss and immittances as functions of frequency.

	CF 03	R-NMR
	SF 04	NORM. ABCD
	XEQ "NMR"	
NORM. ABCD		NRM. TO SPEC? SF 02
		OR TO VL? SF 03
		TO FO? CFLGS
		SF 03
	RUN	RUN
A		A
		R24= 0.0140
R24= -0.0011		R25= 0.0011
R25= -0.0060		R26= 0.2704
R26= -0.0008		R27= 0.1002
R27= -8.1432-07		
		B/RL
B/RL		R24= 0.8548
R24= -0.0682	◀	R25= 0.2486
R25= -0.0183		R26= 0.0978
R26= -0.0003		R27= 0.2592
R27= -2.1061-06		
		CRG
CRG		R24= 0.0033
R24= -0.0003		R25= 0.2334
R25= -0.0172		R26= 0.4551
R26= -0.0013		R27= 0.1601
R27= -1.3661-06		
		DRG/RL
DRG/RL		R24= 0.1279
R24= -0.0102	◀	R25= 0.4370
R25= -0.0322		R26= 0.1767
R26= -0.0005		R27= 0.4724
R27= -3.8380-06		
		TOTAL
TOTAL		R28= 1.0000
R28= -0.0797	◀	R29= 1.0000
R29= -0.0736		R30= 1.0000
R30= -0.0028		R31= 1.0000
R31= -8.1246-06		
		DONE

Figure 8.4. The *ABCD* polynomial coefficients normalized to the source and load immittances. In part *b* these values are divided by the loss polynomial coefficients to give sensitivities of loss to each *ABCD* polynomial coefficient.

neglected. Thus the sensitivity of dc loss to r_E is 0.85, or 85%; the sensitivity to δ is 0.13. The remaining terms add up to 0.02 and may be neglected. At higher frequencies the sensitivity of loss to τ_T is 0.44. All first-degree $ABCD$ coefficients contribute to the high-frequency loss, with A contributing least (0.08).

In Figs. 8.3 and 8.4 we have shown an example of the type of circuit calculation that relates performance of the network to the device characteristics. We have shown the calculation for a simple transistor stage, considering it as a hybrid feedback amplifier. More complex two-port circuits can be analyzed in the same way, as we see later.

Effect of Feedback on Port Immittances

The effect of the four types of unitary feedback on the input and output immittances is seen directly in eqs. (8.2-4) and (8.2-9). A feedback reduces both Y_i and Z_o ; when it acts alone, it drives both to zero. D feedback, on the other hand, drives both the input *impedance* and the output *admittance* toward zero. B feedback drives both input and output *admittances* toward zero, and C feedback does the same for input and output *impedances*. These statements express ideal relationships. Thus if an amplifier with low-input admittance and low-output impedance is desired, A feedback should be used. Such an amplifier is characterized by its voltage loss; its input admittance can be viewed as an input conductance in parallel with an input capacitance. In simple cases these will be constant with frequency, so that the description in terms of voltage loss and input admittance has direct physical meaning. The output impedance will (again in simple cases) be the series combination of a resistance and an inductance, so that again there is a simple physical picture of the output port.

Similar descriptions of the other three feedback types and their port immittances and input-output characteristics can also be made; they are summarized in Fig. 8.5a-d. We use the terms L_A and L_D to designate the voltage and current losses in A - and D -feedback amplifiers, respectively. The term Z_B is e_G/i_o and is the reciprocal of the transadmittance in the conventional formulation. Similarly, $Y_C = i_G/v_o$ is the reciprocal of the transimpedance in a C -feedback amplifier.

Expressions like that in eq. (8.2-3) can be written for each of the other three amplifier types. Thus for the B -feedback network, we relate the input voltage to the output *current*:

$$Z_B = AR_L + B + R_G CR_L + R_G D \quad (8.2-10)$$

$$= L_A R_L \quad (8.2-11)$$

For C and D feedback, we replace the Thevenin source by its Norton equivalent i_G, G_G ; we then have

$$Y_C = \frac{i_G}{v_o} = G_G L_A \quad (8.2-12)$$

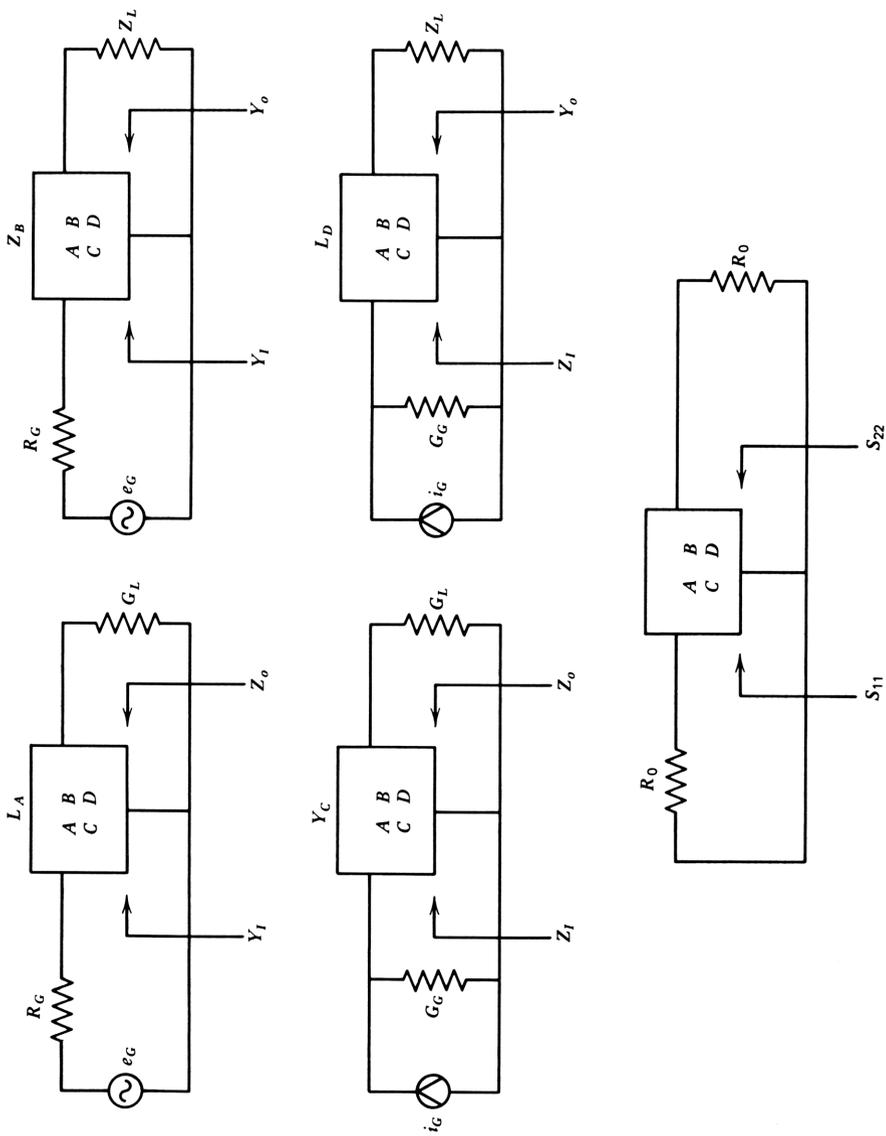


Figure 8.5. Four sets of network functions corresponding to the four feedback types. Network functions corresponding to the S parameters are shown in part e.

and

$$L_D = \frac{i_G}{i_o} = G_G L_A R_L \tag{8.2-13}$$

The port immittances in each case are given by eqs. (8.2-4) and (8.2-9) or their reciprocals.

The voltage and current losses L_A and L_D are dimensionless; the Z_B has dimensions in kilohms and Y_C has dimensions in millisiemens. We use the generic term “loss” for any of the four functions where a statement can be made about the four without distinguishing among them. Equations derived for L_A can be translated to Z_B , Y_C , or L_D through eqs. (8.2-11) to (8.2-13).

Figure 8.5e describes the S -parameter formulation of network loss; it is appropriate where A , B , C , and D are in proper relationship to drive S_{11} and S_{22} to zero, that is, to make the input and output impedances equal to R_o . This is discussed further in the paragraphs that follow.

Effect of Series and Shunt Port Immittances

The effect of a shunt admittance at the input of a network is easily found by premultiplying the $ABCD$ matrix of the network by that for the shunt admittance:

$$\begin{aligned} T_{YI} &= \begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \\ &= \begin{bmatrix} A & B \\ C+AY & D+BY \end{bmatrix} \end{aligned} \tag{8.2-14}$$

A conductance G added from base to emitter of a transistor, for example, will add Gr_E to D of the transistor. Figure 8.6 shows the results of adding immittances at the output or input of a network in all four possible ways.

We now have occasion to express the four operations depicted in Fig. 8.6 in a compact, operational matrix notation. We call the four operations $Z_i(T)$, $Z_o(T)$, $Y_i(T)$, and $Y_o(T)$ for adding an impedance to the input, an impedance to the output, an admittance to the input, and an admittance to the output, respectively. The argument of the function $Z_i(T)$ is the original $ABCD$ matrix. Thus we adopt the notation

$$Z_i(T) = \begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{8.2-15}$$

Also

$$Z_o(T) = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix} \tag{8.2-16}$$

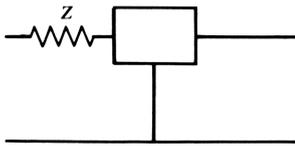
Note carefully that in the latter equation Z_o post multiplies the $ABCD$ matrix. Similarly

$$Y_i(T) = \begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad (8.2-17)$$

and

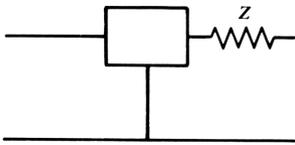
$$Y_o(T) = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix} \quad (8.2-18)$$

Since part of the process of designing networks will be to tailor $ABCD$ parameters to secure desired characteristics, it is well to be familiar with these relationships. We illustrate their use in Section 8.3.



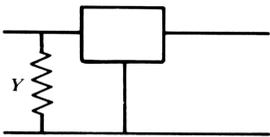
$$\begin{bmatrix} A+CZ & B+DZ \\ C & D \end{bmatrix}$$

(a)



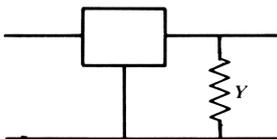
$$\begin{bmatrix} A & B+AZ \\ C & D+CZ \end{bmatrix}$$

(b)



$$\begin{bmatrix} A & B \\ C+AY & D+BY \end{bmatrix}$$

(c)



$$\begin{bmatrix} A+BY & B \\ C+DY & D \end{bmatrix}$$

(d)

Figure 8.6. Effects of series and shunt port immittances.

8.3 FEEDBACK TYPES AND THEIR COMBINATIONS: CONTROL OF IMMITANCES

An ideal A -feedback amplifier can be represented as shown in Fig. 8.7a, in which an ideal transformer with turns ratio n_A adds a voltage $n_A v_o$ to the ideal amplifier input voltage (zero for the ideal amplifier). An equivalent circuit for the ideal transformer is shown in Fig. 8.7b, which is recognizable from Fig. 6.4 as the h -parameter equivalent circuit of the transformer taken as a two-port. Since the transformer is lossless, h_{11} and h_{22} are zero. Since the input current to the ideal amplifier is zero, the input current to the amplifier with feedback is also zero. The input voltage consists entirely of the feedback voltage $-n_A v_o = -h_{12} v_o$ and is not a function of the output current. Hence, the transmission matrix has only one nonzero element:

$$\begin{bmatrix} v_i \\ i_i \end{bmatrix} = - \begin{bmatrix} n_A & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_o \\ i_o \end{bmatrix} \tag{8.3-1}$$

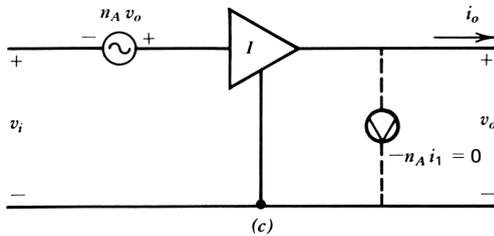
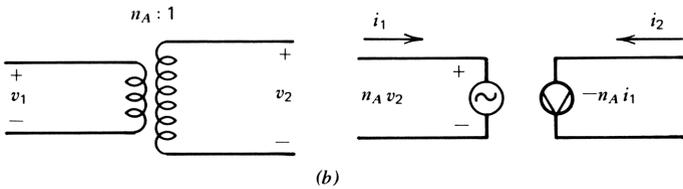
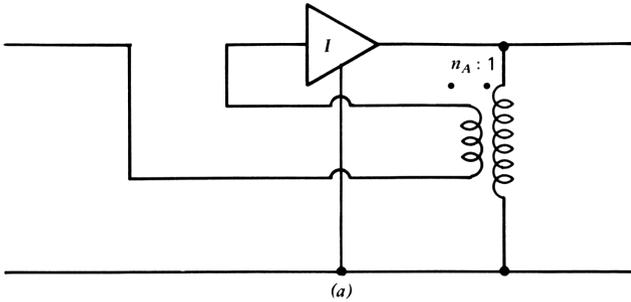


Figure 8.7. Ideal A feedback.

so that from eqs. (8.2-4) and (8.2-9), Y_i and Z_i are both zero. The feedforward current generator at the output is null because the input current to the ideal amplifier is zero; the transformer current is therefore constrained to be zero.

A Lossless Hybrid Feedback Amplifier (Symmetric)

Suppose now that we add a second transformer as shown in Fig. 8.8 that has a secondary winding in series with the output so that it senses the output current; its primary winding augments the input current by $n_D i_o$. The equivalent circuit generators for this transformer are added to the circuit as shown in Fig. 8.8b. Now, both the input current and the voltage are augmented so that the transmission matrix for the circuit becomes

$$\begin{bmatrix} v_i \\ i_i \end{bmatrix} = - \begin{bmatrix} n_A & 0 \\ 0 & n_D \end{bmatrix} \begin{bmatrix} v_o \\ i_o \end{bmatrix} \tag{8.3-2}$$

Note that the feedforward voltage across the secondary of the n_D transformer is not zero; it is $n_D n_A v_o$. It faces the output of an ideal amplifier, however, so there is no effect on circuit operation when the amplifier is ideal.

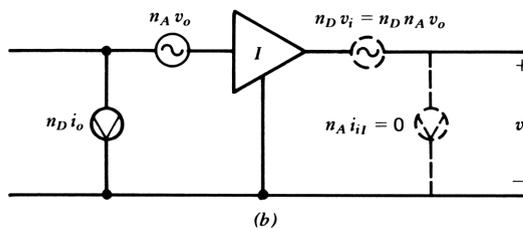
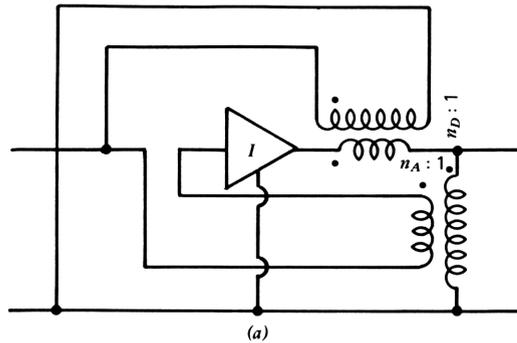


Figure 8.8. Combined ideal A and D feedback.

The input admittance of this circuit is given by (8.2-4) as

$$Y_i = \frac{n_D}{n_A} G_L \quad (8.3-3)$$

and the output impedance, from (8.2-9), is

$$Z_o = \frac{n_D}{n_A} R_G \quad (8.3-4)$$

Where $n_D = n_A$, the input impedance is equal to the load impedance and the output impedance is equal to the source impedance. This circuit, with transistor(s) replacing the ideal amplifier and using nonideal transformers, is of practical significance in providing a resistive termination at the amplifier input without adding the thermal noise of a resistor. We explore this in more detail later. Both A and D feedback are termed *symmetric feedback*.

This brings us to the fifth diagram in Fig. 8.5. If we load the output of the hybrid feedback amplifier that has $n_D = n_A$ with R_0 , its input impedance will likewise be R_0 ; insofar as the amplifier is ideal, $S_{11} = S_{22} = 0$, and the S -parameter description of the network becomes appropriate. The *insertion loss* $1/S_{21}$ is given by

$$\frac{1}{S_{21}} = \frac{1}{2} [A + B/R_0 + R_0C + D] \quad (8.3-5)$$

as seen from Table 6.1. The value of $1/S_{21}$ is thus one-half L_A (or L_D) when $R_G = R_0$ and $G_L = 1/R_0$. The factor of 2 arises because $1/S_{21}$ compares the loss when the amplifier is in place to the loss when the amplifier is removed and the source is connected directly to the load, in which case $e_G/v_o = 2$.

Where S_{11} and S_{22} depart from zero, reflections appear at input and output. We can characterize the reflection coefficients by use of Table 6.2 which gives

$$S_{11} = \frac{A + B/R_0 - R_0C - D}{A + B/R_0 + R_0C + D} \quad (8.3-6)$$

and

$$S_{22} = \frac{-A + B/R_0 - R_0C + D}{A + B/R_0 + R_0C + D} \quad (8.3-7)$$

Clearly, with $A = D$ and $B = C = 0$, both S_{11} and S_{22} must be zero. We can also see that if we make $B/R_0 = R_0C$ nonzero, we can still retain zero reflection coefficients.

An Antisymmetric Hybrid Feedback Amplifier

To understand this case, consider the ideal amplifier with B feedback shown in Fig. 8.9a. Its input voltage is $-i_o R$; there is no input current, so its transmission matrix is

$$T_B = - \begin{bmatrix} 0 & R \\ 0 & 0 \end{bmatrix} \tag{8.3-8}$$

Similarly, the C -feedback amplifier in Fig. 8.9b with $R=0$ has the transmission matrix

$$T_C = - \begin{bmatrix} 0 & 0 \\ G & 0 \end{bmatrix} \tag{8.3-9}$$

What happens when both types of feedback are applied simultaneously? We develop a simple method of calculating such cases in the following paragraphs. For the present, we use the method developed to obtain eqs. (3.2-7) and (3.2-10) in Section 3.2 for the circuit of Fig. 3.3. An equivalent ladder circuit was given in Fig. 3.4. If we set $\tau_3=0$ there and replace r_3 by R , G_2 by G , and $G_L v_o$ by i_o , we obtain

$$T_{BC} = \frac{-1}{1-RG} \begin{bmatrix} RG & R \\ G & RG \end{bmatrix} \tag{8.3-10}$$

If we make $RG=1$, the loss of this circuit goes to infinity; it becomes a

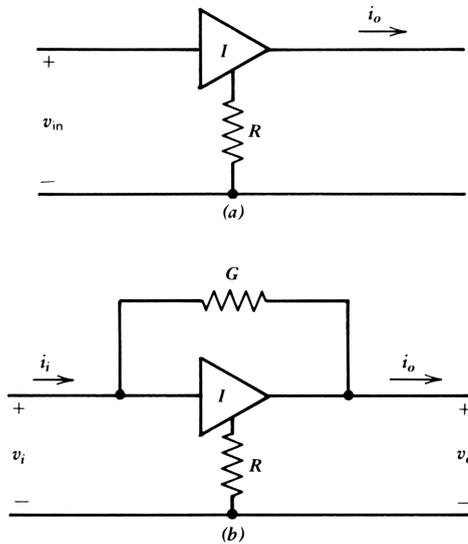


Figure 8.9. Antisymmetric feedback: (a) B feedback; (b) combined B and C feedback.

balanced bridge. If $RG \ll 1$, however, it is the *antisymmetric* version of the circuit in Fig. 8.8a. (We could make the analogy exact by use of lossless gyrators to provide B and C feedback.) In any case, if we make R and G *complementary*, that is $R/G = R_0^2$, we obtain zero values for S_{11} and S_{12} as shown by (8.3-6) and (8.3-7).

If we ignore the RG products in (8.3-10), we obtain the sum of the matrices of (8.3-8) and (8.3-9);

$$T_{BC} = - \begin{bmatrix} 0 & R \\ G & 0 \end{bmatrix} \quad (8.3-11)$$

whereupon the input admittance is, from (8.2-4)

$$Y_i = \frac{G}{R} \frac{1}{G_L} = \frac{1}{R_0^2 G_L} \quad (8.3-12)$$

and the output impedance is, from (8.2-9)

$$Z_o = \frac{R}{G} \frac{1}{R_G} = \frac{R_0^2}{R_G} \quad (8.3-13)$$

Thus a higher source impedance gives a lower output impedance, and similarly for the input admittance. The antisymmetrical hybrid feedback amplifier tends to *gyrate* the impedance at the opposite port, whereas the symmetrical hybrid feedback amplifier tends to *transform* it.

A Unilateral Feedback Amplifier

These opposing tendencies can be canceled by making $AD = BC$. When this is done, the determinant of the $ABCD$ matrix vanishes and the inverse goes to infinity. From eq. (8.2-8) the reverse transmission vanishes, and the amplifier becomes *unilateral*. In this important special case, the input admittance is independent of the load and the output impedance is independent of the source. As an example, suppose that

$$T_{ABCD} = - \begin{bmatrix} 0.1 & 0.01 \\ 1.0 & 0.1 \end{bmatrix} \quad (8.3-14)$$

Then $AD - BC = 0$, and for any G_L , the numerator and denominator factors containing G_L cancel, so that

$$Y_i = \frac{1 + 0.1G_L}{0.1 + 0.01G_L} = 10 \text{ mS} \quad (8.3-15)$$

and similarly for any R_G ,

$$Z_o = \frac{0.01 + 0.1R_G}{0.1 + 1.0R_G} = 0.1 \text{ K}\Omega \quad (8.3-16)$$

Separate control of each of the $ABCD$ parameters can be achieved by combining the feedback paths in Figs. 8.8(a) and 8.9(b), as shown in Fig. 8.10. Note that of the four *feedforward* paths, only one, the Gv_i generator, is effective in modifying the transmission matrix where the amplifier is ideal. There is no input current to the ideal amplifier, so that the n_A transformer current is constrained to be zero. The feedforward voltages Ri_i and $n_D v_i$ are in series with the output of an ideal amplifier and hence have no effect. The C -feedback conductance G , however, is connected to the input where there is a nonzero voltage; this creates the only effective feedforward current, $v_i G$.

An equivalent circuit that takes into account only nonzero effects in this circuit (with ideal amplifier) is shown in Fig. 8.10b. An exact analysis of this network's transmission matrix would show that

$$T_{ABCD} = -\frac{1}{1-RG} \begin{bmatrix} n_A + RG & R \\ G(1+n_A+n_D-\Delta_A) & n_D + RG \end{bmatrix} \quad (8.3-17)$$

This matrix is the sum of those of eqs. (8.3-2) and (8.3-10), except for the term

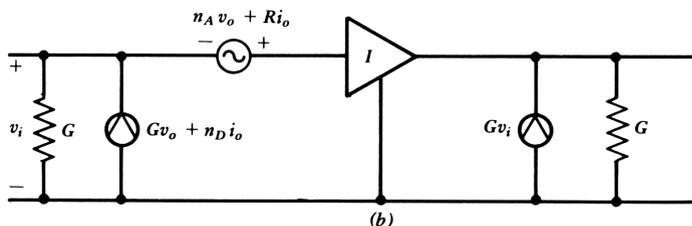
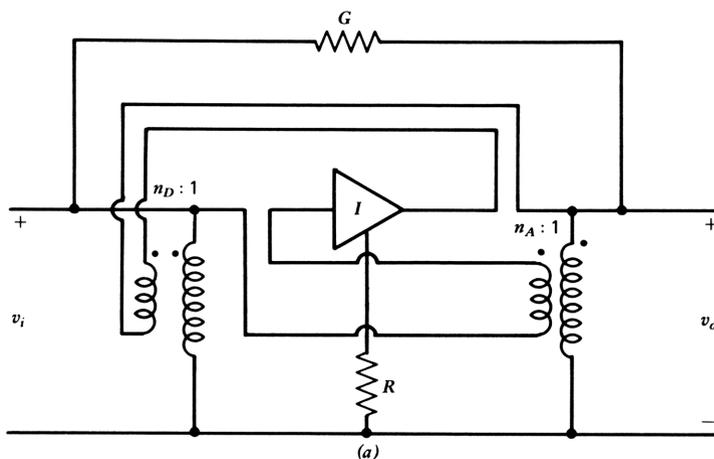


Figure 8.10. Application of all four feedback types to an ideal amplifier.

multiplying G in C . It is

$$\Theta = 1 - A - D + \Delta_A \approx 1 \tag{8.3-18}$$

and arises from the nonzero feedforward current through G . This circuit shows in principle that any combination of the four feedback types can be realized physically. The transformers make this a simple matter since we can arrange the polarities of the A - and D -feedback circuits arbitrarily. When the constraint is introduced that no transformers be used, the problem of combining symmetric and antisymmetric feedback types becomes more difficult.

The reason is that pure A or D feedback types tend to have no phase reversal, whereas B and C feedback generally include a phase reversal. Four generic feedback types are shown in Fig. 8.11 without transformers. In the

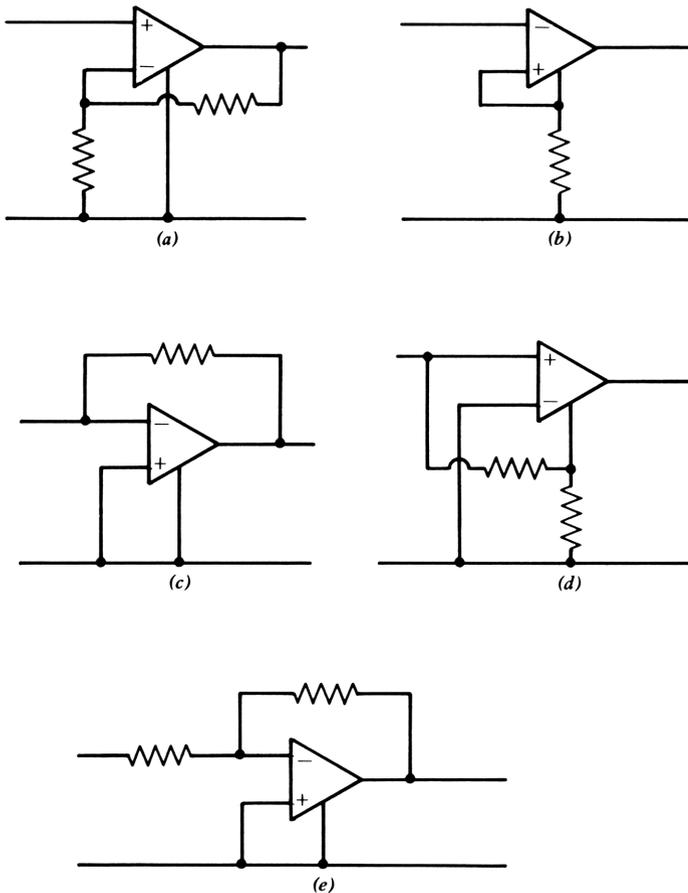


Figure 8.11. Four generic feedback types; symmetric types have no phase reversal. A C -feedback amplifier with series input resistance is shown in part e .

symmetric types, separate positive and negative inputs, as in an operational amplifier, are required and no phase reversal is included. In Fig. 8.11e an A -feedback amplifier is shown with a phase reversal, but the transmission matrix also includes a nonzero $C = -G$. This is basically a C -feedback amplifier to which a series input resistance has been added and must be counted a hybrid feedback amplifier.

Example: A Single-Stage Hybrid Feedback Amplifier. The bipolar transistor itself is a hybrid feedback amplifier, as noted previously. Suppose that we wish to provide an amplifier stage with a simple cutoff and with both input and output impedances matched to 75Ω source and load. Can we modify or augment the internal feedback processes in the transistor to achieve this result? Design freedom is limited in a single stage, but it is instructive to see what can be done to control the impedances in this case. It will also make the advantages of multistage circuits clearer in a precise way.

We wish the input and output impedances to be equal, so we know that A should be equal to D . With the impedances both $0.075 \text{ k}\Omega$, we know that the ratio B/C should be $(0.075)^2$. Furthermore, this should be true for the $ABCD$ polynomial coefficients of each degree individually. Inspection of the polynomial coefficients in Fig. 8.1c shows that a_0 is an order of magnitude smaller than d_0 and that a_1 is only one-fifth d_1 . Both a_0 and a_1 must be increased. The value of $\sqrt{b_0/c_0}$ is $1.21 \text{ k}\Omega$; hence c_0 must be increased greatly. But $\sqrt{b_1/c_1}$ is $0.077 \text{ k}\Omega$, almost exactly right. To increase c_0 , we add a feedback resistor from collector to base as shown in Fig. 8.12, increasing the magnitude of c_0 by about G_F . Hence since $b_0/(c_0 + G_F)$ must equal R_0^2 , we obtain

$$\begin{aligned} G_F &= \frac{b_0}{(0.075)^2} - c_0 \\ &= 0.903 \end{aligned}$$

This brings c_0 into approximate balance with b_0 ; c_1 already balances b_1 .

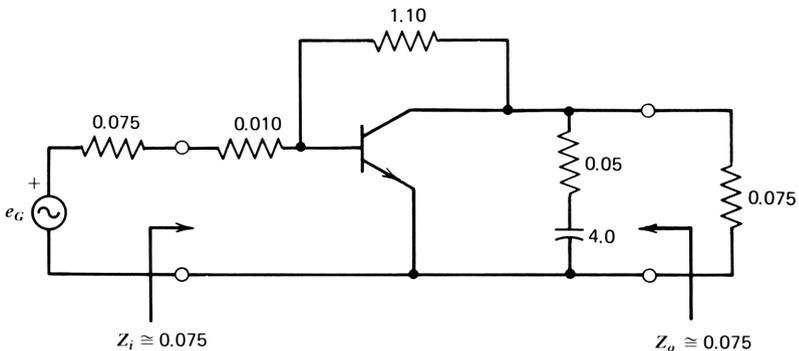


Figure 8.12. A single-stage hybrid feedback amplifier.

To increase the magnitude of a_0 , we could add a shunt conductance to the output as in Fig. 8.6*d* or add a series resistance to the input, as in Fig. 8.6*a*. Choosing the latter, we have

$$-R_B G_F \simeq d_0 - a_0$$

or

$$\begin{aligned} R_B &= (0.0102 - 0.0011) / 0.903 \\ &= 0.010 \text{ k}\Omega \end{aligned}$$

To increase a_1 by a series impedance at the input, we should add an inductance. Instead, we add a shunt capacitance C_o at the output. Its value is given approximately by

$$c_0 b_0 = d_1 - a_1$$

so that

$$C_o = \frac{0.0322 - 0.0060}{0.00511} = 5.1 \text{ pF}$$

In these equations we have ignored interaction terms; the series input resistor affects B (slightly) and the output capacitance affects C . In particular, the output capacitor increases c_2 . To reduce this effect, a resistor is connected in series with C_o , and the value of C_o is reduced to 4.0 pF, as in Fig. 8.12. (This was done in a second iteration.)

The analysis of the resulting circuit is shown in Fig. 8.13. In Fig. 8.13*a* the $ABCD$ parameters of the circuit in Fig. 8.12 are shown (the means for calculating these for this and other circuits are discussed in Section 8.1 and Appendix C). The loss polynomial of the circuit is also given; in Fig. 8.13*b* the cutoff frequency is 0.24 GHz and the dc loss is -15.4 dB. The variation of loss Y_i and Z_o are given as a function of frequency; the cutoff is simple and the impedances are roughly 0.075 k Ω up to the cutoff frequency.

Departures from matched impedances can be analyzed by use of the list of coefficients normalized to the loss polynomial coefficients as shown in Fig. 8.13*c*. This is a list of sensitivities of loss to the polynomial coefficients of the $ABCD$ parameters and is useful in discovering why the impedances depart from the desired values. Each coefficient of A (or B) should equal the corresponding coefficient of D (or C).

In Fig. 8.13*c*, for example, $a_0 = 0.0880$ and $d_0 = 0.0882$, giving close agreement. Similarly, $b_0 = 0.4166$ and $c_0 = 0.4072$, also good agreement. Thus we were successful in making $a_0 = d_0$ and almost successful in making $b_0 = R_0^2 c_0$. For the first-degree coefficients, c_1 is a bit too large. Since c_1 is primarily caused by the collector capacitance, it cannot be reduced easily except by raising the dc collector voltage. Also, a_1 is less than d_1 , indicating that C_o

POLYS	F0= 0.2402	NORM. ABCD
A-POLY:	a0	A
R00= -0.0149	R20= -0.1692	R24= 0.0000
R01= -0.0311	-15.4338dB	R25= 0.2773
R02= -0.0032	LH	R26= 0.2466
R03= 0.0003		R27= 0.2004
B	F L, dB PH	B/RL
R04= -0.0053	0.010 -15.43 -177.62	R24= 0.4166
R05= -0.0017	0.022 -15.41 -174.87	R25= 0.2047
R06= -2.8133-05	0.046 -15.33 -168.98	R26= 0.0291
R07= -2.6705-07	0.100 -14.96 -156.65	R27= -0.0022
	0.215 -13.42 -133.06	
C	0.464 -8.59 -99.23	CRG
	1.000 1.27 -72.95	
R08= -0.9185	2.154 15.50 -68.42	R24= 0.4072
R09= -0.3265	YIN	R25= 0.2185
R10= -0.1169	F RE IM	R26= 0.6000
R11= 0.0174	0.010 13.008 0.050	R27= 0.0050
D	0.022 13.002 0.100	DRG/RL
	0.046 13.056 0.240	
R12= -0.0149	0.100 12.963 0.592	R24= 0.0002
R13= -0.0336	0.215 12.907 1.736	R25= 0.2996
R14= -0.0006	0.464 14.724 4.547	R26= 0.0435
R15= -5.3347-06	1.000 22.659 6.003	R27= -0.0033
	2.154 35.386 1.000	
LA POLY:	Z0	TOTAL
	F RE IM	
R20= -0.1692	0.010 0.076 0.000	R28= 1.0000
R21= -0.1121	0.022 0.077 0.000	R29= 1.0000
R22= -0.0129	0.046 0.077 0.000	R30= 1.0000
R23= 0.0016	0.100 0.080 -0.002	R31= 1.0000
	0.215 0.083 -0.014	
	0.464 0.058 -0.034	
	1.000 0.026 -0.010	
	2.154 0.010 -0.003	
(a)	(b)	(c)

Figure 8.13. Calculated values for the amplifier shown in Fig. 8.12: (a) *ABCD* parameters; (b) loss and immittances as functions of frequency; (c) sensitivities of loss to individual *ABCD* parameter coefficients.

should be raised. We do not raise it because of its effect on c_2 . As seen in Fig. 8.13a, c_2 dominates the quadratic loss coefficient, tending to increase the input admittance and to reduce the output impedance at high frequencies. Raising C_o also reduces the bandwidth of the stage; thus a balance must be sought between loss and impedance performance. The sensitivity list helps to achieve such a balance in the design process.

That the cubic coefficients of B and D change sign is caused by the denominator—the *pole of loss* arising from the resistor in series with C_o . The purpose of the resistor is to reduce the quadratic coefficient, which it does. Since the cubic coefficient of A and C is negligible, when the phase of the denominator is taken into account, the equivalent polynomial coefficient is negative. In the frequency range of interest, the phase introduced by the denominator is equivalent to that of a negative cubic polynomial coefficient. The negative coefficient comes about by attempting to model the rational function—including a pole of loss—by a polynomial. Although close agreement is obtained in the frequency range of interest in this case, an extension of the modeling process to include denominators is often helpful and sometimes essential. This extension is given in Section 9.2.

In this single transistor design the departure from desired performance may be larger than desired. The purpose of the exercise is to show how the $ABCD$ parameter polynomial coefficients can be manipulated to force the performance toward the desired characteristics. Performance can be improved by using a greater number of transistors. To do this easily, we need the means for finding the $ABCD$ parameters of combinations of transistors.

8.4 GENERAL METHOD OF TWO-PORT ANALYSIS¹

In this section we develop a method of two-port analysis that is applicable in general to any linear or quasilinear two-port network. The method is useful not only in developing circuit insight, but in providing an easily programmable procedure for exact network evaluation. It is based on ladder network analysis by means of cascaded networks described by their $ABCD$ (or *cascade*) parameters. The basis for the method is the notion that any two-port network can be represented as an equivalent ladder with the addition of coupling (through dependent voltage or current generators) between nonadjacent circuit nodes of the ladder. For simplicity, we restrict the discussion to networks employing resistors, capacitors, and active devices, although it may be extended to include inductors. We begin the discussion of the method by briefly reviewing simple ladder evaluation; the discussion is then extended to include the coupling mentioned previously, in which the concept of *spanning networks* is introduced. Several examples of its use are then given.

The two-port parameters of ladder networks, passive or active, are particularly easy to evaluate. This is true because each element of the ladder network can be represented by an $ABCD$ matrix: the $ABCD$ matrix of a ladder made up of such elements is simply the product of the $ABCD$ matrices of the constituent

two-ports. Two two-ports, for example, have the product

$$T_1 T_2 = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} = \begin{bmatrix} A_1 A_2 + B_1 C_2 & A_1 B_2 + B_1 D_2 \\ C_1 A_2 + D_1 C_2 & C_1 B_2 + D_1 D_2 \end{bmatrix} \quad (8.4-1)$$

The ladder may be made up of transistors, combinations of immittances, or any other network whose two-port parameters are defined.

The $ABCD$ matrices of transistors can be obtained by measurement, as discussed in Chapter 7. To make one set of measurements applicable to transistors of different sizes and dc operating conditions, an equivalent circuit can be derived along with scaling rules and bias dependencies to obtain transistor $ABCD$ matrices over a wide range of conditions. The relationships were derived in Chapter 7 and programs are given in Appendix C.

The $ABCD$ matrices of series impedances and shunt admittances containing resistances and capacitances are found as illustrated in Fig. 8.14. To make the

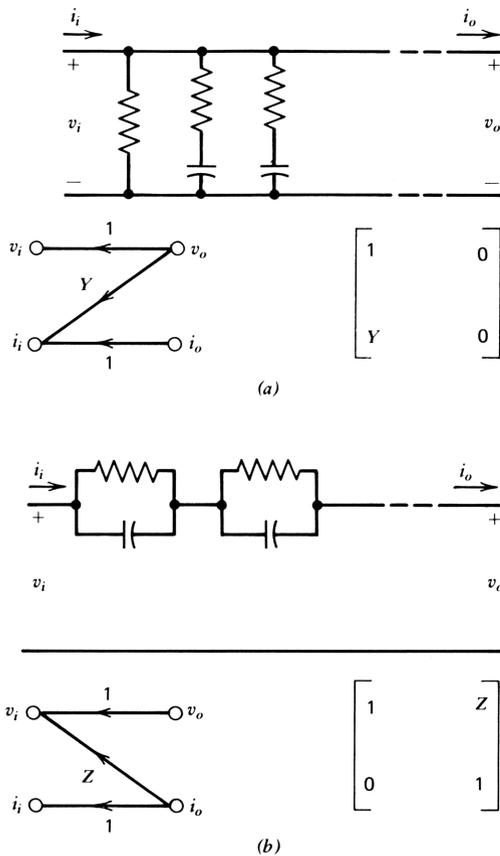


Figure 8.14. Shunt admittances and series impedances in Foster-like form.

immittances general, they are expressed in *Foster-like form*, with series impedances expressed as the parallel combination of a resistor and a capacitor and the dual arrangement for shunt admittances. These immittances are incorporated into a ladder network by multiplication; Fig. 8.6 gives the results.

Evaluation of ladder networks is a simple matter, therefore, although the amount of hand calculations is prohibitive, even for networks of modest size. The operations are easily programmed for calculator or computer. Here and the following section, we assemble a set of equations for matrix operations to be incorporated in a calculator program. With these operations, we can work with matrices on the calculator or computer in the same way that we deal with numbers on a simple four function (programmable) calculator. With practice, the interpretation of the *ABCD* parameters becomes second nature; we have begun the intuitive understanding of the individual *ABCD* parameters in the study of the bipolar transistor and in the hybrid feedback amplifier in the previous section.

Equivalent Ladder Networks

The technique of analyzing ladder networks by multiplying cascaded networks can be extended to any two-port by use of an *equivalent ladder* representation of the two-port.¹ This may seem surprising until it is realized that a wire connection between two circuit nodes can be represented as shown in Fig. 8.15. Two nodes of an arbitrary grounded two-port have a wire connection between them in Fig. 8.15*a*. In Fig. 8.15*b* the wire connection has been replaced by an ideal transformer with unity turns ratio. In Fig. 8.15*c* the transformer has, in turn, been replaced by its *h*-parameter equivalent circuit model, with $h_{12} = 1$ and $h_{21} = -1$. The step represented by Fig. 8.15*b* is unnecessary but may aid understanding. Naturally, there are two ways to show the equivalent circuit for a wire connection—we could have used the *g*-parameter representation of the ideal transformer, as in Fig. 8.15*d*. Which one is correct is determined by the following considerations.

Suppose that the network is a ladder network except for the wire connection. If it is not, other such connections may be pulled out of the network and treated as we are treating this one. Then the equivalent dependent current generator can be connected to the ladder where it *adds* a current to the node to which it is connected. The voltage generator cannot be connected to a node of the ladder, however, since its zero internal impedance would short-circuit the ladder. It must be connected as in Fig. 8.16*a*, *in series with a shunt element of the ladder*. It can be replaced by a pair of voltage generators as in Fig. 8.16*a*, where they add to the voltage at the input and output of the ladder section containing the shunt element.

The transformation shown in Fig. 8.16*a* is a useful one, which we have occasion to use later, particularly in the noise analysis in Chapter 11. Voltage sources can be shifted from one branch to each of the other two branches without affecting the circuit equations at the network terminals. This has been termed the *Blakesley transformation*.²

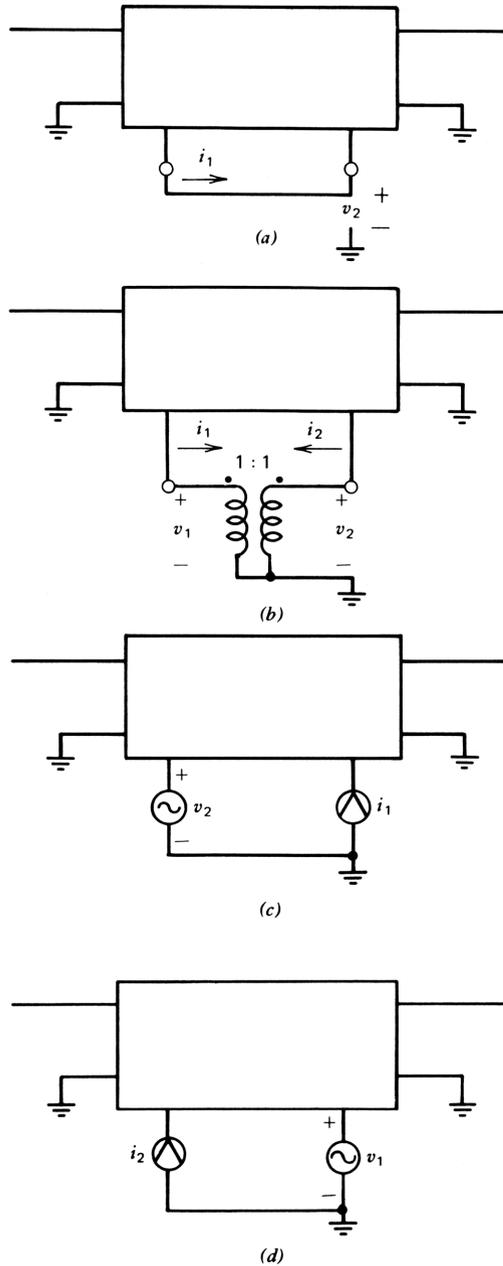


Figure 8.15. Putting arbitrary circuit connections into equivalent ladder form.

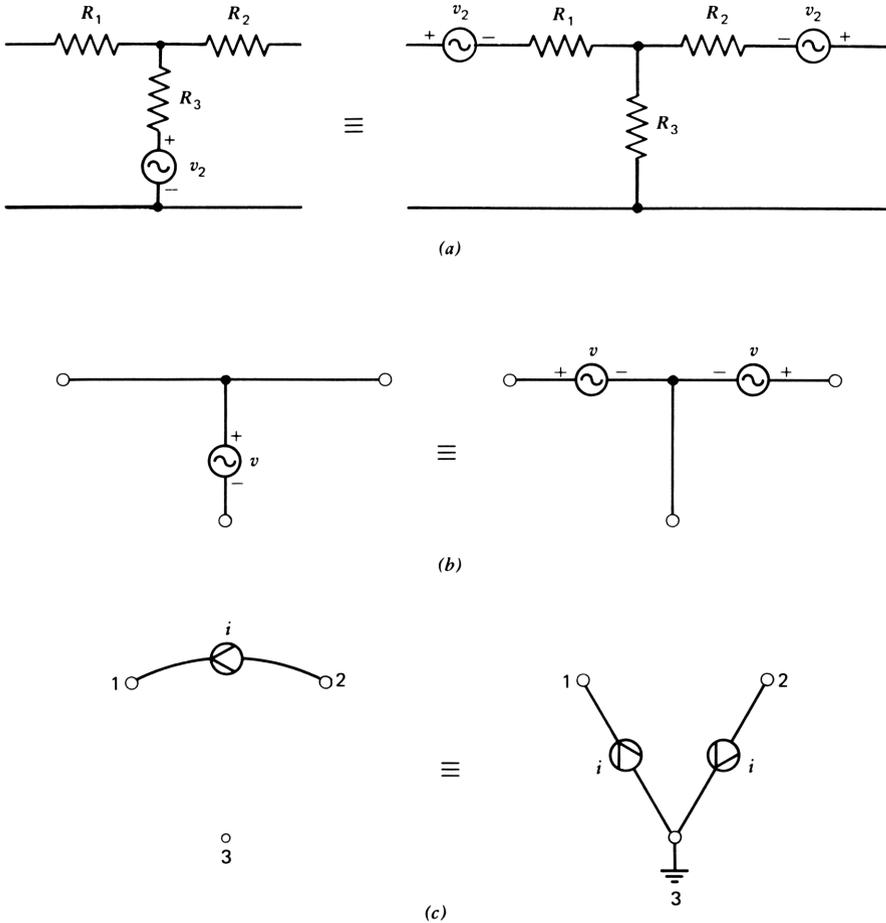


Figure 8.16. Equivalent series-arm voltage generators replace a shunt-arm generator in parts *a* and *b*. Two grounded current generators replace a single generator in part *c*.

An analogous transformation of current sources has also been used previously, as in Fig. 7.14, shown here in Fig. 8.16*c*. Clearly, the two representations in Fig. 8.16*c* are equivalent since the net current at each of the terminals is the same.

Unity Gain Follower

Figure 8.17*a* illustrates the use of an equivalent ladder network for a unity gain (or unity loss) follower circuit. It is a feedback circuit since 100% of the output voltage is returned to the input through a direct connection to the emitter of the first stage; v_o is in series with the input loop consisting of the source, the base-emitter path, and v_o . In Fig. 8.17*b* the wire connection between the emitter of the first transistor and the collector of the second has been replaced

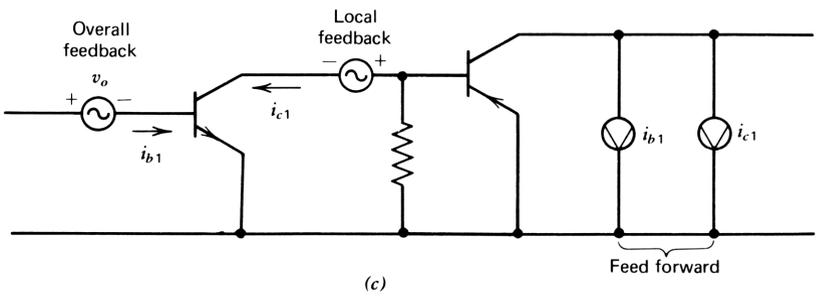
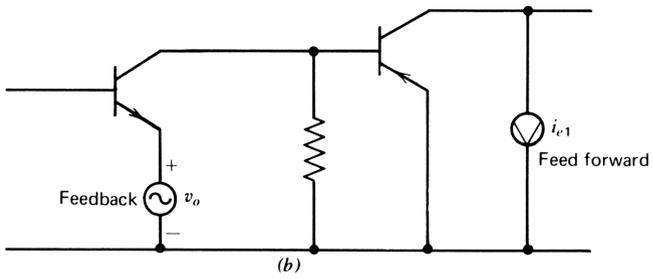
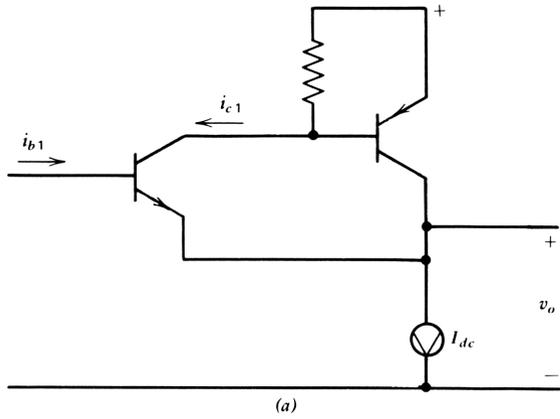


Figure 8.17. Unity gain follower circuit and its equivalent ladder network.

by a voltage generator v_o connected to the emitter and a current generator i_{e1} connected to the output. In Fig. 8.17c the voltage generator has been replaced by a pair of voltage generators connected in series with the base and collector of the first stage. The base voltage generator is recognized as feedback over the whole amplifier; the second generator is local feedback over the output stage. The current generator is the feedforward, or direct feedthrough. The circuit in Fig. 8.17c is an equivalent ladder network for the circuit in Fig. 8.17a.

It is important to remember that the flow of functional dependencies is from output to input—in the anticausal direction. Thus we start with a known output voltage and current and proceed toward the input calculating the currents and voltages at each successive circuit node. When we come to the node to which the voltage generator is connected, we simply *add* the output voltage to the voltage calculated for the network to the right of the voltage generator. We do this for each voltage generator as we come to it.

We can represent the effect of each feedback generator by an $ABCD$ matrix that has only one nonzero element. The input voltage to be added is given by

$$\begin{bmatrix} \Delta v_i \\ \Delta i_i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_o \\ i_o \end{bmatrix} = \begin{bmatrix} h_{12} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_o \\ i_o \end{bmatrix} = \beta_A \begin{bmatrix} v_o \\ i_o \end{bmatrix} \quad (8.4-2)$$

For the unity gain follower, h_{12} of the spanning network is unity, as indicated in eq. (8.4-2). Later in this section we discuss the A -feedback pair, in which h_{12} can be less than unity. The added input current is zero, and the output current has no effect. The matrix in this equation is represented by the symbol β_A , having only one nonzero element. The subscript A indicates A feedback, for which output voltage is observed and input voltage is augmented or controlled. This matrix represents the loss of the circuit when the forward path or active path is in its reference condition. Two β_A matrices, one for the local feedback and one for the overall feedback, must be considered. The matrix for the local feedback generator is the negative of that for the overall feedback matrix.

Calculation of $ABCD$ Parameters of the Circuit

How do we calculate the $ABCD$ parameters of the circuit from the $ABCD$ parameters of the transistors? Let us begin by ignoring the incidental feedforward to make the main thrust clear. Let the $ABCD$ matrices of the first and second transistors be T_1 and T_2 , respectively. The $ABCD$ matrix of the circuit can be written by inspection:

$$T_{UF} = \beta_A + T_1[-\beta_A + T_2] \quad (8.4-3)$$

To T_2 , we *add* the local feedback matrix $-\beta_A$; we then premultiply the combination by T_1 ; finally, we add the overall feedback. This picture of the unity gain follower is complete except for relatively unimportant details considered next.

Incidental Feedforward

The feedforward generator can be taken care of in the same way as the feedback generators, but in a causal direction. We first define a *feedforward matrix* F_A ; it tells us how much current is to be added to the output in response to the emitter current of the first stage. It is advantageous to define the feedforward matrix in terms of the signal at the point(s) in the circuit where the feedback generators are applied, that is, at the base and collector of the first stage rather than at the emitter. Since $i_{e1} = i_b + i_c$, we can split the output current generator into two parts, $F_A i_c$ and $F_A i_b$, where

$$F_A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -h_{21} \end{bmatrix} \quad (8.4-4)$$

for the unity gain follower.

To account for feedforward, we note that for the *ladder signal variables*, that is, the series currents and shunt voltages, there are two points where feedforward takes place in the ladder, corresponding to the addition of the first-stage collector current and the base current to the output. We begin, as before, with T_2 . We note that the local feedback generator is encountered next, so that at this point feedforward of i_{c1} must be added to the output. To do this, we invert T_2 , add $-F_A$ to it, and invert the combination. This operation reflects the fact that feedforward simply adds at the output. We then proceed as before until we come to the next feedback generator. We then invert the combination—the transmission matrix calculated thus far, add F_A , and reinvert. Finally, β_A is added to the result. The equation representing the *ABCD* matrix of the unity gain follower is

$$T_{UF} = \beta_A + \left(F_A + \left\{ T_1 \left[-\beta_A + (-F_A + T_2^{-1})^{-1} \right] \right\}^{-1} \right)^{-1} \quad (8.4-5)$$

Note that the β_A matrix is always followed by an F matrix within a single parenthesis and that the F matrix itself is followed by an inverted expression. Such equations are thus easily checked for accuracy since when written in this form, they express the “rocking back and forth” between inverted and noninverted *ABCD* matrices. When feedforward can be ignored, the inversions drop out.

A program to calculate the *ABCD* matrix including feedforward requires a subroutine for calculating the matrix inverses. Using eq. (8.4-5), the program would start with T_2 , invert it, subtract F_A , invert the combination, subtract β_A , multiply the result by T_1 , invert the combination, add F_A , invert, and finally add β_A . Note that as stages of gain are added, the elements of the *ABCD* matrix become smaller and smaller. Hence the inverse grows larger and larger so that the effect of adding F_A becomes smaller and smaller. This is illustrated by the calculations shown in Fig. 8.18 of the *ABCD* parameters of the circuit with the first stage operating at 1.0 mA and the second at 8.0 mA. The first

ABCD MATRIX OF UNITY FOLL.	ABCD MATRIX OF UNITY FOLL.	ABCD MATRIX OF UNITY FOLL.
FREQ., GHZ = 0.3160	FREQ., GHZ = 0.3160	FREQ., GHZ = 0.3160
ABCD, MAG.+PH:	ABCD, MAG.+PH:	ABCD, MAG.+PH:
A:	A:	A:
R01= 0.9918	R01= 0.9928	R01= 0.9932
R02= 1.8849	R02= 1.8926	R02= 1.9008
B:	B:	B:
R03= 0.0023	R03= 0.0023	R03= 0.0023
R04= 94.8316	R04= 91.0296	R04= 90.8726
C:	C:	C:
R05= 0.1833	R05= 0.1854	R05= 0.1863
R06= 111.2534	R06= 110.6894	R06= 110.5325
D:	D:	D:
R07= 0.0060	R07= 0.0059	R07= 0.0059
R08= 156.0958	R08= 152.2938	R08= 152.1368
(a)	(b)	(c)

Figure 8.18. The ABCD matrix of the unity gain follower: (a) ignores direct feedthrough; (b) adds feedthrough of output-stage local feedback; (c) is the exact calculation. Input transistor is small npn at 3 V, 1 mA; output is large transistor at 3 V, 8 mA.

calculation, (Fig. 8.18a) ignores all feedforward; the second, (Fig. 8.18b) takes into account only the output stage local feedforward, and the third (Fig. 8.18c) gives the exact result. The overall feedforward is entirely negligible, but the local feedforward *reduces* the phase of *D* by about 4°. The equation for the second calculation, ignoring only the overall feedforward, is

$$T_{UF} = \beta_A + T_1 \left[-\beta_A + (-F_A + T_2^{-1})^{-1} \right] \quad (8.4-6)$$

Spanning Networks: Feedback and Feedforward

The return of a portion of the output of a system to the input (feedback) by practical circuit means usually carries with it the feeding of a portion of the input to the output (feedforward). It may not be clear at the outset whether the feedforward or the feedback is dominant, so we need a term to designate such a network without reference to its function in the network. We choose the term

“spanning network” to designate any network whose function is to *add* signals at a pair of nonadjacent circuit vector nodes. Our object in this section is to show (1) how the spanning network can be expressed by a set of $ABCD$ matrices and (2) how these matrices can be combined with that of the original network to give the $ABCD$ matrix of the combination.

Where the spanning network is complicated and contains several circuit elements (active elements can be included) a formal approach can be taken. The two-port parameters of the spanning network can be found by straightforward $ABCD$ analysis. The $ABCD$ parameters are then converted to the set of parameters suitable to the type of feedback introduced by the spanning network, h parameters for A feedback, z parameters for B feedback, y parameters for C feedback, or g parameters for D feedback. We now illustrate this by analysis of a widely used circuit.

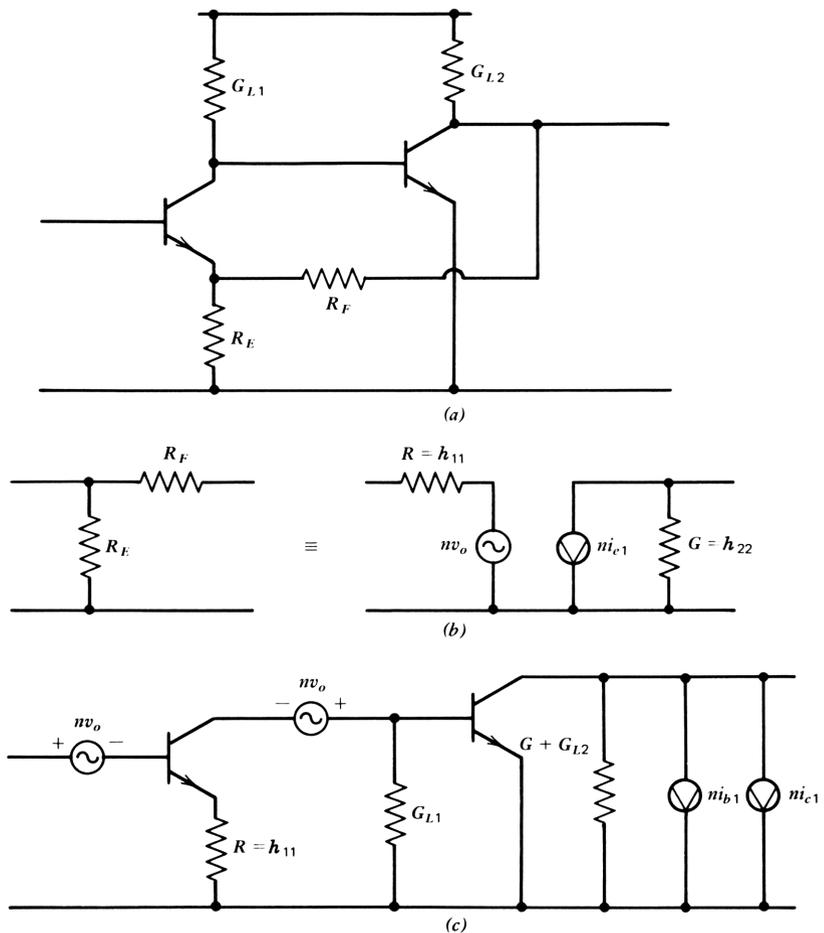


Figure 8.19. An A -feedback pair.

The A-Feedback Pair

As an example, the A -feedback pair in Fig. 8.19 uses a voltage divider consisting of R_F and R_E as a spanning network. The divider is connected across the output port, and the tap is connected to the input emitter. The $ABCD$ parameters of the divider are, from Fig. 8.14

$$T_{sp} = \begin{bmatrix} 1 & 0 \\ \frac{1}{R_E} & 1 \end{bmatrix} \begin{bmatrix} 1 & R_F \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & R_F \\ \frac{1}{R_E} & 1 + \frac{R_F}{R_E} \end{bmatrix} \quad (8.4-7)$$

We then convert the $ABCD$ parameters to h parameters, using Table 6.1:

$$h_{\text{div}} = \begin{bmatrix} \frac{R_E R_F}{R_E + R_F} & \frac{R_E}{R_E + R_F} \\ -\frac{R_E}{R_E + R_F} & \frac{1}{R_E + R_F} \end{bmatrix} \quad (8.4-8)$$

The spanning network can be interpreted as an ideal transformer with turns ratio $R_E/(R_E + R_F)$ with a series loading resistance h_{11} at the input and a shunt loading conductance h_{22} at the output, as shown.

The analysis of the amplifier then proceeds directly with the loading elements considered part of the forward path; the generators are then treated as described previously in the unity gain case. The β_A matrix is obtained from h_{12} , the F matrix, from h_{21} ; the input and output loading matrices—call them H_A and J_A —are obtained from h_{11} and h_{22} , so that we have for these four $ABCD$ matrices:

$$\beta_A = \begin{bmatrix} \frac{R_E}{R_E + R_F} & 0 \\ 0 & 0 \end{bmatrix} \quad (8.4-9)$$

$$F_A = \begin{bmatrix} 0 & 0 \\ 0 & \frac{R_E}{R_E + R_F} \end{bmatrix} \quad (8.4-10)$$

$$H_A = \begin{bmatrix} 1 & \frac{R_E R_F}{R_E + R_F} \\ 0 & 1 \end{bmatrix} \quad (8.4-11)$$

$$J_A = \begin{bmatrix} 1 & 0 \\ \frac{1}{R_E + R_F} & 1 \end{bmatrix} \quad (8.4-12)$$

The β_A and F_A matrices are treated as in the unity gain follower case, replacing unity there by the ratio $R_E/(R_E+R_F)$. The output loading matrix J_A simply postmultiplies T_2 , thereby adding a shunt conductance $1/(R_E+R_F)$ to its output. The H_A matrix is that for a series impedance, but this impedance is in series with the emitter of the first stage, where it introduces local B feedback on that stage.

Thus for a given nominal amplifier loss, if the impedance of the spanning network (voltage divider) is made high to reduce the output loading, the input stage local feedback increases, and vice versa.

The characteristics of the first stage with B feedback are easily obtained in the text that follows. For the present, we represent the $ABCD$ matrix of the first-stage transistor with B feedback as T_{1B} , so that we can write the transmission matrix equation for the circuit as

$$T_{Apr} = \beta_A + T_{1B}(-\beta_A + T_2 J_A) \quad (8.4-13)$$

in which we have ignored the direct feedthrough. When direct feedthrough is not ignored, the expression is similar to that of (8.4-5):

$$T_{Apr} = \beta_A + \left(F_A + T_{1B} \left\{ -\beta_A + \left[-F_A + (T_2 J_A)^{-1} \right]^{-1} \right\}^{-1} \right)^{-1} \quad (8.4-14)$$

in which the $ABCD$ matrix T_{1B} remains to be found from T_1 and H_A .

Exact evaluation of the effect of a spanning network involves four matrices, illustrated for the case of the A -feedback pair by eqs. (8.4-9) to (8.4-12). The four types of feedback can all be treated in the way we treated the A -feedback pair by finding the β and F matrices to be added at the appropriate equivalent ladder network nodes and finding the input and output loading immittance matrices. Table 8.2 shows the four matrices for each type of feedback. Because

Table 8.2 Feedback, Feedforward, Input Loading, and Output Loading $ABCD$ Matrices for the Four Feedback Types

Type of Feedback	β	F	H	J
A	$\begin{bmatrix} h_{12} & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & -h_{21} \end{bmatrix}$	$\begin{bmatrix} 1 & h_{11} \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ h_{22} & 1 \end{bmatrix}$
B	$\begin{bmatrix} 0 & -z_{12} \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & -z_{21} \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & z_{11} \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & z_{22} \\ 0 & 1 \end{bmatrix}$
C	$\begin{bmatrix} 0 & 0 \\ y_{12} & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ y_{21} & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ y_{11} & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ y_{22} & 1 \end{bmatrix}$
D	$\begin{bmatrix} 0 & 0 \\ 0 & -g_{12} \end{bmatrix}$	$\begin{bmatrix} g_{21} & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ g_{11} & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & g_{22} \\ 0 & 1 \end{bmatrix}$

of the sign difference in the output current convention between the $ABCD$ and mixed-epoch parameter sets, the signs of all entries in the second column of the β and F matrices are negative.

Whereas Table 8.2 shows four separate matrices generated from a single h , z , y , or h matrix, the information content is that of the original matrix. The matrices in Table 8.2 serve to direct the computation of the effect of the spanning network. Program “SP” in Appendix C shows a compact way of implementing the computation; only the original $ABCD$ matrix of the original network and the h , z , y , or g parameters of the spanning network are stored. Equations such as (8.4-14) or (8.4-16) are programmed directly as they appear, as illustrated in Appendix C. Program “SP” is used in conjunction with programs “CNV” and “ABCD”, both described in Appendix C. Program “CNV” converts two-port matrices by the equations in Table 6.1. Program “ABCD” comprises a set of subroutines that enable the user to find the $ABCD$ parameters of two-port networks from their constituent parts, converting the calculator into a programmable network calculator. The basis for some of the subroutines of “ABCD” remain to be developed in Chapter 9; with this work completed, the combination of programs (all in program memory) enables us to find the $ABCD$ parameters of any two-port network.

We may use Table 8.2 to generate the $ABCD$ matrix of the first stage of the A -feedback pair with B feedback. The spanning network is just $h_{11} = R_E R_F / (R_E + R_F) = R$, so that the four spanning network matrices are, from Table 8.2

$$\beta_B = \begin{bmatrix} 0 & -R \\ 0 & 0 \end{bmatrix}, \quad F_B = \begin{bmatrix} 0 & -R \\ 0 & 0 \end{bmatrix} \quad (8.4-15)$$

$$H_B = \begin{bmatrix} 1 & -R \\ 0 & 1 \end{bmatrix}, \quad J_B = \begin{bmatrix} 1 & -R \\ 0 & 1 \end{bmatrix} \quad (8.4-16)$$

and the $ABCD$ matrix of the stage is

$$T_{1B} = \beta_B + [F_B + (H_B T_1 J_B)^{-1}]^{-1} \quad (8.4-17)$$

An equivalent ladder circuit is shown in Fig. 8.20 to aid in the interpretation of this equation. For circuits as simple as this (and for many circuits of greater complexity), permutative feedback (Section 9.1) offers a computationally simpler means of generating the $ABCD$ matrix of the network.

Each element of the $ABCD$ matrix of a network can be interpreted as a feedback element or a network connected around an amplifier of zero loss. For the bipolar transistor, this expresses the physical interactions more naturally than do conventional equivalent circuit representations. For the B feedback in Fig. 8.20, this becomes particularly clear. When an external emitter resistance is added, the feedback merely adds to the internal feedback of the device; $B = -(R_E + r_E)$. Certainly, if R_E is to be considered feedback, then r_E can also

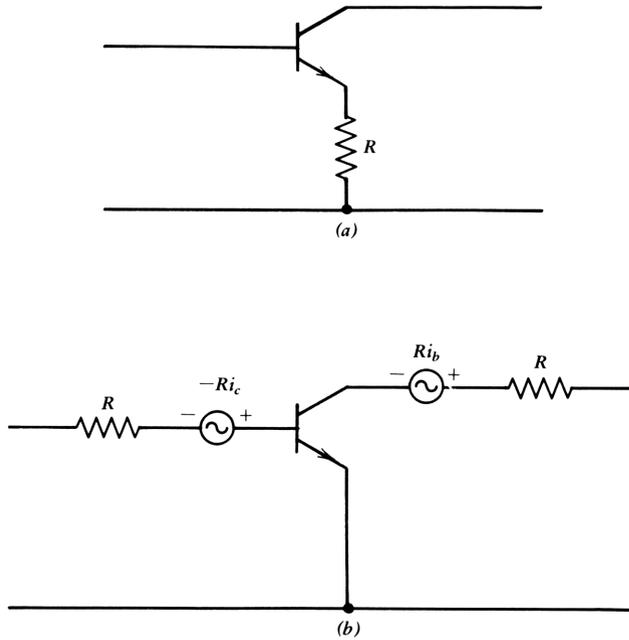


Figure 8.20. Circuit and equivalent ladder for B -feedback stage.

be considered feedback. There is no nonarbitrary way to define one as feedback and the other as part of the forward path.

PROBLEMS

- 1 For the D -feedback circuit in Fig. 8.21, write expressions for the β , F_D , H_D , and J_D matrices, and write the equation for the transmission matrix of the circuit.
- 2 Repeat Problem 8.21 for the B -feedback circuit in Fig. 8.22.

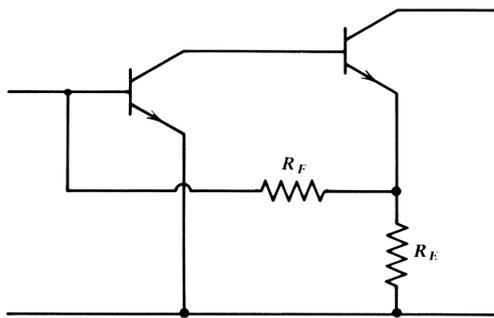


Figure 8.21.

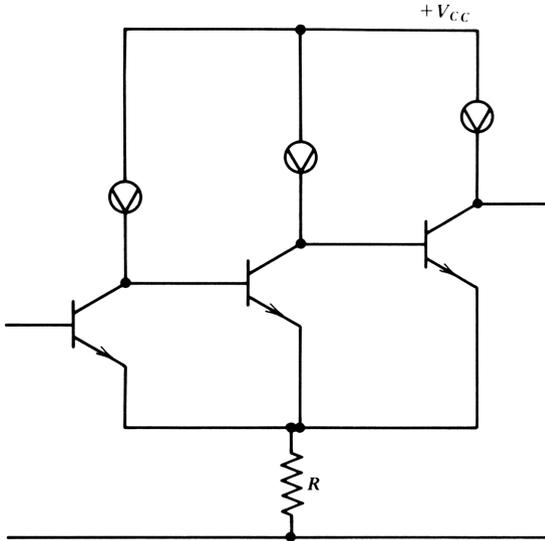


Figure 8.22.

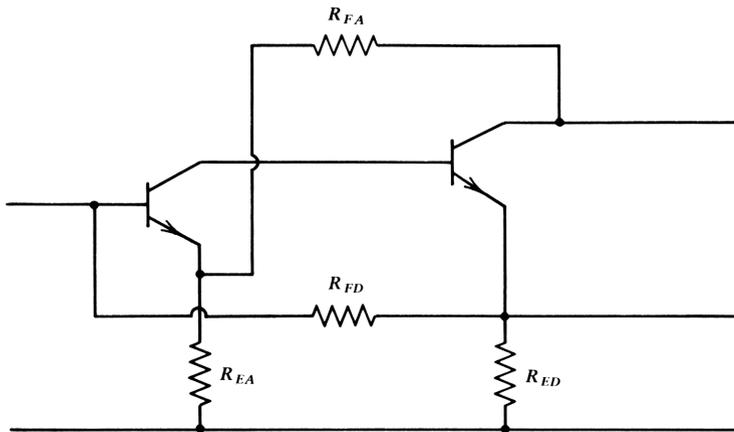


Figure 8.23.

- 3 The circuit in Fig. 8.23 combines A and D feedback. Write the transmission matrix equation for this circuit ignoring direct feedthrough.

REFERENCES

- 1 F. D. Waldhauer, "Anticausal Analysis of Feedback Amplifiers," *BSTJ* **56** (8), 1337–1386 (1977).
- 2 E. H. Nordholt, "Classes and Properties of Multiloop Negative-Feedback Amplifiers," *IEEE Trans. CAS-28* (3), 203–211 (March 1981).

Chapter 9

Analog Integrated Circuit Design: Feedback and Feedforward

When we apply the concepts of feedback and feedforward to two-port networks, we find that they represent nothing more than coupling between nonadjacent circuit nodes of a ladder network—either passive or active. Therefore, the subject of this book encompasses all circuits that have one input and one output port. Feedback cannot be defined more restrictively than this in a nonarbitrary way. Thus feedback theory can be made a foundation of two-port network theory. What can be done for the two-port network can be extended to networks that have more than two ports—we explore examples of this in Chapter 12.

This would not be of particular interest, except that by anticausal methods, feedback becomes easier to analyze and becomes a powerful circuit analysis tool in its own right. In this chapter we apply feedback concepts to simplify the analysis and design of elements of analog integrated circuits, a subject that has been treated elsewhere.^{1,2} The coverage here is not as extensive as that in the references, and the reader is urged to consult them for information on further areas in this field. Applying the methods of this book to the problems treated in the references may well yield new and worthwhile results.

We begin by introducing a markedly useful technique for finding the $ABCD$ parameters of combinations of transistors and other circuit elements. This technique consists of finding the $ABCD$ matrix of a three-terminal network (e.g., a transistor or a combination of transistors) when the input or output leads are permuted, or interchanged. Many circuits can be analyzed quite simply this way, including the case study Design B amplifier in Chapter 3.

We then analyze the transmission and output admittance characteristics of current mirror circuits, which are forms of D -feedback amplifiers. The problem of accurately sensing the output current in B - and D -feedback amplifiers concludes the chapter.

9.1 PERMUTATIVE OR LEAD-INTERCHANGE FEEDBACK³

A feedback transformation of a three-terminal network arises from interchanging the output and common leads of the network, as shown in Fig. 9.1a, where the original network is taken for illustrative purposes as a common emitter transistor whose output lead is the collector. After the permuting process, the collector is the common lead between the input and the output. This is a feedback process because the original output voltage is preserved (except for a sign change) and the output voltage is connected in series with (or is returned to) the input. The process is termed *output permutative feedback*.

Similarly, we may interchange the input leads of the original network as shown in Fig. 9.1b. In this case, termed *input permutative feedback*, the input voltage is reversed in sign, and the output current is returned to the input. Where the original network is a transistor, input permutative feedback yields a common base stage.

The function of this section is to find the $ABCD$ parameters of the input or output permuted network from those of the original network. These two transformations are extremely useful in understanding feedback networks and in calculating their characteristics.

A Composite Matrix for a (Grounded) Two-Port

To write the expressions for the $ABCD$ parameters of any network whose input or output leads are permuted, we begin by writing a *composite* matrix for the network—one that gives the voltage and current of all three possible ports in terms of the voltage and current at one of the ports: with the sign convention given in Fig. 9.1c, we can write

$$\begin{bmatrix} v_1 \\ i_1 \\ v_2 \\ i_2 \\ v_3 \\ i_3 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \\ 1 & 0 \\ 0 & 1 \\ 1-A & -B \\ -C & 1-D \end{bmatrix} \begin{bmatrix} v_2 \\ i_2 \end{bmatrix} \quad (9.1-1)$$

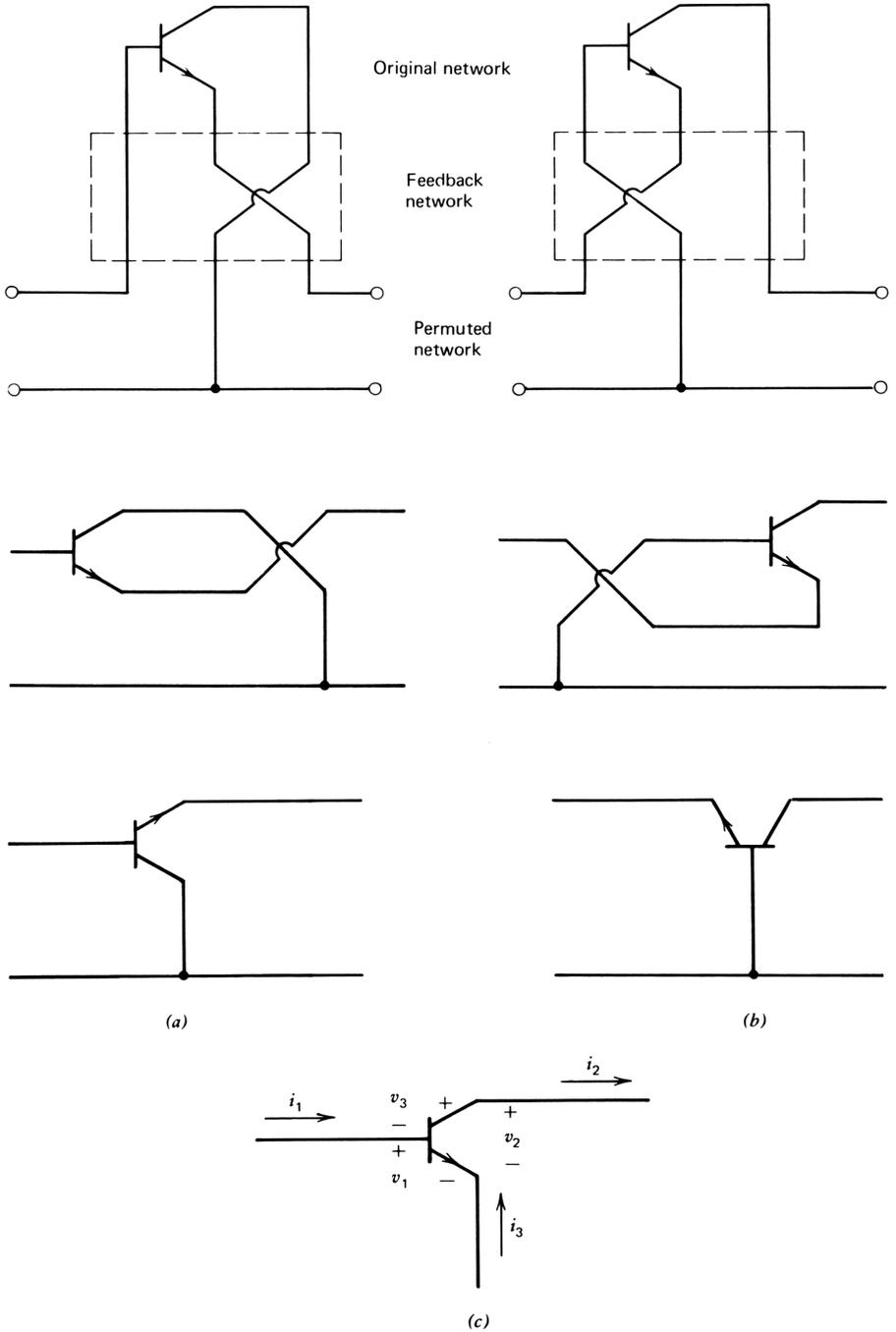


Figure 9.1. Feedback by permutation of leads: (a) common collector stage; (b) common base stage; (c) sign convention used for composite matrix in eq. (9.1-1).

where $v_3 = v_2 - v_1$ and $i_3 = i_2 - i_1$. With this composite matrix, we can derive the $ABCD$ parameters of the permuted networks very easily. If we permute the output leads, for example, Fig. 9.1 shows that the input voltage is $-v_3$ and the input current is i_1 , so that from the fifth row of the matrix we have

$$-v_3 = v_{bc} = (A - 1)v_2 + Bi_2 \tag{9.1-2}$$

Note that v_2 and i_2 on the right side of the equation come from the column on the right side of (9.1-1); eq. (9.1-2) is simply a literal reading of the fifth line of (9.1-1) with the sign changed to obtain v_{bc} . Similarly, from the second line of (9.1-1), we obtain

$$i_1 = i_b = Cv_2 + Di_2 \tag{9.1-3}$$

Combining (9.1-2) and (9.1-3) into a matrix, we obtain

$$\begin{bmatrix} -v_3 \\ i_1 \end{bmatrix} = \begin{bmatrix} A - 1 & B \\ C & D \end{bmatrix} \begin{bmatrix} v_2 \\ i_2 \end{bmatrix} \tag{9.1-4}$$

With the output leads permuted, the output signal variables are $-i_3$ and $-v_2$, so that from the third and sixth rows, we find the output in terms of v_2 and i_2 :

$$\begin{bmatrix} -v_2 \\ -i_3 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ C & D - 1 \end{bmatrix} \begin{bmatrix} v_2 \\ i_2 \end{bmatrix} \tag{9.1-5}$$

Inverting this matrix as in eq. (8.5-2), we have

$$\begin{bmatrix} v_2 \\ i_2 \end{bmatrix} = \frac{1}{1 - D} \begin{bmatrix} 1 - D & 0 \\ C & 1 \end{bmatrix} \begin{bmatrix} -v_2 \\ -i_3 \end{bmatrix} \tag{9.1-6}$$

Substituting this in (9.1-4) and performing the indicated multiplication, we obtain

$$\begin{bmatrix} -v_3 \\ i_1 \end{bmatrix} = \frac{1}{1 - D} \begin{bmatrix} \theta & -B \\ -C & -D \end{bmatrix} \begin{bmatrix} -v_2 \\ -i_3 \end{bmatrix} \tag{9.1-7}$$

where

$$\theta = 1 - A - D + \Delta_A \tag{9.1-8}$$

as in eq. (8.3-18). For the common collector transistor, this equation is written

$$\begin{bmatrix} v_{bc} \\ i_b \end{bmatrix} = \frac{1}{1 - D} \begin{bmatrix} \theta & -B \\ -C & -D \end{bmatrix} \begin{bmatrix} v_{ec} \\ i_e \end{bmatrix}$$

in which θ is close to unity over the entire useful frequency range of the transistor.

By an exactly similar process, we find the $ABCD$ parameters of a network with its *input* leads permuted. The input signal consists of $-v_1$ and i_3 and the output signal, of v_3 and i_2 , so that we use rows 1 and 6 of the composite matrix for the first two equations and rows 4 and 5 for the second two equations. The result is

$$\begin{bmatrix} -v_1 \\ i_3 \end{bmatrix} = \frac{1}{1-A} \begin{bmatrix} -A & -B \\ -C & \theta \end{bmatrix} \begin{bmatrix} v_3 \\ i_2 \end{bmatrix} \quad (9.1-9)$$

where θ is given in (9.1-8).

Common Base and Common Collector Parameters

The pair of permuting equations, (9.1-7) and (9.1-9), are useful for both insight and calculations involving feedback. Where the $ABCD$ parameters of a transistor are available, as from the equivalent circuit discussed in Chapter 7, the common collector and common base parameters can be found directly. Figure 9.2 shows the common emitter parameters of a transistor at 0.4 GHz, with the common base and common collector parameters on either side; B is in kilohms, and C in millisiemens.

Note that the magnitude of D for the common emitter case is 0.0811; since the measurement was made at 0.4 GHz, the value of f_T for this transistor is

ABCD MATRIX OF COMMON BASE	ABCD MATRIX OF COMMON EMIT.	ABCD MATRIX OF COMMON COLL
FREQ., GHZ = 0.4000	FREQ., GHZ = 0.4000	FREQ., GHZ = 0.4000
ABCD, MAG.+PH:	ABCD, MAG.+PH:	ABCD, MAG.+PH:
A:	A:	A:
R01= 0.0155	R01= 0.0154	R01= 0.9987
R02= 102.9604	R02= -76.1781	R02= 0.7062
B:	B:	B:
R03= 0.0061	R03= 0.0061	R03= 0.0060
R04= 33.8066	R04= -145.3320	R04= 30.0830
C:	C:	C:
R05= 0.5867	R05= 0.5846	R05= 0.5786
R06= 99.3983	R06= -79.7403	R06= 95.6748
D:	D:	D:
R07= 1.0126	R07= 0.0811	R07= 0.0802
R08= 4.4298	R08= -94.9995	R08= 80.4156

Figure 9.2. Comparison of the $ABCD$ parameters of the common emitter, common collector, and common base stages at 400 MHz.

0.4/0.0811, or 4.9 GHz. We can also estimate the collector capacitance from the magnitude of C ; it is $0.58/(0.4 \times 2\pi)$ or 0.23 pF. The rest of the parameters can also be found, as discussed in the modeling section of Chapter 7 and in program “A>E” in Appendix C.

The values of the common base $ABCD$ parameters correspond closely with the common emitter parameters except for D where unity has been added. The approximation given at the beginning of this section is good, then. The phase of D — 4.4° —should give us the value of transit delay including the emitter time constant $r_e C_{je}$. Translating to radians and dividing by the angular frequency, the transit time is 0.031 ns, which corresponds to the value found in Chapter 8. Note that the common base transistor includes this effectively as a delay of the current signal from input to output. (To compare the phase of the common emitter and common base parameters in Fig. 9.2, subtract 180° from the common base parameters.)

The common collector parameters are also similar to those for the common emitter parameters except for A , where unity has been added. Note that the phase of A is only 0.7° , indicating an absence of delay. This also applies to D ; when 180° is subtracted from the phase of D , the common collector stage has 4.6° less phase than that of the common emitter stage.

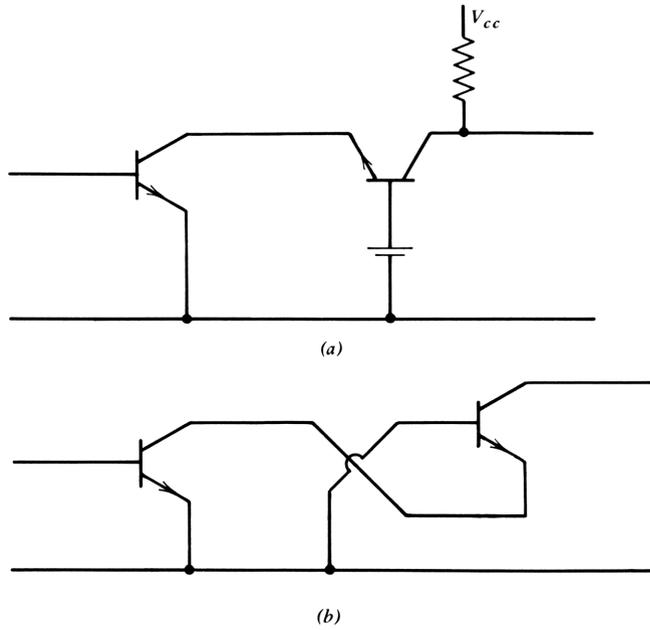
Cascode Stage

The lead permutation process and programs are useful for more than finding the common collector and common base parameters of transistors. Suppose that we wish to calculate the $ABCD$ parameters of a cascode stage. Figure 9.3a shows the circuit, which consists of the tandem combination of a common emitter and a common base stage. We begin with the $ABCD$ parameters, in Fig. 9.3b, of the second stage, permute its input leads to obtain the common base parameters, and then premultiply the resulting matrix by that for the first-stage common emitter parameters. A program for doing this is shown in connection with program “ABCD” in Appendix C. The cascode $ABCD$ parameters at 0.4 GHz are shown in Fig. 9.3c.

As is characteristic of the cascode stage, A and C drop by one order of magnitude. The common base stage removes the output voltage from the collector of the first stage; this is what gives the cascode stage its salient character. An unfortunate characteristic is that the delay is increased by the common base stage. In the example, the phase of D (at 0.4 GHz) is increased over the common emitter stage by 6° and B , by 5° .

Emitter-Coupled Pair

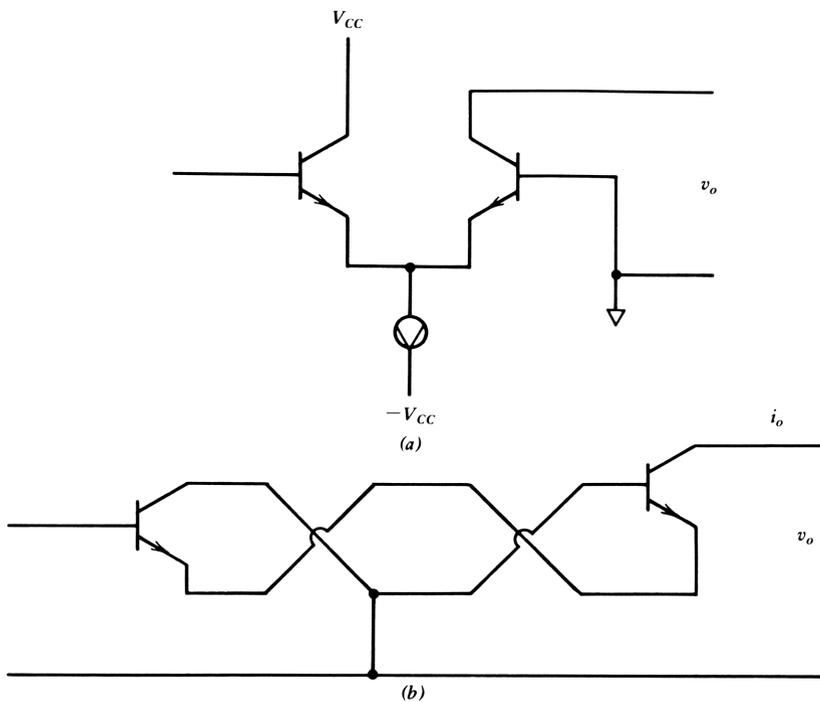
The emitter-coupled pair may be considered the tandem combination of a common collector and a common base stage, as shown in Fig. 9.4a. Other characteristics of this type of stage, including common mode rejection, are considered in Chapter 12. The $ABCD$ parameters of the combination are found by permuting the input leads of the second-stage transistor, permuting the output leads of the first transistor, and multiplying the two matrices with



ABCD MATRIX OF XSTR, 3V, 8MA	ABCD MATRIX OF CASCODE STGE
FREQ., GHZ = 0.4000	FREQ., GHZ = 0.4000
ABCD, MAG.+PH:	ABCD, MAG.+PH:
A:	A:
R01= 0.0154	R01= 0.0036
R02= -76.1781	R02= -42.3333
B:	B:
R03= 0.0061	R03= 0.0061
R04= -145.3320	R04= -140.0339
C:	C:
R05= 0.5846	R05= 0.0562
R06= -79.7403	R06= 7.3769
D:	D:
R07= 0.0811	R07= 0.0847
R08= -94.9995	R08= -88.8795

(c)

Figure 9.3. A cascode stage, with ABCD parameters.



ABCD MATRIX
OF
XSTR 3V, 1MA

FREQ., GHZ = 0.3160

ABCD, MAG.+PH:

A:

R01= 0.0189
R02= -82.7726

B:

R03= 0.0320
R04= -164.1096

C:

R05= 0.1751
R06= -80.3330

D:

R07= 0.0738
R08= -96.1132

ABCD MATRIX
OF
EM. CPLD PR.

FREQ., GHZ = 0.3160

ABCD, MAG.+PH:

A:

R01= 0.0244
R02= 99.9482

B:

R03= 0.0641
R04= 15.5924

C:

R05= 0.0160
R06= -178.9771

D:

R07= 0.0790
R08= 85.4022

(c)

Figure 9.4. An emitter-coupled pair.

results at 0.3 GHz given in Fig. 9.4c. Note that B doubles, since the emitter resistances of the two transistors are in series, with the output current passing through both. Also, C virtually disappears for the same reason as in the cascode case. Except for those two changes (and the lack of a phase reversal), the characteristics are similar to a common emitter stage.

Darlington Pair

The Darlington pair in Fig. 9.5a provides an interesting example of the use of lead permutation for determining network characteristics. As shown in Fig. 9.5b, the circuit can be considered the tandem combination of two common collector stages that, in combination, have their output leads permuted.

At this point we introduce some notation; let us call $P_i[T]$ the result of permuting the input leads of $ABCD$ matrix T and similarly $P_o[T]$ for permuting the output leads. If T_1 and T_2 are the common emitter $ABCD$ matrices for the first and second transistors of a Darlington pair, we can write the $ABCD$ matrix of the first common collector stage as $P_o[T_1]$ and similarly for the second stage, $P_o[T_2]$; thus we can write

$$T_{\text{Dar}} = P_o[P_o[T_1] * P_o[T_2]] \quad (9.1-10)$$

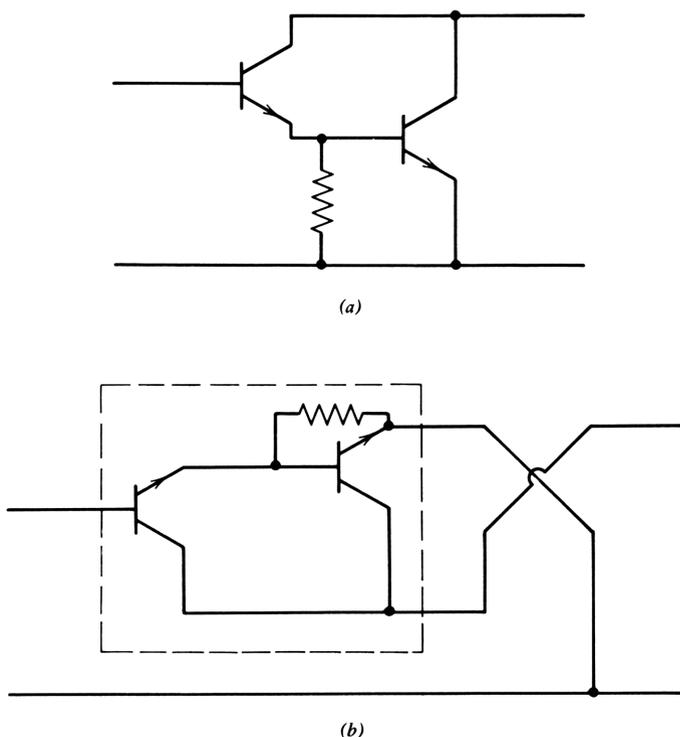


Figure 9.5. A Darlington pair.

ABCD MATRIX
OF
DARL. PAIR

FREQ., GHZ = 0.3160

ABCD, MAG.+PH:

A:

R01= 0.0386
R02= -00.9603

B:

R03= 0.0069
R04= -138.8008

C:

R05= 0.1825
R06= -73.3007

D:

R07= 0.0058
R08= -31.6964

(c)

Figure 9.5. Continued.

in which $*$ is the matrix multiplication operation. The result of this computation at 0.3 GHz is the transmission matrix shown in Fig. 9.5c.

B and C Feedback

The permuting operation gives us a particularly simple way to calculate the $ABCD$ parameters of many circuits whose circuit equations are otherwise complex. Consider, for example, a network with both B and C feedback, shown in Fig. 9.6. If we begin the analysis by permuting the output leads of the network before applying the feedback, we find that the C feedback path becomes a shunt admittance across the input terminals of the permuted network. Similarly, Fig. 9.6b shows that the B -feedback impedance is now an impedance in series with the output. Both feedback paths are then incorporated by the operations of Fig. 8.6—simple cascade multiplications. After

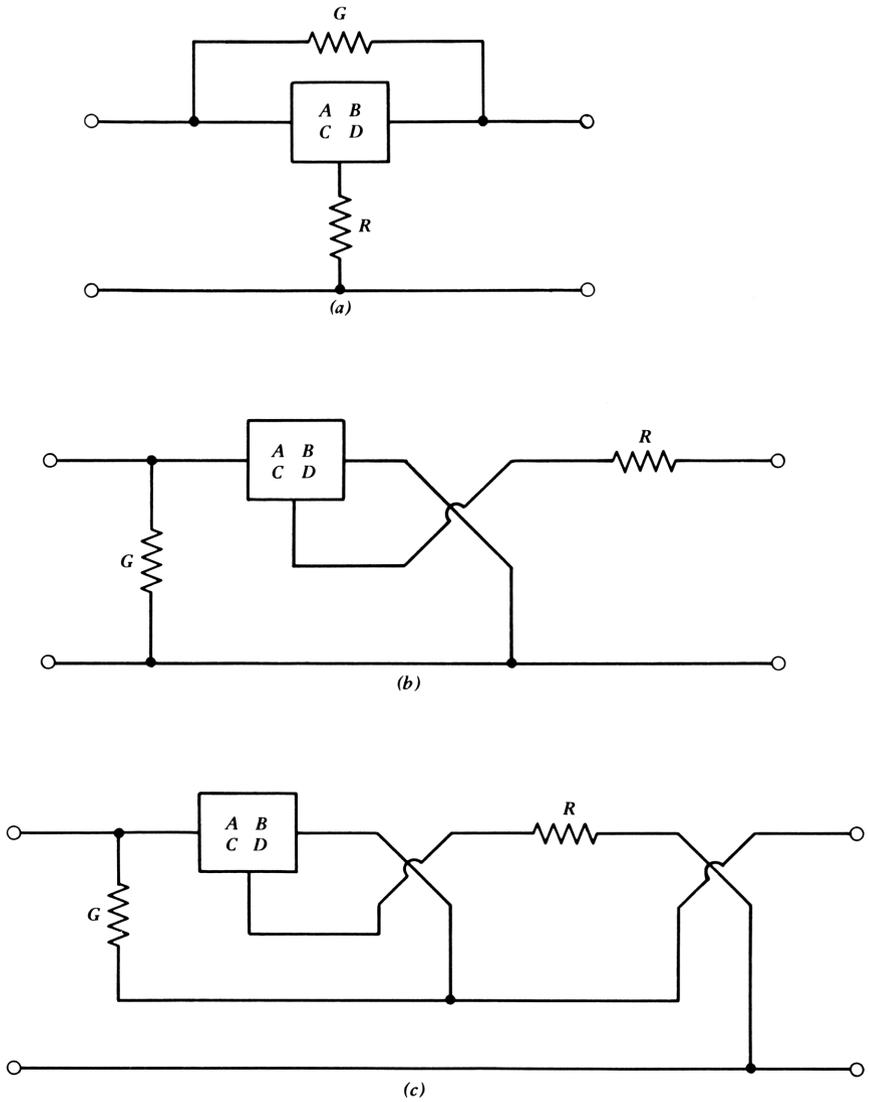


Figure 9.6. Combination B and C feedback calculated by topological manipulation.

performing these operations, the output leads are again permuted as in Fig. 9.6c, giving the $ABCD$ parameters with B and C feedback. All four effects of feedback, feedforward, and input and output loading are thus taken care of in this simple set of operations.

In operational notation the transmission matrix of the network with B and C feedback can be expressed compactly as

$$T_{BC} = P_o [Z_o (Y_i (P_o [T_a]))] \tag{9.1-11}$$

where T_a is the original network. The operation $Y_i(T)$ denotes premultiplication by the $ABCD$ matrix of a shunt admittance, as illustrated in Fig. 8.6c, and Z_o denotes postmultiplication by the $ABCD$ matrix of a series impedance as in Fig. 8.6b. This expression is a compact way of writing the two-port matrix of any network that has either C or B feedback or both in which the feedback path is representable as a two-terminal immittance. It is also a computational algorithm for finding the matrix. Compare eq. (9.1-11) with the algebraic expression for the same circuit found by applying feedback B and C successively to the original network:

$$T_{BC} = \frac{1}{1 + CR + BG - RG\theta} \begin{bmatrix} A + CR + BG - RG\theta & B - R\theta \\ C - G\theta & D + CR + BG - RG\theta \end{bmatrix} \tag{9.1-12}$$

where θ is given by eq. (9.1-8). Equation (9.1-11) is simpler than (9.1-12), a result of moving up one step in hierarchical level to the two-port formulation.

Design B of Our Case Study

Design B in the Chapter 3 case study can be described as the combination of an output stage with C feedback in tandem with two common emitter stages and having overall C feedback. We can thus write the $ABCD$ matrix of the Design B amplifier by using the permutation algorithm. The permuting operations are shown in sequence in Fig. 9.7 and can be written

$$T_{Des B} = P_o [Y_{iF} (P_o [T_1 * T_2 * P_o [Y_{i2} (P_o [T_3])]])] \tag{9.1-13}$$

This expression constitutes a program for the evaluation of the transmission matrix of the Design B amplifier. The evaluation begins by starting with the innermost expression T_3 , permuting its output leads, premultiplying by the admittance G_2 , and again permuting the output leads. This resulting $ABCD$ matrix is then premultiplied by T_2 , and then T_1 . The output leads of the resulting $ABCD$ matrix are permuted, and this matrix is premultiplied by the matrix of $Y_F = (G_F + C_F s)$. Finally, the output leads of the combination are permuted, giving the $ABCD$ matrix of the amplifier.

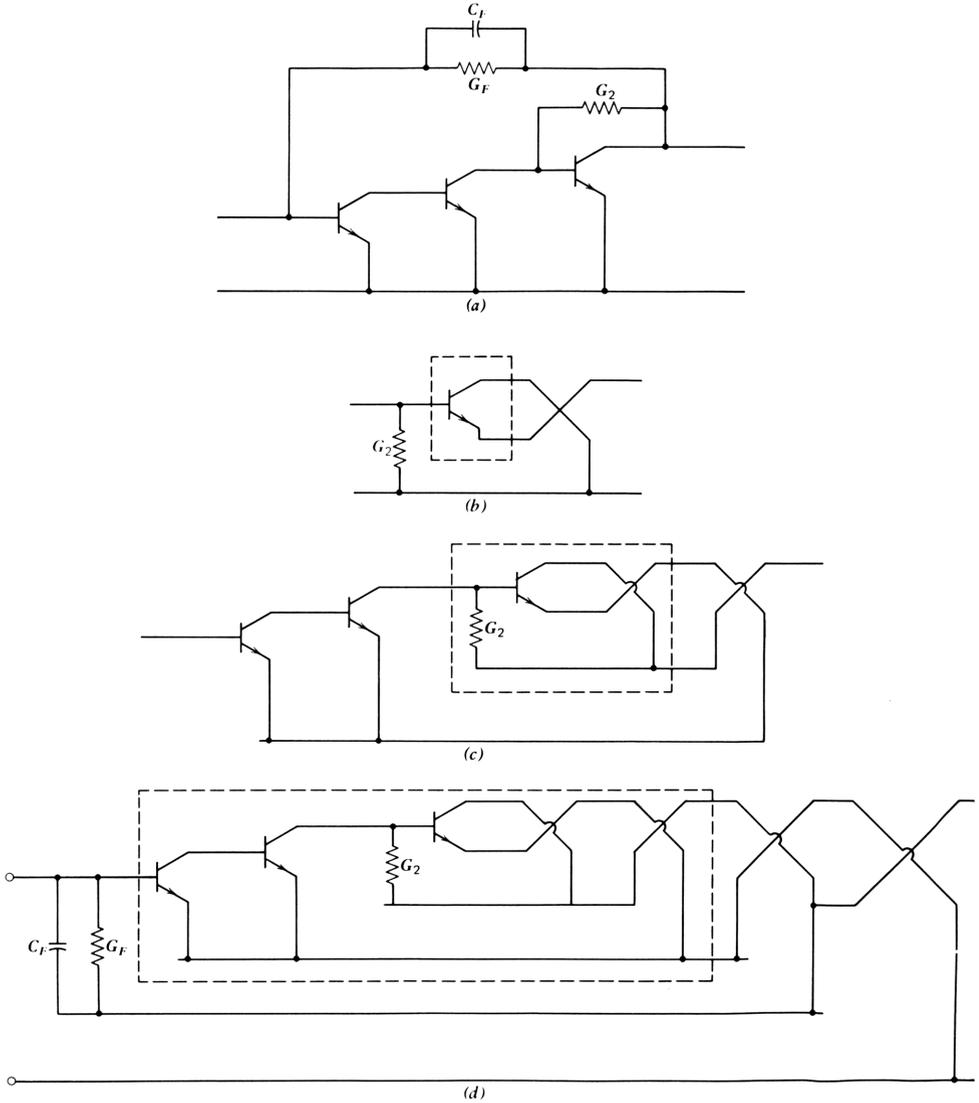


Figure 9.7. Analysis of Design B by output lead permutation.

The analysis of our Design B case study amplifier is finally reduced to a one-line calculator program, where the calculator is the special type given in program "ABCD" in Appendix C.

The permuting operations provide a way of analyzing a surprisingly general class of feedback structures by repeated application of eqs. (9.1-7) and (9.1-9). Yet it cannot always be applied. The A - and D -feedback pairs of the previous chapters, for example, cannot be analyzed in this way; thus the general method

must be employed. (Note, however, that the local B feedback in these circuits *can* be analyzed by permutative feedback.) One general rule can be stated: permutative feedback analysis can be done if either the feedback network or the active path is ungrounded. In the Design B example both the G_2 and the Y_F feedback paths had no connection to ground, so that the permutative analysis could proceed. We see an example of permutative feedback applied to a grounded feedback path with ungrounded active path in Section 9.2.

9.2 CURRENT SOURCES AND CURRENT MIRRORS

In several circuits discussed so far we have incorporated dc current sources—ideal devices that provide dc bias current without otherwise affecting circuit operation. In this section we show how such sources can be realized and how we can estimate their departure from ideal performance. These circuits can also be used as amplifying circuits; for this application, we are interested not only in how their output admittance departs from zero, but also in their *transmission characteristics*. We see that these circuits can also serve as broadband amplifiers, which are useful in feedforward applications.

An ideal current source has the transmission matrix

$$T_{CS(\text{ideal})} = \begin{bmatrix} 0 & 0 \\ 0 & -n \end{bmatrix} \quad (9.2-1)$$

Thus the input current is n times the output current, and from eqs. (8.2-4) and (8.2-9), the input impedance and the output admittance are zero. These are the characteristics of an ideal D -feedback amplifier.

Simple Current Source (Mirror)

A simple current source is shown in Fig. 9.8a and includes a diode-connected transistor and an output transistor. If the resistors are made zero, the base-emitter voltage of the two transistors is the same, thus enabling us to write

$$V_{be} = \frac{kT}{q} \ln \frac{I_{c1}}{I_{S1}} = \frac{kT}{q} \ln \frac{I_{c2}}{I_{S2}} \quad (9.2-2)$$

The ratio I_{c1}/I_{c2} is equal to I_{S1}/I_{S2} . If the areas of the two transistors are equal and the temperature is the same, $I_{S1} = I_{S2}$, so that $I_{c1} = I_{c2}$.* The output current is I_{c1} , taken as unity in Fig. 9.8a; the input current is I_{c1} plus the base currents of the two transistors; thus at low frequencies, ignoring collector voltage effects for the moment, we obtain

$$D = -(1 + 2\delta) \simeq -1 \quad (9.2-3)$$

If the diode area is r times the transistor emitter area, the equation becomes

$$D = -r - (1 + r)\delta \simeq -r \quad (9.2-4)$$

*The doping density is also assumed to be the same.

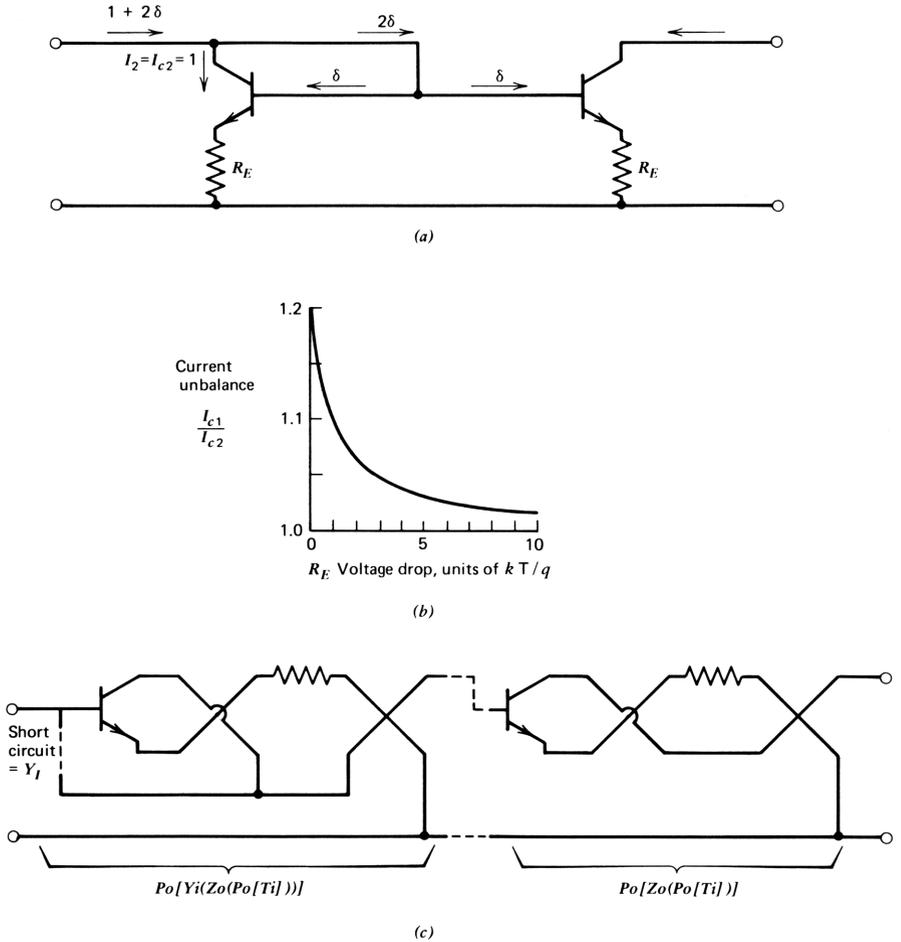


Figure 9.8. Simple current mirror: (a) circuit; (b) effect of equal emitter resistors on current unbalance (with $R_E = 0$, current unbalance is assumed to be 20%); (c) and (d) finding the ABCD matrix of the current mirror by topological manipulation.

Thus if the diode area is smaller than the transistor area, the input current is smaller than the output current by the factor r . The value of n in eq. (9.2-1) is about equal to the diode: transistor area ratio.

The current mirror is almost always used with resistors in series with the emitters of Q_1 and Q_2 , for several reasons: (1) any unbalance in the transistors is reduced greatly; (2) the circuit tends to be quite noisy, and the resistors reduce the noise (as we see in Chapter 11); and (3) the output admittance at all frequencies is reduced. With resistors, equating base voltages, we obtain

$$I_{c1} R_{E1} + \frac{kT}{q} \ln \frac{I_{c1}}{I_{S1}} = I_{c2} R_{E2} + \frac{kT}{q} \ln \frac{I_{c2}}{I_{S2}} \quad (9.2-5)$$

This transcendental equation can be solved iteratively on the calculator; where the voltage drop across the resistors is (nominally) greater than kT/q , convergence is rapid in the logarithmic form given. For R_E voltage drops less than kT/q , the equation should be converted to exponential form, raising e to the expressions on either side of the equation. The improvement in current balance is shown in Fig. 9.1*b* as a function of the voltage drop across R_{E1} and R_{E2} , assuming an unbalance in saturation currents of 20%. For a resistive drop of only kT/q (0.026 V), the unbalance is halved, and for a drop of $10 kT/q$, the unbalance is reduced to less than one-tenth the saturation current unbalance.

At low frequencies, the circuit input resistance is that of the diode in series with R_{E2} , since it is much smaller than that of the transistor. The output conductance is essentially g_{ce} for $R_E = 0$, and decreases as R_E is increased. In most applications the source resistance for the circuit is high, and the load conductance is high compared with g_{ce} , so that the circuit is characterized by its D parameter.

We can estimate the high-frequency transmission of the circuit by adding τ_{T-S} to δ wherever it appears in Fig. 9.8*a*. Thus, approximately, $-D = 1 + 2\delta + 2\tau_{T-S}$, giving a high-frequency cutoff at approximately half the unity loss frequency of the transistor. The situation is actually somewhat more complicated, as we see later.

To obtain the transmission properties of the circuit over the complete frequency range, we start with the $ABCD$ parameters of the transistors. We then operate on these parameters with the immittance and permuting operators previously defined, as indicated in Fig. 9.8*c* for the diode and Fig. 9.8*d* for the transistor. For the diode, we permute the output leads $P_o[T_2]$; postmultiply by the $ABCD$ matrix for the series resistor R_{E2} , giving $Z_o(P_o[T_2])$; and then premultiply by the Y matrix for the collector-to-base short circuit, giving $Y_i(Z_o(P_o[T_2]))$. We use a large capacitance (e.g., 10^{12} pF) in series with zero resistance for the short circuit. The final step is to permute the output leads, giving the $ABCD$ matrix of a shunt diode. This procedure is repeated for the transistor, omitting the Y_i operation, as shown in Fig. 9.8*d*. The $ABCD$ matrix of the current mirror is obtained by premultiplying that of the transistor by that for the shunt diode. Although this procedure is elaborate to describe, each operation represents only one or two button pushes on the calculator (either to calculate by hand or to program the calculator).

Suppose that we wish to find the parameters of a current mirror operating at 1.0 mA and a collector voltage of 1.5 V. We assume $R_{E1} = R_{E2} = 0.10$ k Ω . The $ABCD$ parameters of the output transistor at this bias condition for a $95 \mu\text{m}^2$ transistor are given in Fig. 9.9*a*, as obtained from the equivalent circuit in Chapter 7 by using program "E>A" in Appendix C. The diode transistor parameters are similar, except that they are translated to a collector voltage of 0.75 V. Following the procedure outlined previously, we find the $ABCD$ parameters of the current mirror, shown in Fig. 9.9*b* at a frequency of 0.316 MHz.

Note in Fig. 9.9*b* that the magnitude of D is $1.02 = 1 + 2\delta$ and that the phase has departed from 180° by only 12.7° at this frequency. The output admittance

ABCD MATRIX OF TRANSISTOR	ABCD MATRIX OF CURR. MIRROR	POLYS
FREQ., GHZ = 0.3160	FREQ., GHZ = 0.3160	A-POLY:
ABCD, MAG.+PH:	ABCD, MAG.+PH:	R00= -0.0013
A:	A:	R01= -0.0245
R01= 0.0257	R01= 0.0490	R02= -0.0028
R02= -80.4671	R02= -78.3615	R03= -0.0001
B:	B:	B
R03= 0.0324	R03= 0.1320	R04= -0.1323
R04= -161.8031	R04= -169.8698	R05= -0.0118
C:	C:	R06= -0.0006
R05= 0.2369	R05= 0.6079	R07= -2.2803-05
R06= -78.8343	R06= -80.1385	C
D:	D:	R08= -0.0101
R07= 0.0046	R07= 1.0217	R09= -0.3044
R08= -95.2124	R08= -167.3354	R10= -0.0290
(a)	(b)	R11= -0.0007
		D
		R12= -1.0229
		R13= -0.1140
		R14= -0.0066
		R15= -0.0002
		(c)

Figure 9.9. The $ABCD$ parameters of current mirror (b) from those of the transistor (a) and diode. In part c these parameters are converted into polynomial coefficients.

for a high-source impedance is, according to eq. (8.2-9)

$$Y_o = \frac{A + CR_G}{B + DR_G} \approx \frac{C}{D} \quad (9.2-6)$$

Since D is approximately -1 , the output admittance is just $-C$, or 0.60 mS at 0.316 GHz. Since the phase is near 90° , the output admittance is that of a capacitor of $0.6/(2\pi \cdot 0.316) = 0.3$ pF.

By repeating the calculation at two frequencies, we can obtain the polynomial coefficients of the $ABCD$ parameters by use of program "T>P3" in Appendix C. The results shown in Fig. 9.9c were obtained by using 0.0316 GHz as the second frequency; several other pairs of frequencies were tried with negligibly different results.

Turning now to the current loss expressed by the D polynomial in Fig. 9.9c, we normalize the polynomial to its cubic coefficient with the use of eq. (2.2-3)

to obtain the character of the cubic. The D polynomial can be written

$$D = -1.023(1 + 0.111s + 0.065s^2 + 0.0002s^3) \quad (9.2-7)$$

$$= -1.023(1 + 1.92p + 1.92p^2 + p^3) \quad (9.2-8)$$

where $p = 0.0584s$. The surprising result is that the loss is essentially Butterworth cubic, with a cutoff frequency of $1/(2\pi \cdot 0.0584) = 2.7$ GHz. The result is surprising because of its cubic character and because of its wide bandwidth. Consider the bandwidth first. The unity loss frequency of the transistor is obtainable from the magnitude of D in Fig. 9.9a as $f/|D| = 0.316/0.0846 = 3.7$ GHz. The bandwidth of the current mirror is more than half the unity loss frequency of the transistor. On the other hand, the first term of the (unnormalized) cubic is approximately $2\tau_T$, taking into account the larger value of τ_T of the diode (because of its lower operating voltage). Hence the simple model discussed earlier is only good up to modestly high frequencies (e.g., 0.5 GHz).

The transmission of the current mirror is "broadbanded" by an equivalent inductance of the diode arising from r_b of the diode transistor and an equivalent effect in the transistor. We need not be concerned with the details here; the exact calculation picks up these small but significant effects. As Fig. 2.18 shows, the cubic Butterworth polynomial can be represented accurately up to half its cutoff frequency by a pure delay, so that, up to a frequency of about 1.5 GHz, we can represent the current loss of this current mirror by

$$D = -1.023e^{0.111s} \quad (9.2-9)$$

The delay term in the exponent is d_1/d_0 from Fig. 9.9c. The low-frequency value of D is dominant over a wide frequency range. Where the source impedance to the mirror is high and the load impedance is low, the performance of the circuit is close to that of the ideal current mirror matrix of (9.2-1).

Wilson Current Source⁴

The simple current source can be improved, as to both input to output current ratio accuracy and output conductance by use of the Wilson current source, shown in Fig. 9.10. In this circuit the simple current source Q_1 , Q_2 is incorporated as a spanning network on an output transistor Q_3 . To analyze this circuit, we can make simple estimates of the $ABCD$ parameters and check them by an exact calculation, as in the previous case.

To estimate D and B , assume that a unit collector flows in the output transistor. The base current is then δ , as shown. The base voltage is $2r_E$, including the drop across the diode, so that B is twice that of the simple current source. The reason for the accurate current ratio of unity for the Wilson current source can be seen by inspection of the currents in Fig. 9.10b. Assuming a unit output current into the collector of Q_3 , the emitter current of

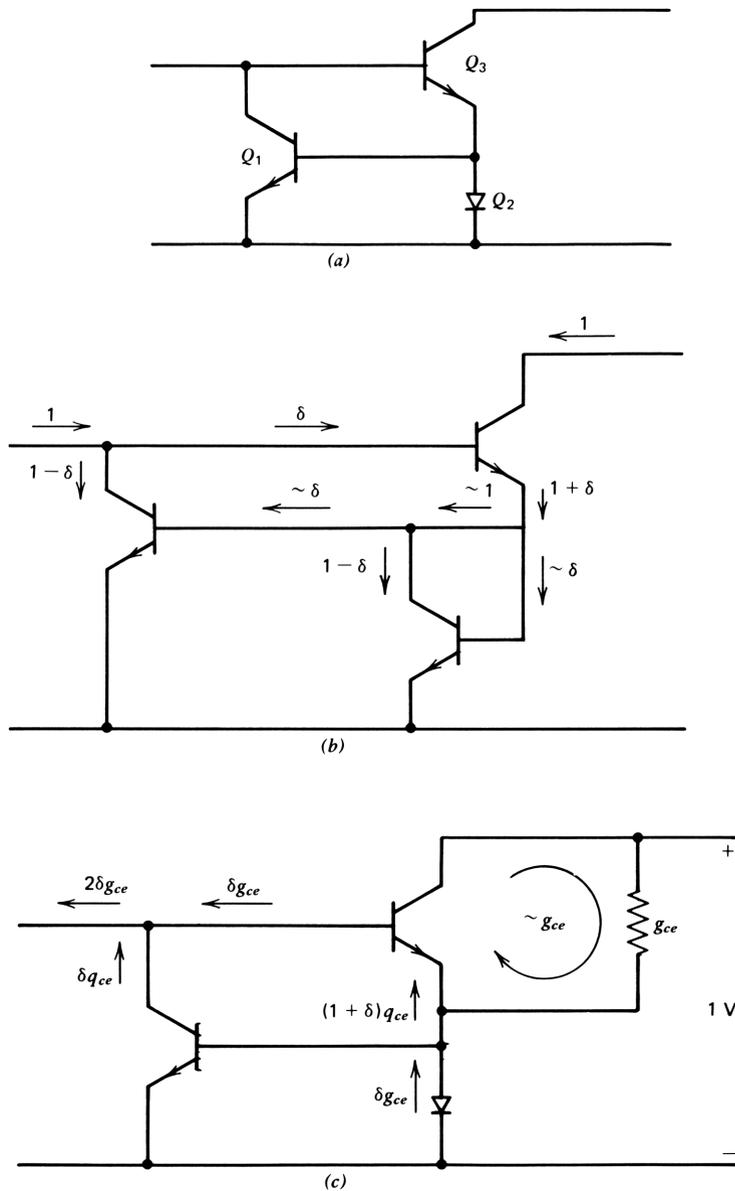


Figure 9.10. Wilson current source: (a) circuit; (b) circuit analysis of dc current transmission; (c) circuit analysis of A and C .

Q_3 is $1 + \delta$. The base currents of Q_1 and Q_2 (the diode transistor) are both approximately δ , so that the current in the collector of Q_2 is $1 - \delta$. This current is mirrored at the collector of Q_1 . When the base current of Q_3 is added to this, we obtain the input current as unity, with all terms linear in δ having canceled. (Second-order terms do not cancel: the base currents of Q_1 and Q_2 are both slightly smaller than δ since the collector currents of both transistors are slightly smaller than unity.)

The exact value is found by analysis of the simple current source spanning network. The spanning network is characterized by its g parameters, of which only g_{12} is of prime significance. But g_{12} is the current *gain*, the reciprocal of D of the simple current source, which we found in (9.2-3) to be $1 + 2\delta$. Hence

$$g_{12} = \frac{1}{1 + 2\delta} \quad (9.2-10)$$

The input current to the diode is $1 + \delta$, so the total input current to the circuit is

$$D = \delta + \frac{1 + \delta}{1 + 2\delta} = 1 + \frac{2\delta^2}{1 + 2\delta} \quad (9.2-11)$$

Since $\delta \ll 1$, current accuracy is much better for this circuit than for the simple current source. Thus B is twice that of the simple current mirror, and D is closely unity. This correction extends at least partially to high frequencies since the collector current of the spanning network transistor falls with frequency as the base current of the output transistor rises, giving some cancellation.

To estimate A and C , we use the open-circuited output in Fig. 9.10c and assume 1 V at the output. The incremental current through g_{ce} of the output transistor flows upward through r_e of the output transistor as shown, giving $A = -r_e g_{ce}$ (this circulatory current does not flow through the spanning network). Hence the base current of the output transistor is $-\delta g_{ce}$, and this current also flows into the diode of the spanning network, as shown. The same current flows out of the collector of the spanning network, so that the input current of the circuit for unit output voltage is $-2\delta g_{ce}$; this is the estimate of C .

The estimate at low frequencies of the $ABCD$ matrix of the Wilson current source from the preceding considerations is about

$$T_{\text{Wlsn}} = - \begin{bmatrix} r_e g_{ce} & 2r_E \\ 2\delta g_{ce} & 1 \end{bmatrix} \quad (9.2-12)$$

The output conductance is C/D as before and is thus $2\delta g_{ce}$, a factor of 2δ smaller than that of the simple current source. This is the chief advantage of the Wilson source and to a great extent eliminates Early effect from the circuit.

The high-frequency transmission of the circuit can be estimated with the aid of the simplified equivalent circuit in Fig. 9.11. In it, the current mirror

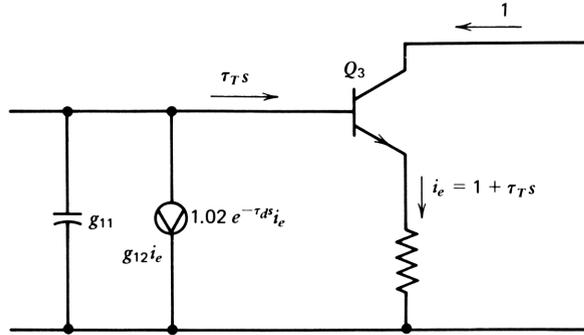


Figure 9.11. Circuit for estimating the high-frequency performance of the Wilson current source.

spanning network has been replaced by its g -parameter, dependent generator equivalent circuit. The output of the mirror spanning network is connected to the base of Q_3 , whereas the input is connected to the emitter circuit, so that in finding the g parameters from the $ABCD$ parameters, the latter matrix must be inverted before the conversion to g parameters is effected. The feedback generator is $-1/D$, where D is given by (9.2-7), or as approximated by (9.2-8). Parameter g_{11} is the output admittance of the current mirror, which we found to be that of a 0.3 pF capacitor. Parameter g_{22} is the input impedance of the current mirror. For this initial estimate, we take g_{22} as b_0 of the polynomial coefficients in Fig. 9.9c. The feedforward is neglected for the present.

Assuming an output current of unity, we find the input current by inspection of Fig. 9.11. The input base current to Q_3 is $\tau_T s$ (ignoring δ of the transistors for this high-frequency calculation). The spanning network adds a current to the input equal to the emitter current of Q_3 delayed by the delay of the spanning network as found in the current mirror analysis. The input current, ignoring the 0.3 pF output capacitance of the spanning network, is thus

$$i'_i = \tau_T s + (1 + \tau_T s) e^{-\tau_d s} \quad (9.2-13)$$

We then add the current through the output capacitance of the spanning network. To do this, we must estimate the voltage across it: it is approximately $R_E + r_{E3} + r_{E2} = R'_E$ multiplied by the emitter current $1 + \tau_T s$. The total input current for unit output current into the collector of Q_3 is $-D_{\text{Wlsn}}$, the D parameter of the Wilson current source, and is given by

$$\begin{aligned} -D_{\text{Wlsn}} &= R'_E C_i s (1 + \tau_T s) + \tau_T s + (1 + \tau_T s) e^{-\tau_d s} \\ &= (R'_E C_i + \tau_T) s + R'_E C_i \tau_T s^2 + (1 + \tau_T s) e^{-\tau_d s} \end{aligned} \quad (9.2-14)$$

Multiplying both sides by $e^{\tau_d s}$ has the effect of removing the external (closed-loop) delay, as noted in Chapter 5. Letting $R'_E C_i + \tau_T = \tau_1$, and changing the

order of the terms, we obtain

$$-e^{\tau_d s} D_{\text{Wilson}} = 1 + \tau_T s + (\tau_1 s + R'_E C_i \tau_T s^2) e^{\tau_d s} \quad (9.2-15)$$

Except for the $\tau_T s$ term and the $R'_E C_i \tau_T s^2$ term in the brackets, this is the classic feedback equation discussed in Section 5.2. To ensure stability, we can infer that τ_1 must be made large enough to ameliorate the effect of delay. Clearly, this may be done by increasing either R_E or by adding capacitance across the input of the circuit.

The values of the rational function coefficients describing the circuit can be found by evaluating the $ABCD$ parameters at a number of frequencies and converting these measurements to coefficients by use of the rational function evaluation procedure given in Chapter 2. Since a transistor is included in the spanning network, we must include the effect of the denominator as well as the numerator. The procedure described in Chapter 2 is expanded in Section 9.3 to rational functions, including denominators. First, we find the $ABCD$ parameters of the Wilson current source.

To find the $ABCD$ parameters, we may first evaluate the spanning network parameters and convert them to g parameters. These are added to the equivalent ladder network consisting of the output transistor. It is often simpler and more interesting to use topological manipulation to find the $ABCD$ parameters. Figure 9.12 describes such a method using lead permutation, inversion, and multiplication (cascading) of $ABCD$ matrices. Five operations are needed, each available on the calculator (using program “ABCD” in Appendix C) at the push of one or two buttons.

The procedure begins with the $ABCD$ parameters of the simple current mirror such as those given in Fig. 9.9*b*. In Fig. 9.12*a* the output leads of the current mirror are permuted. In Fig. 9.12*b* the resulting $ABCD$ matrix is inverted and the result stored. (In each part of the figure, the result of a previous operation is enclosed in a box drawn with dashed lines.) In Fig. 9.12*c* the $ABCD$ parameters of Q_3 are read in and the input leads permuted. In Fig. 9.12*d* the result of the step in Fig. 9.12*b* is cascaded with that in Fig. 9.12*c*. Finally, the input leads of the combination are permuted, giving the $ABCD$ parameters of the Wilson current source. To prove that it is indeed that of the desired circuit, note that the input terminal of the circuit is connected to the base of Q_3 and the collector of Q_1 ; the output is connected to the collector of Q_3 ; and the two resistors are connected to the common lead.

Although this procedure is lengthy to describe, the algorithm is compact. It is in a way similar to putting a picture puzzle together, turning pieces to see how they may fit together. Perhaps a better analogy is to chemical bonding; the transistors (the molecules) fit together according to various rules, which relate to directions of biases and current and voltage levels, rather than to attractive forces.

Once we have a program, the $ABCD$ parameters can be evaluated for as many frequencies as we choose. We can obtain two rational function coeffi-

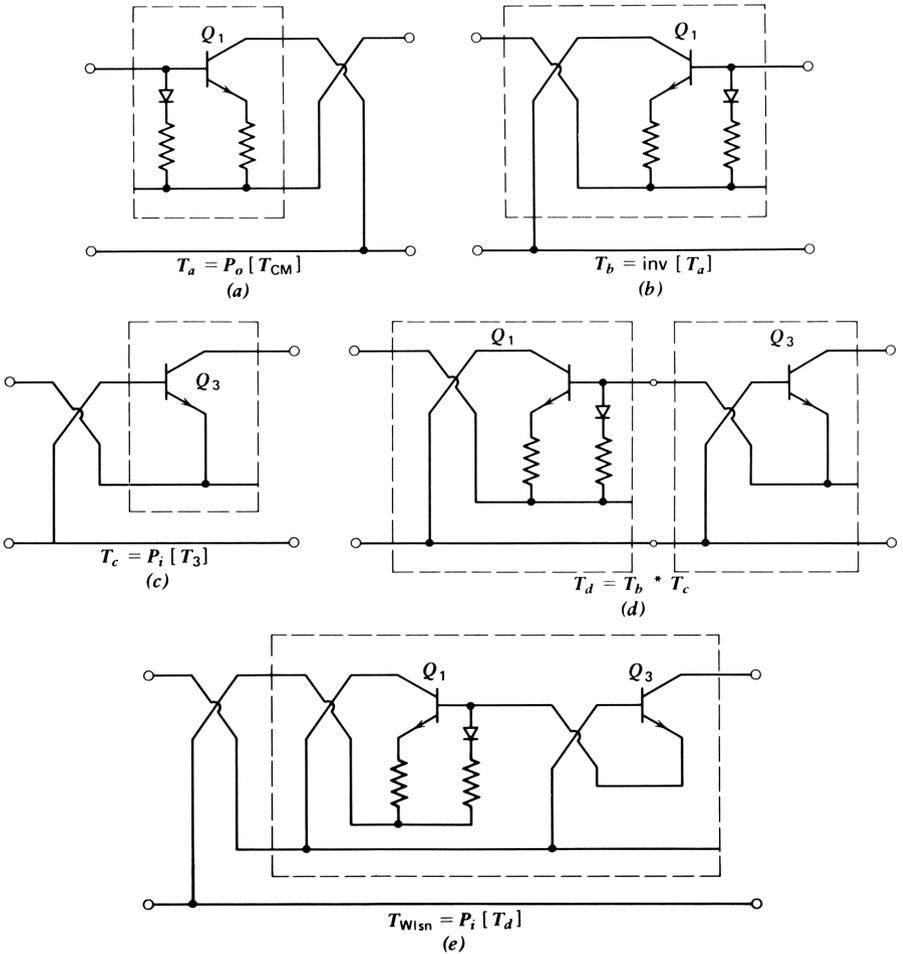


Figure 9.12. Analysis of Wilson current source by topological manipulation.

cients for each such evaluation and for each parameter. Since we have been concerned primarily with polynomial evaluation, it is worthwhile at this point to review what is involved in finding the numerator polynomial when a denominator is present.

9.3 EVALUATION OF THE NUMERATOR POLYNOMIAL COEFFICIENTS OF A RATIONAL FUNCTION (WITH KNOWN DENOMINATOR COEFFICIENTS)

The feedback current of the Wilson current source includes a denominator (in fact, the numerator is unity). Thus the loss of the circuit includes this denominator, whose coefficients were found in Section 9.2. They are listed as

the D polynomial coefficients in Fig. 9.11c. The rationale for this is that the current loss of the circuit is the sum of the feedback current and the D parameter of Q_3 , so that the current loss must include this denominator. This provides the immediate motivation for this section, but the application is quite general. We can often analyze the spanning network, find its denominator coefficients (if any), and use this knowledge to evaluate the numerator coefficients from calculations (or measurements) of the loss and phase of the circuit at several frequencies.

In this section we find the coefficients of a cubic numerator in the presence of a cubic denominator; the extension to functions of any degree will be clear. It will be found that the evaluation is merely a slight extension to the method of polynomial coefficient evaluation given in Section 2.5. The extension is easily applied to programs "RCU" and "RQU" in Appendix A and to "T>P3" in Appendix C (it is included in the latter program).

Let the loss of a circuit be expressed by the rational function

$$L(s) = \frac{a_0 + a_1s + a_2s^2 + a_3s^3}{1 + d_1s + d_2s^2 + d_3s^3} \quad (9.3-1)$$

The term $L(s)$ may be expressed in terms of two even functions of s :

$$L(s) = M(s) + sN(s) \quad (9.3-2)$$

Thus

$$L(j\omega) = M + j\omega N \quad (9.3-3)$$

where M and N are real (for $s=j\omega$). The denominator can also be split into real and imaginary parts:

$$D(j\omega) = 1 - d_2\omega^2 + j\omega(d_1 + d_3\omega^2) \quad (9.3-4)$$

The product $L(j\omega)D(j\omega)$ is equal to the numerator polynomial $A(j\omega)$, so that

$$A(j\omega) = a_0 - a_2\omega^2 + j\omega(a_1 - a_3\omega^2) \quad (9.3-5)$$

Since both $L(j\omega)$ and $D(j\omega)$ are presumed known (at two frequencies in this case), we can equate the real and imaginary parts of the product with the real and imaginary parts of $A(j\omega)$, respectively.

We may then proceed exactly as in Section 2.5, replacing $L(j\omega)$ in that section by the product $L(j\omega)D(j\omega)$. Thus the real part of $L(j\omega)D(j\omega)$ is

$$\begin{aligned} \operatorname{Re}[L(j\omega)D(j\omega)] &= M(1 - d_2\omega^2) - N\omega^2(d_1 - d_3\omega^2) \\ &= a_0 - a_2\omega^2 \end{aligned} \quad (9.3-6)$$

and the imaginary part is

$$\begin{aligned} \frac{\text{Im}[L(j\omega)D(j\omega)]}{\omega} &= M(d_1 - d_3\omega^2) + N(1 - d_2\omega^2) \\ &= a_1 - a_3\omega^2 \end{aligned} \quad (9.3-7)$$

By evaluating the circuit at two frequencies, we obtain four equations, allowing us to solve for the four a_i coefficients; the equations are given by eqs. (2.5-6) to (2.5-9), in which we must replace $L(j\omega)$ by $L(j\omega)D(j\omega)$. The change is

CMN DEN: d0=1 d1, d2, d3:	CMN DEN: d0=1 d1, d2, d3:
R44= 0.1140 R45= 0.0063 R46= 0.0002	R44= 0.1140 R45= 0.0063 R46= 0.0002
POLYS	POLYS
A-POLY:	A-POLY:
R00= -0.0006 R01= -0.0281 R02= -0.0048 R03= -0.0006	R00= -0.0006 R01= -0.0281 R02= -0.0048 R03= -0.0006
B	B
R04= -0.1618 R05= -0.0322 R06= -0.0029 R07= -0.0002	R04= -0.1618 R05= -0.0322 R06= -0.0029 R07= -0.0002
C	C
R08= 0.0041 R09= -0.2323 R10= -0.0356 R11= -0.0042	R08= 0.0133 R09= -0.2329 R10= -0.0403 R11= -0.0090
D	D
R12= -0.9994 R13= -0.1663 R14= -0.0174 R15= -0.0016	R12= -0.9958 R13= -0.3282 R14= -0.0404 R15= -0.0044
(a)	(b)

Figure 9.13. Polynomial coefficients for the $ABCD$ parameters of the Wilson current source, with known common denominator coefficients: (a) without input capacitor; (b) with added input capacitor; (c) with current transmission loss.

equally applicable to matrix equations for the quintic polynomial given in eqs. (2.5-12) and (2.5-13).

The modification to the programs consists of replacing M —the real part of $L(j\omega)$ —by $M(1-d_2\omega^2)-N\omega^2(d_1-d_3\omega^2)$ and replacing N by $M(d_1-d_3\omega^2)+N(1-d_2\omega^2)$. This has been done in program “T>P3” in Appendix C.

The operations of Fig. 9.12 were carried out at two frequencies (0.1 and 1.0 GHz), and the results were translated into polynomial coefficients by using program “T>P3”, as shown in Fig. 9.13a. The current transmission loss of the circuit is shown plotted in the lower curve in Fig. 9.13c and is a Bode plot of the rational function given by the D polynomial divided by the common denominator polynomial, all listed in Fig. 9.13a.

A hazardous dip in loss in this plot indicates that the circuit is close to instability in the vicinity of 1.4 GHz. (A repetition of the entire calculation using frequencies of 1.0 and 1.5 gave negligibly different results, but the low-frequency loss was slightly in error.) The complex roots of the numerator cubic polynomial are at 1.43 GHz at an angle of 100.4° , giving a damping factor of only 0.18. (The real root is at -1.26 GHz.) As noted in Section 9.2, the circuit stability margin can be improved by adding capacitance across the input terminals. The effect of adding 1 pF at the input is shown in the polynomial in Fig. 9.13b, and the resulting Bode plot is shown in the upper curve in Fig. 9.13c. Although stability is improved, the margin is still rather small: the complex roots of the numerator have moved to 1.16 GHz at an angle of $\pm 109.6^\circ$, giving a damping factor of 0.34. The real root moves to -0.67 GHz.

The output admittance of the circuit for high source impedance is given by C/D , as noted previously. Since $D = -1$, the output admittance polynomial

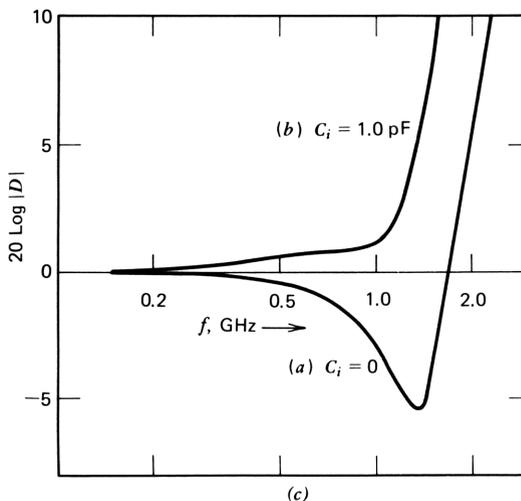


Figure 9.13. Continued.

coefficients are the negative of the C polynomial in Fig. 9.13a or 9.13b. In both cases the output conductance (c_0) is shown in negative and very small. A calculation at a lower frequency would show that it is actually slightly positive. The important coefficient is c_1 , which is the negative of the output capacitance, 0.23 pF. This is a bit smaller than that found previously for the simple current mirror (0.3 pF). The Wilson current mirror provides a drastic improvement only in the output conductance, not the capacitance. This improvement, as well as the more accurate dc current ratio, is bought at the price of a potential stability problem, so the circuit should be used with care.

9.4 ACTIVE LOADS

An important application for current mirror circuits is the provision of an active load for signal circuits such as the emitter-coupled pair. In this service the current mirror acts as a spanning network whose prime function is that of feedforward rather than feedback; hence this technique is of interest in developing the theory of spanning networks.

A differential pair with a current mirror active load is shown in Fig. 9.14a. Our purpose is (1) to give a general description of its operation and then (2) to find the $ABCD$ parameters of the network including it.

The collector signal currents of the two transistors of an emitter-coupled pair are opposite in direction and almost exactly equal in magnitude. By mirroring the output of one and adding it to the other, the output current of the pair is doubled for a given input; put the other way around, D of the pair is halved. The basic operation of the circuit is to *add* a signal to the output from a point earlier in the circuit; hence the purpose of the circuit is to add feedforward to (1) raise the current gain by a factor of 2 and (2) double the output current capability of the emitter-coupled pair.

As we have noted, any spanning network provides both feedback and feedforward; usually one or the other is incidental to the purposes of the network. For this circuit, the feedback is incidental and is for practical purposes negligible.

The circuit is redrawn in Fig. 9.14b, where Q_1 is shown as an emitter follower and Q_2 is a common base stage. The mirror is drawn as a spanning network that has the collector current of Q_1 as its input; this current is mirrored by Q_4 and added to the output current of Q_2 .

The current relationships are shown in Fig. 9.14b. Assuming a unit collector current of Q_2 , the emitter current is $1 + \delta$, which is also the emitter current of Q_1 . (We ignore any admittance of the emitter current source; its effect is described in Chapter 12.) The collector current of Q_1 is also unity, and this current forms the input of the current mirror. The current *gain* of the mirror is $-1/D$, or $1/(1+2\delta)$; the output current of the mirror is added to the circuit output, so that the output current of the circuit is $i_{c2} + i_{c4}$, or $1 + 1/(1+2\delta)$, approximately 2.

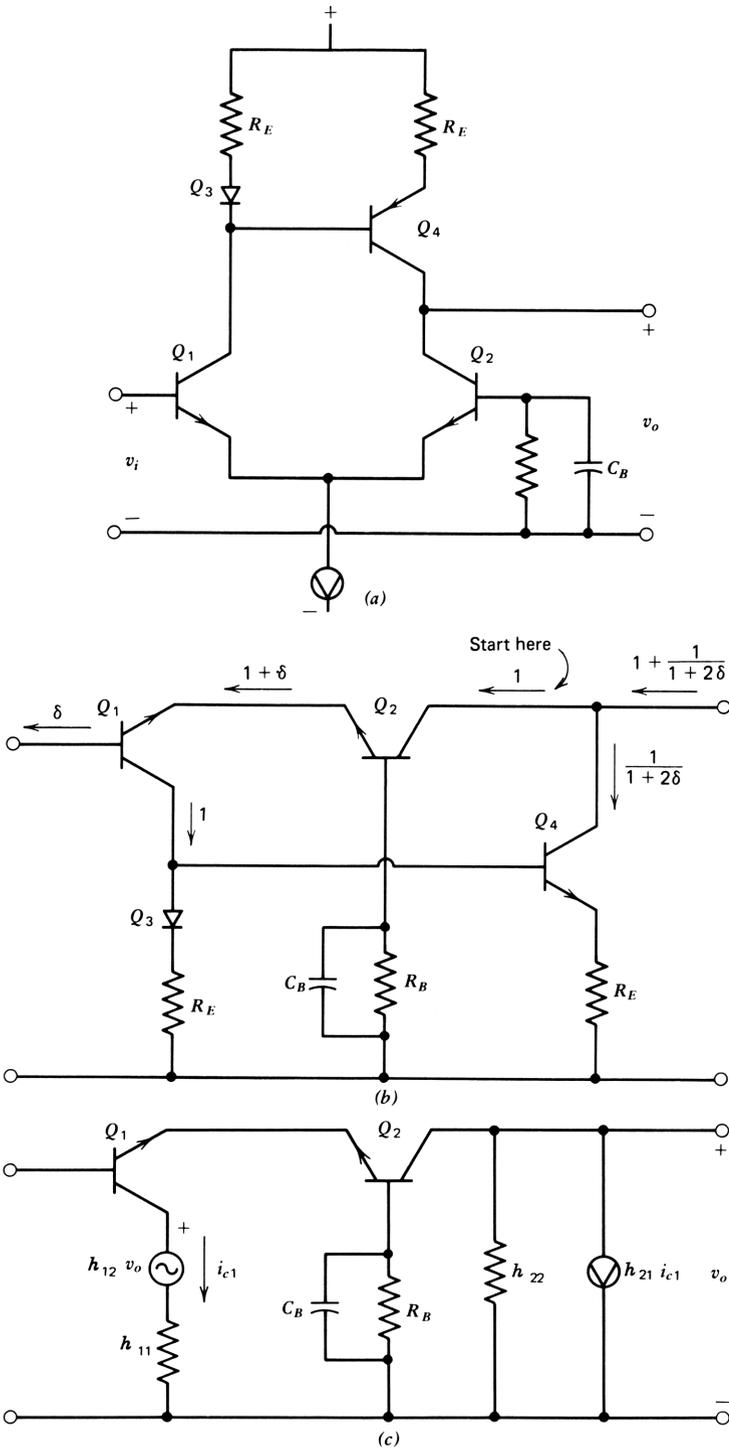


Figure 9.14. Use of a simple current mirror as an active load for a (noninverting) differential pair. Active load is treated as a feedforward spanning network.

These relationships are formalized in Fig. 9.14c in which the current mirror spanning network is represented by its h parameters. The desired effect is represented by the output generator $h_{21}i_{c1}$; h_{22} represents output loading by the current mirror and is small. At the input side h_{11} is the impedance of the diode and current mirror resistor in series; $h_{12}v_o$ is a feedback voltage, the current mirror input voltage: output voltage ratio. Its effect on circuit operation is negligible since it faces a high-impedance collector and is very small in the first place.

Since both the differential pair and the spanning network are grounded, we cannot calculate the $ABCD$ parameters by lead permutation exclusively; the general method described in Chapter 8 must be employed. A program for this was written, and the results are shown in Fig. 9.15a for a differential pair operating at 1.0 mA with a supply of 5 V. In this calculation the parallel combination of 0.1 k Ω and 1.0 pF was connected between the base of Q_2 and ground to account for a feedback network to be applied later. The frequency is 0.316 GHz.

ABCD MATRIX OF NON-INV D.P.	ABCD MATRIX OF INV D.P.
FREQ., GHZ = 0.31623	FREQ., GHZ = 0.31623
ABCD, MAG.+PH:	ABCD, MAG.+PH:
A:	A:
R01= 0.04936	R01= 0.03610
R02= 100.32621	R02= -72.29854
B:	B:
R03= 0.03447	R03= 0.03491
R04= 27.12102	R04= -151.60009
C:	C:
R05= 0.03377	R05= 0.14978
R06= 181.58088	R06= -74.49825
D:	D:
R07= 0.04650	R07= 0.03687
R08= 93.77776	R08= -86.96787
(a)	(b)

Figure 9.15. The $ABCD$ parameters of differential pairs with active load: (a) noninverting; (b) inverting.

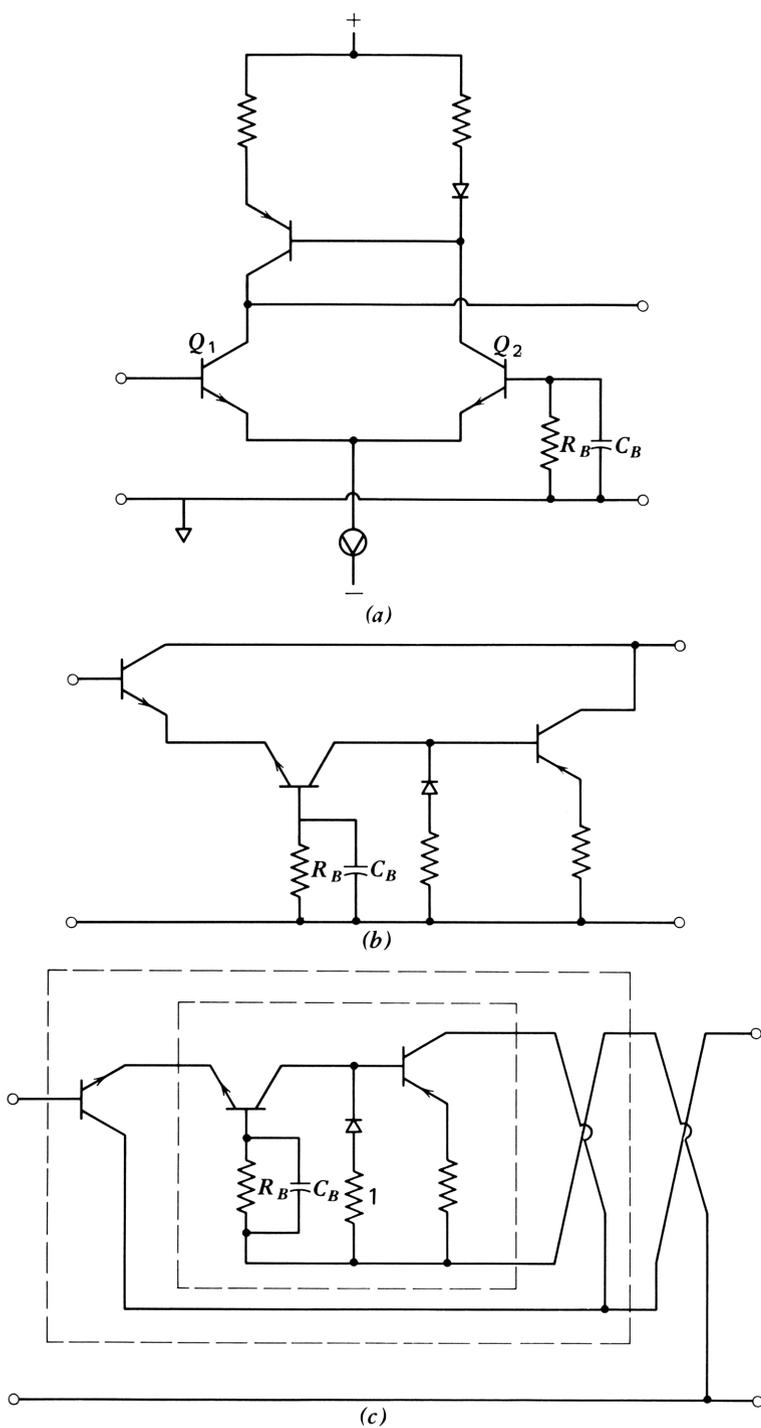


Figure 9.16. Inverting differential pair with active load: (a) circuit; (b) circuit redrawn; (c) analysis by topological manipulation.

The value of B is approximately that for a single transistor since only half the output current flows through the emitter resistances of Q_1 and Q_2 . The value of C is extremely small and is a negative conductance (with a 12 dB per octave positive slope with frequency). The value of D is likewise small, corresponding to a unity loss time constant of $0.0465/(2\pi \times 0.316) = 0.23$ ns. This value is a little more than half the τ_T of the transistors; the feedforward makes the differential pair broader in bandwidth than the common emitter stage.

The circuit in Fig. 9.14 is a noninverting stage. A differential pair that has a phase reversal is obtained by exchanging the roles of the bases of Q_1 and Q_2 , as shown in Fig. 9.16a. This circuit may be redrawn as in Fig. 9.16b; the spanning network can now be taken as the conductor from the collector of Q_1 to the output. Since this spanning network is ungrounded, we can calculate the $ABCD$ parameters of the circuit by a series of lead permutations, as shown in Fig. 9.16c. The characteristics of the circuit were calculated in this way, with the results shown in Fig. 9.15b. The transistors and circuit are the same as for the noninverting circuit except for the exchange of bases of Q_1 and Q_2 .

Note that B remains essentially unchanged except for the phase reversal. This is to be expected since B is essentially the sum of the emitter resistances of Q_1 and Q_2 divided by 2. The value of C increases since the collector capacitance of Q_1 is now connected between circuit input and output. However, D becomes smaller, primarily because h_{11} of the spanning network is no longer in series with the collector of Q_1 , thus eliminating the collector time constant $h_{11}C_{jc1}$ from D .

The $ABCD$ matrices of the inverting and noninverting stages are not identical. This leads to imperfect *common mode rejection* at high frequencies, a subject we discuss in Chapter 12.

9.5 ACCURATE OUTPUT CURRENT SENSING: B AND D FEEDBACK

The current mirror circuits are examples of D feedback in which a circuit that is basically a B -feedback amplifier is converted to D feedback by addition of a shunt conductance at the input, as in Fig. 8.6c. The effect of B feedback is minimized by operating the circuit from a high-source impedance. Other configurations in which D feedback is provided alone—without incidental B feedback—are also possible. One of the simplest of such configurations is shown in Fig. 9.17, in which the output current is sensed at the emitter of the second stage. This emitter voltage causes a current to flow through G_F to the input, where it augments the input current. Where the internal feedback paths of the individual transistors can be ignored (i.e., the transistors are in their reference condition), the $ABCD$ matrix of the combination contains only one nonzero element, in the D position. Therefore, it is a unitary D -feedback amplifier.

Sensing the output current at the emitter of the output stage has serious limitations for critical applications. Since the emitter current contains not only

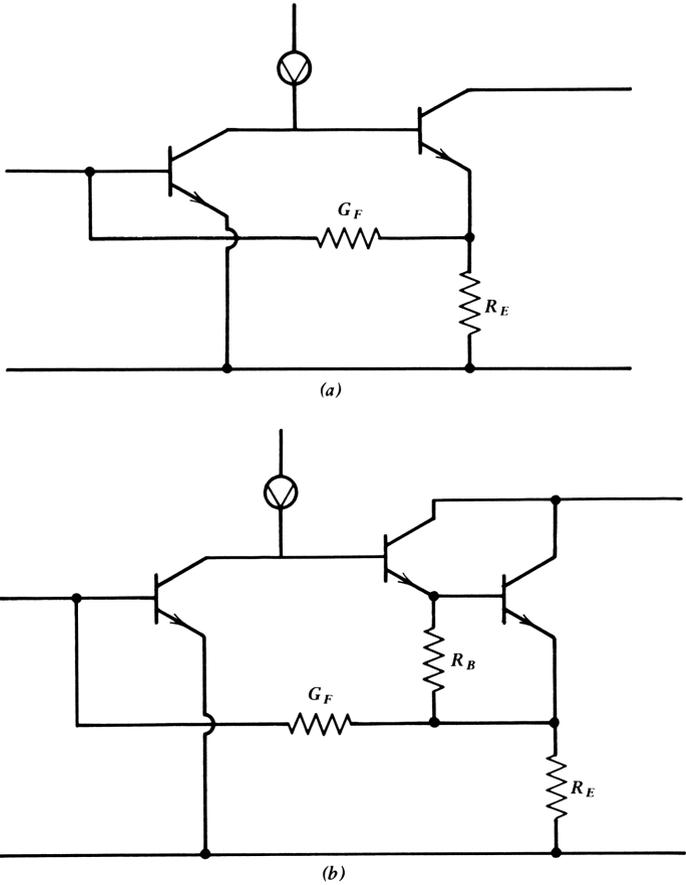


Figure 9.17. A *D*-feedback pair is shown in part *a*. In part *b* a Darlington pair is substituted for the output transistor to obtain accurate output current sensing.

the output current but the base current as well, the resulting feedback current is proportional not only to the output current, but to the base current. Thus if the base current is nonlinearly related to the collector current, the feedback current, and hence the input current, will contain a nonlinear component when the output current is perfect. The problem here is the imperfect sensing of the output current and may be all but eliminated by making the emitter current more closely approximate the output current. One method is to substitute a Darlington pair for the output transistor as shown in Fig. 9.17*b*, in which bias current for the first transistor of the pair is provided by the resistor between the emitters. According to Kirchoff's current law, the difference between the (composite) emitter current and the output current is the base current of the first transistor of the pair, a current that is a factor δ smaller than the base current of the simple stage. Almost perfect output current sensing can be achieved in this way.

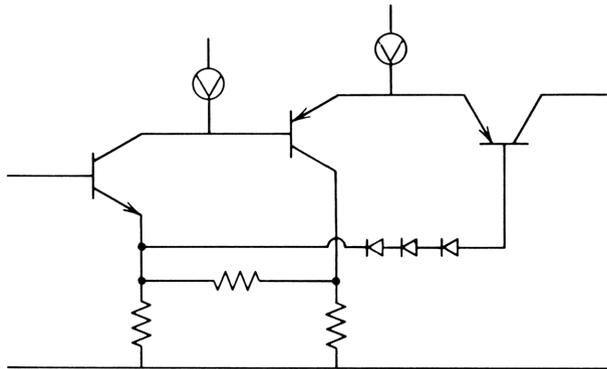


Figure 9.18. A B -feedback amplifier with accurate output current sensing.

Adding the transistor that converts the output stage to a Darlington pair raises the degree of the characteristic polynomial by 1. For unit output current, the base current of the simple output stage contains a term proportional to s (and some delay due to transit time). When the Darlington pair transistor is added, its base current contains a term proportional to s^2 . To retain control of the intermediate polynomial coefficient, a capacitor may be connected from collector to base of the output transistor, as discussed in Section 5.5.

Another method for accurately sensing the output current is shown in the B -feedback amplifier in Fig. 9.18. In this circuit Q_2 and Q_3 operate as an emitter-coupled pair, so that the collector signal currents of the two transistors are almost exactly equal and opposite, as are the base currents. The three series diodes provide collector bias voltage for Q_1 . The collector current of Q_2 is divided by the resistive feedback network and causes a feedback voltage to appear at the emitter of Q_1 . It follows that this component of feedback is almost perfect. The base current of Q_2 , however, also flows through the external emitter resistance of Q_1 (through the collector-emitter path of this transistor), giving imperfect output current sensing. This component is canceled by the equal and opposite base current of Q_3 , so that the emitter voltage of the first stage is solely proportional to the output current.

PROBLEMS

- 1 Find expressions for the input resistance and output conductance of the simple current mirror with a diode that has an emitter area r times that of the transistor.
- 2 Find the approximate low-frequency input current and output conductance for the current mirror stage in Fig. 9.19.
- 3 The circuit of Fig. 9.20 is a cascode current source. Find its input current for unit output current and its output conductance.

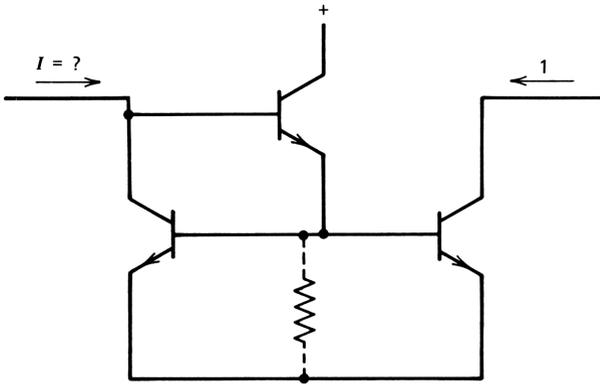


Figure 9.19.

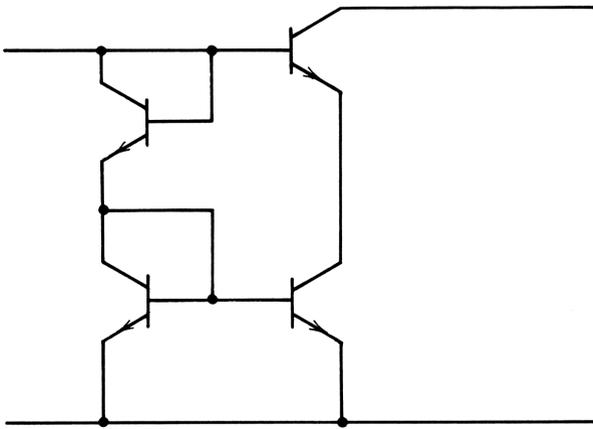


Figure 9.20.

REFERENCES

1. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Wiley, New York, 1977.
2. A. B. Grebene, *Analog Integrated Circuit Design*, Van Nostrand Reinhold, New York, 1972.
3. S. J. Mason, "Power Gain in Feedback Amplifiers," *Trans. IRE CT-1* (1), 20–25 (1954).
4. G. R. Wilson, "A Monolithic Junction FET-NPN Operational Amplifier," *IEEE J. Solid-State Circuits SC-4*, 341–348 (December 1968).

Chapter 10

Output-Stage Design

Output stages of operational amplifiers, power supply regulators, and in general any analog signal processing circuit may be called on to provide signals at a power level that causes the signal to vary over the complete bias range of the output device or devices. As we saw in Chapter 7, the transmission parameters vary with bias, so that the loss of the circuit will be dependent on the signal level. It follows that the input signal is a nonlinear function of the output current and voltage. Three sources of nonlinearity discussed in Chapter 7 strongly affect output stages and are dealt with here: (1) emitter junction nonlinearity; (2) avalanche multiplication; and (3) saturation.

We begin by considering a single transistor used as an output stage; operation is restricted to the single quadrant of the I_c-V_{ce} plane in which the transistor is in its forward-active region of operation. The collector current can vary above and below its quiescent value, generally from nearly zero to twice its quiescent value. Where the signal varies above and below a quiescent value, operation is termed *Class A*. The power input from the power supply is constant in this class of operation and is independent of signal level. Power dissipation in the output stage tends to be high and is highest in the absence of signal since none of this constant power is delivered to the load. Hence Class A operation requires considerable *standby power* in the absence of input signals. The discussion is then extended to include push-pull operation with the use of

two devices. This adds one extra quadrant to the signal range and improves efficiency by roughly a factor of 2, still retaining Class A operation.

In the next section *Class B* operation, in which the standby power is reduced or removed, is considered. In this class of operation the power input from the power supply depends on the input signal; considerably less power is dissipated in the devices. It is a natural choice for higher-power integrated circuits.

In both classes of operation nonlinearities are much more easily analyzed as predistortion at the input to produce a desired output waveform—a sinusoid, for example. Since nonlinearities occur in both input current and voltage, the nonlinearities can be added at the input (through the source impedance) to give the total nonlinearity, as discussed in Chapter 1. Many problems of the conventional formulation, such as defining and eliminating transient distortion, for example, do not arise in this approach. Drive requirements for the output stage are formulated compactly, making the design of the driver stage simple and direct.

The final section of this chapter is devoted to a description of the variation of transistor dc characteristics with temperature. Means are developed to design circuits that have a prescribed variation of currents and voltages with temperature. As examples, the design of band-gap voltage reference circuits¹ is described, as well as the design of the temperature-dependent quiescent current for Class B output stages to maintain low-distortion operation over the operating temperature range.

10.1 A COMMON EMITTER OUTPUT STAGE

In the circuit in Fig. 10.1a the upper transistor Q_2 supplies a quiescent current I_Q to the lower transistor Q_1 , a common emitter amplifier. Current I_{in} is adjusted in the absence of signals to give a quiescent output voltage of zero. (The adjustment may be made by feedback from the output terminal and is not shown in Fig. 10.1a.) Transistor Q_2 is a current source transistor as discussed in Section 9.2. The I_c - V_{ce} plane for Q_1 (the collector characteristics) is shown in Fig. 10.1b; the light lines show I_c and V_{ce} for various values constant of base current, but these lines are immaterial for the present discussion. From Fig. 10.1a, we can write

$$I_{c1} = I_Q - I_o \quad (10.1-1)$$

and

$$V_{ce1} = V_o + V_{CC} \quad (10.1-2)$$

Since $V_o = i_o R_L$, we can write the equation for the *load lines* shown in Fig. 10.1b

$$V_{ce} = -I_c R_L + (V_{CC} + I_Q R_L), \quad V_{ce} > 0 \quad (10.1-3)$$

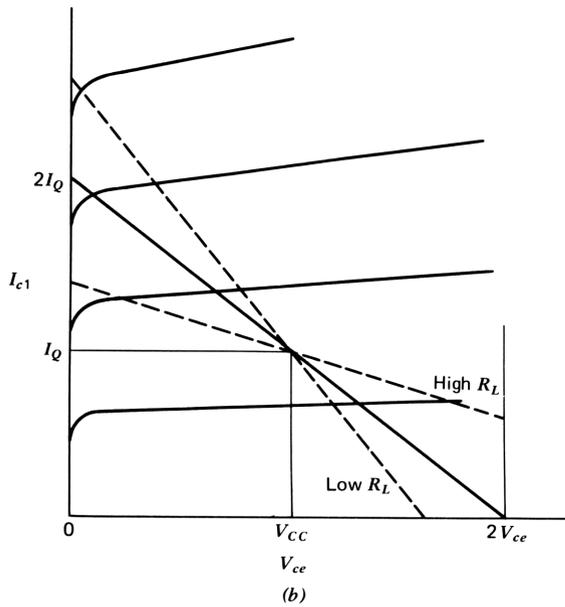
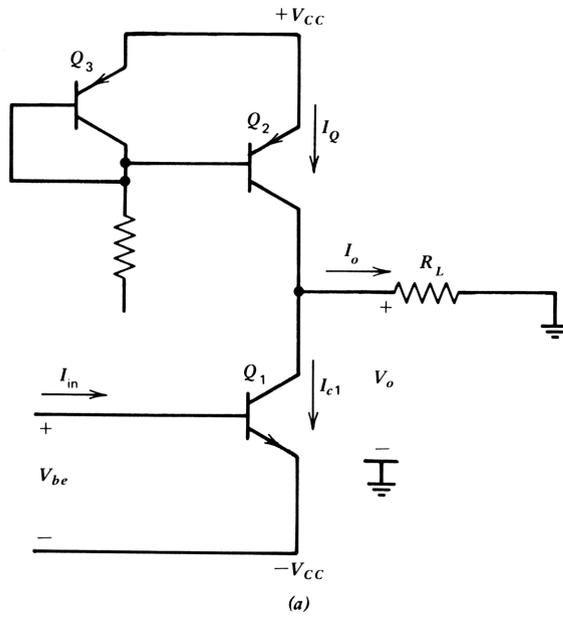


Figure 10.1. Common emitter output stage with current source: (a) circuit; (b) output characteristics and load lines.

or

$$I_c = \frac{V_{CC} - V_{ce}}{R_L} + I_Q \quad (10.1-4)$$

Three load lines are shown in Fig. 10.1*b*; the solid line is drawn for $R_L = V_{CC}/I_Q$, and the dashed lines are for values of R_L above and below this value. The slope of the load line is $1/R_L = G_L$. For values of R_L equal to or greater than V_{CC}/I_Q , the output voltage can swing from a little above $-V_{CC}$ to a little below $+V_{CC}$. At the lower limit, *saturation* of Q_1 prevents a further excursion, whereas at the upper limit, saturation of the current source transistor prevents a further positive excursion. These limits can be reached only if the collector junction breakdown voltage V_{BR} is much higher than V_{CC} . If it is not, avalanche breakdown of the transistors will restrict operation to a smaller range and may even lead to destruction of the transistors. This case is studied in the text that follows. For the present discussion, we assume high breakdown voltages for Q_1 and Q_2 . For values of R_L less than V_{CC}/I_Q , the positive voltage limit is a little less than $I_Q R_L$, whereas the lower limit remains at a little greater than $-V_{CC}$, so that the positive and negative limits are unsymmetrical.

Maximum (symmetrical) power output is obtained for $R_L = V_{CC}/I_Q$. The maximum sinusoidal output signal, ignoring the small saturation voltage, is

$$V_{o\max} = V_{CC} \sin \omega t \quad (10.1-5)$$

so that the maximum output power is

$$P_{o\max} = \frac{V_{CC}^2}{R_L} \sin^2 \omega t \quad (10.1-6)$$

$$= \frac{V_{CC}^2}{2R_L} (1 - \cos 2\omega t) \quad (10.1-7)$$

which has an average value of $V_{CC}^2/2R_L$. The efficiency, defined as the ratio of output power to power input from the power supplies, may be obtained by finding the total power from the positive and negative supplies. The former is $I_Q V_{CC}$ and is constant. The latter is $I_c V_{CC}$, where I_c varies sinusoidally from zero to $2I_Q$ and has an average value of I_Q ; the average power from the lower supply is also $V_{CC} I_Q$. Since $I_Q = V_{CC}/R_L$, the efficiency is

$$= \frac{V_{CC} I_Q / 2}{2V_{CC} I_Q} = \frac{1}{4} \quad (10.1-8)$$

or 25%.

Anticausal Distortion Analysis

The input drive required to secure a desired output is fundamental to the design of output stages; both input voltage and current must be taken into account. The input voltage to the common emitter stage is

$$V_{be} = \frac{kT}{q} \ln \frac{I_{c1}}{I_S} = \frac{kT}{q} \ln \frac{I_Q - I_o}{I_S} \quad (10.1-9)$$

$$= \frac{kT}{q} \ln \frac{I_Q}{I_S} \left(1 - \frac{I_o}{I_Q} \right)$$

$$= V_{BE} + \frac{kT}{q} \ln \left(1 - \frac{I_o}{I_Q} \right) \quad (10.1-10)$$

where V_{BE} is the quiescent value of V_{be} and is $(kT/q)\ln(I_Q/I_S)$. The second term is the (nonlinear) input signal required to obtain an output current of I_o . The second term, plotted in Fig. 10.2, exhibits considerable curvature. The input signal for a perfectly sinusoidal output signal is shown in Fig. 10.2, and is seen to be distorted. It is plotted for an output signal amplitude of 90% of V_{CC} ; for illustrative purposes, V_{CC} is 20 V, R_L is 1.0 k Ω , and I_Q is 20 mA.

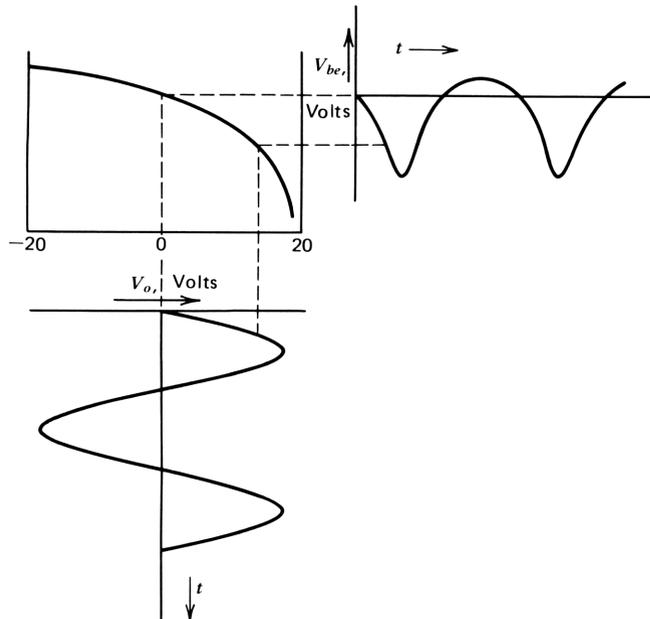


Figure 10.2. Transfer characteristic for input voltage as a function of output voltage, showing nonlinearity of BG_L .

Note that in a “good” design the output signal *will* be almost sinusoidal and the input voltage of the output stage will be predistorted as in the example here. The predistortion may be obtained automatically by overall feedback; regardless of the means used, the base voltage of the output stage, observed on an oscilloscope, will resemble the wave shown in the Fig. 10.2. The importance of this concept becomes increasingly clear as we proceed and becomes essential in the design of Class B amplifiers.

To obtain numerical values for the effects of this nonlinearity, we can obtain the harmonic distortion of the input waveform for a sinusoidal output signal. We begin (as earlier, in Section 1.2) by expanding (10.1-10) in a power series: from the expansion of $\ln(1-\gamma)$, we obtain

$$V_{in} = -\frac{kT}{q} \left(\gamma + \frac{\gamma^2}{2} + \frac{\gamma^3}{3} + \frac{\gamma^4}{4} + \dots \right) \tag{10.1-11}$$

where $\gamma = I_o/I_Q$. If we then set

$$I_o = AI_Q \cos \omega t \tag{10.1-12}$$

then

$$-\frac{qV_{in}}{kT} = A \cos \omega t + \frac{A^2 \cos^2 \omega t}{2} + \frac{A^3 \cos^3 \omega t}{3} + \dots \tag{10.1-13}$$

Each power of the cosine function $\cos^k \omega t$ can be written

$$\cos^k \omega t = \frac{\cos k \omega t}{2^{k-1}} + f(\omega t) \tag{10.1-14}$$

The first term on the right gives the k th harmonic of the wave. The second term contains terms of lower harmonics. Where k is even (odd), the second term contains all even (odd) harmonics of lower order than k .

Using standard trigonometric identities, we can write for (10.1-13)

$$\begin{aligned} -\frac{qV_{in}}{kT} &= \frac{A^2}{4} \left(1 + \frac{3A^2}{32} + \frac{5A^4}{96} + \dots \right) + A \left(1 + \frac{A^2}{4} + \frac{A^4}{8} + \dots \right) \cos \omega t \\ &+ \frac{A^2}{4} \left(1 + \frac{A^2}{2} + \frac{5A^4}{16} + \dots \right) \cos 2\omega t + \frac{A^3}{12} \left(1 + \frac{3A^2}{4} + \dots \right) \cos 3\omega t \\ &+ \frac{A^4}{32} (1 + A^2 + \dots) \cos 4\omega t + \dots + \dots \end{aligned} \tag{10.1-15}$$

Note that A is a normalized amplitude; when $A=1$, the peak value of the output current is I_Q . When A is considerably less than unity, the higher-degree terms of A become negligible, and the second term on the right of (10.1-14) can

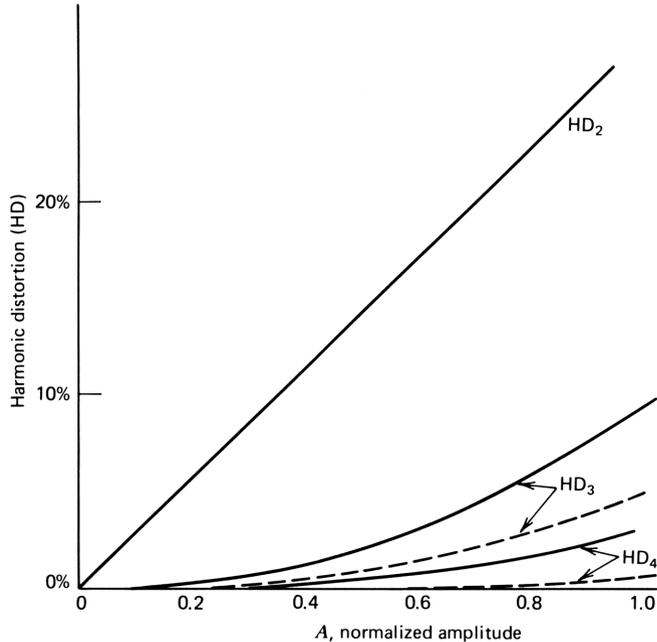


Figure 10.3. Harmonic distortion of the *input* signal to obtain a sinusoidal *output* for the emitter junction nonlinearity. The dashed lines show the third and fourth harmonics of the *output* for a sinusoidal *input* excitation.

be ignored. The harmonic products of the predistorted signal voltage are shown as a function of level by the solid lines of the plots in Fig. 10.3. They are shown as a percentage of the fundamental for the second, third, and fourth harmonics. The second harmonic is proportional to the signal level, and the third harmonic is proportional to the square of the signal level.

Comparison with Causal Distortion Analysis

Anticausal distortion analysis finds the input predistortion required to produce a desired output waveform. We used a sinusoid as the desired waveform in the preceding analysis, but the principle applies to any desired waveform. If the gain falls to zero (or the loss rises to infinity), as in the case of cutoff or saturation, the required input signal rises to infinity or becomes unbounded. Similarly, the input predistortion becomes unbounded. On the other hand, if we apply a high-level sinusoidal input signal to an amplifier that saturates at high levels, the output distortion is not infinite, or unbounded. The two ways of specifying distortion are quite different, therefore.

In general, the input predistortion as a percentage of the fundamental contains considerably greater high-order harmonics than does the output signal for sinusoidal input excitation. Hence the drive requirement for a distorting output stage at high frequencies is greater than one might expect from the

distortion products expressed as a percentage of output signal for input sine-wave excitation.

As an example, consider the emitter junction nonlinearity and compare the power series coefficients for the two representations. If we define a normalized input variable x' as qV_{in}/kT and an output variable $I_o/I_Q = \gamma$, we can write the input-output relation in the anticausal case as in (10.1-11):

$$\begin{aligned} x' &= \ln(1 - \gamma) \\ &= - \sum_{i=1}^{\infty} \frac{\gamma^i}{i} \end{aligned} \quad (10.1-16)$$

The amplitude of the coefficients of the power series decrease as the order of the term. On the other hand, the causal representation can be expressed as

$$\begin{aligned} \gamma &= 1 - e^{-x'} \\ &= - \sum_{i=0}^{\infty} \frac{x'^i}{i!} \end{aligned} \quad (10.1-17)$$

In this direction the amplitudes of the coefficients decrease as the *factorial* of the order of the term, a much more rapid decrease. The ratio of the amplitudes of the power series coefficients for the two descriptions is thus $(n-1)!$. Thus the second harmonic term is the same for either direction of analysis, the third harmonic is twice as large in the anticausal direction as in the causal, and the fourth harmonic is six times as large. The third and fourth harmonics of the *output* distortion as a percentage of the fundamental are shown by the dashed lines in Fig. 10.3.

Evidently, analysis of distortion in the anticausal direction is considerably more sensitive than in the causal direction. The reason for this difference is that in the forward direction, any predistortion is itself distorted in passing from input to output. If the incremental gain is low, for example, the predistortion signal will suffer attenuation by the very distortion it is trying to correct. Consequently, it must be larger to overcome this effect. This is the "endless chain of dependencies" that is characteristic of analysis in the forward direction and is avoided by proceeding in the anticausal direction.

Input Current Drive Requirements

The base current required to secure a given output voltage and current includes both static and dynamic components; both are important to output circuit design. In the simplest case the static input current is given by $-\delta I_c$. The transfer characteristic I_{in} versus V_o (or I_o) for this case is nearly linear since δ is only mildly nonlinear over the range of signals. In this case current drive

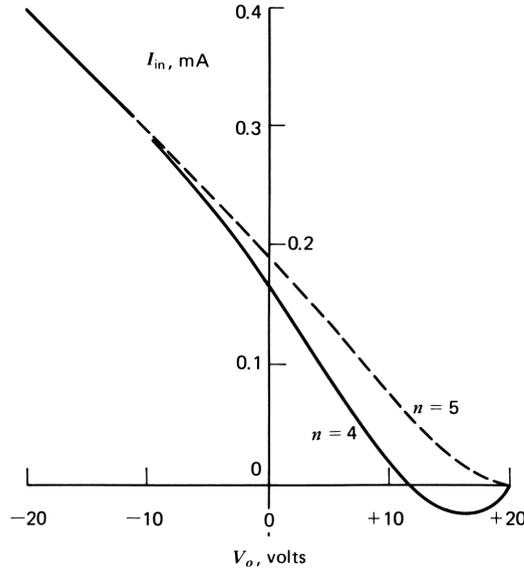


Figure 10.4. Transfer characteristic for input current as a function of output voltage, showing nonlinearity introduced by avalanche multiplication for a collector junction breakdown voltage of 100 V.

provides more linear operation than does voltage drive, and a high-impedance source gives lower distortion than does a low-impedance source.

Output stages are often called on to provide considerable output voltage. If the avalanche breakdown voltage is not sufficiently high, avalanche multiplication can change the static drive requirements significantly. This is illustrated in Fig. 10.4, which shows the transfer characteristic for currents for the case in which the positive and negative supplies are each 20 V and the avalanche breakdown voltage is 100 V. Two characteristics are shown; the solid line is drawn for an avalanche multiplication exponent of $n=4$ in the equation for the base current (see Section 7.5):

$$I_b = - \left(\delta - \frac{V_{ce}}{V_{BR}} \right)^n \quad (10.1-18)$$

The base current becomes zero for $V_{ce} = V_{sus}$; with $\delta=0.01$, we obtain

$$\begin{aligned} V_{sus} &= V_{BR}(\delta)^{1/n} = 100(0.01)^{0.25} \\ &= 32 \text{ V} \end{aligned} \quad (10.1-19)$$

The base current of the lower transistor in Fig. 10.1 thus goes to zero for an output voltage of +12 V and becomes negative for greater voltages. Hence current drive will cause high distortion and, if overall feedback is used, can

cause instability since the incremental loss undergoes a phase reversal at high collector voltages of Q_1 . Voltage drive removes the difficulty but causes its own distortion as described previously.

The dashed curve in Fig. 10.4 is drawn for $n=5$ in eq. (10.1-18); the sustain voltage in this case is 40 V, removing the reversal in the slope of the transfer characteristic, but causing considerable curvature for positive voltages. Note that the curvature of the characteristic is of sign opposite from that for the emitter junction nonlinearity, so that by choice of a suitable source impedance, some distortion cancellation can be effected. Such cancellation has been used in submarine cable repeaters where the transistors are manufactured with extremely well-controlled characteristics (and are consequently costly). In ordinary design, such cancellation should be avoided because although the emitter junction nonlinearity is well controlled, avalanche multiplication seldom is.

The conclusion from this discussion is that the sustain voltage should be equal to or greater than the total supply voltage $2V_{CC}$ to be free of the input current reversal. It is possible to design satisfactory circuits with sustain voltages less than this value, simply by degrading the effective defect current with added shunt input conductance. This increases the current drive requirement, however, so that the overall tradeoffs must be examined.

When the sustain voltage is made adequately large—greater than the total supply voltage—the input current as a function of output voltage becomes reasonably linear.

The dynamic current drive requirement is set primarily by the collector capacitance of Q_1 . This capacitance must be charged and discharged through the output voltage for each cycle of the output signal. In many designs, from audio to microwave frequencies, the charging of this capacitance establishes the current level at which the driver must operate. Failure to provide this amount of driver capability will lead to *slope overload*, or slew rate limiting, as shown in Fig. 10.5. The driver current capability is easily found for any given worst-case output signal. Suppose, for example, that the output stage is to provide a full amplitude signal at a frequency of 25 MHz and that the collector capacitance is 2 pF. The maximum slope of a 20 V sinusoid at 25 MHz is found by differentiating

$$V_o = 20 \sin 2\pi(0.025)t \quad (10.1-20)$$

$$\frac{dv_o}{dt} = 20(2\pi)(0.025)\cos 2\pi(0.025)t \quad (10.1-21)$$

The maximum slope occurs at the zero crossing and is π V/ns. The maximum current to be supplied by the driver is thus

$$i_{d\max} = C \left. \frac{dv_o}{dt} \right|_{t=0} = 6.3 \text{ mA} \quad (10.1-22)$$

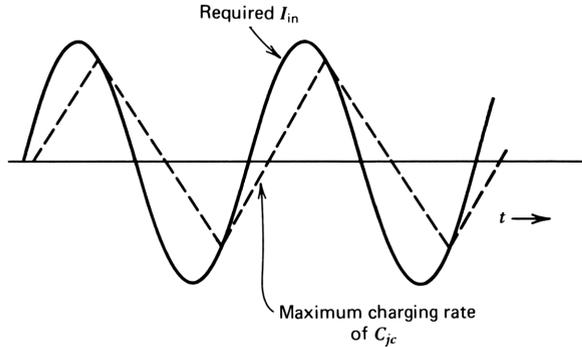


Figure 10.5. Slope overload (slew limiting) introduced by inadequate driver output current capability.

The quiescent current of the driver stage must be equal to or greater than 6.3 mA to meet the requirement without slope overload. Failure to provide this drive capability results in *transient overload*. Conversely, with adequate drive capability, transient distortion products do not arise.*

The main sources of distortion are usually those discussed previously. Other sources may be of importance in special circumstances. Under high-level injection conditions, the defect current ratio rises, (Webster effect). At high frequencies Kirk effect may cause a sudden increase in the current loss time constant in transistors too small for the application. Early effect can cause nonlinearity, but low-load impedances usually swamp this effect.

10.2 LINEARITY IMPROVEMENT

Linearity of the output stage may be improved by adding B feedback. Figure 10.6a shows a resistor in series with the emitter of Q_1 that augments the input voltage by $v_o R_E / R_L$, and this component of input voltage is linearly related to the output voltage and current. *Note that the input distortion voltage is unchanged by the feedback* (actually, the voltage drop across R_E subtracts from the output voltage capability, thus causing the distortion for a given output level to increase slightly). The transfer characteristic is linearized by the *addition* of a linear component to V_{in} , so that the distortion component tends to be swamped out.

Another method for linearizing the voltage drive requirement is to use an emitter follower output stage, as shown in Fig. 10.6b. In this case the full output voltage is added to the input, and the linearity of the output stage is greatly improved. Again, the distortion products at the input for sine-wave

*It has been (erroneously) stated in many sources (particularly in audio publications) that feedback "causes" transient intermodulation. This is clearly not the case, as the preceding discussion indicates. The cause is lack of driver stage capability.

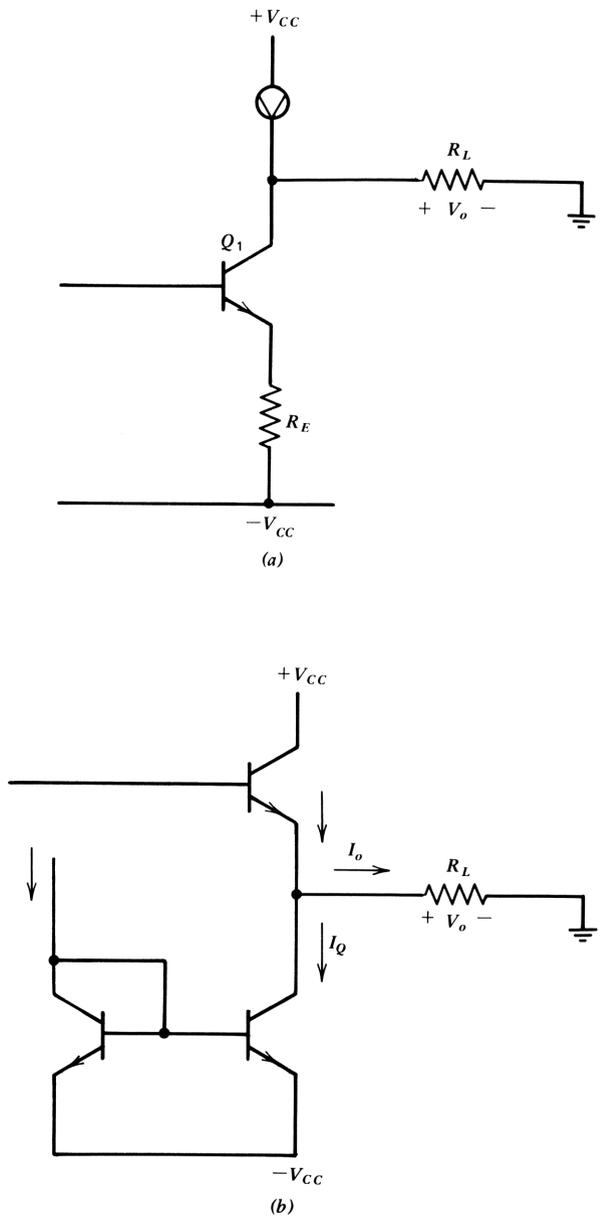


Figure 10.6. Introduction of feedback to swamp out nonlinearity of input voltage: (a) B feedback; (b) emitter follower (output permutative feedback).

output *do not change* from those of the common emitter configuration; they are swamped out by the addition of the output voltage. The input voltage is given by

$$v_{\text{in}} = (A + BG_L)v_o \quad (10.2-1)$$

Except for the lack of a phase reversal, B is the same as in the common emitter configuration. However, A is changed from zero to unity. The harmonic products of the input waveform for sinusoidal output can be estimated simply by multiplying the values given in Fig. 10.3 by the sensitivity to B , which is $BG_L/(A + BG_L)$. Since $BG_L \ll A$, the harmonics are greatly reduced. In a sense, however, the linearity problem has been transferred to the driver stage since the latter stage now must provide slightly more than the output voltage to drive the output stage. It does this at much lower current, however, so the method is viable.

The dynamic current drive requirement also becomes greater. The reason for this is that the driver now has to charge its own collector capacitance to the output voltage as well as that of the output transistor. In addition, in an integrated circuit, the driver must also charge its own collector-to-substrate capacitance through the same voltage change. The current drive requirement imposed by these capacitances is found as in the above slew rate calculation, substituting the total capacitance at the base node of the output stage for the collector capacitance of the output stage.

With its large input voltage drive requirement, it would seem that the emitter follower is a poor choice for an output stage. Yet it is widely used and thus must provide a further benefit over a common emitter stage with feedback over its driver stage(s). Indeed, it does. As we saw earlier, the common collector stage is unique because it does not include transit time delay from input to output. Thus it is easier to stabilize and to realize a prescribed response than is the common emitter stage. This can be of major importance particularly where the current loss time constant (and hence delay) of the output transistor is larger than those of the earlier stage transistors.

Push-Pull Operation

The *single-ended* common emitter or common collector stages discussed previously need rarely be used because, with few extra parts, the circuit can be rearranged to be twice as efficient. By converting the current source to an active load, as discussed in Section 9.4, the current source transistor can be made to contribute signal current to the load. In Fig. 10.7a the emitter follower circuit in Fig. 10.6b has been thus converted, with the emitter of the driver stage arranged to drive the active load. Such an arrangement is termed a *push-pull circuit* because when the upper transistor conducts heavily (negative signal applied to the base of the driver), the lower transistor conducts lightly, so that more of the current of the upper transistor reaches the load. The upper transistor pushes current into the load. When positive signals are applied to the

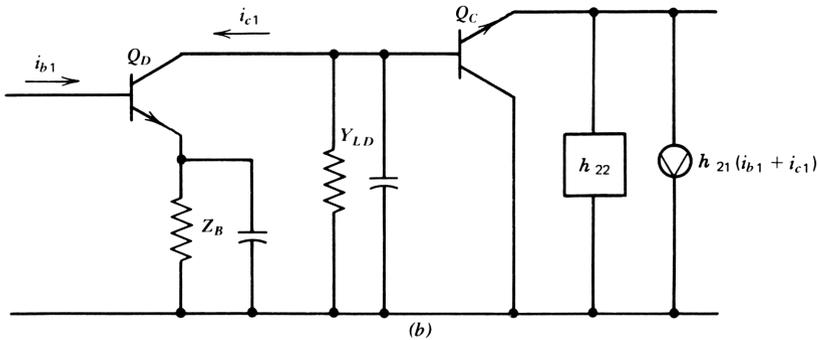
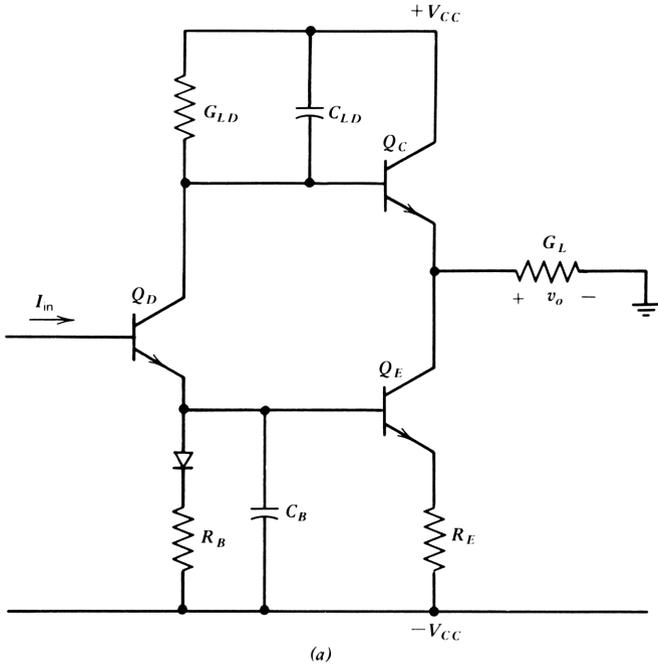


Figure 10.7. Push-pull output stage with the use of an active load: (a) circuit; (b) equivalent ladder circuit.

driver, the lower transistor conducts heavily and the upper transistor conducts lightly. The lower transistor pulls current from the load.

With the same quiescent current as in the single-ended case, the load current can rise almost to $2I_Q$ and fall to $-2I_Q$ for the same voltage change as in the single-ended case. The load resistance for maximum output power is thus $V_{CC}/2I_Q$, half that of the single-ended case. This assumes that the two output transistors contribute equally to the load, a circumstance that does not occur automatically but that must be made to happen by appropriate arrangement of

the driving signals to the two transistors. Since both transistors contribute to the load, the sine-wave efficiency becomes 50%, twice that of the single-ended case. The actual efficiency is less than this because the full output voltage of V_{CC} is not obtainable. At the upper limit the V_{be} drop of Q_C subtracts from the maximum obtainable driver voltage, as does the drop in the load conductance G_{LD} of the driver. At the lower limit the voltage drops across R_E and V_{be} of Q_E limit the swing.

The combination of output stage and driver is capable of providing broadband voltage loss at high-power efficiency for Class A operation. For best efficiency, the active load should supply half the output signal current. This establishes the relationship among the four immittances in Fig. 10.7, namely, G_L , Y_{LD} , Z_B , and R_E . The signal current through Y_{LD} is $v_o Y_{LD}$ since the output voltage appears across it. This current flows through Q_D and into the active load spanning network input. To provide half the output current $v_o G_L/2$, the current gain h_{21} of the spanning network must be

$$h_{21} = \frac{G_L/2}{Y_{LD}} \quad (10.2-2)$$

With Q_E in its reference condition, h_{21} is given by

$$h_{21} = \frac{Z_B}{R_E} \quad (10.2-3)$$

so we can write

$$Y_{LD} Z_B = \frac{R_E G_L}{2} = n \quad (10.2-4)$$

where $-n$ is the voltage loss of the whole circuit in its reference condition. The magnitude of the voltage loss is chosen to give the bandwidth required and may be in the range of 0.1–0.4. The bandwidth obtainable is somewhat less than nf_T . For a given G_L and n , (10.2-4) gives the value of R_E required to obtain equal current contributions from Q_C and Q_E . It also gives the relation between Y_{LD} and Z_B .

The value of Y_{LD} is determined by the output-stage drive requirement and consists of the dc load conductance of the driver stage in parallel with the total parasitic capacitance to ground at the base node of Q_C . This capacitance includes the collector-to-base capacitance of Q_C and Q_D , the collector-to-substrate capacitance of Q_D , and any wiring capacitance or other parasitics. With unit output voltage, the voltage at the base of Q_D is $-n$, so that the current flowing through C_{jcD} is $(1+n)C_{jcD}$, giving a total effective node capacitance of

$$C_{LD} = C_{jcC} + (1+n)C_{jcD} + C_{csD} + \frac{G_L \tau_{TC}}{2} \quad (10.2-5)$$

The last term on the right accounts for the base charge of Q_C ; it is divided by 2 because half the output current is supplied by the lower transistor. Base charge is given by $I_C\tau_F$; to this we can add the charge on the emitter capacitance, thus rendering the total base charge roughly $I_C\tau_T = v_o G_L\tau_T$. The term C_{LD} is the dominant element in the design of this output stage. The dc quiescent current of the driver is established by C_{LD} , and the slew rate requirement as

$$i_{CD} = C_{LD} \frac{dV_o}{dt} \quad (10.2-6)$$

With a total C_{LD} of 1.5 pF, for example, and a required slew rate of 4 V/ns, the driver quiescent current should be 6.0 MA.

For a given supply voltage, we find G_{LD} from the dc relation

$$G_{LD} = \frac{I_{CD}}{V_{CC} - V_{beC}} \quad (10.2-7)$$

Hence the slew rate requirement (and the supply voltage) determines the essential features of the design. The variable Z_B has the same time constant as Y_{LD} and is the parallel combination of resistor R_B and capacitor C_B ; the capacitance should include the input capacitance of Q_E . The variable R_E is purely resistive (for zero load capacitance). In the actual realization, a purely resistive R_E reduces the bandwidth of Q_E , so that a small capacitance is connected in parallel with R_E to compensate the stage.

This output stage has been used in a broadband integrated operational amplifier.² In that application $n=0.30$, and the bandwidth is about 1.0 GHz.

Other push-pull arrangements for Class A amplifiers are possible; we next discuss Class B circuits, for which Class A versions are always possible simply by raising the quiescent current sufficiently that the transistors remain in conduction at all times.

10.3 CLASS B OUTPUT STAGES

The chief failing of Class A stages is that they consume the same power under idling conditions as when they deliver signals to the load. In Class B circuits, on the other hand, the quiescent current is nominally zero; the collector current rises in response to input signal. Consequently, Class B circuits consume little standby power. Where the signal is of an intermittent type, such as in analog speech, the energy saving of Class B circuits can be much larger than a comparison of efficiencies under large-signal conditions would suggest. The large-signal efficiency of a Class B circuit (under conditions of maximum sine-wave excitation) approaches 78.6% ($100 \times \pi/4$), found by integrating the total power into the circuit and dividing it into the output power. This is to be compared with 50% for the ideal Class A push-pull stage.

The design of Class B circuits is controlled by the same considerations of drive requirements as in the Class A case, but in addition, the switching on and

off of the output transistors must also be considered. The circuit in Fig. 10.7 could be used in Class B, for example, but when the upper transistor is conducting, the output impedance is considerably lower than when the lower transistor is conducting. It is advantageous in Class B designs to use circuits that exhibit symmetry under positive and negative signal excitation.

With their greater efficiencies, Class B circuits can be used to provide considerably greater power than Class A circuits by using transistors of the same heat-dissipating ability and size; consequently, temperature variations in the Class B circuits under variations in signal conditions will be much greater than those of Class A circuits. Therefore, temperature stability and prevention of transistor burnout become more important.

Crossover Distortion

A central problem in Class B circuits arises from the turning off of one or the other transistor in the vicinity of zero output signal. In effect, the nonlinearity of input voltage that occurred in Fig. 10.2 at the positive extreme of output voltage now occurs for each transistor in the vicinity of zero output current, giving rise to *crossover distortion*. The effect is shown in extreme form in the circuit in Fig. 10.8, in which *npn* and *pnp* emitter followers are connected in series for dc currents and in parallel for signals. When the driver transistor conducts heavily, the lower transistor conducts current from the load; when the driver transistor conducts less heavily than the current source, the upper transistor conducts current from the positive supply to the load. When the driver conducts a current equal to the current source, neither transistor conducts. Figure 10.8*b* shows the transfer characteristic. To provide a sine-wave *output*, the input voltage must switch suddenly as shown. If the driver is incapable of this, considerable distortion will result, no matter how much feedback is provided around the whole amplifier, including the driver. The driver is required to charge the collector capacitances of the output transistors as well as the base charge $I_c\tau_F$, and to do it rapidly.

The capacitive charging current is given by $C(dV_{in}/dt)$; the derivative of the input waveform in Fig. 10.8 is shown in Fig. 10.9 and is reasonably sinusoidal except at the transitions, where spikes of current are necessary to switch the transistor. If the driver cannot provide these spikes, feedback will not correct the problem. This is equivalent to the slope overload problem of Class A amplifiers discussed previously. There, the solution was to increase the driver capability; here, the problem can be eliminated by the combination of two techniques: the first (and most important) is to provide a nonzero quiescent current I_Q for the output transistors by adding biasing diodes or diode-connected transistors between the bases of the output transistors as shown in Fig. 10.10. This technique alone goes a long way toward solving the problem and allows overall feedback to reduce distortion to manageable amounts.

When forward bias is added, both transistors contribute to the output over a signal range near zero.

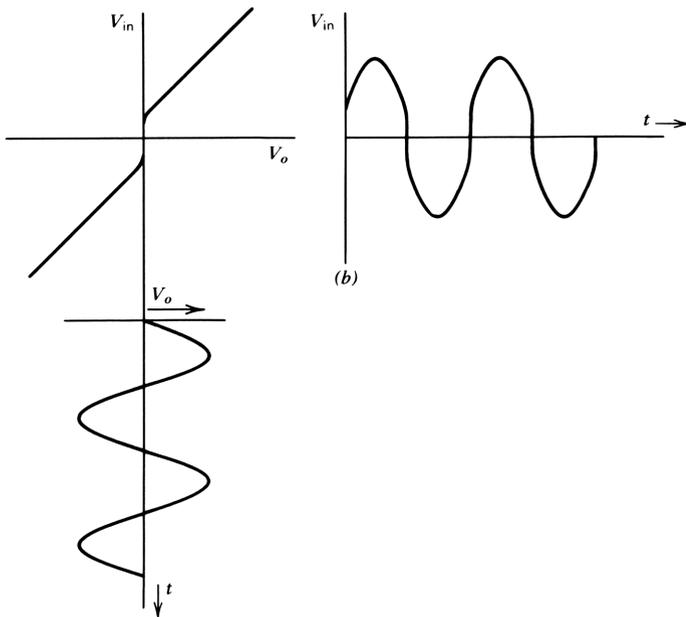
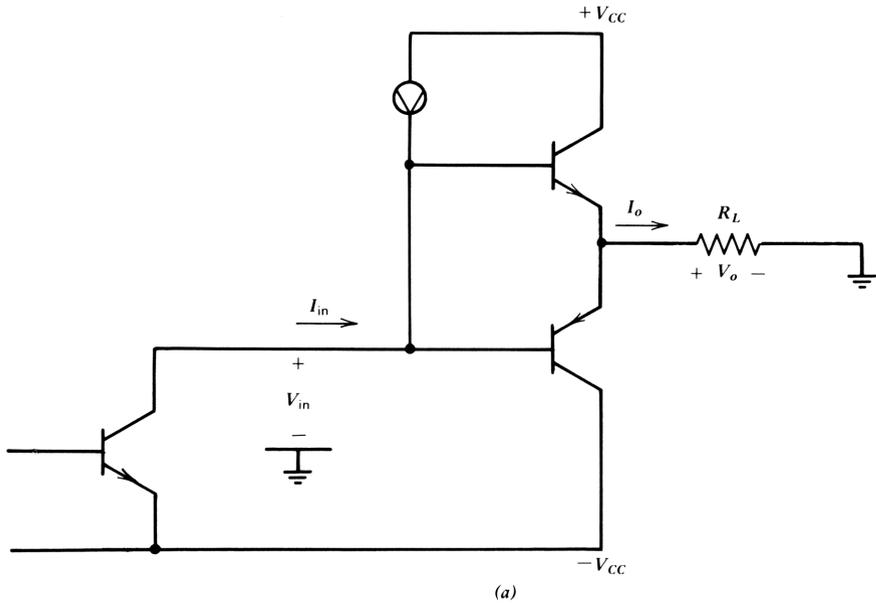


Figure 10.8. Class B output stage with severe crossover distortion. The voltage transfer characteristic and required predistortion are shown in part *b*.

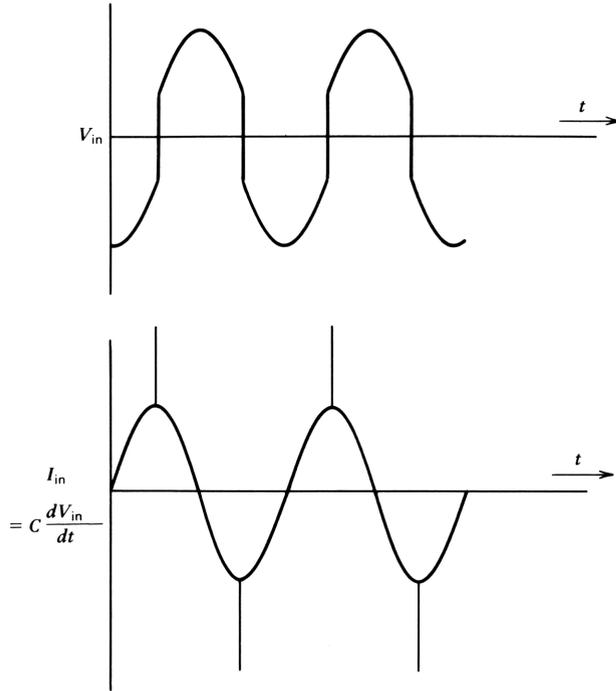


Figure 10.9. Input voltage and current waveforms for the Class B stage shown in Fig. 10.8. Where this stage is used in a feedback amplifier, these waveforms can be observed with an oscilloscope.

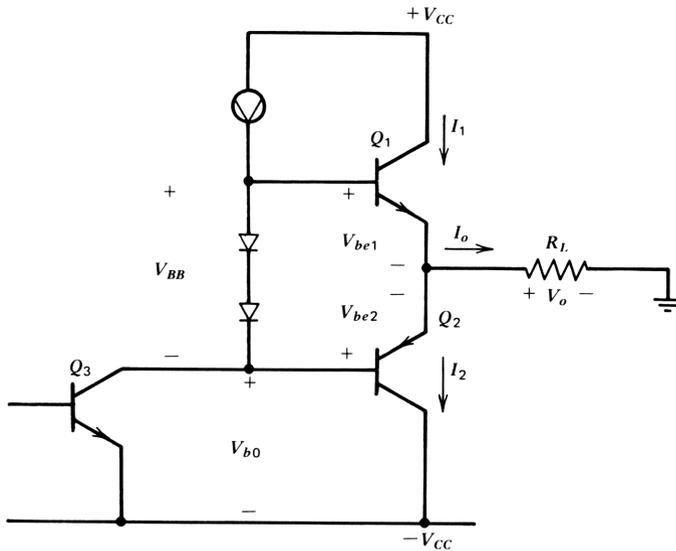


Figure 10.10. Use of forward bias to reduce crossover distortion.

The second technique is to add emitter resistors to the output transistors, thus allowing minimization of crossover distortion for a given value of I_Q .

Rather than developing the transfer characteristic itself, we treat the *slope* of the transfer characteristic. Hence linearity is measured by the constancy of the derivative. This is a more sensitive indicator of nonlinearity. From (10.2-1) the slope is given by

$$\frac{dV_{in}}{dV_o} = A + BG_L \quad (10.3-1)$$

In this equation A is unity (for the emitter follower) and is linear. The nonlinearity is concentrated in B ; the A feedback helps to swamp out this nonlinearity. Therefore, we need only investigate the variation of B with output current to determine the input voltage nonlinearity.

In the circuit shown in Fig. 10.10, transconductances add, so that with $q_m = qI/kT$, we can write

$$B = \frac{1}{g_{m1} + g_{m2}} = \frac{kT}{q} \frac{1}{I_1 + I_2} \quad (10.3-2)$$

Under quiescent conditions, $I_1 = I_2 = I_Q$, so that $B = kT/2I_Q$, half the value for either transistor. On the other hand, under maximum signal conditions one transistor is cut off, so that $B = kT/qI_{max}$.

Between these extremes of current, we can find B by finding I_1 and I_2 and substituting them in (10.3-2). This is easily done by assuming a value for I_1 and finding the value of I_2 from it, given a desired value of I_Q . When I_2 is found, we can obtain I_o , the output current, from

$$I_o = I_1 - I_2 \quad (10.3-3)$$

which allows us to plot B as a function of I_o . This is an example of using an internal variable in the circuit and finding the input and output from it, rather than starting at either the input or the output. The procedure allows us to write a sequential series of equations.

Given I_1 , we can write

$$V_{be1} = \frac{kT}{q} \ln \frac{I_1}{I_S} \quad (10.3-4)$$

The diodes in Fig. 10.10 provide a bias voltage V_{BB} that biases the output transistors to the desired quiescent current I_Q . The design of the bias network will be considered later; the present analysis does not require us to know V_{BB} . The base emitter voltage V_{be2} of the second transistor is given by $V_{be1} - V_{BB}$, so we can write

$$I_2 = I_S \exp \frac{q}{kT} (-V_{be2}) \quad (10.3-5)$$

where the negative sign arises from the direction assumed for I_2 (V_{be} is a negative quantity for a *pn*p transistor). Hence

$$I_2 = I_S \exp \frac{q}{kT} \left(V_{BB} - \frac{kT}{q} \ln \frac{I_1}{I_S} \right) \quad (10.3-6)$$

$$= \frac{I_S^2}{I_1} \exp \frac{qV_{BB}}{kT} \quad (10.3-7)$$

We eliminate V_{BB} from this equation by noting that when $I_2 = I_1 = I_Q$, we have

$$I_Q^2 = I_S^2 \exp \frac{qV_{BB}}{kT} \quad (10.3-8)$$

so that by dividing (10.3-7) by (10.3-8), we obtain

$$I_2 = \frac{I_Q^2}{I_1} \quad (10.3-9)$$

and

$$I_o = I_1 - \frac{I_Q^2}{I_1} \quad (10.3-10)$$

Multiplying by I_1 and solving the quadratic, we can obtain I_1 in terms of I_o and I_Q . Substituting this back into (10.3-2), we obtain B as

$$B = \frac{kT}{q\sqrt{I_o^2 + (2I_Q)^2}} \quad (10.3-11)$$

This relation is plotted in Fig. 10.11 for $I_Q = 5$ mA, showing the variation of B over a current range of -100 to $+100$ mA. The linearizing effect of adding AR_L is also shown in the plot, with $R_L = 0.1$ k Ω (giving a maximum output voltage range of ± 10 V). Although the percentage variation of B is large, the stage is reasonably linear. The voltage loss—found by multiplying the ordinate by R_L —varies from a minimum of about unity for large output to a maximum of 1.026 at zero output, a total variation of 0.22 dB.

This is the *differential loss* over the operating range; its reciprocal is the *differential gain*, which is 0.22 dB *lower* at zero output than at maximum output. Differential gain is a linearity specification often used for video amplifiers.³ Clearly, the differential gain in decibels is equal to the negative of the differential loss. Since the latter is a measure of input predistortion and the former is one of output distortion, the two specifications in decibels are equal and opposite. Where the nonlinearity is large, the harmonic content of the

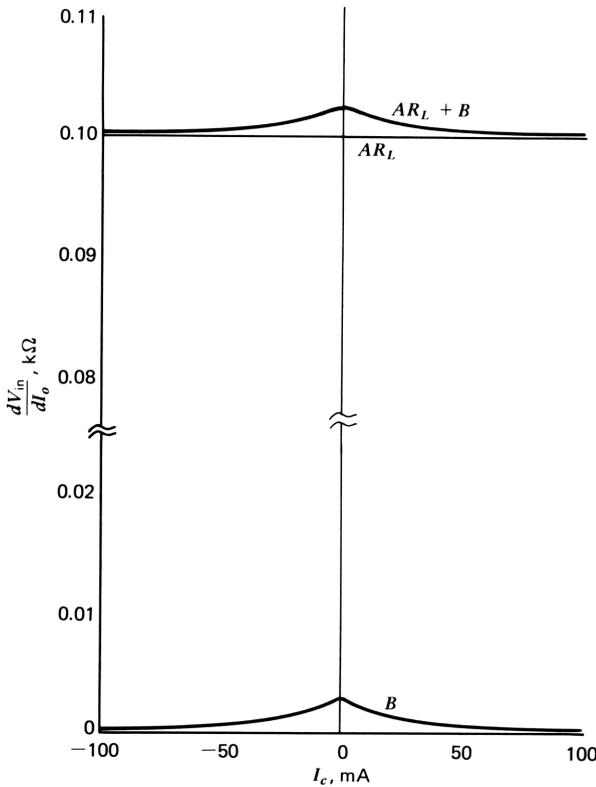


Figure 10.11. Terms AR_L and B as functions of output current for the circuit shown in Fig. 10.10. Their sum is the voltage loss of the circuit and shows differential loss variation.

input predistortion may be significantly different from the output distortion, as we have seen in Section 10.1.

The remaining nonlinearity can be reduced by adding resistors in series with each emitter. Suppose, for example, that we add resistors nominally equal to

$$R_E = \frac{kT}{qI_Q} \tag{10.3-12}$$

Under maximum signal conditions, the collector current is large on one side, and r_E approaches zero, so that B for the stage approaches R_E . At zero signal, B of each transistor is $2R_E$, and the combined B is also R_E ; thus there is some correction of the nonlinearity.

To obtain the variation of B with output signal when emitter resistors are added, we proceed as before, adding the emitter resistor voltage drops; thus

$$V_{be2} = V_{be1} + (I_1 + I_2)R_E - V_{BB} \tag{10.3-13}$$

By the same substitutions used above, the equation corresponding to (10.3-9)

becomes

$$I_2 = \frac{I_Q^2}{I_1} \exp \frac{qR_E}{kT} (2I_Q - I_1 - I_2) \quad (10.3-14)$$

Since I_2 appears in the right-side exponential, this transcendental equation is best solved by iteration on the calculator. It converges quickly for $I_2 \leq I_Q$. Since the equation is symmetrical in I_1 and I_2 , the solution for one side is the same as the one for the other, and we have the solution for the complete signal range.

We can then find B as the parallel combination of B for the individual transistors:

$$B = \left[\left[\frac{1}{\frac{kT}{qI_1} + R_E} \right]^{-1} + \left[\frac{1}{\frac{kT}{qI_2} + R_E} \right]^{-1} \right]^{-1} \quad (10.3-15)$$

The output current is found from (10.3-3); thus we select a value of I_1 and calculate I_2 , B , and I_o . The results for various values of R_E are plotted in Fig. 10.12, which clarifies the selection of R_E . The curves are drawn for the same currents as those in Fig. 10.11. When R_E takes on its nominal value $R_{En} = kT/qI_Q$, differential loss is minimized for the situation in which the quiescent current is small compared with the maximum current. For the case shown, with $I_{\max} = 20(I_Q)$, a smaller value of R_E gives slightly better distortion performance. Large values of R_E give poorer distortion performance. As I_Q is increased, the curves retain their shape, with the value of B reduced and the abscissa stretched by the ratio of the new I_Q to the value (5 mA) used in the plots.

In general, a value of R_E between the nominal value kT/qI_Q and half this value should be used. Whatever value is chosen, the optimum will vary with absolute temperature as well as I_Q . By making I_Q proportional to absolute temperature, a fixed value of R_E can be used to attain the optimum. We show in Section 10.4 how a prescribed variation of I_Q with temperature can be obtained.

Output-Stage Bias Circuits

The output-stage quiescent current is chosen by considerations such as those described for the optimum design for low distortion. The driver quiescent current is established by consideration of the input current drive requirements as discussed for both Class A and Class B circuits. Since these two quiescent currents are determined in the design by distinctly separate considerations, the designer should have the circuit means for separately determining them. Figure 10.13 gives several such means. The transistor: diode area ratio is a primary

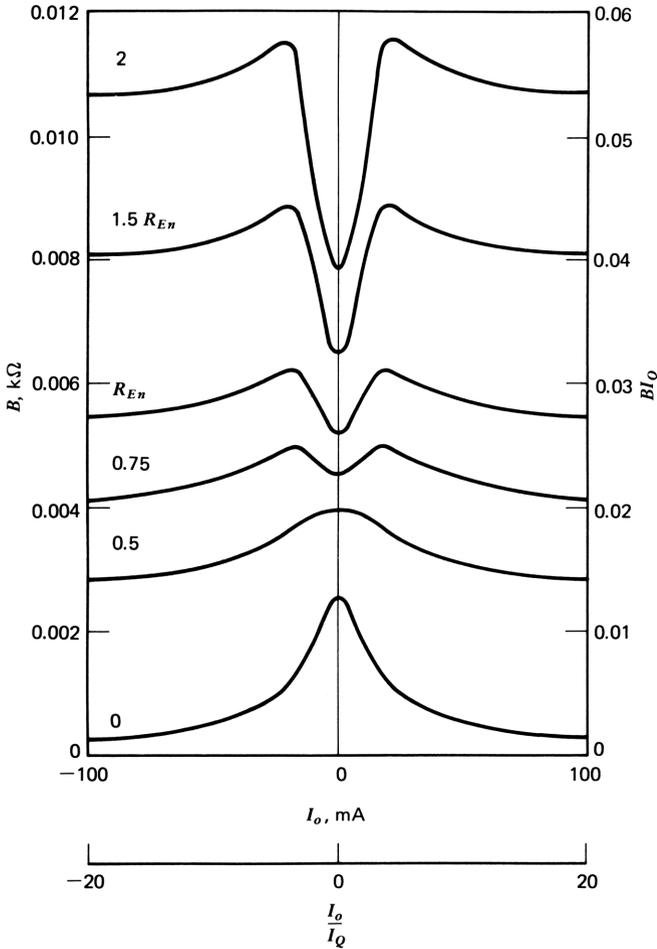


Figure 10.12. Term B as a function of output current for various values of linearizing resistors R_E . $R_{En} = kT_0/qI_Q$.

way to set I_Q/I_D , but this may be inconvenient. If the output transistors are to handle high power, for example, they will be large and will take up much room on an integrated circuit chip. To have the diodes take up comparable area to provide their simple biasing function would be wasteful of silicon area. The solution is to use smaller diodes (that have a higher voltage at a given current) and to reduce the effect of their higher voltage by one of the techniques illustrated in Fig. 10.13. Driver current can also be increased by adding a conductance G_{BB} between the output stage bases as shown in Fig. 10.13a. This conductance can also be applied to any of the other circuits in Fig. 10.13.

In Fig. 10.13a–c a resistive drop is added to or subtracted from the base-emitter drops of the transistors, changing the ratio of I_Q/I_d by the factor

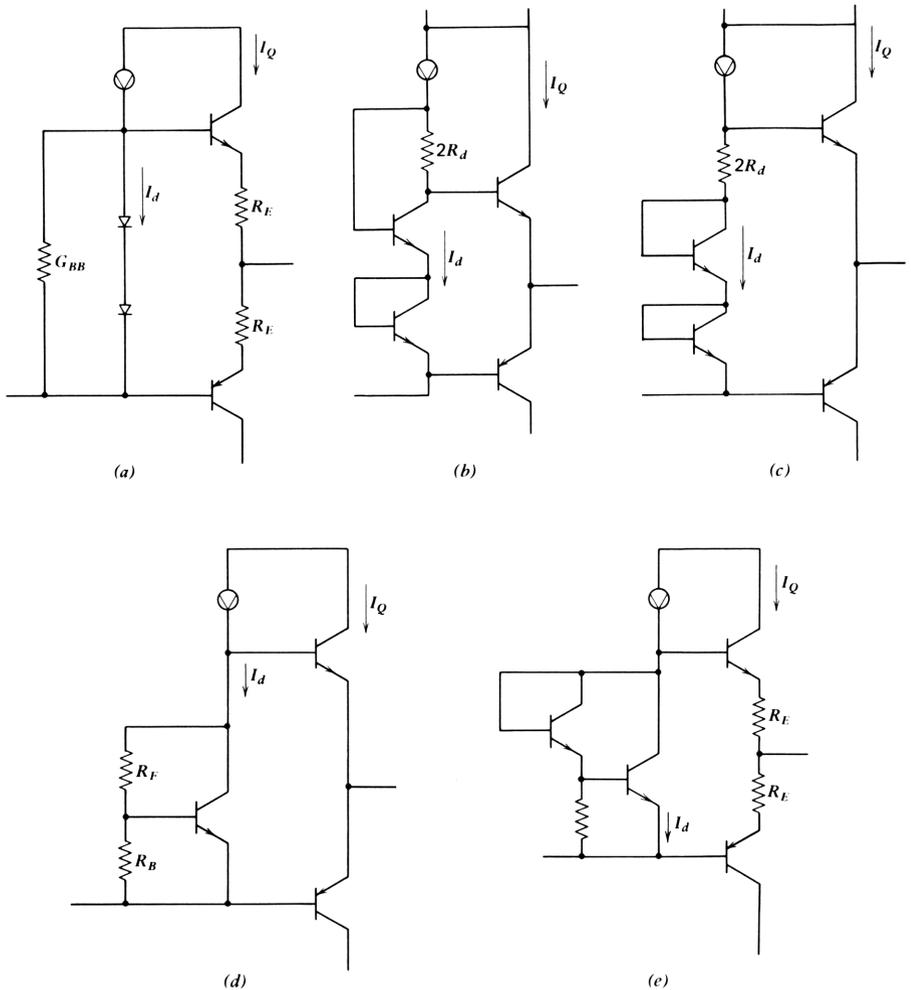


Figure 10.13. Circuits for obtaining separate control of output-stage and driver-stage quiescent currents. The circuit shown in part *d* should be avoided (see text).

qV_r/kT :

$$\frac{I_Q}{I_d} = \frac{I_S}{I_{Sd}} \exp \frac{qV_r}{kT} \quad (10.3-16)$$

In Fig. 10.13*a* $V_r = -2R_E I_Q$; in Fig. 10.13*b*, $V_r = -2R_d I_d$; and in Fig. 10.13*c*, $V_r = 2R_d I_d$.

Two other circuits often used are given in Figs. 10.13*d* and 10.13*e*. The circuit in Fig. 10.13*d* multiplies V_{BE} of the transistor by the factor $n = 1 + R_F/R_B$

(ignoring base current) where n is near 2. For this case,

$$V_{BB} = \frac{nkT}{q} \ln \frac{I_d}{I_{Sd}} = \frac{2kT}{q} \ln \frac{I_Q}{I_S}$$

From which we obtain

$$I_Q = I_S \left(\frac{I_d}{I_{Sd}} \right)^{n/2} \quad (10.3-17)$$

This circuit is sensitive to the exact value of n ; the sensitivity is

$$S_n^{I_Q} = \frac{n}{2} \ln \frac{I_d}{I_{Sd}}$$

With $n \approx 2$, the sensitivity is typically 20–30. This circuit, therefore, does not give good control of I_Q and should be avoided. Much better sensitivity is exhibited by the circuit in Fig. 10.13e, in which the base-emitter drops of a Darlington pair provides bias for the output transistors.

The design of these bias circuits must take temperature effects into account; the matter is discussed in detail in the following section.

Note that when the Class B output stage is designed for low distortion, dynamic crossover distortion effects vanish along with the static effects that caused them. The requirements on driver current capability are then established, as in the Class A case, by the slew requirements of the stage—the input capacitance (including base charge or diffusion capacitance) and the maximum required value of dV_{in}/dt .

Avalanche multiplication effects are reduced in Class B operation compared with Class A because the transistor with high instantaneous V_{ce} across it is reverse biased; therefore, the collector current to be multiplied by the multiplication factor M is much smaller than in the Class A case. Input current linearity is thus improved by Class B operation. The conclusion is that where the optimum value of R_E is used, Class B circuits can exhibit favorable linearity performance compared with Class A circuits.

Burn-Out Protection

If the output of the amplifier is inadvertently short-circuited to one of the supply voltages, the collector current may rise to destructive levels. To avoid this, protector transistors Q_{p1} and Q_{p2} are connected as in Fig. 10.14. When the output current becomes high enough, the protecting transistor conducts, shunting some base current drive from the output transistor and preventing the damaging increase. Both output transistors can be protected in this way. When R_E is kT/qI_Q , the voltage drop across R_E under quiescent conditions is kT/q volts (about 0.026 V), so that the protector transistor is turned off. Not until

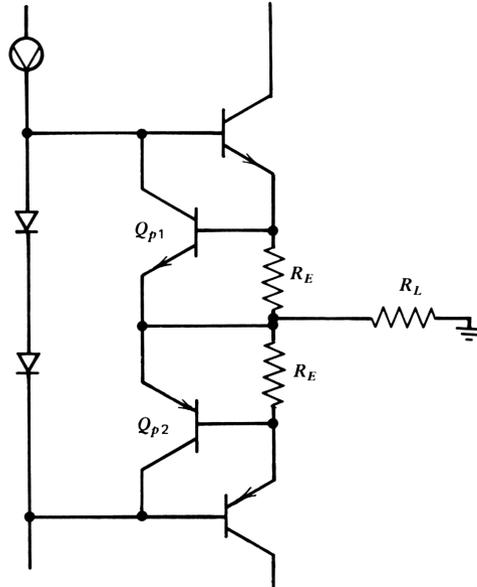


Figure 10.14. Output-stage burn-out protection.

this voltage rises to the vicinity of 0.7 V (depending on the ratio of the areas of the output and protector transistors) does the protector transistor conduct significant current relative to the base drive current. The ratio of the maximum value of the output current to the quiescent current is thus approximately $0.7/0.026=27$ and is about right to protect the output transistor without interfering with normal operation.

10.4 EFFECTS OF TEMPERATURE IN THE BIPOLAR TRANSISTOR

The base-emitter voltage of the transistor is strongly temperature dependent. Fortunately, it is also well controlled, so that circuits can be designed for predictable temperature performance. Base current is also temperature dependent but less predictable. In newer integrated circuits this dependence is small. The effect of base current variations can be reduced by use of low-impedance designs. Where precise temperature characteristics are desired, the combination of low-impedance design and recognition of the temperature dependence of V_{be} in the design is required. In this section we derive the temperature dependence of V_{be} and investigate its effects on circuit designs.

To find V_{be} as a function of temperature, we begin by dividing the equation for V_{be} by the absolute temperature:

$$\frac{V_{be}}{T} = \frac{k}{q} \ln \frac{I_c}{I_S} \quad (10.4-1)$$

In the expression on the right, I_S is proportional to $D_n n_i^2$, from eq. (7.1-8). The temperature dependence of both of these quantities is discussed in the text that

follows; the most important temperature dependence is that of n_i^2 , which is an exponential function of temperature. Thus I_S is also an exponential function of temperature. When the logarithm is taken in (10.4-1), V_{be} becomes an essentially *linear* function of temperature.

From eq. (7.3-9) in Chapter 7 we can express the saturation current I_S at temperature T in terms of its value I_{S0} at a reference temperature T_0 thus:

$$\frac{I_S}{I_{S0}} = \frac{n_i^2 D_n}{n_{i0}^2 D_{n0}} \quad (10.4-2)$$

where n_i and n_{i0} are the intrinsic carrier concentrations at T and T_0 , and D_n and D_{n0} are the diffusion constants at the two temperatures. It is the intrinsic carrier concentration that is an exponential function of temperature. Its variation causes the major portion of the V_{be} temperature dependence. Physical considerations give the intrinsic concentration in terms of V_{go} , the band-gap voltage. For silicon, this voltage, extrapolated to zero absolute temperature, equals 1.205 V. The ratio of intrinsic concentrations is given by⁴

$$\frac{n_i^2}{n_{i0}^2} = \left(\frac{T}{T_0} \right)^3 \exp \frac{gV_{go}}{k} \left(\frac{1}{T_0} - \frac{1}{T} \right) \quad (10.4-3)$$

The diffusion constant is also a (relatively slight) function of temperature, approximated as

$$\frac{D_n}{D_{n0}} = \left(\frac{T}{T_0} \right)^{-\sigma} \quad (10.4-4)$$

where σ depends on the doping density in the base and the transistor geometry and is typically about 0.8. Thus I_S/I_{S0} can be written

$$\frac{I_S}{I_{S0}} = \left(\frac{T}{T_0} \right)^{3-\sigma} \exp \frac{qV_{go}}{k} \left(\frac{1}{T_0} - \frac{1}{T} \right) \quad (10.4-5)$$

In (10.4-1) I_c may also be a function of temperature; this depends on the requirements of the particular circuit design. We may, for example, require that I_c remain constant with temperature. In the Class B design for low distortion described in the previous section we require that the quiescent current be a linear function of temperature to minimize the crossover distortion. In many designs we can approximate the desired temperature dependence by

$$\frac{I_c}{I_{c0}} = \left(\frac{T}{T_0} \right)^\xi \quad (10.4-6)$$

where ξ takes the values 0 and 1 in the two examples cited.

The change in V_{be}/T from its value at the reference temperature is found from (10.4-1) as

$$\frac{V_{be}}{T} - \frac{V_{be0}}{T_0} = \frac{k}{q} \left(\ln \frac{I_c}{I_{c0}} - \ln \frac{I_S}{I_{S0}} \right)$$

which from (10.4-5) and (10.4-6) is

$$\frac{V_{be}}{T} - \frac{V_{be0}}{T_0} = -\frac{\eta k}{q} \ln \frac{T}{T_0} + V_{go} \left(\frac{1}{T} - \frac{1}{T_0} \right) \quad (10.4-7)$$

where $\eta = 3 - \sigma - \xi$. Multiplying by T and rearranging, we can write

$$V_{be} = V_{go} - (V_{go} - V_{be0}) \frac{T}{T_0} - \frac{\eta k T}{q} \ln \frac{T}{T_0} \quad (10.4-8)$$

Differentiating this expression with respect to temperature, we obtain

$$\frac{dV_{be}}{dT} = -\frac{V_{go} + (\eta k T_0 / q) - V_{be0}}{T_0} - \frac{\eta k}{q} \ln \frac{T}{T_0} \quad (10.4-9)$$

The first term on the right-hand side of the equation, which is much larger than the second, is *constant* with temperature and characterizes the temperature dependence of V_{be} accurately over a wide operating range centered at T_0 . The second is a correction term that is temperature dependent. It is zero at T_0 . A plot of V_{be} versus temperature is shown in the upper portion of Fig. 10.15 for two values of collector current, 100 and 1.0 mA. Both curves are a linear approximation to the actual variation of V_{be} with temperature; the dashed line gives the actual curve in the 1 mA case, and the deviation from linearity is shown (expanded) as a correction curve below. This correction is essentially independent of collector current and applies to both curves.

The curves are plotted for a value of η of 2.2. Note that the intercept voltage V_I is the same for both straight-line approximations. From Eq. (10.4-8) the intercept voltage is given by

$$V_I = V_{go} + \frac{\eta k T_0}{q} \quad (10.4-10)$$

With $\eta = 2.2$ and $T_0 = 300^\circ\text{K}$, the intercept voltage is $1.205 + 2.2(0.026) = 1.262$ V.

For purposes of general circuit design, the slight curvature of the temperature characteristic of V_{be} can be ignored, so we can write

$$V_{be} \doteq V_I - (V_I - V_{be0}) \frac{T}{T_0} \quad (10.4-11)$$

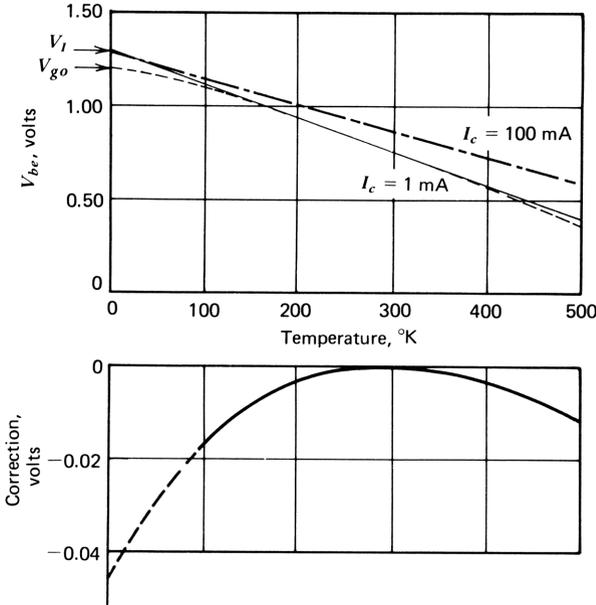


Figure 10.15. Base-emitter voltage as a function of temperature for a transistor operated at two currents two orders of magnitude different. Variation is linear within the correction given in the lower diagram.

This is an extremely useful equation for the design of networks to have a prescribed temperature performance, as we shall see in the text that follows. It shows the negative variation of V_{be} with temperature in terms of quantities that are readily ascertained. We can choose any desired value of T_0 ; V_I will be a slight function of T_0 and V_{be} a strong function. Equations (10.4-8) and (10.4-10) may be used to find values of V_I and V_{be0} for any new reference temperature. The importance of selecting a reference temperature appropriately is that the curvature correction is zero at the reference temperature. Hence T_0 should be selected at the center of the temperature range over which a design is to work.

The V_{be} difference between two transistors has a *positive* variation with temperature and is found by subtracting their logarithmic expressions:

$$V_{be1} - V_{be2} = \frac{kT}{q} \ln \frac{J_{c1}}{J_{c2}} \tag{10.4-12}$$

where $J_c = I_c/A_e$ is the collector current density. The saturation current density I_S/A_e is assumed to be the same for the two transistors. By operating two transistors at different current densities, we have a readily available voltage source with a positive temperature coefficient. The curvature of the V_{be} variation of the two transistors cancels, so that this positive variation is more linear than the negative variation of V_{be} of (10.4-11). By combining both

effects—negative and positive—much design freedom exists for obtaining desired temperature performance of circuits.

A Band-Gap Voltage Reference⁴

Widlar’s band-gap reference provides an excellent illustration of a practical application of the relationships derived previously. The object is to obtain a voltage reference whose output is independent of temperature. The principle is to add a voltage proportional to the difference between two base-emitter voltages (with positive temperature coefficient) to the base-emitter voltage of a transistor (with negative temperature coefficient) such that the temperature variations cancel. Many circuits employing the principle have been devised; perhaps the simplest is shown in Fig. 10.16.⁵ It employs an operational amplifier that will be taken as ideal for purposes of explanation. The effect of nonzero input offset voltage is discussed later.

In this circuit diode voltage V_A is established by feeding a current from the output through R_A to the positive input of the operational amplifier. A smaller current is fed to diode B through R_B and R_C . Since the voltages across R_A and R_B are the same, the currents are in inverse relation to the resistances. Since diode voltage V_B is smaller than V_A , a small voltage appears across R_C . The

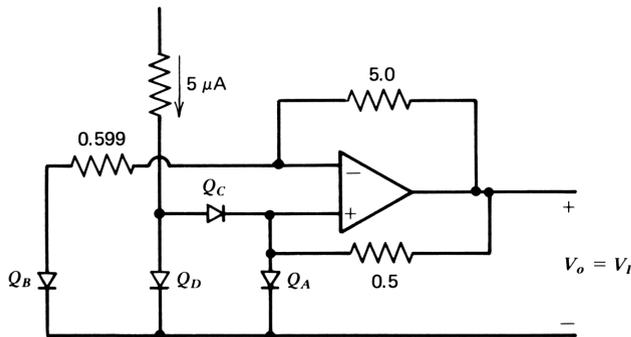
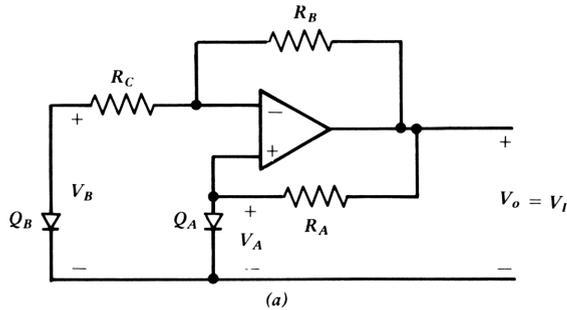


Figure 10.16. Basic band-gap reference circuit. A start-up circuit has been added in part *b*.

current through R_C , which also flows through R_B , is given by

$$I_B = \frac{V_A - V_B}{R_C} \quad (10.4-13)$$

The output voltage of the circuit is given by the sum of V_A and the drop across R_B and is

$$V_o = V_A + \frac{R_B}{R_C}(V_A - V_B) \quad (10.4-14)$$

In this equation V_A has a negative temperature coefficient, and $V_A - V_B$ has a positive temperature coefficient. The resistor ratio multiplies $V_A - V_B$ and also multiplies its positive temperature coefficient. For one value of this ratio, the temperature coefficients will cancel, and the output voltage will not be a function of temperature. We can find this ratio by substituting the diode voltage of eq. (10.4-11) for the two diodes in (10.4-14):

$$\begin{aligned} V_o &= V_I \left(1 - \frac{T}{T_0}\right) + V_{A0} \left(\frac{T}{T_0}\right) + \frac{R_B}{R_C}(V_{A0} - V_{B0}) \left(\frac{T}{T_0}\right) \\ &= V_I - \frac{T}{T_0} \left[V_I - V_{A0} - \frac{R_B}{R_C}(V_{A0} - V_{B0}) \right] \end{aligned} \quad (10.4-15)$$

This voltage, V_o , will be constant with temperature if the second term is made zero. Under these conditions the output voltage is V_I , the intercept voltage. Because V_I is closely equal to the band-gap voltage, this voltage reference circuit is termed a *band-gap reference*. The value of R_C to give an output voltage of V_I is found by setting the second term to zero and is

$$R_C = \frac{V_{A0} - V_{B0}}{V_I - V_{A0}} R_B \quad (10.4-16)$$

If the diode saturation currents are the same, then

$$V_{A0} - V_{B0} = \frac{kT_0}{q} \ln \frac{I_{A0}}{I_{B0}} \quad (10.4-17)$$

Note that when the output voltage $V_o = V_I$,

$$I_A = \frac{1}{R_A}(V_I - V_A) \quad (10.4-18)$$

$$\begin{aligned} &= \frac{1}{R_A} \left[V_I - V_I \left(1 - \frac{T}{T_0}\right) - V_{A0} \frac{T}{T_0} \right] \\ &= \frac{1}{R_A}(V_I - V_{A0}) \frac{T}{T_0} \end{aligned} \quad (10.4-19)$$

and hence is a linear function of temperature. Thus the value of ξ in eq. (10.4-6) is unity, so that $\eta = 2 - \sigma$. Therefore, from eq. (10.4-10) the value of V_I to be used in this circuit design is,

$$V_I = 1.205 + \frac{(2 - \sigma)kT_0}{q} \quad (10.4-20)$$

For $\sigma = 0.8$, $V_I = 1.236$ V for $T_0 = 300^\circ\text{K}$.

Design Example

An example of the design of a band-gap reference is given in Fig. 10.16*b*, in which two diodes that have $V_{B0} = 0.736$ V at 1.0 mA (at 300°K) are used ($I_S = 5 \times 10^{-13}$ mA). With $V_I = 1.236$ V, the voltage across R_A is 0.500 V, and the current through it is 1.0 mA, making $R_A = 0.5$ k Ω . If we make $R_B = 5.0$ k Ω , $I_B = 0.1$ mA; $V_{A0} - V_{B0}$ is then

$$V_{A0} - V_{B0} = \frac{kT_0}{q} \ln 10 = 0.0599 \text{ V} \quad (10.4-21)$$

and from (10.4-16)

$$R_C = \frac{0.0599}{0.500} R_B = 0.599 \text{ k}\Omega \quad (10.4-22)$$

thus completing the design.

The design example given here is sensitive to the input offset voltage of the operational amplifier. The offset voltage adds to V_A . To see its effect, we assume that the offset is proportional to temperature. The reason for this assumption is that the offset arises because of the V_{be} difference of the amplifier differential input transistors. This difference is thus proportional to temperature. We can represent the offset as

$$V_e = V_{e0} \left(1 + \frac{T}{T_0} \right) \quad (10.4-23)$$

where V_{e0} is the value of V_e at T_0 . The offset is in series with V_A , so that we find its effect by adding it to V_A where it appears in eq. (10.4-14). The terms to be added to (10.4-15) to account for the offset are obtained from (10.4-23). They can be expressed as the sum of a temperature-independent term and a temperature-sensitive term:

$$\Delta V_0 = \left(1 + \frac{R_B}{R_C} \right) V_{e0} + \frac{T}{T_0} \left(1 + \frac{R_B}{R_C} \right) V_{e0} \quad (10.4-24)$$

Thus both the output voltage and its temperature variation are affected by the offset voltage of the amplifier.

With an input offset voltage of 1.0 mV, for example, the error in the output voltage would be $1 + R_B/R_C = 9.4$ mV, and the temperature coefficient would be $9.4/300 = 0.031$ mV/°C. The problem here is that the difference voltage $V_A - V_B$ is too small; when it is amplified, the error voltage is also amplified by the same amount.

To minimize the effect, the difference voltage can be increased by reducing the current density in diode B relative to diode A by increasing the (1) current ratio I_A/I_B and (2) area of diode B relative to diode A , thereby reducing the current density in diode B . A third method is to use several diodes in series for both A and B , thereby reducing the relative effect of the offset. Clearly, a low-drift amplifier should be used. To evaluate the improvement realized by the first two methods, suppose that the area of diode B is increased by a factor of 10 over that of A and that the current ratio is increased to 40:1. Then, from (10.4-12), we obtain

$$V_{A0} - V_{B0} = 0.1558 V \quad (10.4-24)$$

With $V_I - V_{A0} = 0.5$ V, R_B/R_C then becomes

$$\frac{R_B}{R_C} = \frac{0.500}{0.1558} = 3.21 \quad (10.4-25)$$

The output error for 1 mV offset then becomes $1 + R_B/R_C$, or 4.21 mV, about a factor of 2 improvement. For improved accuracy, see Kujik⁵.

The circuit in Fig. 10.16a is *bistable* as a result of the positive feedback path from the output through R_A to the positive amplifier input. One stable state is the desired one, with an output voltage of V_I . In this state the feedback signal to the negative input dominates; the positive feedback is minimal because of the low impedance of diode A . With zero output voltage, however, the low resistance of R_A and the high impedance of diode A serve to keep the voltage at zero. To prevent the circuit from remaining in this state, a start-up circuit has been added in Fig. 10.16b. If the voltage at the positive input is zero, current flows through diode C into the positive input, turning the circuit on. As soon as this occurs, the voltage across diode A diverts the start-up current from diode C to diode D , where it has no effect on circuit operation.

10.5 CLASS B OUTPUT STAGE BIAS DESIGN FOR PRESCRIBED TEMPERATURE DEPENDENCE

We found in Section 10.3 that crossover distortion can be reduced in a Class B amplifier by adding emitter resistors having the value

$$R_E = \frac{kT}{qI_Q} \quad (10.5-1)$$

This equation shows that the added resistors should have a positive temperature coefficient. Alternatively, we can use fixed resistors if the quiescent current is increased with temperature, that is, if we impose the condition that

$$\frac{I_Q}{I_{Q0}} = \frac{T}{T_0} \quad (10.5-2)$$

where I_{Q0} is the quiescent current at a reference temperature. Then

$$R_E = \frac{kT_0}{qI_{Q0}} \quad (10.5-3)$$

In the discussion that follows, we show how this prescribed variation of I_Q can be obtained with a driver current source that is independent of temperature.

Consider the circuit in Fig. 10.17a operating under quiescent conditions. It includes resistors R_E as given by (10.5-3). Thus the base-to-base output stage input voltage can be written

$$V_{BB} = 2(V_{be} + I_Q R_E) \quad (10.5-4)$$

so that from (10.4-11) we can write

$$\frac{V_{BB}}{2} = V_I - \left((V_I - V_{be0}) - \frac{kT_0}{q} \right) \frac{T}{T_0} \quad (10.5-5)$$

in which the term kT_0/q is the drop across the emitter resistor at T_0 .

The base-to-base voltage appears across the diode network. In this network we have added resistor $2R_d$ for convenience in adjusting the driver current relative to the output quiescent current. Thus we can write

$$\frac{V_{BB}}{2} = V_I - [(V_I - V_{d0})] \frac{T}{T_0} + I_d R_d \quad (10.5-6)$$

Subtracting this equation from (10.5-5), we obtain

$$0 = \left(V_{be0} - V_{d0} + \frac{kT_0}{q} \right) \frac{T}{T_0} - I_d R_d \quad (10.5-7)$$

The quantity in the parentheses is not a function of temperature; we also assume that R_d is temperature insensitive. Hence I_d must be proportional to temperature:

$$I_d = I_{d0} \frac{T}{T_0} \quad (10.5-8)$$

It follows from this equation and (10.5-2) that the diode current to quiescent

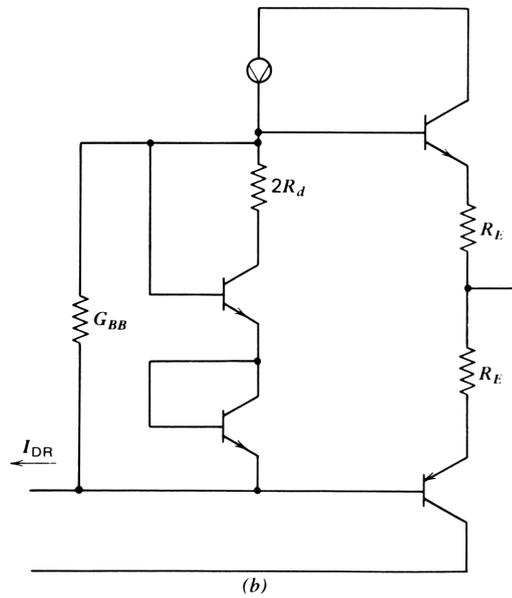
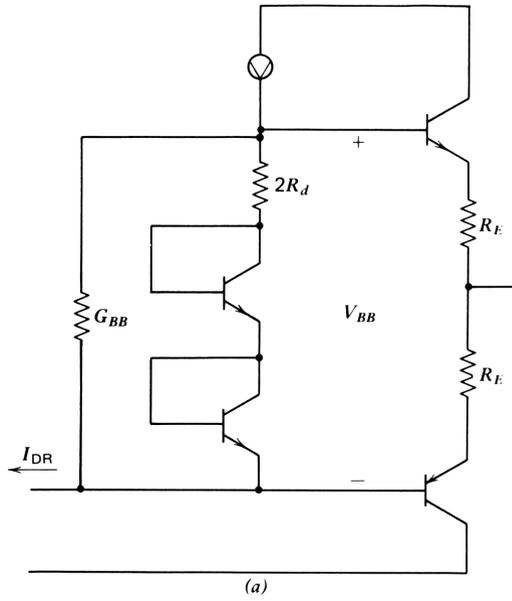


Figure 10.17. Circuit illustrating the bias circuit design of a Class B amplifier for a given temperature performance.

current ratio is independent of temperature. This ratio is given by

$$\frac{I_d}{I_Q} = \frac{I_{d0}}{I_{Q0}} = \exp \frac{q}{kT_0} (R_E I_{Q0} - R_d I_{d0})$$

Since $R_E I_{Q0} = kT_0/q$, the ratio becomes

$$\frac{I_d}{I_Q} = \exp \left(1 - \frac{qR_d I_{d0}}{kT_0} \right) \quad (10.5-9)$$

Thus for $R_d = 0$, the diode current is e times the output-stage quiescent current (for equal emitter areas of diode and transistor); the diode current (the driver quiescent current) *decreases* as R_d is increased.

We may wish to *increase* the driver current relative to the quiescent current. If the area of the diodes is smaller than the transistors to save room on an integrated circuit chip, for example, the driver current may become too small to satisfy the drive requirement. The circuit in Fig. 10.17*b* is a method for increasing I_{DR}/I_Q . The equations for this circuit are the same as those developed previously, with $-R_d$ replacing R_d .

We can alternatively obtain a driver current that is *independent* of temperature, by connecting a conductance in parallel with the diodes, as shown in Fig. 10.17. Since the diode current increases with temperature and V_{BB} falls with temperature, there exists some value of G_{BB} for which the total driver-stage current is independent of temperature. It is found by writing the expression for the driver current:

$$I_{DR} = I_{d0} \frac{T}{T_0} + G_{BB} V_{BB} \quad (10.5-10)$$

Applying (10.5-5), we have

$$I_{DR} = 2G_{BB} V_I + \left[I_{d0} - 2G_{BB} \left(V_I - V_{be0} - \frac{kT_0}{q} \right) \right] \frac{T}{T_0} \quad (10.5-11)$$

For the driver current to be independent of temperature, the second term must be zero. The driver current is then given simply by the first term:

$$I_{DR} = 2G_{BB} V_I \quad (10.5-12)$$

so that

$$G_{BB} = \frac{I_{DR}}{2V_I} \quad (10.5-13)$$

Since the driver current is known at the outset (from current drive requirements), we obtain G_{BB} immediately.

To find I_{d0} and R_d , we set the second term in (10.5-11) to zero, from which

$$I_{d0} = 2G_{BB} \left(V_I - V_{be0} - \frac{kT_0}{q} \right) \tag{10.5-14}$$

With I_{d0} and I_{Q0} known, we obtain R_d from (10.5-9) as

$$R_d = \frac{kT_0}{qI_{d0}} \left(1 - \ln \frac{I_{d0}}{I_{Q0}} \right) \tag{10.5-15}$$

Note that $2R_d$ is to be added to the circuit. Where R_d becomes negative, the circuit in Fig. 10.17b must be used, for which the equations are slightly different, since the conductance is placed across the base-emitter drops of the diodes rather than the transistors. The equations for this case are left as an exercise.

The principles illustrated by the two examples—the band-gap reference and the Class B bias design—can be applied to many situations in which a desired variation of current or voltage is to be obtained. The basic equations are (10.4-11) and (10.4-12).

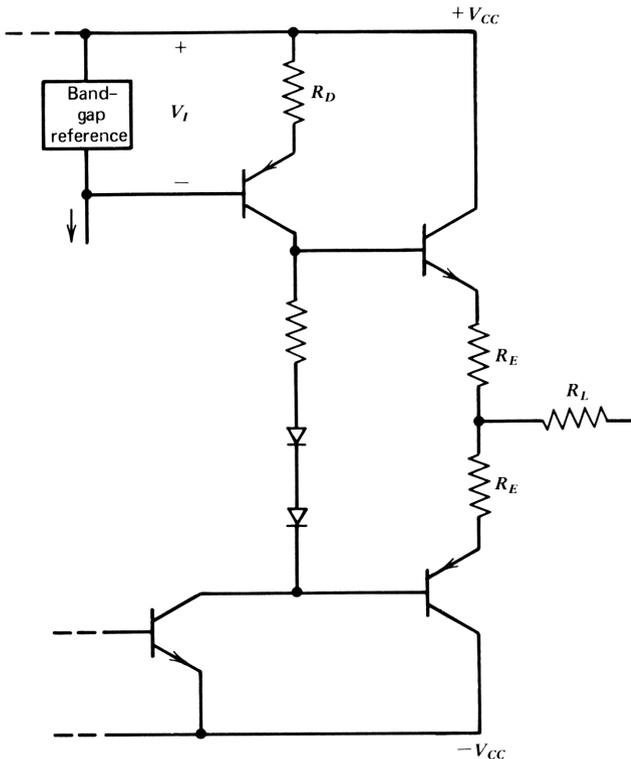


Figure 10.18. Circuit designed to provide both output-stage and driver-stage quiescent currents proportional to temperature.

Another way to provide a quiescent current that is proportional to temperature is to ensure that the driver quiescent current is also proportional to current; in this case G_{BB} would be removed from the circuit. A simple way to obtain this driver current temperature variation is shown in Fig. 10.18, in which the base of the driver current source is connected to a (temperature-independent) band-gap reference source. When this is done, we can write

$$V_I = V_{be0} + I_D R_D$$

or

$$V_I = V_I - \left(V_I - V_{be0} - \frac{kT_0}{q} \right) \frac{T}{T_0} + I_D R_D$$

so that

$$I_D = \frac{V_I - V_{be0}}{R_D} \frac{T}{T_0}$$

which has the desired temperature dependence.

PROBLEMS

- 1 In the circuit in Fig. 10.19, find the value of R_B if I_{in} is 1.0 mA, constant with temperature, for the output current to be proportional to absolute temperature, given by

$$I_C = 1.0 \frac{T}{300}$$

Find R_E .

- 2 In the circuit of Problem 1, find the values of R_B and R_E if the output collector current is to vary with temperature as

$$I_C = 1.0 \left(1 + k \frac{T}{300} \right)$$

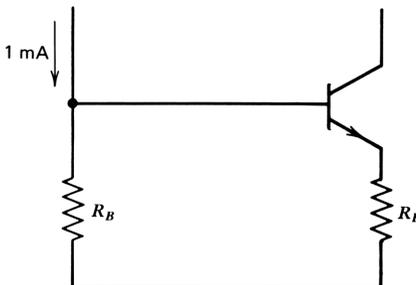


Figure 10.19.

where $k=0.1$. Repeat for $k=-0.1$. What is the temperature coefficient in each case, expressed in percent per degree Celsius?

- 3 Find the output voltage of the band-gap reference in Fig. 10.16*b* with an ideal operational amplifier at temperatures of 250, 300, and 350°K, taking curvature of V_{be} with temperature into account.

REFERENCES

- 1 R. J. Widlar, "New Developments in IC Voltage Regulators," *IEEE J. Solid State Circuits* **SC-6**, 2–7 (February 1971).
- 2 W. Kruppa and F. D. Waldhauer, "A UHF Monolithic Operational Amplifier," *ISSCC Digest of Technical Papers*, Vol. XXI, IEEE Cat. No. 78CH 1298-9 SSC 1978, p. 74.
- 3 H. P. Kelly, "Differential Phase and Gain Measurements in Color TV Systems," *Electr. Eng.* 799–802 (September 1954).
- 4 R. J. Widlar, "An Exact Expression for the Thermal Variation of the Emitter Base Voltage of Bipolar Transistors," *Proc. IEEE (Lett.)*, **55**, 96–97 (January 1967).
- 5 K. E. Kujik, "A Precision Reference Voltage Source," *IEEE J. Solid State Circuits* **SC-8**, 222–226 (June 1973).

Chapter 11

Noise and Input Stages

Noise imposes a lower limit on the signal level that can be processed by systems. In an information transmission system a signal is transmitted through some transmission medium (space, copper cable, glass fiber, etc.) to a remote receiver whose maximum usable distance from the transmitter is controlled by noise introduced by its own amplifying devices, in addition to external noise. To maximize this distance, we are interested in minimizing the noise. In this chapter we investigate the basic sources of noise in devices and circuits and shown how the noise can be calculated.

The standard method of specifying noise in circuits and devices is by an input noise network containing a series noise voltage generator and a shunt noise current generator that together represent the contributions of all the noise sources in the circuit or device.^{1,2} The translation of any one internal noise source to the input is accomplished through use of the $ABCD$ matrix of the network between it and the input. However, since noise signals from independent physical noise sources in a network must be added on a power basis, care must be exercised when combining noise from several sources.

Calculation of the total noise is found by integrating the noise over the frequency range of interest. The spectral density of noise referred to the network input is often a simple function of frequency, simplifying the integration. The effect of equalization on the input noise network generators is developed.

A separate section is devoted to the effects of feedback and feedforward on noise.

11.1 SOURCES OF NOISE IN ELECTRICAL CIRCUITS

Electrical noise may be defined generally as an unwanted signal. We usually think of noise in somewhat more restricted terms, as a random variation in the voltage or current of a circuit. Such variations arise in many ways, the more important of which are listed in Table 11.1. In this chapter we focus primarily on thermal and shot noise that arises in resistors, diodes, and transistors.

Shot Noise^{4,5}

Shot noise in pn junctions arises by the discrete nature of charge flow since charge is quantized into units of q . As it crosses the junction, each electron (or hole) constitutes an impulse of current. The sum of all such impulses constitutes the dc current I . The time of each event is random, giving rise to a noise current superimposed on the dc that can be shown to be the mean square value of the fluctuation of direct current I , given by

$$\overline{i_n^2} = 2qI\Delta f \quad (11.1-1)$$

where Δf is the frequency band over which the fluctuation (or noise) is evaluated. In our consistent set of units, $q = 1.6 \times 10^{-7}$ pC, I is in milliamperes, and Δf is in gigahertz. Thus in a 100 MHz band, the noise associated with a current of 10 mA will be

$$\begin{aligned} i_{n\text{rms}} &= \sqrt{3.2 \times 10^{-7} \times 10 \times 0.1} \\ &= 0.566 \times 10^{-3} \text{ mA} \end{aligned}$$

Table 11.1 Summary of Major Electrical Noise Sources³

Natural sources
Circuit noise
Thermal noise
Shot noise
$1/f$ or excess noise
Received noise
Atmospheric noise
Galactic noise
Synthetic sources
Received interference
Hum

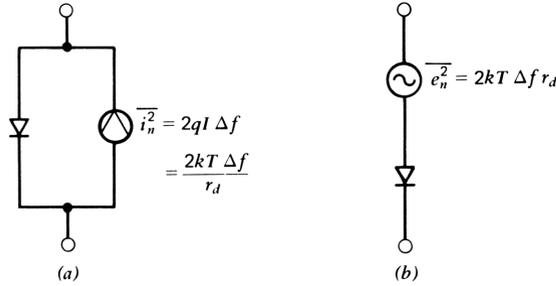


Figure 11.1. Representation of diode noise.

or $0.566 \mu\text{A}$. Equation (11.1-1) implies that the spectral response of shot noise is flat with frequency; this holds up to extremely high frequencies, roughly the reciprocal of the transit time of carriers across the diode depletion region. The spectral distribution is given the name “white noise,” by analogy with visual light in which all frequencies are present.*

Since shot noise arises from the occurrence of a large number of independent events, the amplitude distribution of the noise is Gaussian by the central limit theorem. The distribution has a mean value given by the direct current I , and eq. (11.1-1) gives the variance of the distribution.

The dynamic resistance r_d of a diode is given by kT/qI . Equation (11.1-1) can be written in terms of this resistance by replacing qI by kT/r_d :

$$\overline{i_n^2} = \frac{2kT\Delta f}{r_d} \quad (11.1-2)$$

The mean square noise voltage across the diode junction can be found by multiplying the mean square diode noise current by r_d^2 :

$$\overline{e_n^2} = 2kT\Delta f r_d \quad (11.1-3)$$

The noise equivalent circuit for the forward-biased diode can be drawn as a noiseless diode in parallel with a mean square noise current represented as a current generator, as in Fig. 11.1a, or alternatively as the noiseless diode in series with a mean square voltage generator, as in Fig. 11.1b.

Thermal Noise⁶

Fluctuating electrical charge in a circuit—Brownian motion of electrons—generates *thermal noise*, first observed by Johnson at the output of a vacuum tube amplifier in 1927. In a resistor these fluctuations give rise to noise power (at temperatures above absolute zero) that can be extracted from the resistor. If

*The analogy is incorrectly drawn since white light consists of equal amounts of light in each *wavelength* interval rather than in each frequency interval. We adopt conventional terminology here.

two resistors at the same temperature are connected together, there will be an exchange of noise power between them with power flow equal in the two directions. If the temperatures are different, there will be a net power flow from the hotter resistor to the cooler one.

The circuit representation of thermal noise in a resistance R is a Thevenin voltage generator in series with a noiseless resistance that has the mean square value

$$\overline{e_{nR}^2} = 4kT\Delta f R \quad (11.1-4)$$

where Δf is the band over which the noise is of interest. In terms of the thermal voltage, $V_T = kT/q = 0.026$ V at room temperature,

$$\overline{e_{nR}^2} = 4V_T q \Delta f R \quad (11.1-5)$$

A resistor of 1 k Ω will exhibit an rms thermal noise voltage in a 100 MHz band of

$$\begin{aligned} e_{nR\text{rms}} &= \sqrt{0.104 \times 1.6 \times 10^{-7} \times 0.1 \times 1} \\ &= 40.7 \times 10^{-6} \text{ V} \end{aligned}$$

A Norton source can also be used to represent the noise, in which case the mean square noise *current* is given by

$$i_{nG}^2 = 4kT\Delta f G \quad (11.1-6)$$

$$= 4V_T q \Delta f G \quad (11.1-7)$$

For a 1 k Ω resistor, $G = 1$ mS, so the rms value of the noise current over the same 100 MHz band will be 40.7×10^{-6} mA.

The spectrum of thermal noise is flat and may be considered white noise. As the amplitude distribution is Gaussian, thermal and shot noise are virtually indistinguishable as observed in a circuit. However, there are differences; thermal noise is temperature dependent whereas shot noise is not and shot noise depends on current, whereas thermal noise does not.

Thermal noise is used as a standard of comparison or measure of noise from various sources. Thus the *noise equivalent resistance* R_n of a noise source is defined for a voltage noise source e_n as

$$R_n = \frac{\overline{e_n^2}}{4kT\Delta f} \quad (11.1-8)$$

and is the resistance whose thermal noise would give a mean square voltage of $\overline{e_n^2}$. The noise equivalent resistance of a resistor exhibiting thermal noise is the

resistance itself. The noise equivalent resistance of a shot noise source such as the diode described by eq. (11.1-3) would be

$$R_{nd} = \frac{r_d}{2} \quad (11.1-9)$$

Noise equivalent conductance G_n is similarly defined for a current noise source i_n as

$$G_n = \frac{\overline{i_n^2}}{4kT\Delta f} \quad (11.1-10)$$

Thus the noise equivalent conductance of the diode described by eq. (11.1-2) is

$$G_{nd} = \frac{1}{2r_d} \quad (11.1-11)$$

Note that these noise equivalent immittances do not denote a physical circuit immittance. They provide a means for expressing the mean square value of a noise voltage or current and are given in kilohms or millisiemens. Their chief value is in conveying the magnitude of noise of a circuit in familiar terms and also in simplifying noise equations.

One-Over- f Noise

A type of noise less well understood than those described previously is $1/f$ noise, also called *excess noise*, *fluctuation noise*, or *flicker noise*. It arises in such a broad range of physical phenomena that it may in some way mirror a subtle statistical property of the world.⁷ From the record of annual flood levels of the Nile River to the wobbling of the earth's axis or the flickering of a flame, $1/f$ statistics are ubiquitous.⁸ In electric circuits and devices, $1/f$ noise may be caused by fluctuations in the conductivity of the medium, as in fluctuations in the number of carriers available to take part in the conduction process.⁹ In transistors, traps associated with defects and heavy-metal atoms in the emitter junction depletion layer capture conduction carriers and release them; the time constants involved in this process may be extremely long, giving a noise that is concentrated in the low-frequency portion of the spectrum.¹⁰

Flicker noise current has a mean square value given by

$$\overline{i_{nf}^2} = K \frac{I^\gamma}{f^\alpha} \Delta f, \quad 1 < \gamma < 2 \quad (11.1-12)$$

where K is a constant, I is the direct current raised to a power γ , and α is an exponent near unity. We take both γ and α as unity in what follows.

Unlike shot or thermal noise, the spectrum of $1/f$ noise is not flat with frequency; thus the noise equivalent conductance for this noise will depend on

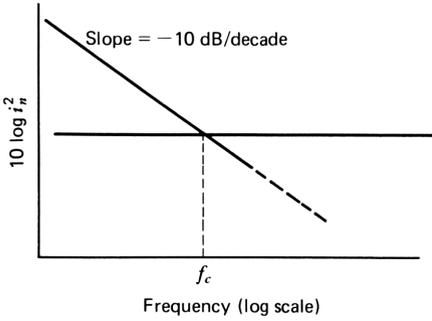


Figure 11.2. Spectral characteristic of $1/f$ noise.

the frequency range over which the noise is evaluated. In the earliest transistors, $1/f$ noise was the dominant noise source over the entire useful frequency range. As processing improved, $1/f$ noise dropped drastically; now it is of little practical consequence in the bipolar transistor. Typically, $1/f$ noise may exceed the sum of shot and thermal noise below a frequency of 0.1–1 kHz.

The equivalent input mean square noise current of a typical transistor is shown in Fig. 11.2 as a function of frequency; the contributions of (shot and thermal) white and $1/f$ noise are equal at 1 kHz (10^{-6} GHz) in this example, so that

$$\frac{KI^\gamma}{10^{-6}} = 4kTG_{nS} \quad (11.1-13)$$

where G_{nS} is the noise equivalent conductance due to shot noise in the transistor. If we take $G_{nS} = 0.2$ mS, for example (we evaluate noise in the transistor in the text that follows), we can evaluate KI :

$$KI = 4 \times 4 \times 10^{-9} \times 0.2 \times 10^{-6} = 3.2 \times 10^{-16} \text{ mA}^2$$

Suppose that we wish to find the total noise in the audio band from 20 to 20,000 Hz. We integrate over this range:

$$\begin{aligned} \overline{i_{nT}^2} &= \int_{f_1}^{f_2} \left(3.2 \times 10^{-9} + \frac{3.2 \times 10^{-15}}{f} \right) df \quad (11.1-14) \\ &= 3.2 \times 10^{-9} \left(f_2 - f_1 + 10^{-6} \ln \frac{f_2}{f_1} \right) \end{aligned}$$

We can express the total as a noise equivalent conductance by dividing by $4kT\Delta f$, where Δf is the total frequency range, or 1.998×10^{-5} GHz (20 kHz), so that

$$G_{nT} = 0.2 \left(1 + \frac{10^{-6}}{f_2 - f_1} \ln \frac{f_2}{f_1} \right) \quad (11.1-15)$$

for the numbers assumed, $G_{nT}=0.27$ mS, a modest increase over that for shot noise alone.

If the lower limit of integration is extended to zero to find the total noise for a dc amplifier, the integral becomes infinite. Zero frequency, on the other hand, exists only conceptually. If we extend the lower limit to one cycle per year (3×10^{-17} GHz), the noise conductance rises only to 0.47 mS because the \ln function approaches $-\infty$ so slowly. Reducing the lower limit by another factor of 20×10^9 gives us one cycle since the beginning of the universe and raises G_n to 0.71 mS.

The point here is that we can express noise as an equivalent conductance (or resistance) even if it is not flat with frequency. Such an expression may not be meaningful, however, as in the latter two preceding examples. Temperature variations will swamp out such small noise currents at low frequencies. One theory of the origin of $1/f$ noise (by Voss and Clarke) suggests that it is due to thermal equilibrium temperature variations.¹¹ The single number G_{nT} fails to take into account the character of the noise; in the audio band, $1/f$ noise sounds different—more “rumbly”—than shot noise; below the audio band, it behaves as “dc drift.” Often, weighting curves are used in noise measurements to equalize the subjective importance of noise over a band of frequencies, emphasizing noise in frequency bands where it is most annoying, for example.

We have devoted more space here to $1/f$ noise than to the other, more important sources of noise. This is because we do not treat it in any further detail in the text that follows since its effects are so small in silicon transistors. New devices often exhibit considerable $1/f$ noise, which is reduced as device processing is better understood and refined. For silicon devices, the reduction took place rapidly. The gallium arsenide FET still exhibits high $1/f$ noise at the time of this writing.

11.2 NOISE IN TRANSISTORS, DIODES, AND CURRENT SOURCES

Noise arises in the bipolar transistor from three primary sources, all uncorrelated: (1) shot noise associated with the minority carrier electron flow (or the collector current); (2) shot noise of the hole current injected into the emitter from the base, or base current; and (3) thermal noise of the base resistance. (This is for an *nnp* transistor. For a *pnnp*, interchange “hole” and “electron.”) These three sources are shown in the equivalent circuit in Fig. 11.3. (Where applicable, a $1/f$ noise generator is connected in parallel with the base current noise generator; an avalanche noise generator would be connected from collector to base. We ignore these here.)

In the equivalent circuit collector shot noise is represented by a current generator i_{nCS} between the collector and emitter terminals; the base current shot noise is represented by a current generator i_{nBS} between the internal base and emitter terminals, and the base resistance thermal noise is represented by a series voltage generator e_{nBT} . The values of the generators are expressed as

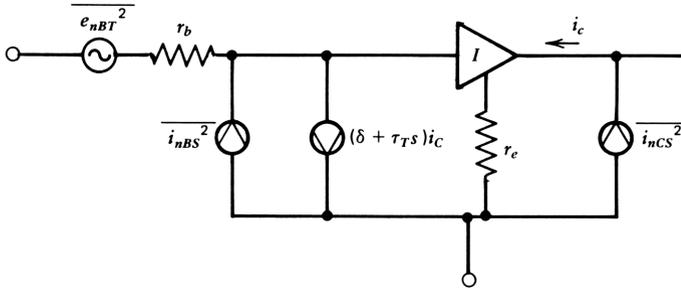


Figure 11.3. Three primary noise sources in the bipolar transistor.

mean square values:

$$\overline{i_{nCS}^2} = 2qI_C\Delta f = \frac{2kT\Delta f}{r_e} \tag{11.2-1}$$

$$\overline{i_{nBS}^2} = 2qI_B\Delta f = 2q\delta I_C\Delta f = \frac{2kT\Delta f\delta}{r_e} \tag{11.2-2}$$

$$\overline{e_{nBT}^2} = 4kT\Delta f r_b \tag{11.2-3}$$

Each of these three noise generators may be expressed as equivalent noise immittances—the value of resistances or conductances that would have the same thermal noise as exhibited by the actual noise generator. Thus by dividing each generator by $4kT\Delta f$, we obtain

$$G_{nCS} = \frac{1}{2r_e} \tag{11.2-4}$$

$$G_{nBS} = \frac{\delta}{2r_e} \tag{11.2-5}$$

$$R_{nBT} = r_b \tag{11.2-6}$$

Note that at 300°C , $4kT = 1.66 \times 10^{-8}$ pJ. Multiplying this factor by the bandwidth Δf in gigahertz and by the noise equivalent resistance in kilohms (or conductance in millisiemens) gives the *square* of the noise voltage (or current) in volts (or milliamperes). Collector shot noise, therefore, is equivalent to connecting the noise current generator of a resistor of $2r_e$ from collector to emitter; for $I_C = 1$ mA, for example, it is equivalent to connecting the *noise current* of a 52Ω resistor from collector to emitter. In this example base shot noise is equivalent to connecting the *noise current* of a $5.2 \text{ k}\Omega$ resistor from base to emitter, assuming $\delta = 1/h_{FE} = 0.01$.

We can illustrate the use of equivalent noise immittances and the calculation of noise at various points in a circuit by a few simple examples. In what

follows, we find the noise associated with a diode-connected transistor and also the simple current source discussed in Chapter 9.

Diodes exhibit shot noise that have a noise equivalent conductance of $1/2r_d$, where r_d is the dynamic resistance of the diode. Where the diode is formed from a transistor with the base and collector shortcircuited, it also exhibits series thermal noise voltage; this is most easily seen from the equivalent circuit in Fig. 11.4a, in which the three transistor noise sources are shown explicitly. We now introduce a series of generator transformations like those shown in Fig. 8.14b (the Blakesley transformation) and Fig. 8.14c (the current generator splitting transformation). The object is to place these noise sources at the diode terminals, taken here as the base and emitter terminals of the transistor. As a first step, we replace the base shot noise generator by one connected directly across the terminals, as shown in Fig. 11.4b. This figure shows only the portion of the circuit in the dotted box of Fig. 11.4a. The transformation is effected by splitting the base shot noise generator in two with one across the input terminals and one across the base resistance. This latter generator can be replaced by its Thevenin equivalent as shown in the third diagram of Fig. 11.4b. Note that in obtaining the noise voltage in this way, we must multiply the (squared) current by r_b^2 to obtain the (squared) voltage.

In Fig. 11.4c the results of this transformation are shown in the complete circuit, with the two (uncorrelated) series voltage generators representing the thermal noise of r_b and the shot-noise-induced voltage across r_b . Note that in labeling these generators, we have dropped the factor $4kT\Delta f$, which multiplies them in Fig. 11.4a. We then obtain the final circuit in Fig. 11.4d by moving these voltage generators through the input node, giving a series generator at the input and another in series with the collector of the transistor. This latter generator may be ignored since it is a voltage generator in series with the high-impedance collector of the transistor. All significant noise sources have been moved to the terminals, and the transistor with its collector-to-base short circuit can be represented as a noiseless diode, as shown in the second diagram in Fig. 11.4c. In this diagram we have replaced $r_e/(1+\delta)$ by its equivalent diode resistance r_d .

As a final step, we can convert the current source and diode resistance to a Thevenin (squared) voltage by multiplying the current $1/2r_d$ by the square of the diode resistance r_d^2 , so that the total noise voltage can be represented by a total noise equivalent resistance

$$R_{nD} = r_b + \frac{\delta r_b^2}{r_e} + \frac{r_d}{2} \quad (11.2-7)$$

At low currents the second term can be dropped. In a *pn* junction diode the r_b term is absent since the total thermal noise is associated only with contact resistance, which is usually negligible.

It is usually convenient to represent all sources of noise in a network by noise generators at the network input. This procedure is formalized in the

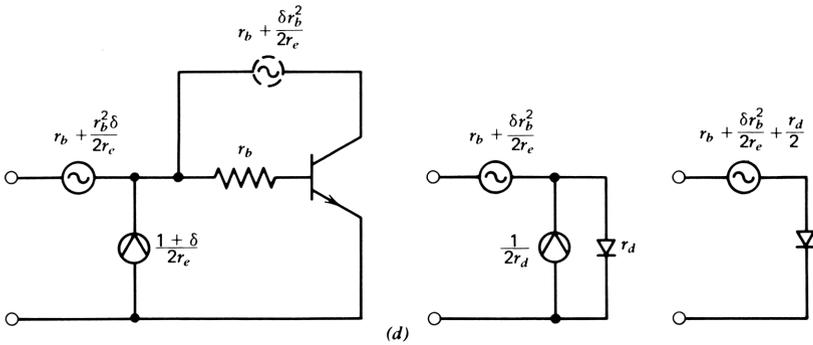
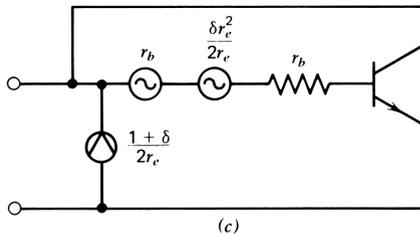
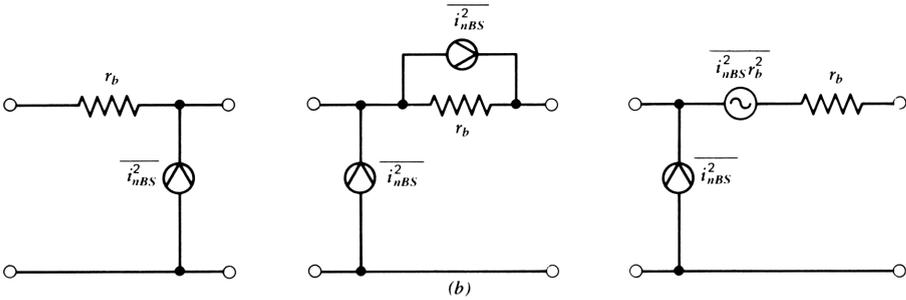
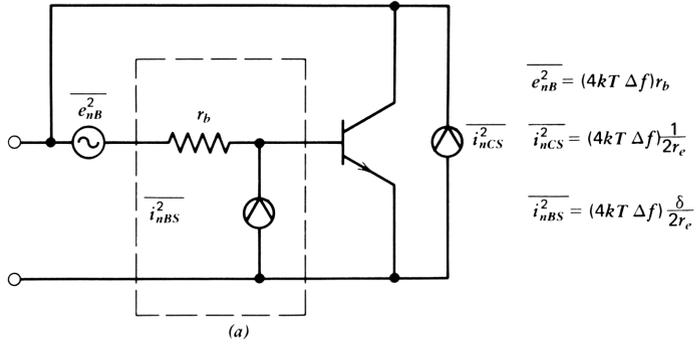


Figure 11.4. Transistor connected as a diode.

following section. Here, we adopt a less formal approach to find the equivalent input noise generators (at low frequencies) for a transistor and then use these results to find the input noise as well as the output noise of the simple unity loss current source discussed in Chapter 9.

Figure 11.5a shows a transistor with its three physical noise sources as discussed previously. Two of these sources, the base resistance thermal noise and the base current shot noise, are already at the input (we ignore the small voltage contribution of $r_b^2/2r_e$ caused by the base shot noise current flowing through r_b discussed immediately above). It remains to transform the collector shot noise generator to the input. This output noise causes both an input noise voltage and noise current. The input noise voltage (squared) is found by multiplying the output noise current (squared) by r_e^2 (parameter B^2 of the

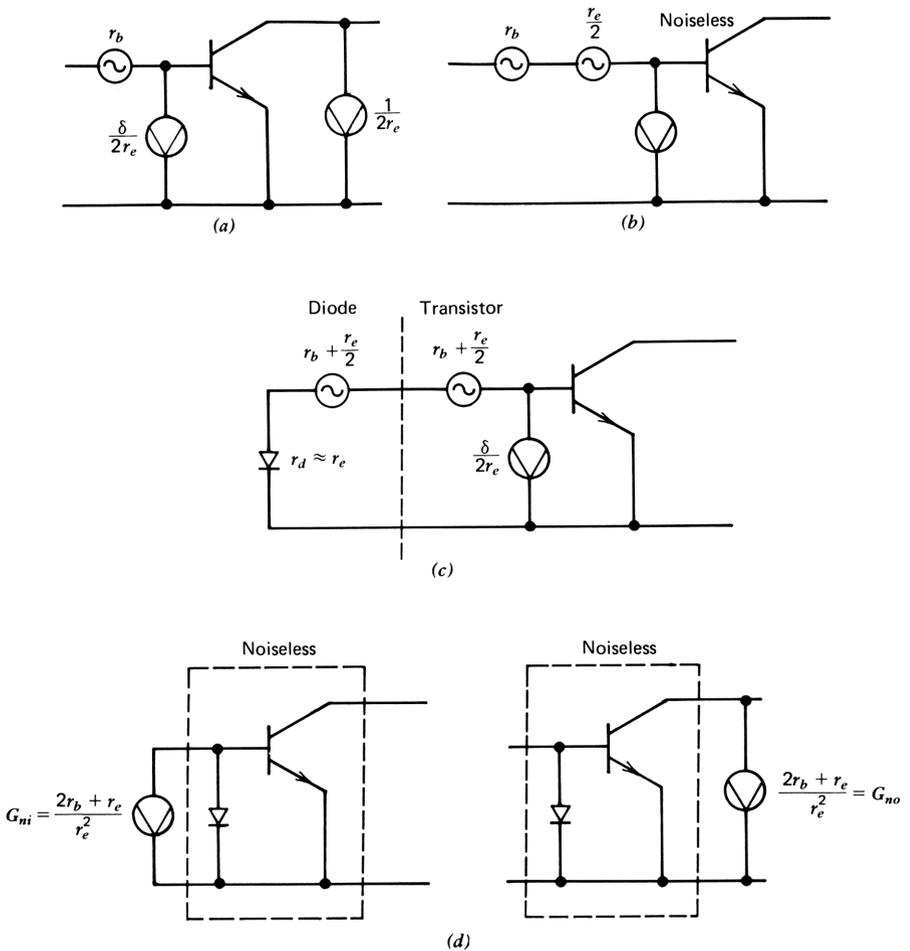


Figure 11.5. Noise equivalent circuit for a current source.

transistor); thus

$$R_{nCS} = r_e^2 \cdot \frac{1}{2r_e} = \frac{r_e}{2} \quad (11.2-8)$$

The input noise current is similarly found by multiplying the (squared) collector shot noise current by δ^2 (parameter D^2 of the transistor); thus

$$G_{nCS} = \frac{\delta^2}{2r_e} \quad (11.2-9)$$

Note that this contribution is less than that of the base current shot noise by a factor of δ and can be ignored. Therefore, the effect of collector shot noise at low frequencies is equivalent to an input noise voltage represented by $r_e/2$, as shown in Fig. 11.5*b*.

In Fig. 11.5*c* the diode is connected across the transistor input terminals, with its noise represented by the series noise voltage generator (squared). The noise voltage is approximated as $r_b + r_e/2$; for clarity, we have dropped terms containing δ as a factor. The voltage contributions of the diode and transistor are about equal. The contribution of the transistor base current shot noise is found by multiplying the current generator (squared) by the diode resistance $r_d = r_e$, giving $\delta r_e/2$; this is a factor of δ below the noise voltage contribution of the collector shot noise and is also ignored.

In Fig. 11.5*d* the series Thevenin noise voltages are summed and represented as a Norton current source across the input; the total noise voltage (squared) is divided by the squared resistance of the diode r_e^2 . The equivalent input noise current (squared) is represented by the equivalent input conductance G_{ni} :

$$\begin{aligned} G_{ni} &= \frac{2r_b + r_e}{r_e^2} \\ &= \frac{1}{r_e} + \frac{2r_b}{r_e^2} \end{aligned} \quad (11.2-10)$$

for $\delta \ll 1$. The term $1/r_e$ is the total collector shot noise contribution of the diode and transistor (twice the value of either one). The term $2r_b/r_e^2$ is the base resistance thermal noise of the transistor and diode, translated to current through the diode resistance. This latter term can be large: at a collector current of 1.0 mA and a base resistance of 0.1 k Ω , the total noise conductance is, for example

$$\begin{aligned} G_{ni} &= \frac{1}{0.026^2} + \frac{0.2}{0.026} \\ &= 38.5 + 295.9 = 334.3 \text{ mS} \end{aligned}$$

Notice that the base resistance thermal noise contribution is eight times that of the collector shot noise in this example (on a squared current, or power, basis).

For the unity loss current mirror, the equivalent noise current can be placed at either the input or the output without a change in value since the current loss is unity. Hence the equivalent output noise G_{no} is equal to G_{ni} .

Figure 11.6a shows a way to quiet the current mirror by adding equal resistors in series with the diode and the transistor emitter. On the surface it seems paradoxical that we can quiet the current mirror by adding resistors that are themselves noisy. The resistors add $2R_E$ to the total equivalent noise voltages of the input loop in Fig. 11.5c. Representing all noise sources at the input including thermal noise in both added resistors, we obtain for the total noise equivalent resistance at the input

$$R_{ni} = 2(r_b + R_E) + r_e$$

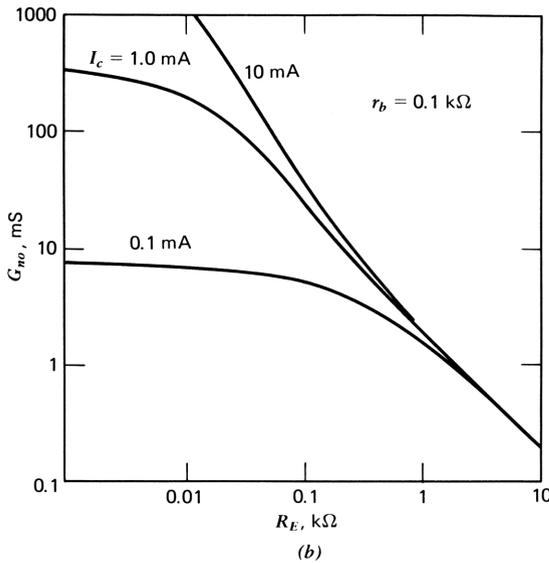
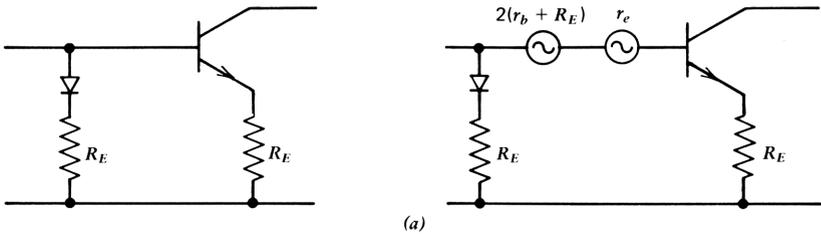


Figure 11.6. Noise equivalent conductance of a simple current source as a function of added series resistance.

where the emitter series resistor noise has been added in series with the diode. The total input mean square noise current can be expressed as a shunt input conductance, obtained by dividing R_{ni} by the square of the diode-resistor series impedance:

$$G_{ni} = \frac{2(r_b + R_E) + r_e}{(r_e + R_E)^2} \tag{11.2-11}$$

This noise current can be moved directly to the output since the current gain is unity; hence $G_{ni} = G_{no}$. The variation of G_{no} with R_E is shown in Fig. 11.6*b*; clearly, even a small value of R_E reduces the noise significantly.

Noise in Field Effect Transistors

The equivalent circuit of a field effect transistor (FET) with equivalent noise sources is shown in Fig. 11.7. The latter comprise thermal noise in the channel, which can be represented as an output current generator connected from drain to source, and shot noise from gate leakage current. The thermal noise component has been shown to be¹²

$$\overline{i_{nD}^2} = 4kT\Delta f\left(\frac{2}{3}\right)g_m \tag{11.2-12}$$

where

$$g_m = \frac{1}{r_{ch}} \tag{11.2-13}$$

in which r_{ch} is the channel resistance and g_m is the transconductance. The shot noise associated with the gate leakage current is

$$\overline{i_{nGS}^2} = 2qI_g\Delta f \tag{11.2-14}$$

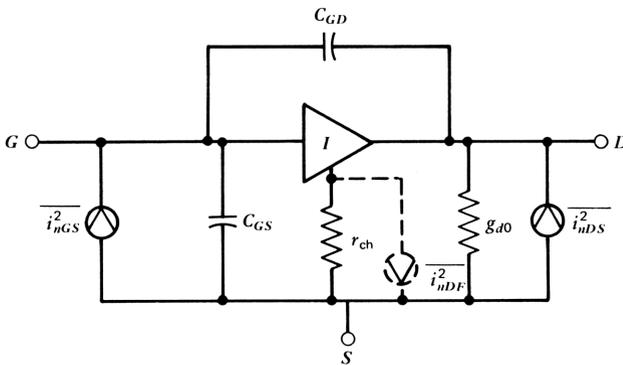


Figure 11.7. Noise equivalent circuit for an FET.

Flicker noise is of some consequence in FETs and may be represented by a current generator in parallel with the thermal noise generator. Its value is

$$\overline{i_{nDF}^2} = K \frac{I_d}{f} \Delta f \quad (11.2-15)$$

As in the case of the bipolar transistor, circuit calculations are more conveniently handled by representing all noise at the input of the device. We show next that this is a simple application of anticausal analysis.

11.3 EQUIVALENT INPUT NOISE NETWORK

Each resistor in a two-port network will exhibit thermal noise; each transistor will exhibit shot noise and thermal noise. To characterize the network, we are interested in the total noise, summed from all these sources. If the network is characterized by its z parameters, for example, we might express the total by a pair of equivalent series noise voltages in the equation

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} + \begin{bmatrix} e_{n1} \\ e_{n2} \end{bmatrix} \quad (11.3-1)$$

To find these noise voltages conceptually, at least, we need only set i_1 and i_2 to zero by open-circuiting the ports (for signals in the frequency range of interest) and measuring the noise voltages at input and output. An equivalent circuit corresponding to (11.3-1) is given in Fig. 11.8a. Three other sets of equations and equivalent circuits similar to these can be found using the h , y , and g parameters for the network. The y parameters, for example, would require

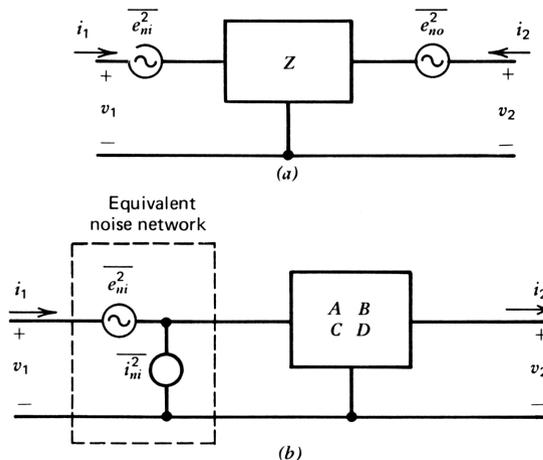


Figure 11.8. Equivalent noise models of two-ports: (a) impedance parameter formulation; (b) $ABCD$ parameter formulation.

equivalent shunt noise current generators at input and output whose noise currents would flow in short circuits connected to the input and output terminals.

When the network is described by its $ABCD$ parameters, the noise is represented by equivalent input noise voltage and current generators connected as shown in Fig. 11.8*b*. As in Fig. 11.8*a*, the network itself is noiseless, with all noise sources represented by the equivalent input generators, given by the equations

$$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} v_2 \\ i_2 \end{bmatrix} + \begin{bmatrix} e_n \\ i_n \end{bmatrix} \tag{11.3-2}$$

This equation is the standard way of specifying noise of a two-port network because of several advantages that become apparent later.

When the noise of a network is specified in this way, we can separate the noise from the rest of the network in an *equivalent noise network*, as shown in Fig. 11.8*b*. If two noisy networks are connected in tandem, as shown in Fig. 11.9*a*, the equivalent noise network of the combination is given in Fig. 11.9*b* as

$$\begin{aligned} \overline{e_{nT}^2} &= \overline{e_{n1}^2} + |A_1|^2 \overline{e_{n2}^2} + |B_1|^2 \overline{i_{n2}^2} \\ \overline{i_{nT}^2} &= \overline{i_{n1}^2} + |C_1|^2 \overline{e_{n2}^2} + |D_1|^2 \overline{i_{n2}^2} \end{aligned} \tag{11.3-3}$$

in which we have added the noises on a power basis and have (temporarily) ignored correlations among them. This equation can be generalized by

$$\overline{u_n^2} = \overline{u_{n1}^2} + |T_1|^2 \overline{u_{n2}^2} + |T_1 T_2|^2 \overline{u_{n3}^2} + \tag{11.3-4}$$

where $\overline{u_n^2}$ is a column vector of the total mean square noise referred to the

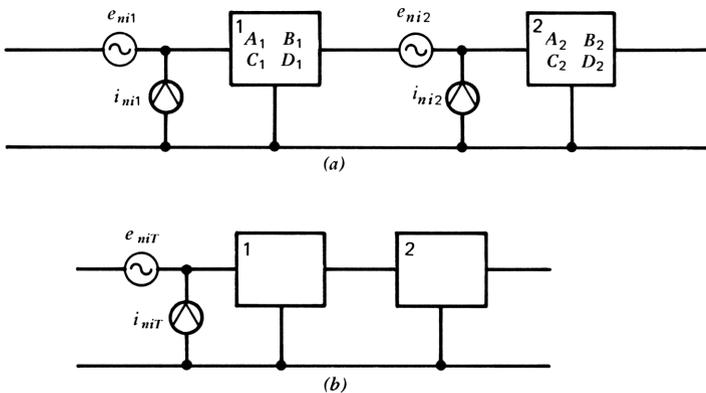


Figure 11.9. Cascaded noise networks.

input; the $\overline{u_{ni}^2}$ similarly represent the noise network currents and voltages of the individual networks of the combination. The T_i are the $ABCD$ matrices of the individual networks. Clearly, if the gain of the first network is high, that is, if the $ABCD$ parameters are sufficiently small, then $e_{nT} \simeq e_{n1}$ and $i_{nT} \simeq i_{n1}$; we can ignore the noise contribution of the second network in this case.

Correlation Between e_n and i_n

An individual physical noise source in a network will in general contribute to both e_n and i_n of the equivalent noise network. In summing the total noise at the input, these components representing the source should be added on a voltage or current basis rather than on a power basis. This complicates the bookkeeping involved since we must keep track of n correlations where there are n independent physical noise sources in the network.

A convenient way to include the effects of correlations is to calculate the total effect at the input of each physical source independently and to add the contributors (all uncorrelated) on a power basis. If the source impedance is Z_G , for example, we would obtain from noise source A the total input noise $e_{nA} + i_{nA}Z_G$ (added on a voltage basis because they are completely correlated), similarly for noise source B , and so on. We then square each contribution and add them for all sources to obtain the total input noise power.

The equivalent input noise resistance from a single mean square *current* noise source contained in a network is found as follows:

$$R_{inj} = \left| \begin{bmatrix} 1 & Z_G \\ & \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right|^2 \begin{bmatrix} 0 \\ G_{nj} \end{bmatrix} \quad (11.3-5)$$

$$= |B_j + Z_G D_j|^2 G_{nj} \quad (11.3-6)$$

Similarly, the equivalent input noise resistance from a single mean square *voltage* noise source is

$$R_{ink} = |A_k + Z_G C_k|^2 R_{nk} \quad (11.3-7)$$

By adding the input noise voltage components *within* the magnitude brackets, correlation between input noise current and voltage is taken into account. The complete input noise of a network is found by summing over all current and voltage noise sources in the network:

$$R_{inT} = \sum_k |A_k + Z_G C_k|^2 R_{nk} + \sum_j |B_j + Z_G D_j|^2 G_{nj} \quad (11.3-8)$$

where the $ABCD$ parameters in the summations are those between the individual noise source R_{nk} or G_{nj} and the input of the network, and R_{inT} is the total noise equivalent input resistance of the network, including all sources of noise in the network.

This formulation is general and allows us to find the noise characteristics for any network and source impedance. It may, of course, also be formulated on an input noise conductance basis by using the relation

$$G_{inT} = R_{inT} (Y_G)^2 \quad (11.3-9)$$

where $Y_G = 1/Z_G$.

As an example, let us find the input noise equivalent resistance of a bipolar transistor from the three primary noise sources in Fig. 11.5a expressed by eqs. (11.2-4) to (11.2-6). The three contributions are analyzed separately.

Collector Shot Noise

The collector shot noise generator is connected to the output of the transistor so that the *ABCD* parameters of the transistor itself are the relevant ones: for the core transistor plus r_b , we have

$$-B = r_e + r_b(\delta + j\tau_T\omega) \quad (11.3-10)$$

$$-D = \delta + j\tau_T\omega \quad (11.3-11)$$

To combine the two terms, we multiply D by Z_G , add it to B and take the squared magnitude. For example, Z_G might be $R_G + jL_G\omega$, where L_G is a source inductance. For simplicity, we assume a resistive source; adding L_G gives a quadratic term in frequency and can be included where necessary. Thus from (11.3-8) we obtain the noise equivalent input resistance due to collector shot noise:

$$R_{inCS} = |r_e + r_b(\delta + \tau_T s) + R_G(\delta + \tau_T s)|^2 G_{nCS} \quad (11.3-12)$$

To find the squared magnitude of the first term, we square the dc term and the s term and add them:

$$\begin{aligned} R_{inCS} &= \left\{ [r_e + \delta(r_b + R_G)]^2 + (r_b + R_G)^2 \tau_T^2 \omega^2 \right\} \frac{1}{2r_e} \\ &= \frac{r_e}{2} + \delta(r_b + R_G) + \frac{\delta^2(r_b + R_G)^2}{2r_e} + \frac{(r_b + R_G)\tau_T^2\omega^2}{2r_e} \end{aligned} \quad (11.3-13)$$

The second and third collector shot noise terms are found to be negligible if $\delta \ll 1$ from considerations discussed in the text that follows, so that

$$R_{inCS} = \frac{r_e}{2} + (r_b + R_G)^2 \frac{\tau_T^2 \omega^2}{2r_e} \quad (11.3-14)$$

Note that r_e appears in the numerator of the first term (the dc term) and in the

denominator of the second (high-frequency) term. Thus high quiescent collector current reduces low-frequency noise at the expense of high-frequency noise, considering only the collector shot noise contribution.

Base Current Shot Noise

The base current shot noise generator appears at the input; only the base resistance separates it from the input terminals. Hence the relevant $ABCD$ matrix is that for the series resistance r_b :

$$T_{r_b} = \begin{bmatrix} 1 & r_b \\ 0 & 1 \end{bmatrix}$$

so that

$$B = r_b \quad (11.3-15)$$

$$D = 1 \quad (11.3-16)$$

From (11.2-5) and (11.2-8), we obtain the noise equivalent input resistance from base current shot noise:

$$R_{inBS} = (r_b + R_G)^2 \frac{\delta}{2r_e} \quad (11.3-17)$$

The base shot noise term is $1/\delta$ larger than the third collector shot noise term of (11.3-13), thus justifying dropping the latter term.

Flicker noise, where not negligible, should be added at this point and is also multiplied by $(r_b + R_G)^2$ to find its contribution to the noise equivalent input resistance.

Base Resistance Thermal Noise

The noise voltage associated with the base resistance is already at the input. In terms of eq. (11.3-7), $A = 1$ and $C = 0$ since the $ABCD$ matrix is the unit matrix. Hence

$$R_{inBT} = r_b \quad (11.3-18)$$

The source resistance will also have noise associated with it with equivalent resistance R_G . The thermal noise contributions of r_b and R_G justify dropping the second collector shot noise term of (11.3-13).

The total noise equivalent input resistance for the transistor is given by the sum of the three contributions:

$$R_{inT} = \underbrace{\frac{r_e}{2} + r_b}_{\text{CS BT}} + (r_b + R_G)^2 \underbrace{\frac{\delta + \tau_I^2 \omega^2}{2r_e}}_{\text{BS CS}} \quad (11.3-19)$$

This equation shows that the input noise *voltage* is dominated at low frequencies by r_b thermal noise and collector current shot noise; which is greater depends on the collector current. The contributions will be equal for $I_C = kT/2qr_b$. The input noise *current* is dominated by base current shot noise at low frequencies and by collector shot noise at high frequencies. The crossover between the two is at

$$\omega_c = \frac{\sqrt{\delta}}{\tau_T} = 2\pi f_T \sqrt{\delta} \quad (11.3-20)$$

The mean square input noise voltage is thus

$$\overline{e_{inT}^2} = 4kTR_{inT}(f) df \quad (11.3-21)$$

where we have replaced Δf by df ; since the noise voltage is proportional to frequency, it is clear that to find the total noise, we must integrate over the desired frequency range. The frequency characteristics of the network following the noise network must be taken into account when the integration is performed, and this is the subject of Section 11.4.

Input Noise Network for FETs

Exactly the same procedure described for the bipolar transistor can be applied to the FET. The approximate *ABCD* parameters of the FET are, by inspection of the equivalent circuit in Fig. 11.7,

$$A = -r_{ch} g_{d0} \quad (11.3-22)$$

$$B = -r_{ch} \quad (11.3-23)$$

$$C = -C_{Gd}s \quad (11.3-24)$$

$$D = -r_{ch}(C_{gs} + C_{gd})s \quad (11.3-25)$$

We can define a total input noise equivalent resistance R_{inT} as before; the contribution from drain shot noise is given by eq. (11.3-6) as

$$\begin{aligned} R_{inDS} &= (B + DR_G)^2 G_{nDS} \\ &= r_{ch}^2 + [r_{ch}R_G(C_{gs} + C_{gd})\omega]^2 G_{nDS} \end{aligned} \quad (11.3-26)$$

From eq. (11.2-12), it follows that $G_{nDS} = 2/(3r_{ch})$, and we obtain

$$R_{inDS} = \frac{2}{3}r_{ch} \left[1 + R_G^2 (C_{gs} + C_{gd})^2 \omega^2 \right] \quad (11.3-27)$$

To this we add the noise voltage generated by the gate leakage flowing through

the source resistance R_G , giving the total input noise equivalent resistance R_{inT} :

$$R_{inT} = \frac{2}{3} r_{ch} \left[1 + R_G^2 (G_{gd} + C_{gs})^2 \omega^2 \right] + R_G^2 G_{nGS} \quad (11.3-28)$$

The FET is often used with high-impedance sources. It is usually more convenient to use a noise conductance formulation for this case; such sources are often capacitive, and the source capacitance can be incorporated conveniently into the total noise conductance. Letting the source admittance be $Y_G = G_G + C_{GS}$ (the subscript G refers to the generator rather than the gate), we can use (11.3-9) to write the total input equivalent noise conductance

$$G_{inT} = (BY_G + D)^2 G_{nDS} + G_{nGS} \quad (11.3-29)$$

$$\begin{aligned} &= \left[r_{ch}^2 G_G^2 + r_{ch}^2 (C_G + C_{gs} + C_{gd})^2 \omega^2 \right] G_{nDS} + G_{nGS} \\ &= \frac{2}{3} r_{ch} \left[G_G^2 + (C_G + C_{gs} + C_{gd})^2 \omega^2 \right] + G_{nGS} \end{aligned} \quad (11.3-30)$$

The capacitance thus also contributes to noise and is incorporated in parallel with the gate-source capacitance and the input loading of the gate-drain capacitance.

A conductance formulation of bipolar transistor input noise is given in Section 11.4.

Noise Figure and Noise Temperature

Noise figure is a measure of noise performance of a network defined as

$$F = \frac{\text{Total noise}}{\text{Noise engendered by input termination}} \quad (11.3-31)$$

It can be found by considering only the source and the equivalent noise network in Fig. 11.8*b* since the remainder of the network is noiseless. The noise figure is unaffected by the input immittance of the network, for example. Where the source is resistive and e_n and i_n are uncorrelated, for example, we can write

$$F = 1 + \frac{\overline{e_n^2} + \overline{i_n^2} R_G^2}{4kT\Delta f R_G} \quad (11.3-32)$$

Where e_n and i_n are correlated, the correlation is incorporated as shown in Section 11.3. In terms of R_{inT} , we can write from (11.3-21)

$$F = \frac{R_G + R_{inT}}{R_G} = 1 + \frac{R_{inT}}{R_G} \quad (11.3-33)$$

Since R_{inT} is in general a function of frequency, we can define F as a *spot noise figure* at a single frequency. To find the noise figure for a network over a band of frequencies, the noise must be integrated over the band.

In terms of the noise equivalent resistance and conductance, the noise figure can be written

$$F = 1 + \frac{R_{ni}}{R_G} + G_{ni} R_G \quad (11.3-34)$$

where R_{ni} is proportional to $\overline{e_n^2}$ and G_{ni} is proportional to $\overline{i_n^2}$. Clearly, this expression reaches a minimum when the second and third terms are equal; we can adjust the source resistance to make them equal, in which case we find the optimum source resistance for noise as

$$R_{Gopt} = \sqrt{\frac{R_{ni}}{G_{ni}}} \quad (11.3-35)$$

When R_{ni} and G_{ni} are correlated the situation becomes more complicated. Furthermore, if they are functions of frequency, the optimum source impedance may not be resistive. This case is covered by Rothe and Dahlke.¹

Note that the noise figure is not a function of the input impedance (or any other circuit property) of the (noiseless) network following the equivalent input noise network.

The *noise temperature* is related to the noise figure through the equation

$$T_n = T_0(F - 1) \quad (11.3-36)$$

where T_0 is the standard temperature (usually taken as 293°K. It is the equivalent temperature of the source whose available noise power is given by $kT\Delta f$. Using (11.3-34), (11.3-36) can be written

$$T_n = T_0 \left(\frac{R_{ni}}{R_G} + G_{ni} R_G \right) \quad (11.3-37)$$

11.4 INTEGRATION TO FIND THE TOTAL NOISE

As we have seen, the noise generated by devices is frequency sensitive, either because of the noise source itself, as in $1/f$ noise, or because of the frequency characteristics of the device. The noise extends over a wide frequency range, well beyond the frequency range of interest in most cases. To obtain a meaningful measure of the noise, we must apply a weighting function to the noise to filter out the irrelevant components—the components outside the frequency range of interest.

To do this correctly, we need to know the characteristics of the detection process. Where the detector is the ear/brain, for example, noise above 20 kHz

is irrelevant to noise evaluation; even in the auditory range, large differences in sensitivity at different frequencies exist, so that noise at 4 kHz, for example, is more important than noise at 100 Hz. If the ear receives the message through a telephone receiver the characteristics of the receiver must also be taken into account. Standard weighting curves have been developed to account for the varying sensitivity of the ear and telephone receiver with frequency.

One further example: if the detector is a decision circuit of a regenerator in a digital transmission system, the noise on the signal applied to the decision circuit can cause errors, converting ones to zeros and vice versa. Often, a noise filter is placed ahead of the decision circuit to minimize the noise while having as little degrading effect on the signal pulse as possible.

In either of these examples we must combine the frequency sensitivity of the input noise network generators with the weighting function to find the total noise. We must integrate the product of the frequency-sensitive noise and the weighting function over the complete frequency range. Representation of the noise at the network input makes this a fairly straightforward procedure. This is because the input noise generators are usually polynomials in the frequency variable, allowing us to combine the noise with the weighting function on a coefficient-by-coefficient basis. We can then sum up the total noise power from each term of the noise polynomial. In this section we illustrate the process by finding the noise produced by transistors in typical situations; the technique can be extended to any network, as is clear later.

We let $H(f)$ represent the weighting function for the noise. Since we are interested only in the magnitude of the noise, we multiply the mean square noise contribution by $|H(f)|^2$ so that the total equivalent noise *at the input* becomes, from eq. (11.3-21)

$$\overline{e_{inT}^2} = 4kT \int_0^\infty |H(f)|^2 R_{inT}(f) df \quad (11.4-1)$$

It is helpful in calculations to normalize the frequency variable. Letting $\nu = f/f_0$, $d\nu = df/f_0$, and we can write

$$\overline{e_{inT}^2} = 4kTf_0 \int_0^\infty |H(\nu)|^2 R_{inT}(\nu) d\nu \quad (11.4-2)$$

where f_0 is a bandwidth that may be defined in several ways depending on the application. For an ideal low-pass filter, it is the cutoff frequency. In a digital transmission system it can be taken as the Nyquist frequency (half the signaling rate). The basis of the choice is to give intuitive meaning to the calculations; as long as we are consistent, the calculations will be correct for any such choice.

Often, $R_{inT}(\nu)$ has a relatively simple variation with frequency, as in the bipolar transistor, that can be expressed as

$$R_{inT} = R_0 + R_1\nu^2 \quad (11.4-3)$$

For eq. (11.3-19), for example, we obtain

$$R_0 = \frac{r_e}{2} + r_b + \frac{(r_b + R_G)^2 \delta}{2r_e} \quad (11.4-4)$$

and

$$R_1 = \frac{(r_b + R_G)^2 \tau_T^2 (2\pi f_0)^2}{2r_e} \quad (11.4-5)$$

Thus (11.4-2) becomes

$$e_{inT}^2 = 4kTf_0(R_0\Sigma_0 + R_1\Sigma_1) \quad (11.4-6)$$

where

$$\Sigma_0 = \int_0^\infty |H(\nu)|^2 d\nu \quad (11.4-7)$$

and

$$\Sigma_1 = \int_0^\infty |H(\nu)|^2 \cdot \nu^2 d\nu \quad (11.4-8)$$

For any required channel shape $H(\nu)$, these integrals are readily evaluated numerically on the calculator or computer and are incorporated in (11.4-6) to find the total noise. Three examples are given here.

If $H(\nu)$ is the transmission of an ideal low-pass filter, as in Fig. 11.10a.

$$H(\nu) = \begin{cases} 1, & 0 < \nu < 1 \\ 0, & \text{otherwise} \end{cases} \quad (11.4-9)$$

we have

$$\Sigma_0 = \int_0^1 d\nu = 1$$

and

$$\Sigma_1 = \int_0^1 \nu^2 d\nu = \frac{1}{3}$$

If $H(\nu)$ is a cubic Butterworth low-pass function, as in Fig. 11.10b, then

$$H(\nu)^2 = \frac{1}{1 + \nu^2} \quad (11.4-10)$$

We can most easily evaluate Σ_0 and Σ_1 by numerical integration; when this is

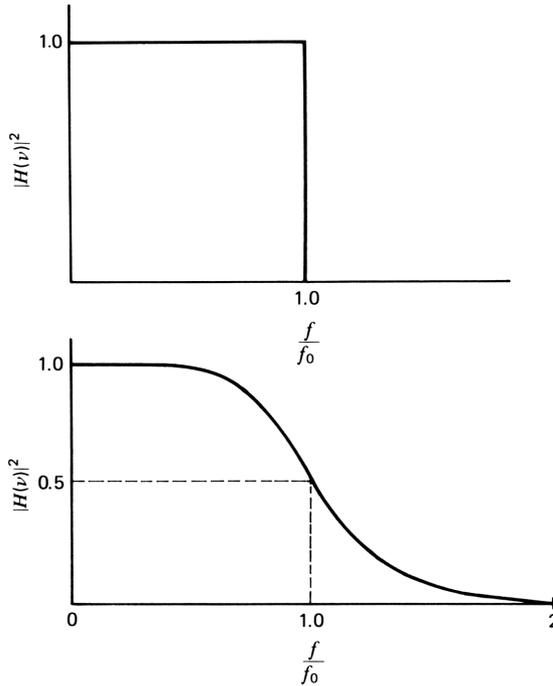


Figure 11.10. Power response of weighting networks for noise: (a) ideal filter; (b) cubic Butterworth response.

done, we obtain

$$\Sigma_0 = 1.04$$

$$\Sigma_1 = 0.482$$

Comparing these integrals with those for the ideal filter, the low-frequency noise is almost the same, but high-frequency noise is noticeably larger.

One final example of noise integration shows the effect of including high-frequency equalization in $H(\nu)^{13}$. In a digital fiber optic transmission system, shown schematically in Fig. 11.11, a laser generates light signal pulses. When a one is transmitted, the laser is on for the full pulse period, and for a zero it is off for the same period. The fiber is assumed here not to disperse the pulses significantly, but attenuation of the signal is large; thus noise of the receiver preamplifier may cause errors at the detector. Equalization is provided with the requirement that it cause (ideally) no *intersymbol interference* (the pulse amplitude at the center of any given time slot is independent of any other pulses in the pulse stream) and that it minimize the noise by minimizing the bandwidth; more accurately, the equalizer is to minimize an integral of the type in eq. (11.4-2).

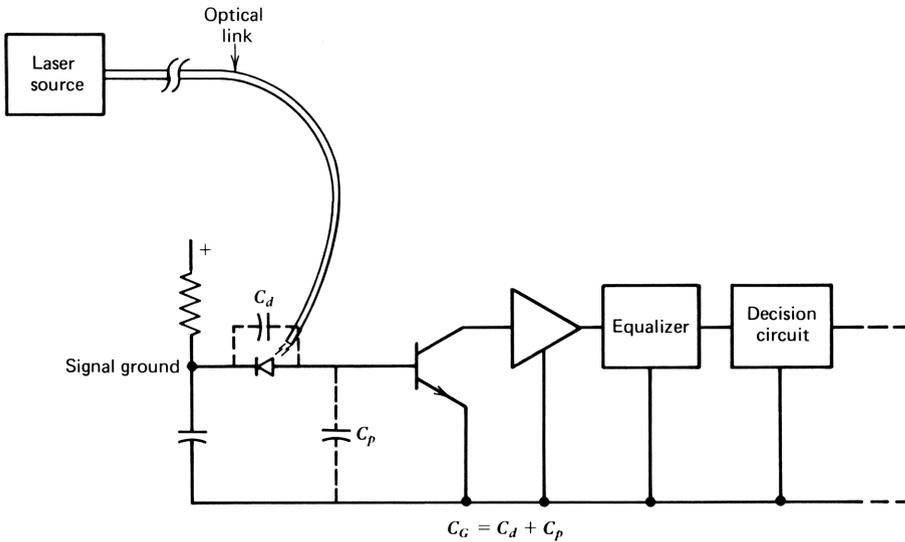


Figure 11.11. Block diagram of digital fiber optic transmission system receiver.

Several equalizer functions give zero intersymbol interference; an often-used function is the cosine rolloff characteristic:

$$\begin{aligned}
 H(\nu) &= 1, & 0 < |\nu| < 1 - \beta \\
 &= \frac{1}{2} \left[1 - \sin \frac{\pi}{2\beta} (\nu - 1) \right], & 1 - \beta < |\nu| < 1 + \beta \\
 &= 0 & \text{otherwise}
 \end{aligned}
 \tag{11.4-11}$$

The frequency is normalized to the Nyquist frequency (half the signaling rate). The constant β characterizes the channel; for $\beta=0$, the cutoff is that of an ideal filter with cutoff at the Nyquist frequency; for $\beta=1$, we obtain a *full cosine rolloff*. This type of function tends to minimize noise as well as giving zero intersymbol interference. In the absence of equalization, the effect of noise is minimized with $\beta=0$ since this minimizes the high-frequency noise of the devices.

To obtain this frequency characteristic at the detector, we must take into account the shape of the transmitted pulse, which has the spectrum

$$P(\nu) = \frac{\sin(\pi\nu/2)}{\pi\nu/2}
 \tag{11.4-12}$$

The channel following the preamplifier must incorporate an equalizer that has a compensating characteristic, a $\sin x/x$ or *sinc* correction. Therefore, the

overall transmission characteristic of the channel following the input noise network is given by

$$\begin{aligned}
 H_T(\nu)^2 &= \left[\frac{\pi\nu/2}{\sin(\pi\nu/2)} \right]^2, & 0 < |\nu| < 1 - \beta \\
 &= \frac{\{(\pi\nu/2)[1 - \sin\pi(\nu-1)/2\beta]\}^2}{4[\sin(\pi\nu/2)]^2}, & 1 - \beta < |\nu| < 1 + \beta \\
 &= 0 \text{ otherwise} & & (11.4-13)
 \end{aligned}$$

This is the required channel shape and is shown plotted for various values of β in Fig. 11.12. Notice that for $\beta=0$, considerable equalization near the Nyquist frequency is required; for larger values of β , less equalization is required below the Nyquist frequency, but there is more response above it. Hence there will be a value of β intermediate 0 and 1 that gives minimum noise, depending on the relative values of T_0 and R_1 . Although the expressions for the channel are complicated, they are easily integrated numerically on the calculator, with values of Σ_0 and Σ_1 given in Table 11.2. For $\beta=0.7$, for example, $\Sigma_0=1.13$, so that the effect of the equalization is to increase the flat (white) noise contribution by about 0.5 dB. The value of $\Sigma_1=0.49$ is comparable to that for the cubic Butterworth cutoff. To find the total noise, of course, the characteristics of the noise network must be taken into account.

Where high-impedance sources must be considered, an admittance formulation of device noise is considerably more convenient than the impedance formulation given previously. Such sources are often limited in noise performance by their capacitance, so it is helpful to incorporate the source capacitance in such a formulation. Applying eq. (11.3-9) to eqs. (11.4-4) and (11.4-5) and letting $Y_G = G_G + j\omega C_G$, we obtain the following equations for the noise of the bipolar transistor in the admittance formulation:

$$G_{inT} = G_0 + G_1\nu^2 \quad (11.4-14)$$

The individual components are evaluated as before, taking correlations into account by adding currents arising from the three noise sources just as we added voltages in eqs. (11.3-5) to (11.3-8):

$$G_{inj} = (Y_G B + D)^2 G_{nj} \quad (11.4-15)$$

$$G_{ink} = (Y_G A + C)^2 R_{nk} \quad (11.4-16)$$

Following the earlier procedure, we obtain

$$G_0 = G_G^2 \left(\frac{r_e}{2} + r_b \right) + (1 + r_b G_G)^2 \frac{\delta}{2r_e} \quad (11.4-17)$$

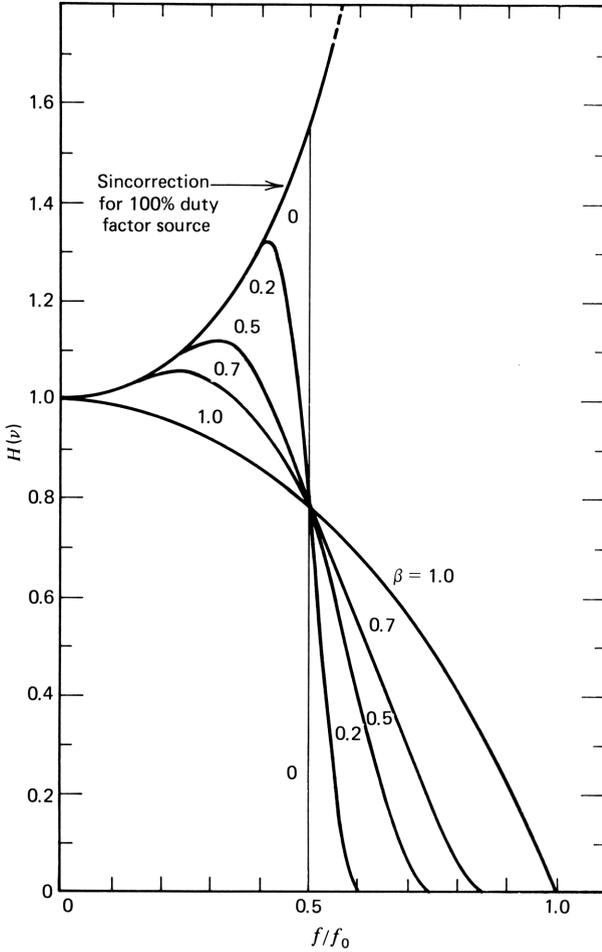


Figure 11.12. Required equalization for digital fiber optic transmission system with raised cosine response for various values of the rolloff factor β ; 100% duty factor transmitted pulse is assumed.

and

$$G_1 = (2\pi f_0)^2 \left[\frac{[\tau_1(1 + r_b G_G)]^2}{2r_e} + C_G^2 \left(r_b + \frac{r_b^2 \delta}{2r_e} \right) \right] \quad (11.4-18)$$

where

$$\begin{aligned} \tau_1 &= \tau_T + C_G r_e \\ &= \tau_F + (C_{j_e} + C_{j_c} + C_G) r_e \end{aligned} \quad (11.4-19)$$

This is a result similar to that for the FET derived in Section 11.3. Because r_b is connected between C_{j_e} and C_G , a quadratic term appears that has been ignored.

Table 11.2 Values of Weighting Function Integrals for a Cosine Rolloff Channel with $\sin x/x$ Correction^a

β	Σ_0	Σ_1
0	1.37	0.567
0.2	1.28	0.490
0.5	1.17	0.452
0.7	1.13	0.486
0.85	1.12	0.555
1.0	1.13	0.695

^aThis table and the associated analysis are largely adapted from Personick.¹³ The frequency normalization in the reference is to twice the Nyquist frequency, whereas we use the Nyquist frequency. Thus values of Σ_0 are twice as large here, and values of Σ_1 are eight times as large.

It is ordinarily of no consequence to the noise; the added root does not affect the magnitude of the response, but only the phase, as discussed in Chapter 5.

The origins of noise sources in eqs. (11.4-17) and (11.4-18) are readily identified. Squared terms all belong to the transmission functions, so the original noise terms appear as the first power. Terms including δ come from base shot noise current. Terms involving r_e (without δ) arise from the collector noise. Terms with r_b (to the first power) are from thermal noise in r_b .

The total mean square input noise current is given by

$$i_{inT}^2 = 4kTf_0(G_0\Sigma_0 + G_1\Sigma_1) \quad (11.4-20)$$

To illustrate the use of these concepts and equations, we now find the optimum collector current at which to operate the first-stage transistor of a preamplifier for a fiber optic digital transmission system with a 274 Mband/s signalling rate. At the receiver the light is detected by a pin photodiode whose output current i_G is related to the incident light power P by the relation¹⁴

$$i_G = R_O P \quad (11.4-21)$$

where R_O is the *responsivity* of the diode in milliamperes per milliwatt, and P is given in milliwatts; R_O is inversely proportional to the photon energy (and thus proportional to λ , the wavelength of the light). It is also proportional to η , the quantum efficiency of the diode, including the fiber-diode interface. It can

be expressed as

$$R_o = 0.81\eta\lambda \tag{11.4-22}$$

Note that the current (not the square of the current) is proportional to the light power.

It is usual to express the noise characteristics of a lightwave receiver by its *sensitivity*, defined as the minimum signal power needed to achieve a certain bit error rate, often taken as one error in 10^9 pulses, or 10^{-9} . The ratio of peak signal to rms noise required to achieve a 10^{-9} bit error rate is 6.0,¹³ so that the sensitivity, expressed in dBm (dB referred to 1 mW) is given by

$$S = 10 \log \frac{6i_{inT}}{R_o}$$

where i_{inT} is interpreted as the rms value of the noise current. The equation can be written

$$S = 5 \log \frac{36(\overline{i_{inT}})^2}{R_o^2} \tag{11.4-23}$$

Through these equations, we can investigate the effect of varying collector current or any of the other parameters controlling the noise performance.

In Figs. 11.13 and 11.14 receiver sensitivity is plotted as a function of collector current for various values of β and current defect ratio. In plotting

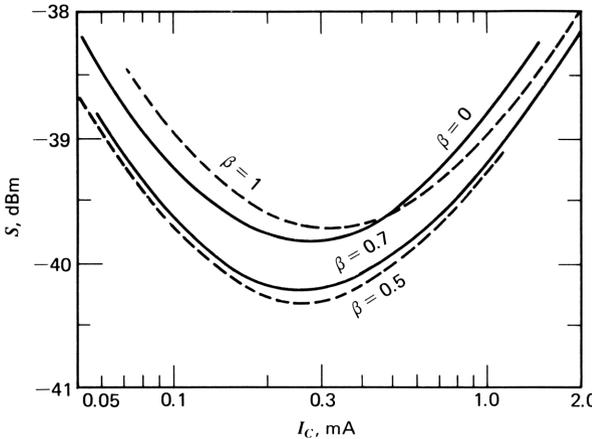


Figure 11.13. Fiber optic receiver for 274 Mbaud/s transmission system; receiver sensitivity as a function of the cosine rolloff factor β . $\delta = 0.005$. $R_o = 1.0$, $C_{01} = 0.8$ pf, $G_G = 0.01$ mS, $G_{I,1} = 0.2$ mS.

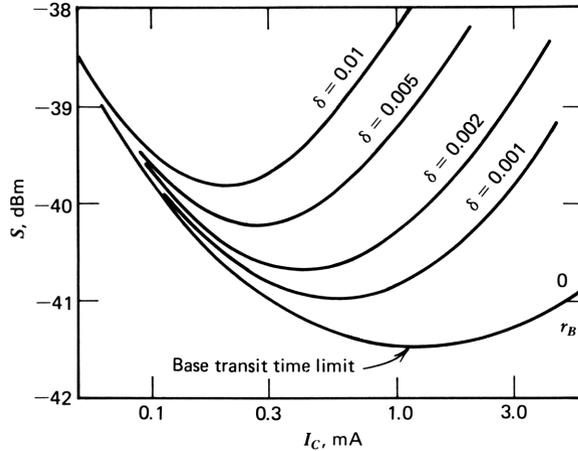


Figure 11.14. Fiber optic receiver sensitivity as a function of first-stage collector current for various values of δ . $C_{01}=0.8$, $G_G=0.01$, $G_{L1}=0.2$, $R_G=1.0$, $\beta=0.7$.

these curves we assumed a transistor having $r_b=0.11$ k Ω and in Fig. 11.13, $\delta=0.005$. The values of r_e and τ_T vary with collector current as described in Chapter 7; the forward transit time was taken as 0.027 ns; also, $C_{je}=0.4$ pF, and $C_{jc}=0.12$ pF. Circuit values were $G_G=0.01$ mS and $C_G=0.8$ pF, including the PIN diode capacitance and input parasitic capacitance. The first-stage load conductance was 0.2 mS. Responsivity was taken as 1.0 mA/mW.

The first plot gives a basis for choosing β ; sensitivity is best for β in the range 0.5–0.7. For small values of β , the peaking near the Nyquist frequency increases the noise. For large values of β , the added bandwidth increases the noise. Larger values give a gentler rolloff that is easier to realize in an equalizer circuit, so we choose a value of β of 0.7. The larger value also makes the timing of the detection process less critical.

The effect of reducing base current shot noise is shown in Fig. 11.14, in which sensitivity is plotted as a function of collector current for various values of δ from 0.01 to zero. The latter value removes the base current and its shot noise, allowing the stage to operate at higher collector currents for minimum sensitivity.

To see the effects of the various individual sources of noise, these sources can be compared at the input by the amount they contribute to the total input noise equivalent conductance. These components are plotted as a function of collector current in Fig. 11.15 and include contributions from both G_0 and G_1 in eqs. (11.4-7) and (11.4-8). (For any individual noise contributor, the components of G_0 and G_1 add on a power basis because they appear in quadrature.) In addition to the three primary sources of transistor noise, the noise contributions of G_G and G_{L1} (the load resistance of the first stage) are also shown. The value of these conductances depends on the choice of f_0 ; they may be interpreted as the conductances that would give an equal noise contribution to the actual sources if the channel were an ideal filter with cutoff at the Nyquist frequency.

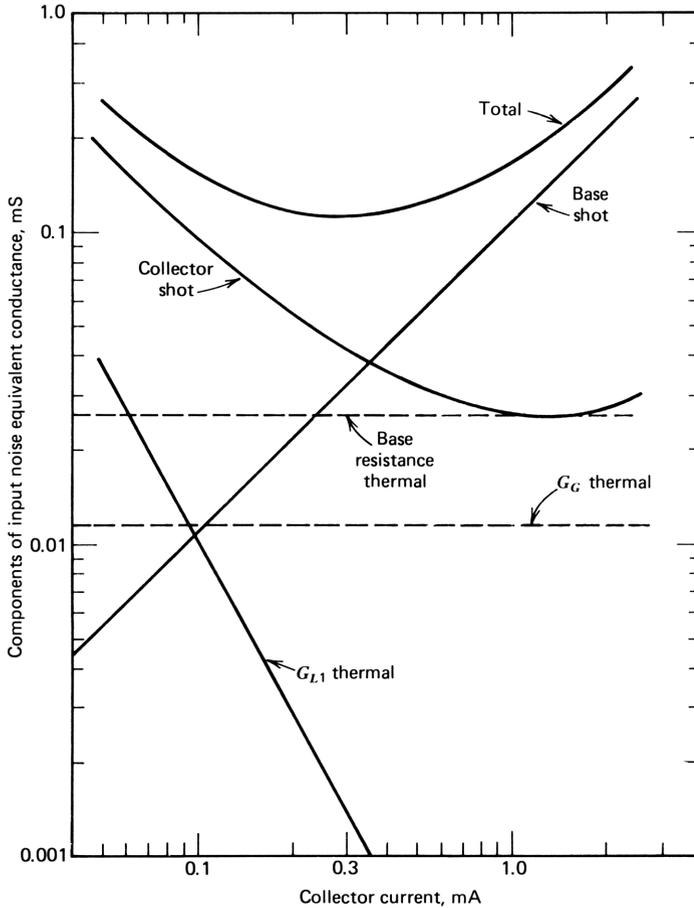


Figure 11.15. Noise equivalent input conductances representing five sources of noise in the first stage of a fiber optic receiver. The total noise is also shown.

As collector current is increased in Fig. 11.15, the base shot noise rises linearly, as we would expect. The contributions from the base and G_G thermal noise is independent of current. Collector shot noise initially falls with increase of current since τ_T is reduced with increased current. As τ_T flattens with current, the noise begins to rise, giving a minimum. Finally, the effect of G_{L1} falls as the square of the current because of the drop of r_e .

The total noise conductance exhibits a minimum for $I_C=0.3$ mA. At this value the noise current is given

$$\begin{aligned}
 i_{inT} &= \sqrt{4kTf_0G_{inT}} \\
 &= \sqrt{(1.66 \times 10^{-8})(0.137)(0.11)} \\
 &= 15.8 \times 10^{-6} \text{ mA}
 \end{aligned}$$

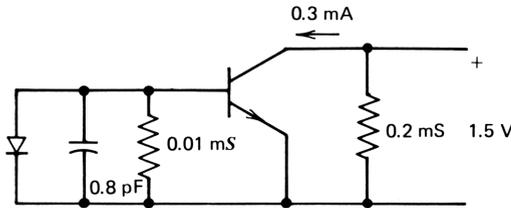
The light power needed for 10^{-9} bit error rate is thus given as

$$P = \frac{6(15.8 \times 10^{-6})}{R_o}$$

$$= 94.9 \times 10^{-6} \text{ mW}$$

and the sensitivity is -40.2 dBm. If I_C were increased to 1 mA, for which $G_f = 0.17$ mS, the sensitivity would be degraded by $10 \log \sqrt{0.17/0.11} = 0.99$ dB.

The PIN diode is back-biased, and its conductance is negligible, as is its noise contribution. The inclusion of G_G is to allow base bias current to be applied to the transistor. The value chosen is small, as is its noise contribution, but care must be taken in providing temperature-stable bias.



ABCD MATRIX OF TRANSISTOR

FREQ., GHZ = 0.1000

ABCD, MAG. +PH:

A:

R01= 0.0182
R02= -90.8076

B:

R03= 0.0917
R04= -176.2901

C:

R05= 0.0740
R06= -84.2968

D:

R07= 0.0429
R08= -103.7694

ABCD MATRIX OF 1ST STAGE

FREQ., GHZ = 0.1000

ABCD, MAG. +PH:

A:

R01= 0.0269
R02= -133.7016

B:

R03= 0.0917
R04= -176.2901

C:

R05= 0.0927
R06= -80.7563

D:

R07= 0.0082
R08= -95.3052

POLYS

A-POLY:

R00= -0.0195
R01= -0.0309
R02= -0.0024
R03= -2.4479-05

B

R04= -0.0916
R05= -0.0094
R06= -0.0002
R07= -2.7321-06

C

R08= 0.0007
R09= -0.1464
R10= -0.0432
R11= -0.0022

D

R12= -0.0115
R13= -0.1398
R14= -0.0092
R15= -0.0002

Figure 11.16. First-stage transmission matrix for the fiber optic receiver.

Analysis of the Optimized Input Stage

This input stage was designed for optimum noise performance without consideration of its transmission performance. It was intended to form the first stage of a feedback preamplifier, with the total preamplifier design being required to give satisfactory transmission performance. The design of the complete preamplifier is discussed in the text that follows after discussion of certain preliminaries. For that design, we need to know the transmission characteristics of the noise-optimized input stage.

The input stage is shown in Fig. 11.16*a* with the shunt admittances assumed in the preceding development. The collector current is 0.3 mA, and the collector voltage is assumed to be 1.5 V. The $ABCD$ matrix of the transistor at 100 MHz is given in Fig. 11.16*b* and the $ABCD$ matrix of the circuit is given in Fig. 11.16*c*. The polynomial coefficients of the circuit, obtained from evaluation of the $ABCD$ matrix at two frequencies (100 and 316 MHz), are shown in Fig. 11.6*d*. The A and B polynomials are of no consequence since the source admittance is already in place; the input current arising from A and B is already taken into account in C and D . When the stage is operated into a low load impedance of the rest of the amplifier (this will be the case), only D is of importance. The dc coefficient of D is unimportant to the response; the linear coefficient is 0.14 ns, and the quadratic coefficient, -0.0092 , implies a delay of approximately $0.0092/0.14=0.066$ ns. These two numbers characterize the essentials of the first-stage response, a unity loss time constant of 0.14 ns, and a delay of 0.066 ns.

11.5 EFFECTS OF FEEDBACK AND FEEDFORWARD ON NOISE

Any network with or without feedback can be analyzed as described in Section 11.4. Since the bipolar transistor is itself a feedback network, as we have shown in Chapter 6, we have already analyzed the noise in a feedback configuration. Nevertheless, it is useful to define the effects of feedback on noise since we may know the noise characteristics of a given network and wish to know its noise characteristics when feedback is applied without going through the complete analysis described in Section 11.4.

We have seen that when feedback is applied to a network, four effects are to be considered: input augmentation, or feedback; output augmentation, or feedforward; and input and output loading. When noise is expressed as an input noise network, it is obvious that input augmentation has no effect on noise since the network itself is noiseless; there is no noise to be fed back. Although this is obvious in one sense, it is also subtle in another, so let us consider examples.

Noise in the Common Collector and Common Base Stages

The technique of moving voltage generators through circuit nodes and splitting current generators discussed in connection with Fig. 8.16 can sometimes be

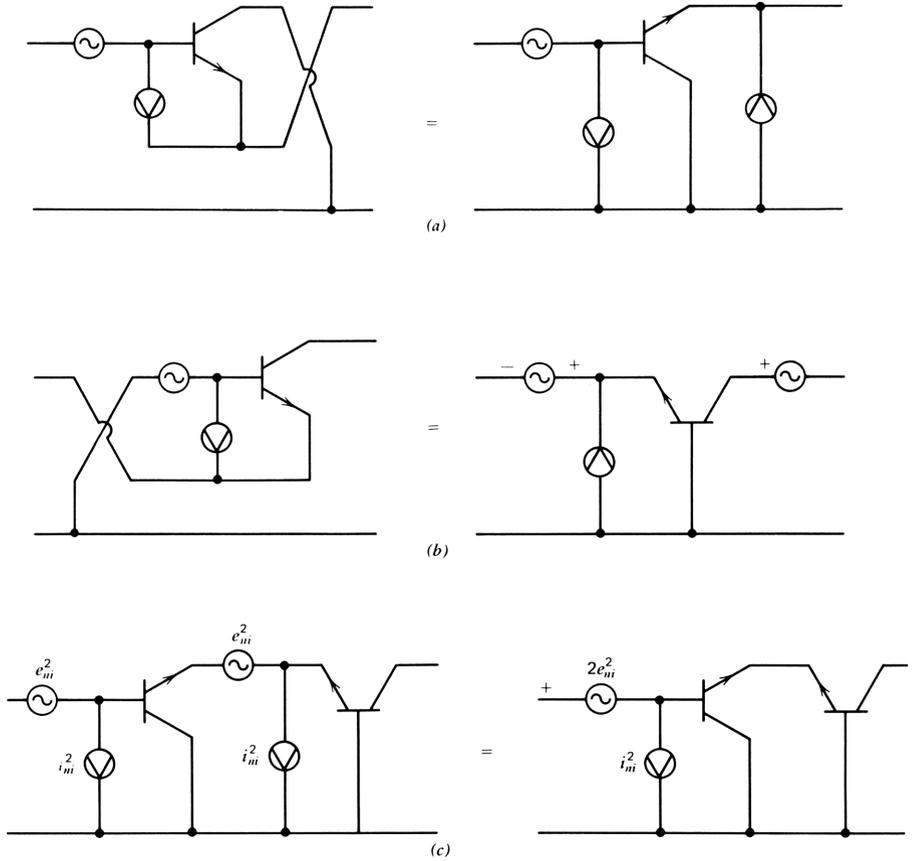


Figure 11.17. Input noise of (a) common collector stage, (b) common base stage, and (c) emitter-coupled pair.

used to evaluate the effects of feedback quickly. Three examples of this technique are shown in Fig. 11.17.

In Fig. 11.7a we obtain the equivalent input noise network for the common collector stage from that of the common emitter noise network. The voltage generator is already at the input, where it remains in the common collector stage. The noise current generator can be split into two generators, one across the input terminals and one across the output. The output generator represents fed-forward noise. Its effect at the input is small; it augments the input noise voltage slightly, after being multiplied by B of the common collector stage, roughly r_e . It also modifies the input noise current slightly, after being multiplied by D . Both effects can be neglected. The chief effect on noise performance of using a common collector stage as an input stage is to make the circuit susceptible to the noise input voltage of a following stage; since the voltage loss is unity, the second-stage noise *voltage* translates to the input undiminished by first-stage loss.

In the case of the common base stage in Fig. 11.17*b*, the input noise voltage generator can be passed through the base of the transistor, resulting in two noise generators, one in series with the input and one in series with the output. Since the common base stage has considerable voltage gain, the (feedforward) generator in series with the output can be neglected; thus the input noise network remains essentially unchanged compared to the common emitter stage. Since the current loss of the common base stage is unity, however, the input noise *current* of a following stage translates to the input of the common base stage without reduction.

Noise in the Differential Pair

The noise of a differential pair with one base grounded is the tandem connection of the common collector and common base stages as shown in Fig. 11.17*c*, so we can immediately find its noise performance. With A of the common collector equal to unity, the input noise voltage is the sum—on a power basis—of the input noise voltage of each transistor. The noise equivalent resistance of the differential pair is thus

$$R_{ni(DP)} = r_e + 2r_b \quad (11.5-1)$$

in which the small term from base shot noise $r_b^2\delta/r_e$ has been dropped for simplicity.

The noise current of the second stage (the common base stage) is reduced by D of the first stage, so the noise current of the combination is essentially that of the first stage alone. Therefore, the noise equivalent input conductance is just that of the transistor:

$$G_{ni(DP)} = \frac{\delta}{2r_e} + \frac{\tau_T^2\omega^2}{2r_e} \quad (11.5-2)$$

where the first term on the right is base current shot noise and the second term is collector shot noise multiplied by $|D|^2$. Where both bases of the pair have signals applied to them, the noise currents of both must be taken into account. They can be multiplied by the squared magnitude of the respective source impedances (for each of the two sources) to obtain the total equivalent input noise resistance. The differential pair is treated as a three-port in Chapter 12, where common mode characteristics are considered.

Lossless Transformer Feedback

The common collector stage is susceptible to the noise input voltage of a following stage, as noted. We can reduce this susceptibility by using lossless transformer feedback, shown in Fig. 11.18. The output voltage is stepped down by the factor n_A (with n_A less than unity) and added to the input voltage,

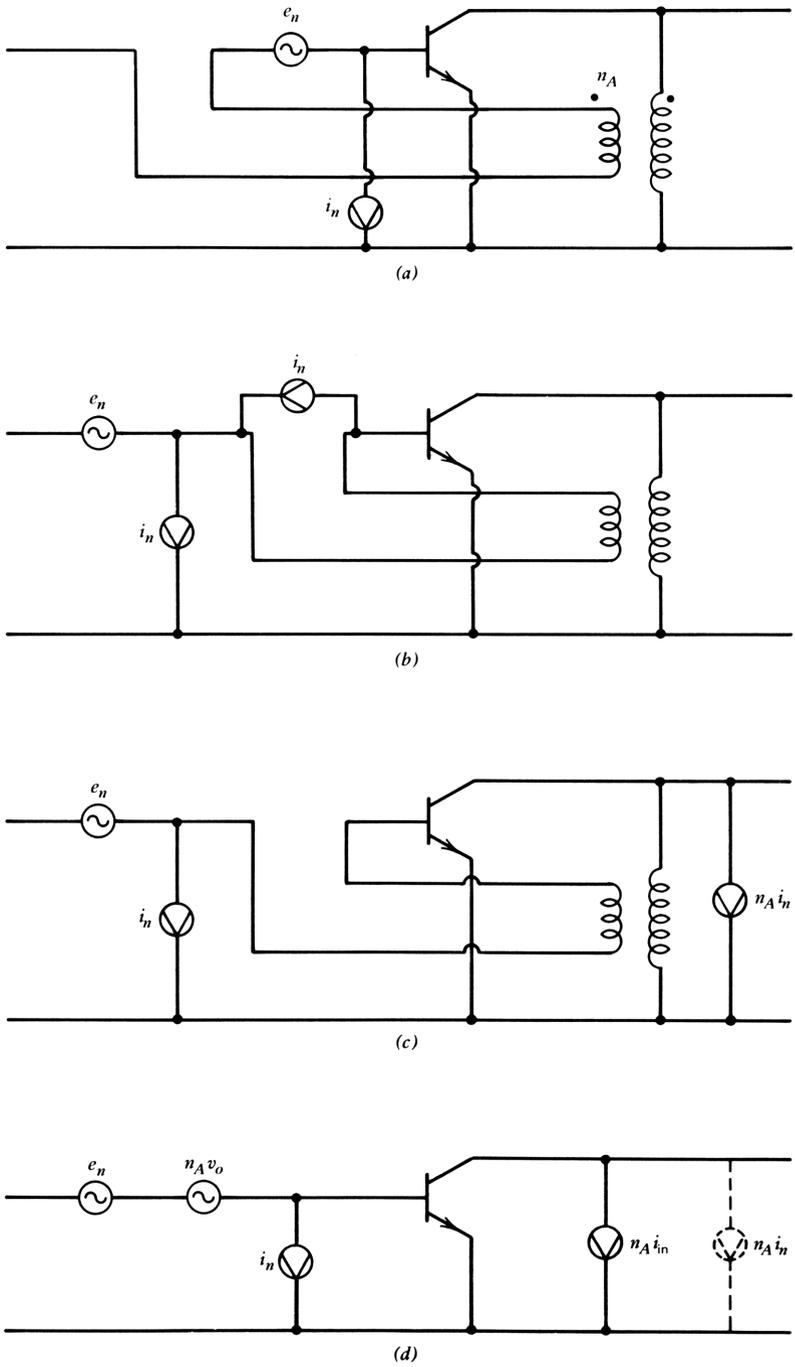


Figure 11.18. The effect of ideal transformer A feedback on noise.

giving the transmission matrix (ignoring direct feedthrough) or

$$T_A = \begin{bmatrix} A^{-n_A} & B \\ C & D \end{bmatrix} \quad (11.5-3)$$

The input noise network of the original transistor is shown inside the transformer winding in Fig. 11.18a; it can be moved outside as shown in Fig. 11.18b by splitting the current generator as shown. The current generator connected across the transformer winding can then be passed through the transformer to the output circuit, as shown in Fig. 11.18c. Its value is $n_A i_n$ and is smaller than that for the common collector stage by the factor n_A . It can be neglected, so the input noise network is the same as that for the common emitter stage.

The input noise voltage of a following second stage can be translated to the input of the first stage through the matrix of (11.5-3); thus it is reduced approximately by the factor n_A , the value of A for the circuit with the transistor in its reference condition.

Lossy Feedback

All examples of feedback in this section have been lossless in the sense that there was no dissipation in the feedback networks. There has also been an additional restriction, namely, that no energy be stored in the feedback networks; such networks are termed *nonenergetic*.¹⁵ The ideal transformer stores no energy, nor do the wires that permute the leads in the common collector or common base connections. As a result, the only effect of the connecting the feedback network was to modify the noise by feedthrough or feedforward. In this section we remove this restriction, first by adding a dissipationless feedback network (by using a feedback capacitor) and then by adding a feedback conductance.

An amplifier with capacitive C feedback is shown in the first stage shown in Fig. 11.19a. Input current signals are integrated by capacitor C_F and the amplifier; capacitor C_C differentiates the signal so that the combination provides flat response. What is the effect of the feedback on the noise represented by generators e_{ni} and i_{ni} ? In Fig. 11.19b the equivalent ladder circuit is drawn. The output loading and feedforward have negligible effect on the noise and are shown dotted. At the input, current generator $C_F s v_{o1}$ represents the feedback and is noiseless; shunt capacitor C_F represents input loading by the feedback capacitor. It contributes no noise itself, but it enhances the noise originating with e_{ni} . Without C_F feedback, the equivalent input noise conductance is, from (11.3-9)

$$G_{inT} = G_G + G_{ni} + G_G^2 R_{ni} \quad (11.5-4)$$

in which correlations between e_{ni} and i_{ni} have been ignored. When C_F is added

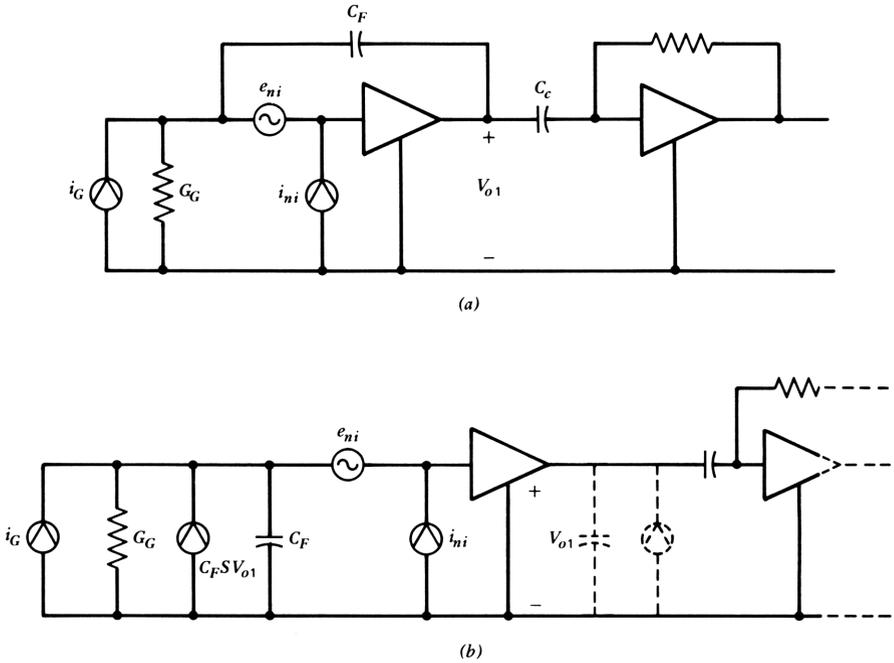


Figure 11.19. Effect of capacitive feedback on noise.

to the circuit, the noise conductance becomes

$$G_{inT} = G_G + G_{ni} + (G_G^2 + D_F^2 \omega^2) R_{ni} \tag{11.5-5}$$

thereby adding $C_F^2 \omega^2 R_{ni}$ to the total.

When a feedback conductance replaces the feedback capacitance, as shown in Fig. 11.20, thermal noise is contributed by the feedback conductance itself; thus

$$G_{inT} = G_G + G_F + G_{ni} + (G_G + G_F)^2 R_{ni} \tag{11.5-6}$$

in which the first G_F term represents the thermal noise contribution and the term in the parentheses gives the enhancement of the effect of R_{ni} by the input loading of the feedback network. Capacitive feedback, therefore, can give better noise performance than can conductive feedback. This can be important in critical applications.

Fiber Optic Preamplifier Incorporating the Optimized First Stage

In the previous section we found the optimum collector current for low noise in a transistor used as the first stage of a preamplifier For a 274 Mband/s fiber optic digital transmission system. The signal source is a PIN photodiode.

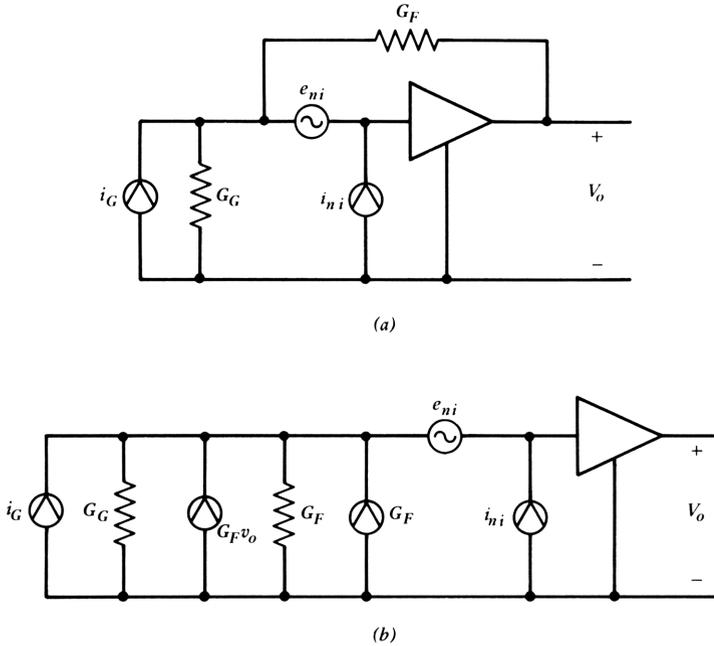


Figure 11.20. Effects of conductive feedback on noise.

The optimization was carried out without regard for transmission requirements, but the stage was analyzed for its transmission properties. Such a stage is of value only if it can be incorporated into a larger system that has satisfactory transmission characteristics; conversely, such a system must not have a significantly adverse effect on noise performance.

What do we mean by “satisfactory transmission performance?” As we saw in Part 1, we require low sensitivity to device and component variations in band and adequate margin against instability out of band. The procedure adopted for ensuring these properties is to select a suitable loss polynomial that has the desired response and sensitivity and to realize it by a circuit that does not introduce high-component-sensitivity problems. For purposes of roughing out a design, we adopt a cubic maximally flat delay polynomial of sufficient bandwidth (400 MHz) to maintain low sensitivity in band to the device parameters that introduce the frequency cutoff.

The design then requires that we synthesize the required polynomial coefficients under the constraint that the first stage operate at the optimum collector current for noise (a collector current at which the device time constants are considerably larger than we would choose from transmission considerations alone).

An additional constraint is that the synthesis of the polynomial coefficients should not introduce a significant increase in noise over that contributed by the optimized first stage.

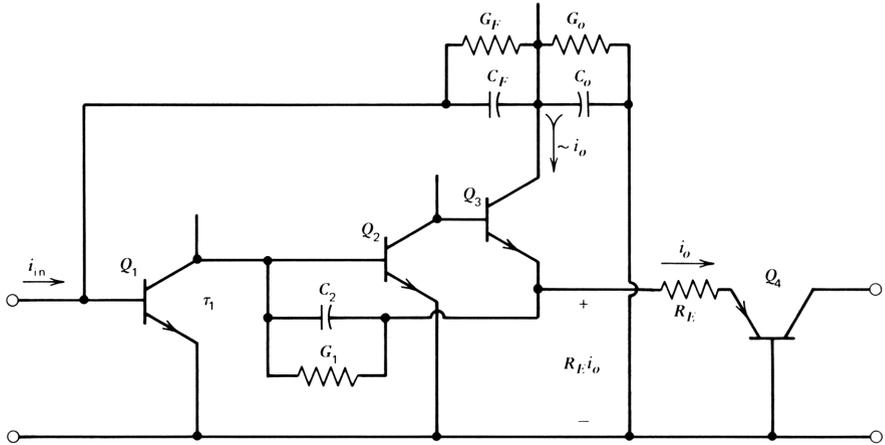


Figure 11.21. Fiber optic receiver preamplifier design with the use of capacitive overall feedback.

The design constraints are satisfied by the circuit shown in Fig. 11.21, in which specific networks are designated to control each of the required polynomial coefficients. We assume the desired loss polynomial to be the ratio of input current to output current:

$$-L(s) = a_0 + a_1 s + a_2 s^2 + a_3 s^3 \quad (11.5-7)$$

The circuit is best described by identifying the elements that determine each coefficient in turn.

Coefficient a_0 is established by the current divider consisting of Y_F and Y_o ; over most of the frequency range it is a capacitive divider, becoming resistive at very low frequencies. (The resistors provide dc base bias for the first stage and limit the dynamic range for low-frequency signals arising from long strings of signal pulses of like polarity.) The mechanism by which a_0 is established by this divider can be understood by noting that the collector current of Q_3 is closely equal to the output current of the amplifier. This collector current is divided by the Y_F , Y_o current divider, giving a value of a_0 of

$$a_0 = \frac{C_F}{C_F + C_o} = \frac{G_F}{G_F + G_o} \quad (11.5-8)$$

If the low-frequency current loss is 0.01 (-40 dB), for example, we would make $C_o = 99C_F$ and $G_o = 99G_F$. The time constants of Y_F and Y_o are equal so as to give flat response to dc. The values of C_F and G_F are typically 0.1 pF and 0.01 mS, respectively.

The remaining three polynomial coefficients all contain τ_1 , the first-stage time constant, as a factor. This time constant, evaluated as 0.14 ns in the previous section, is the constraint introduced by the noise optimization.

Coefficient a_1 is synthesized by use of resistor R_E and conductance G_1 . The voltage at the emitter of Q_3 is $R_E i_o$. (Let R_E include the emitter resistance of the common base stage.) This voltage drives a current $R_E G_1 i_o$ through G_1 , whereupon it is multiplied by D of the first stage, or $\tau_1 s$. Thus a_1 is given approximately as

$$a_1 \approx R_E G_1 \tau_1 \quad (11.5-9)$$

Coefficient a_2 is synthesized in the same way as a_1 , except that we replace G_1 by the susceptance $C_2 s$; therefore, a_2 is given by

$$a_2 \approx R_E C_2 \tau_1 \quad (11.5-10)$$

Coefficient a_3 comes about from the product of the time constants of the three transistors Q_1 , Q_2 , and Q_3 connected in tandem and is given roughly by

$$a_3 \approx \tau_1 \tau_2 \tau_3' \quad (11.5-11)$$

where τ_3' includes not only the current loss time constant of Q_3 , but the time constant $R_E C_{23}$, where C_{23} is the shunt parasitic capacitance of the Q_2-Q_3 interstage, roughly 0.5 pF. These time constants are the basic physical constraints on amplifier gain–bandwidth performance and are not easily changed. Hence, as in Section 3.1, we adjust the low-frequency loss to obtain the required bandwidth of 0.4 GHz. Thus we set

$$a_0 = a_3 \omega_0^3 \quad (11.5-12)$$

Other effects such as $r_b C_{jc}$ products also affect the coefficients and add delay, or, equivalently, increase the degree of the loss polynomial. These effects have been taken into account in a complete design of the circuit in Fig. 11.21 using the methods developed in the earlier chapters; the simple analysis given here gives equivalent results if we take $\tau_1 = 0.16$ ns, $\tau_2 = 0.055$ ns, and $\tau_3 = 0.06$ ns. With $R_E = 0.158$ k Ω , $G_1 = 0.75$ mS, and $C_2 = 0.3$ pF (this latter includes C_{jc} of Q_2 , with which it is essentially in parallel), we obtain for the loss polynomial:

$$-L(s) = 0.0196 + 0.0190s + 0.0076s^2 + 0.00124s^3 \quad (11.5-13)$$

This polynomial in normalized form becomes

$$-L(p) = 0.0196(1 + 2.43p + 2.44p^2 + p^3) \quad (11.5-14)$$

where $p = s/(2 \times 0.4)$. This is close to the maximally flat delay cubic given in Table 2.1, with a dc loss of -34.1 dB.

The design of the preamplifier is thus roughed out with the use of a simple integrator model for each transistor. A more thorough analysis would show

that delay in the transistors and due to parasitics controls the gain that can be achieved for a given required bandwidth and shape. For the design given here, the gain can be increased by lowering the value of R_E . The more complete analysis indicates that R_E must have a minimum value to assure out-of-band stability. These considerations are beyond the scope of this chapter but can be investigated by methods already discussed in previous chapters.

To account for integrated circuit process variations, it is desirable to include means for adjusting the response—adjusting the relative values of the polynomial coefficients—in the completed circuit. This can be accomplished in this circuit by changing the dc collector current of the first stage, which changes τ_1 . This changes a_1 , a_2 , and a_3 , but not a_0 , and changes the relative values of b_1 and b_2 . Since b_1 and b_2 control the in-band response shape, and because the sensitivity of the in-band response shape to the cubic coefficient is much smaller than to b_1 and b_2 , as discussed in Section 2.6, we obtain good control of the response shape by this simple strategy. First-stage dc collector current lower than nominal increases b_1 relative to b_2 , and vice versa.

Noise in Current Sources and Active Loads

Although feedforward is usually inconsequential to noise performance, such is not always the case. In the differential amplifier with active load shown in Fig. 11.22*a*, the collector current of the first transistor is fed forward through the simple current mirror to the output, doubling the current gain of the combination. Development of the equivalent ladder network is shown in Figs. 11.22*b* and 11.22*c*. In Fig. 11.22*b* the circuit is simply redrawn showing only the essential elements. In Fig. 11.22*c* the current mirror is replaced by the $i_o/2$ generator, as described in Chapter 9; the noise of the current mirror is represented by the *output* noise equivalent conductance G_{no} , whose value is given by eq. (11.2-11) as described previously. (Since the current gain of the circuit is unity, the equivalent noise input conductance is equal to the output noise conductance.) Hence

$$G_{no} = \frac{2(r_b + R_E) + r_e}{(r_e + R_E)^2} \quad (11.5-15)$$

The noise equivalent input immittances of the differential pair are given by (11.5-1) and (11.5-2). To express G_{no} by equivalent input immittances, we multiply G_{no} by the squared magnitudes of B and D of the differential pair with the feedforward in place, as indicated in Fig. 11.22*c*, for which

$$B = r_e \quad (11.5-16)$$

and

$$D = \frac{1}{2}(\delta + \tau_T s) \quad (11.5-17)$$

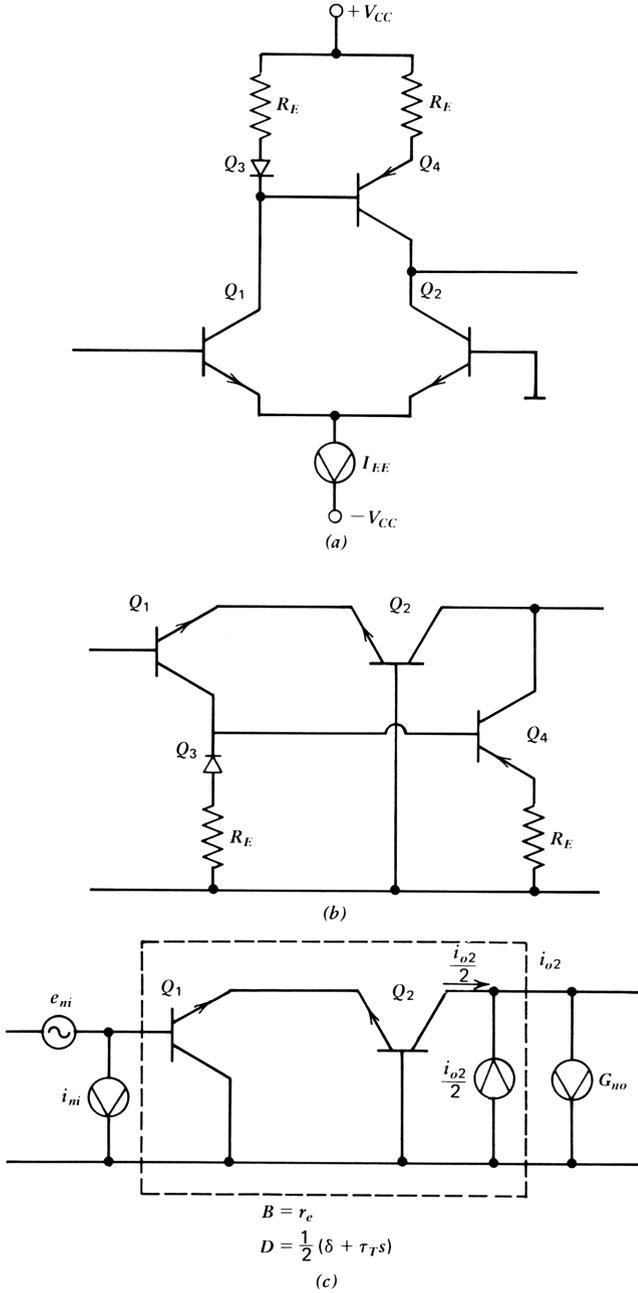


Figure 11.22. Differential pair with active load to show effect of feedforward on noise.

Note that whereas B of the differential pair alone is $2r_e$, the active load supplies half the output current, giving the value shown. Since the transistors of the pair and the active load all have equal collector currents, r_e is the same for all transistors, and the total noise equivalent input resistance for the circuit becomes

$$R_{ni} = r_e + 2r_b + r_e^2 \frac{2(r_b + R_E) + r_e}{(r_e + R_E)^2} \quad (11.5-18)$$

Often, differential pairs are operated at low currents, making r_e dominant over r_b . If we ignore r_b and set $R_E = 0$, the value of R_{ni} reduces to $2r_e$. Hence the active load doubles the input noise equivalent resistance. Usually r_b cannot be ignored, however. As we saw in Section 11.2, addition of R_E helps considerably in reducing the noise contribution of r_b .

Note that it is not the feedforward itself that causes the noise in this circuit. Rather, it is the noisy character of the current mirror. If the same current mirror were used as a simple current source to supply the collector of the second stage, its noise power contribution to the input of the differential pair would be four times as large since B would double, and it enters into R'_{ni} as $|B|^2$. The feedforward actually acts to reduce the noise contribution of the current source by increasing the gain of the pair.

In the 741 operational amplifier, current mirror feedforward is employed as described previously. The output noise of the current mirror is the same as that described previously, but the presence of the *pnp* level shifter transistors doubles the value of B to $2r_e$ and also doubles the input noise equivalent resistance. Hence the noise contribution of the current mirror is four times that of the circuit in Fig. 11.22, and the input contribution is twice as large; for this circuit, therefore, we obtain

$$R_{ni(741)} = 2r_e + 4r_b + 4r_e^2 \frac{2(r_b + R_E) + r_e}{(r_e + R_E)^2} \quad (11.5-19)$$

With collector currents of 0.012 mA, $R_E = 1.0$ k Ω , and $r_b = 0.5$ k Ω , the noise equivalent input resistance is about 16 k Ω . Note that even with r_b and R_E set to zero, the noise equivalent input resistance of the 741 amplifier is $6r_e$; with collector currents of 12 μ A, $r_e = 2.2$ k Ω , and R_{ni} would then be 13 k Ω , of which two-thirds comes from the current mirror.

The noise equivalent input conductance is found by multiplying G_{no} by the square magnitude of D . For either the circuit in Fig. 11.22 or in the 741 amplifier, this contribution is much smaller at low frequencies than the base shot noise contribution of the input stage $\delta/2r_e$. With a collector current of 0.012 mA and with $\delta = 0.01$, the value of G_{ni} at low frequencies is 0.0023 mS. The optimum source resistance for the 741 amplifier is thus 83 k Ω , from eq. (11.3-34), and the noise figure at this source resistance is 1.4 dB, from eq. (11.3-33).

REFERENCES

- 1 H. Rothe and W. Dahlke, "Theory of Noisy Fourpoles," *Proc. IRE* **44**, 811 (June 1956).
- 2 H. A. Haus, Chairman, "IRE Standards on Methods of Measuring Noise in Linear Two Ports," *Proc. IRE* 60–74; (January 1960).
- 3 C. Hall, *Noise in Electronics*, H. Sams, Indianapolis, IN, 1973.
- 4 A van der Zeil, *Noise, Sources, Characterization, Measurement*, Prentice-Hall, Englewood Cliffs, NJ, 1970.
- 5 W. R. Bennett, *Electrical Noise*, McGraw-Hill, New York, 1960, Chapter 5.
- 6 *Ibid.*, Chapter 2
- 7 M. Gardner, "White and Brown Music, Fractal Curves, and One-Over- f Fluctuations," *Sci. Am.*, 16 (April 1978).
- 8 B. Mandelbrot, *Fractals: Form, Chance, and Dimension*, W. Freeman, San Francisco, 1977.
- 9 D. Wolf, "1/ f -Noise," D. Wolf, Ed., in *Noise in Physical Systems*, Springer-Verlag, New York, 1978, p. 122. Gives 77 references and a good overview.
- 10 P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Wiley, New York, 1978, Chapter 10.
- 11 R. F. Voss and J. Clarke, *Phys. Rev.* **B13**, 556 (1976).
- 12 J. T. Wallmark and H. Johnson, *Field Effect Transistors*, Prentice-Hall, Englewood Cliffs, NJ, 1966, Chapter 6.
- 13 S. D. Personick, "Receiver Design for Digital Fiber Optic Communication Systems, Parts I and II," *BSTJ* **52**, 843–886 (July–August) 1973.
- 14 H. Kressel, Ed., *Semiconductor Devices and Optical Communications*, Springer-Verlag, New York, 1979, Berlin, 1979.
- 15 E. H. Nordholt, "Classes and Properties of Multiloop Negative Feedback Amplifiers," *IEEE Trans. CAS-28*, (3), 203–212 (March 1981).

Chapter 12

Differential and Operational Amplifiers

Differential amplifiers have two input ports. Such an amplifier processes the difference in *voltage* between the two ports; it may combine the signals at a single output port, as in most operational amplifiers, or it may present a balanced output signal. Where a single output port is provided, the circuit is a three-port network, so that the two-port *ABCD* characterization previously discussed must be modified. The main purpose of this chapter is to show how this modification should be made. We wish to analyze differential amplifiers in such a way that loop gains disappear from the description; thus the analysis is sequential, straightforward, and appropriate.

The small-signal properties of a two-port can be expressed at any given frequency by four parameters (magnitude and phase) relating the input and output current and voltage. For a three-port there are three currents and voltages to be interrelated and thus nine complex numbers to represent it. The increased complexity of this three-port has been handled in many ways since the introduction of the differential dc amplifier. Several books appearing in the literature have been concerned exclusively with differential amplifiers.^{1, 2} The treatment given here is distinctly different from the earlier treatments and is in

keeping with a principle that can be stated as follows:

In a class of things to be represented by a model, the model parameters should be so chosen that an ideal member of that class shall have no numbers to describe it.

This is the principle of the *null reference matrix* and was encountered for two ports in the ideal two-port amplifier. By choosing to represent two-port amplifiers by their *ABCD* or transmission matrices, an ideal two-port amplifier was found to have all four parameters zero. Nonideal operation can then be analyzed as the sum of several small effects, added to a base or datum of zero. In this chapter we describe the differential amplifier and the operational amplifier such that all nine parameters representing it are zero for an ideal member of the class.

12.1 DEFINITION OF SIGNALS AND CHOICE OF INDEPENDENT VARIABLES

An operational amplifier such as that shown in Fig. 12.1a has a single output port with port voltage v_o and current i_o and two input ports with voltages v_1 and v_2 to ground and currents i_1 and i_2 . The amplifier is arranged to be sensitive to the difference between v_1 and v_2 ; this is the *differential input voltage* v_d :

$$v_d = v_1 - v_2 \tag{12.1-1}$$

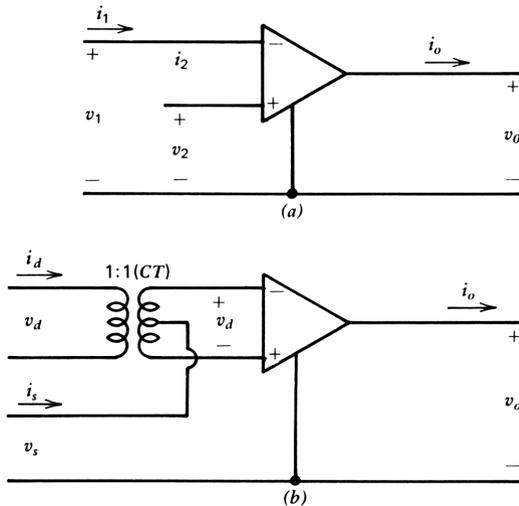


Figure 12.1. Definition of signals in an operational or differential amplifier. In Fig. 12.1b an ideal transformer is used to clarify the definition of differential and common mode signals.

Ideally, this differential voltage is zero for any finite output voltage and current. The amplifier is also arranged to be insensitive to the average value of v_1 and v_2 . This average value is the *common mode input voltage* v_s :

$$v_s = \frac{v_1 + v_2}{2} \quad (12.1-2)$$

We can also define a *differential input current* i_d :

$$i_d = \frac{i_1 - i_2}{2} \quad (12.1-3)$$

and a common mode input signal current i_s :

$$i_s = i_1 + i_2 \quad (12.1-4)$$

Ideally, since i_1 and i_2 are zero, both i_d and i_s are zero.

These definitions can be given a physical interpretation by use of an ideal transformer at the input, as shown in Fig. 12.1b. The differential voltage appears across the primary of the transformer and the common mode voltage appears at the center-tap of the transformer secondary. Note that because i_1 (or i_2) flows through only half of the secondary, the factor of 2 appears in the denominator of eq. (12.1-3).

The differential and common mode voltages given by (12.1-1) and (12.1-2) are the standard IEEE definitions of these terms. To preserve power relationships for the differential and common mode signals, we obtain the differential and common mode current definitions of (12.1-3) and (12.1-4). Unfortunately, these latter definitions are not those commonly used in specifying operational amplifier characteristics in manufacturer's data sheets. The commonly used definitions give twice the value for i_d and half the value of i_s . For technical consistency, we adopt the current definitions of (12.1-3) and (12.1-4), bearing in mind that manufacturer's data must be interpreted in this light.

Null Reference Matrix for a Differential Amplifier

Of the six signal variables of interest, three are ideally zero: in an ideal operational amplifier v_d , i_d , and i_s are all zero. In accordance with the null reference matrix principle, we may immediately write the appropriate matrix description of an operational amplifier by taking the ideally null variables as the dependent variable and the remaining three as the independent variables, so that the null reference matrix is given by the equation

$$\begin{bmatrix} v_d \\ i_d \\ i_s \end{bmatrix} = \begin{bmatrix} A_{do} & B_{do} & A_{ds} \\ C_{do} & D_{do} & C_{ds} \\ C_{so} & D_{so} & Y_{ss} \end{bmatrix} \begin{bmatrix} v_o \\ i_o \\ v_s \end{bmatrix} \quad (12.1-5)$$

An ideal operational amplifier then is described by the null matrix: each of the nine elements is zero. Each of the nine elements is a function of the independent variables in general, so that (12.1-5) is a nonlinear differential description of the operational amplifier where the functional dependencies are known. It is also a complete description of the three-port, which requires nine elements for its full characterization. Many of the elements often can be ignored; elementary operational amplifier circuit analysis often ignores *all* the elements, regarding the amplifier as ideal. In what follows, we define each element and evaluate it in typical cases.

Null Reference Matrix Elements: Definitions

The differential amplifier matrix in eq. (12.1-5) is more easily understood and remembered when partitioned as shown. The submatrix in the upper left position is the *differential ABCD* matrix relating the *differential* input signal vector to the output signal vector, *with the common mode voltage set to zero*. Thus

$$A_{do} = \left. \frac{\partial v_d}{\partial v_o} \right|_{i_o=0, v_s=0} \quad (12.1-6)$$

$$B_{do} = \left. \frac{\partial v_d}{\partial i_o} \right|_{v_o=0, v_s=0} \quad (12.1-7)$$

$$C_{do} = \left. \frac{\partial i_d}{\partial v_o} \right|_{i_o=0, v_s=0} \quad (12.1-8)$$

$$D_{do} = \left. \frac{\partial i_d}{\partial i_o} \right|_{v_o=0, v_s=0} \quad (12.1-9)$$

Note that we adopt the convention that with the variables as defined in Fig. 12.1*b*, these parameters all exhibit a phase reversal.

Common Mode Rejection Ratio and Admittance

The column submatrix in the upper right corner of the matrix in eq. 12.1-5, $\{A_{ds}, C_{ds}\}$, represents the *common mode rejection*, the differential input signal required to make both the output voltage *and* current zero in the presence of a nonzero common mode input voltage. Thus

$$A_{ds} = \left. \frac{\partial v_d}{\partial v_s} \right|_{v_o=0, i_o=0} \quad (12.1-10)$$

$$C_{ds} = \left. \frac{\partial i_d}{\partial v_s} \right|_{v_o=0, i_o=0} \quad (12.1-11)$$

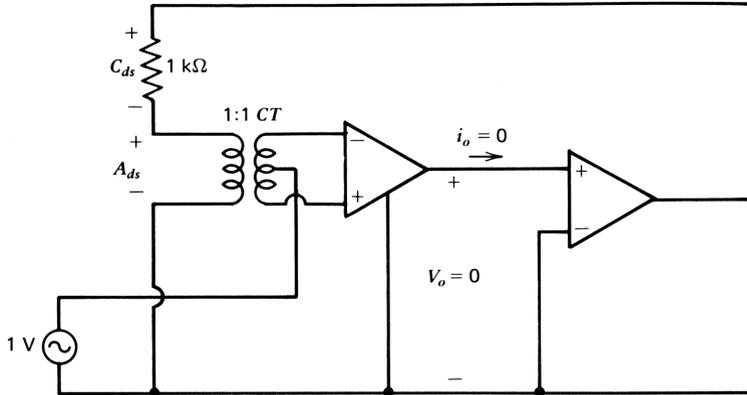


Figure 12.2. Schematic representation of the common mode rejection parameters A_{ds} and C_{ds} .

where A_{ds} is the *common mode rejection ratio* quoted in operational amplifier data sheet specifications (usually expressed in decibels)* and C_{ds} is a *common mode rejection admittance*, usually ignored in operational amplifier data sheets. To gain intuitive insight into these two parameters, recourse to their measurement is useful.

Figure 12.2 shows a conceptual measurement scheme for determining the common mode rejection parameters of a differential or operational amplifier.

Note that in eqs. (12.1-10) and (12.1-11) both v_o and i_o must be set to zero simultaneously. This is done by connecting the input of an ideal amplifier to the output of the amplifier under test. By applying a 1 V input signal to the common mode input terminal of the circuit in Fig. 12.2 and measuring the voltage that appears across the differential input terminals, we obtain a numerical measure of the common mode rejection ratio A_{ds} . Similarly, by measuring the voltage across the 1 k Ω resistor in the feedback lead, we obtain, directly, the value of C_{ds} in millimhos.

As a practical matter, the common mode rejection parameters are usually too small to be measured in this way. By placing attenuation in the feedback path we can make the common mode signal large enough to be measured at the second amplifier output. Numerically, A_{ds} and C_{ds} should be as small as possible, and ideally zero.

Common Mode Input Current and Admittance

The row submatrix in the lower left position of the matrix in eq. (12.1-5), [C_{so} , D_{so}], relates the output signal vector to the common mode input current,

*Common mode rejection is taken here as the reciprocal of that usually used (or the negative if expressed in dB).

with the common mode input voltage set to zero. Thus

$$C_{so} = \left. \frac{i_s}{v_o} \right|_{\substack{i_o=0 \\ v_s=0}} \tag{12.1-12}$$

$$D_{so} = \left. \frac{i_s}{i_o} \right|_{\substack{v_o=0 \\ v_s=0}} \tag{12.1-13}$$

Finally, the scalar Y_{ss} in the lower right corner is the *common mode input admittance*, which can be found numerically as the common mode input current in Fig. 12.2. This is the current engendered solely by the common mode voltage, with the output current and voltage preset to zero.

DC Definitions of the Dependent Variables

The dc values of the three dependent signal variables are also of interest in applications of differential amplifiers and operational amperees. The total (dc and ac) differential input voltage V_d , and the current I_d , and the common mode current I_{CM} can be written as the sum of the ac part given by (12.1-5) and the dc values. This equation is similar in form to that used to express noise in a two-port, as in eq. (11.3-1);

$$\begin{bmatrix} V_d \\ I_d \\ I_{cm} \end{bmatrix} = \begin{bmatrix} V_{OS} \\ I_{OS} \\ I_B \end{bmatrix} + [M_{NR}] \begin{bmatrix} v_o \\ i_o \\ v_s \end{bmatrix} \tag{12.1-14}$$

where M_{NR} is the null reference matrix given in (12.1-5) and the dc values of the dependent signal variables are defined thus:

V_{OS}	<i>input offset voltage</i>	$V_1 - V_2$
I_{OS}	<i>input offset current</i>	$(I_1 - I_2)/2$
I_B	<i>input bias current</i>	$I_1 + I_2$

With V_{OS} , I_{OS} , and I_B simultaneously applied to the amplifier, the output voltage and current and the common mode input voltage are all zero. This defines these three dc quantities.

In the following section we analyze a specific differential stage to give physical significance to each of the nine elements of the M_{NR} matrix and to the dc values of the three dependent variables.

It should be noted that the output current from an operational amplifier flows through a load and returns through the (positive and negative) power supply leads. We assume that the power supply leads are grounded in the following analysis and thus can ignore the effects of signal currents and

voltages at the power supply ports. Without this assumption, an operational amplifier becomes a five-port network, with 25 parameters to characterize it. One group of these parameters that models the power supply rejection is important and is discussed later.

12.2 DC ANALYSIS OF A DIFFERENTIAL AMPLIFIER STAGE

In Section 12.4 we present a general method of analysis of the differential amplifier stage that will enable us to find its characteristics by use of transistors or any other three-terminal amplifying devices or combinations of devices. In this section we consider the analysis of such a stage by use of bipolar transistors with a current mirror active load. Such a stage is of practical importance since it forms the input stage of many operational amplifiers. The analysis brings out many essential features of the stage.

To simplify the analysis we shall initially ignore Early effect and base resistance, and adopt the Ebers-Moll model (Chapter 7) for the transistor. We further restrict the discussion to dc or low frequencies. The circuit is shown in Fig. 12.3. The current mirror is ideal (for the present) because the collector current of Q_1 , I_{c1} , is accurately mirrored as a current source feeding the collector of Q_2 . An ideal current source I_{EE} provides emitter current for the two transistors; its shunt admittance is assumed negligible.

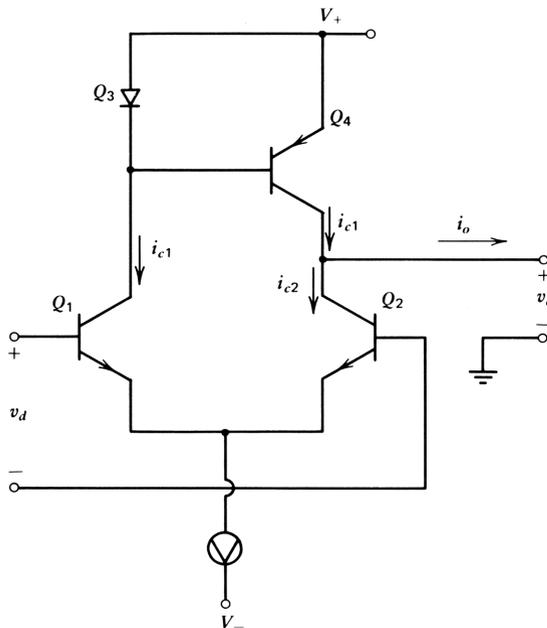


Figure 12.3. Differential pair with current mirror load.

Bias and Output Circuit Constraints

The basic constraint that gives the differential pair its characteristic rejection of common mode signals is that the emitter signal currents of the two emitters are equal and opposite. This constraint can be represented by Kirchoff's current law at the emitter node:

$$I_{c1}(1+\delta_1)+I_{c2}(1+\delta_2)=I_{EE} \quad (12.2-1)$$

We are interested in establishing (1) the relationship between the output current and voltage, and (2) the input differential current and voltage. This relationship depends on the output coupling arrangement, taken here as the current mirror active load circuit in Fig. 12.3. For this arrangement, we can write

$$I_{c2}=I_{c1}-I_o \quad (12.2-2)$$

Substituting this equation in (12.2-1), we find I_{c1} in terms of the output current I_o :

$$I_{c1}=\frac{I_{EE}+(1+\delta_2)I_o}{2+\delta_1+\delta_2}\simeq\frac{1}{2}\left(\frac{I_{EE}}{1+\delta}+I_o\right) \quad (12.2-3)$$

Similarly

$$I_{c2}=\frac{I_{EE}-(1+\delta_1)I_o}{2+\delta_1+\delta_2}\simeq\frac{1}{2}\left(\frac{I_{EE}}{1+\delta}-I_o\right) \quad (12.2-4)$$

In the expressions on the right the approximation has been made that $1+\delta_1=1+\delta_2=1+\delta$.

The magnitude of the quantity $I_{EE}/(1+\delta)=-I_{\text{omax}}$ is the maximum positive or negative current I_{omax} that can be obtained from the differential pair with active load. It is obtained with large positive or negative differential signal applied to the stage. Hence we can write

$$I_{c1}=-\frac{I_{\text{omax}}}{2}(1+\gamma)=I_C(1+\gamma) \quad (12.2-5)$$

and

$$I_{c2}=\frac{I_{\text{omax}}}{2}(1-\gamma)=I_C(1-\gamma) \quad (12.2-6)$$

where I_C is the quiescent collector current of either transistor and $\gamma=I_o/I_{\text{omax}}$ is defined as the *signal intensity*.

Input Differential and Offset Voltages

The input differential voltage is obtained by subtracting V_{be2} from V_{be1} :

$$V_d = \frac{kT}{q} \left(\ln \frac{I_{c1}}{I_{S1}} - \ln \frac{I_{c2}}{I_{S2}} \right) \quad (12.2-7)$$

$$= \frac{kT}{q} \left(\ln \frac{I_{c1}}{I_{c2}} + \ln \frac{I_{S2}}{I_{S1}} \right)$$

$$= \frac{kT}{q} \left(\ln \frac{1+\gamma}{1-\gamma} + \ln \frac{I_{S2}}{I_{S1}} \right) \quad (12.2-8)$$

$$= v_d + V_{OS} \quad (12.2-9)$$

in which the first term in the brackets is identified with the input differential signal voltage and the second term with the input offset voltage.

In this simple model the ratio of saturation currents is the single contributor to the offset voltage. We evaluate other contributors later on.

The input signal voltage can also be written (by trigonometric identity) as

$$v_d = \frac{2kT}{q} \tanh^{-1} \gamma \quad (12.2-10)$$

This function is plotted as a solid line in Fig. 12.4.

Since $\gamma = I_o / I_{o\max}$, it follows that v_d is a function of the output current but is not a function of the output voltage or the common mode input voltage for this simple transistor model. The Ebers-Moll model, with its ideal collector junction, gives complete voltage isolation of the input circuit from the output circuit; the ideal emitter current source likewise completely isolates the input circuit from ground or the power supplies. Hence the common mode rejection is complete, with A_{ds} of eq. (12.1-5) equal to zero.

It is also true that for this model, the input signal and the offset voltage are *independent* of the unbalance between δ_1 and δ_2 since V_{be} of the two transistors depend only on the *collector* current (not the *emitter* current). Since the ratio of collector currents is established by the current mirror, any inequality in these currents will cause an offset voltage to arise since such an inequality is indistinguishable from a (dc) signal. If we should use a simple current mirror, for example, we saw in Chapter 8 that the input current to the mirror is $(1+2\delta)I_{\text{mirror}}$, giving a ratio of the two currents of $1+2\delta$. With $\delta=0.01$, for example, the resulting offset voltage would be

$$\begin{aligned} V_{OS}(\text{mirror}) &= \frac{kT}{q} \ln 1.02 \\ &= (0.026)(0.02) = 0.00052 \text{ V} \end{aligned}$$

or 0.5 mV.

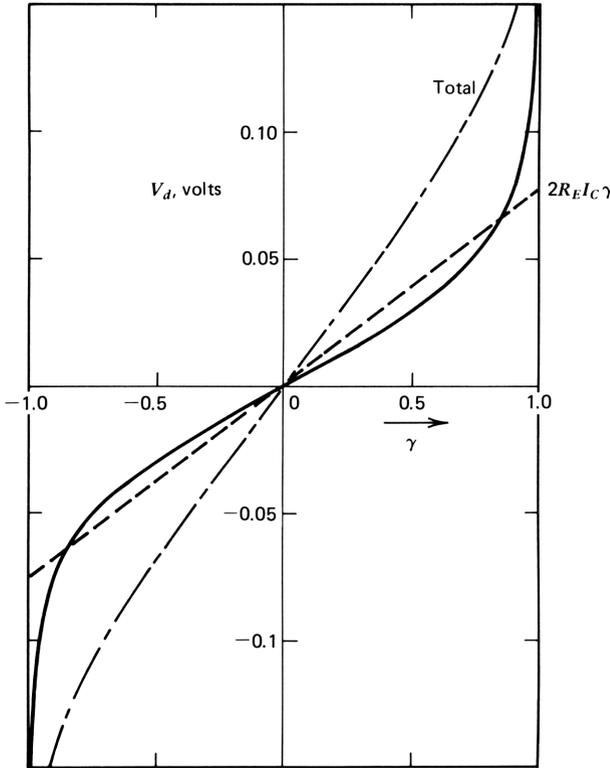


Figure 12.4. Differential input voltage as a function of output current signal intensity. Dashed line represents the differential input voltage contribution from added emitter resistors.

An offset nulling adjustment is often provided that allows the ratio of the current mirror to be modified to bring the total offset voltage to zero.

Null Reference Matrix Parameters

The small-signal parameters of the stage are found by differentiating the expression for v_d with respect to the independent variables of (12.1-5). As we have seen, A_{do} and A_{ds} are both zero since the stage is insensitive to output or common mode voltages. The value of B_{do} is negative of the slope of the curve of Fig. 12.4 and is found by differentiating (12.2-10), giving

$$\begin{aligned}
 B_{do} &= \frac{\partial v_d}{\partial I_o} = \frac{2kT}{q} \frac{d}{dI_o} (\tanh^{-1} \gamma) \\
 &= - \frac{kT}{qI_C} \frac{1}{1-\gamma^2}
 \end{aligned}
 \tag{12.2-11}$$

where I_C is the quiescent current of either transistor, equal to $I_{EE}/2(1+\delta)$. The

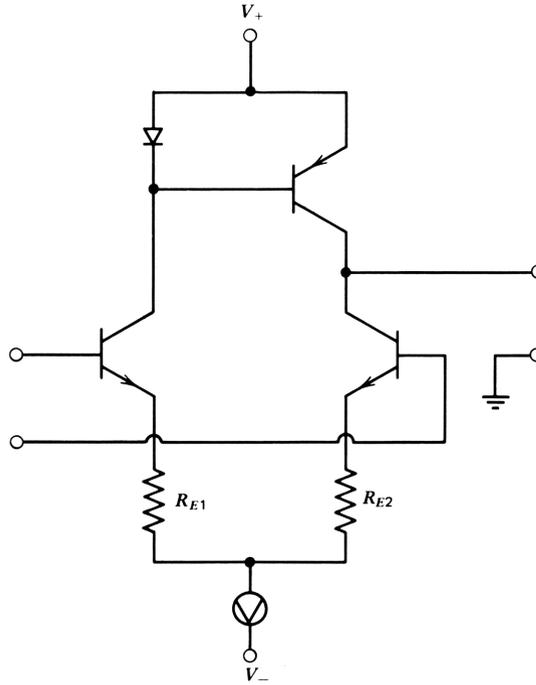


Figure 12.5. Differential amplifier with linearizing emitter resistors.

term $|B_{do}|$ is a minimum for zero signal intensity and rises as the signal intensity increases in either direction.

The nonlinearity of B_{do} can be reduced by adding B feedback—adding emitter resistors in series with each emitter, as shown in Fig. 12.5. This simply augments the input differential voltage by a linear term, shown by the dashed line in Fig. 12.4, giving the total input shown by the broken line. The linearized input voltage becomes

$$V_d = \frac{2kT}{1} \tanh^{-1} \gamma + \ln \frac{I_{s2}}{I_{s1}} + (1 + \delta_1) I_C (1 + \gamma) R_{E1} - (1 + \delta_2) I_C (1 - \gamma) R_{E2} \quad (12.2-12)$$

This expression can be split into signal and offset terms as in the following equations:

$$v'_d = \frac{2kT}{q} \tanh^{-1} \gamma + (1 + \delta) I_C \gamma (R_{E1} + R_{E2}) \quad (12.2-13)$$

and

$$V'_{OS} = \frac{kT}{q} \ln \frac{I_{s2}}{I_{s1}} + I_C [R_{E1} - R_{E2} + (\delta_1 - \delta_2)(R_{E1} + R_{E2})] \quad (12.2-14)$$

in which we have added the saturation current to the offset equation. Note that the offset voltage is now a function of the difference between δ_1 and δ_2 as well as the difference between R_{E1} and R_{E2} since the emitter current includes the base current as well as the collector current.

The input differential voltage and B_{do} similarly become functions of the defect current δ . From (12.2-5), $I_C = I_{omax}/2$, so that the new value of $-B_{do}$, obtained by differentiating (12.2-13), is

$$-B'_{do} = \frac{kT}{qI_C} \frac{1}{1-\gamma^2} + (1+\delta) \frac{R_{E1} + R_{E2}}{2} \quad (12.2-15)$$

Augmentation due to B feedback is essentially $(R_{E1} + R_{E2})/2$. The denominator factor of 2 arises because of the feedforward provided by the active load.

The input differential current as defined in (12.1-3) is simply $(I_{c1}\delta_1 - I_{c2}\delta_2)/2$; from (12.2-5), we have

$$\begin{aligned} I_d &= \frac{I_C(1-\gamma)\delta_1 - I_C(1-\gamma)\delta_2}{2} \\ &= \frac{I_C\gamma(\delta_1 + \delta_2)}{2} + \frac{I_C(\delta_1 - \delta_2)}{2} \end{aligned} \quad (12.2-16)$$

Since $I_C\gamma = -I_o/2$, we have

$$I_d = -\frac{I_o(\delta_1 + \delta_2)}{4} + \frac{I_C(\delta_1 - \delta_2)}{2} \quad (12.2-17)$$

$$= i_d + I_{OS} \quad (12.2-18)$$

As before, we have split I_d into signal and offset components. Neither is a function of either V_o or V_s , so that C_{do} and C_{ds} of (12.1-5) are both zero. We obtain D_{do} by differentiating (12.2-17) with respect to I_o , giving

$$-D_{do} = \frac{\delta_1 + \delta_2}{4} \quad (12.2-19)$$

We now have the first and second rows of the matrix of (12.1-5), and with the offset equations, the first and second rows of (12.1-14).

Common Mode and Bias Current

To complete the dc analysis of the differential amplifier, we find the common mode input current. It is given by

$$I_{cm} = I_{c1}\delta_1 + I_{c2}\delta_2 \quad (12.2-20)$$

$$= I_C(1+\gamma)\delta_1 + I_C(1-\gamma)\delta_2$$

$$I_{cm} = I_o \frac{\delta_1 - \delta_2}{2} + I_C(\delta_1 + \delta_2) \quad (12.2-21)$$

$$I_{cm} = i_s + I_B \quad (12.2-22)$$

from which we conclude that C_{s_o} and Y_{s_s} are both zero, and D_{s_o} is given by

$$-D_{s_o} = \frac{\delta_1 - \delta_2}{2} \quad (12.2-23)$$

and the bias current is given by the second term in (12.2-21):

$$I_B = I_C(\delta_1 + \delta_2) \quad (12.2-24)$$

This completes the dc analysis of the differential amplifier using the Ebers-Moll transistor model, with both all-dc and small-signal terms accounted for. The equations for the stage may be written as in eq. (12.1-14):

$$\begin{bmatrix} V_d \\ I_d \\ I_{cm} \end{bmatrix} = \begin{bmatrix} \frac{kT}{q} \ln \frac{I_{S2}}{I_{S1}} \\ \frac{\delta_1 - \delta_2}{2} I_C \\ (\delta_1 + \delta_2) I_C \end{bmatrix} - \begin{bmatrix} 0 & \frac{kT}{qI_C} \frac{1}{1-\gamma^2} & 0 \\ 0 & \frac{\delta_1 - \delta_2}{2} & 0 \\ 0 & \frac{\delta_1 - \delta_2}{2} & 0 \end{bmatrix} \begin{bmatrix} v_o \\ i_o \\ v_s \end{bmatrix} \quad (12.2-25)$$

in which $I_C = I_{EE}/2(1+\delta)$. This is the matrix equation when no emitter resistors are added. With emitter resistors, the equation is modified as shown previously.

We can conclude from the upper left submatrix that where the Ebers-Moll approximation is valid, the small-signal input voltage is that of a common emitter stage. But the input circuit may have an arbitrary common mode signal applied to it with no effect on the output. The differential input current is half that of a common emitter stage for a given output current.

For critical applications, we next consider the effects of collector voltage.

12.3 EFFECT OF COLLECTOR VOLTAGE ON DC OFFSETS

Integrated circuits allow good matching of transistors so that small offsets result. When this is the case, the unbalance introduced by Early effect can become significant. We illustrate this by determining the dc offset arising from Early effect in the transistors in Fig. 12.6. In this circuit we know intuitively that balanced operation with low offset voltage is secured by making the collector voltages of Q_1 and Q_2 equal.

More particularly, we know that V_{BE} is a function of the saturation current I_S and also that I_S is a function of the Gummel number N_G as given in eq. (7.1-8). But N_G , the number of dopant atoms per square micrometer in the base, depends on the base *width*, which is dependent on collector voltage. hence if the collector voltages are not balanced, the saturation currents will differ in the two transistors, and an offset voltage will arise. Collector voltage

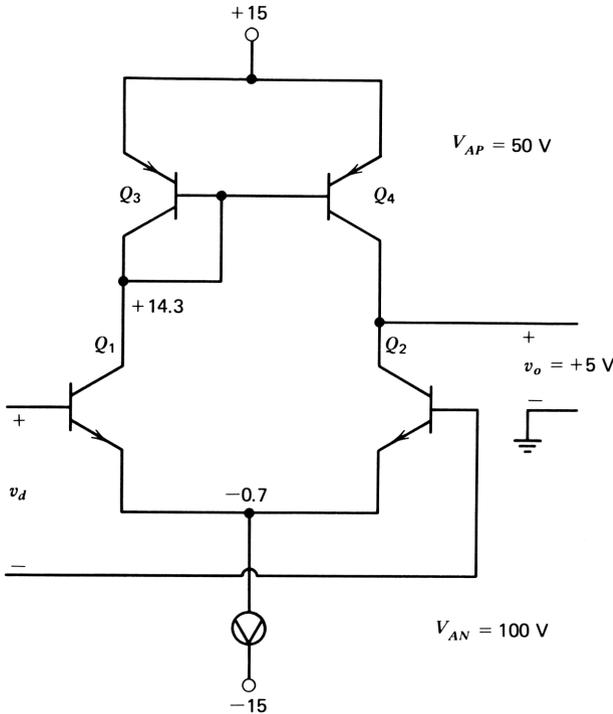


Figure 12.6. Circuit for the analysis of the effect of collector voltage.

balance is often realized in practice, as in the 741 amplifier where the collectors of the differential stage are connected to points that are both two diode drops more positive than the negative power supply.

To illustrate the effect of unequal collector voltages on offset performance, we connect the collector of Q_1 to a point one diode drop less than the positive supply, and the collector of Q_2 to a point 10 V less positive than the supply, as shown in Fig. 12.6.

Early effect is described geometrically in Fig. 7.13, for which we can write

$$\frac{I_C^o}{V_A} = \frac{I_C}{V_{CE} + V_A} \tag{12.3-1}$$

where I_C^o is the collector current in the absence of Early effect and V_A is the Early voltage. The value of I_C^o will be taken as equal in Q_3 and Q_4 . The two current mirror currents are equal to the currents of the differential pair, so we can write

$$I_{C3}^o \left(1 + \frac{V_{CE3}}{V_{AP}} \right) = I_{C1}^o \left(1 + \frac{V_{CE1}}{V_{AN}} \right) \tag{12.3-2}$$

and

$$I_{C4}^{\circ} \left(1 + \frac{V_{CE4}}{V_{AP}} \right) = I_{C2}^{\circ} \left(1 + \frac{V_{CE2}}{V_{AN}} \right) \quad (12.3-3)$$

in which V_{AP} is the Early voltage for the *pn*p current mirror transistors and is negative and V_{AN} is that for the *np*n transistors and is positive. For the circuit in Fig. 12.6, taking the common mode input voltage as zero and $V_E = V_S - V_{BE1} = -0.7$ V, we have

$$V_{CE1} = V_{CC} + V_{BE3} - V_E = 15 \text{ V} \quad (12.3-4)$$

$$V_{CE2} = V_o - V_E = 5.7 \text{ V} \quad (12.3-5)$$

$$V_{CE3} = V_{BE3} = -0.7 \text{ V} \quad (12.3-6)$$

$$V_{CE4} = -(V_{CC} - V_o) = -10 \text{ V} \quad (12.3-7)$$

The input offset voltage contributed by Early effect V_{OSE} is found as the differential input voltage, given by

$$V_d = V_{OSE} = \frac{kT}{q} \ln \frac{I_{C1}^{\circ}}{I_{C2}^{\circ}} \quad (12.3-8)$$

$$= \frac{kT}{q} \left[\ln \left(1 + \frac{V_{CE3}}{V_{AP}} \right) - \ln \left(1 + \frac{V_{CE4}}{V_{AP}} \right) + \ln \left(1 + \frac{V_{CE2}}{V_{AN}} \right) - \ln \left(1 + \frac{V_{CE1}}{V_{AN}} \right) \right] \quad (12.3-9)$$

If we take $V_{AP} = -50$ V and $V_{AN} = 100$ V, the offset changes by

$$\begin{aligned} V_{OSE} &= 0.026(\ln 1.014 - \ln 1.20 + \ln 1.057 - \ln 1.15) \\ &\begin{array}{rcc} & \text{Current mirror} & \text{Differential pair} \\ = & -0.00438 & -0.00219 \\ = & -0.0066 \text{ V} & \end{array} \end{aligned} \quad (12.3-10)$$

A sizable offset (6.6 mV) thus arises from unequal collector voltages of the current mirror and differential pair. The contribution of the current mirror is larger because of the smaller Early voltage of the *pn*p transistors.

If we make the approximation that for small x , $\ln(1+x) \simeq x$, (12.3-9) becomes

$$V_{OSE} \simeq \frac{kT}{q} \left(\frac{V_{CE3} - V_{CE4}}{V_{AP}} + \frac{V_{CE2} - V_{CE1}}{V_{AN}} \right) \quad (12.3-11)$$

Note that the collector voltage differences for the current mirror and the differential pair are equal in magnitude and of a sign to make the two terms in the brackets add, so that we can write

$$V_{OSE} \simeq \left(\frac{kT}{qV_{AN}} - \frac{kT}{qV_{AP}} \right) \Delta V_{CE} \quad (12.3-12)$$

The two terms in the brackets of this equation are recognizable as the A parameters of the nnp and the $pnnp$ transistors (see Section 7.5), sometimes called the Early factor. Hence we can write for the Early component of the offset voltage

$$V_{OSE} \simeq -(A_{nnp} + A_{pnnp}) \Delta V_{CE} \quad (12.3-13)$$

an equation that applies to both offset and small-signal voltages.

In this approximation the emitter voltage, equal to the common mode input voltage less the average base emitter drop of the differential pair, drops out. To a first approximation, therefore, the offset is unaffected by the common mode input voltage.

Effect on Common Mode Rejection

We can also use the preceding development to find the effect of collector voltage differences on the common mode rejection. For this purpose, however, the approximation of (12.3-11) leads to perfect common mode rejection since the latter is a second-order effect. If we use the exact expression [eq. (12.3-9)], however, we find the common mode rejection ratio (CMRR) as V_{OS}/V_S or V_{OS}/V_E , by either differentiating (12.3-9) with respect to V_E or changing V_E by 1 V and finding the change in V_{OSE} . With $V_E = -0.7$ V, the CMRR becomes 20×10^{-6} , or -94 dB. It is a slight function of V_E , as shown in the plot in Fig. 12.7. When the collector voltages are made equal, the CMRR contribution from Early effect drops out.

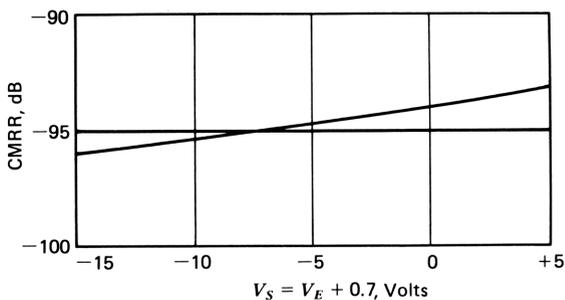


Figure 12.7. Common mode rejection as a function of common mode input voltage for the circuit shown in Fig. 12.6.

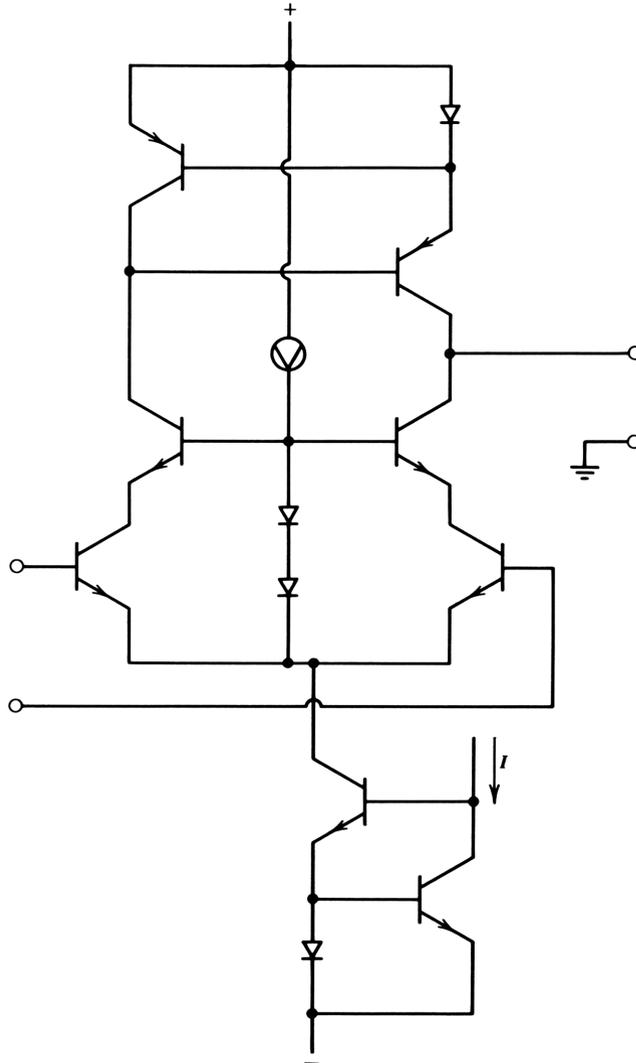


Figure 12.8. Differential amplifier with superior common mode rejection.

Where better common mode suppression is required, more elaborate circuits may be substituted for each of the transistors of the differential pair. To reduce the effect of Early conductance, for example, each transistor may be replaced by a cascode stage, with the bases of the common base transistors connected to the common emitter point for signals, as shown in Fig. 12.8. Note that a Wilson current mirror loads the stage since the advantage of the cascode connection would be largely lost with a simple current mirror. Why? A Wilson current source is used in the emitter circuit to further improve common mode rejection.

12.4 GENERAL ANALYSIS OF THE DIFFERENTIAL STAGE

There are many possible variations of the basic differential stage that allow superior performance over the basic stage discussed in Section 12.3. A Darlington pair, for example, might be used to reduce the input bias current or to reduce the differential and common mode input currents. A cascode stage might replace the transistors of the pair, as noted previously, or a pair of FET transistors may be used to reduce input currents.

A general analysis method that enables us to evaluate the performance of these and other alternatives can be formulated according to the method shown in Fig. 12.9. Each element of the differential pair is represented by its *ABCD* parameters, and input and output coupling networks are connected to the input and output to represent the source and load arrangements. Each *ABCD* matrix might be a single transistor as discussed previously or may be a collection of transistors whose configuration is intended to provide some desired performance features.

The input coupling network may involve merely a change in the definition of input variables as discussed in Section 12.1. Such a transformation may be represented by an ideal transformer, shown in Fig. 12.1b. The output coupling network might be a current mirror, for example, as discussed previously, or it might simply be a single-ended output from one collector of the pair.

Generation of the Null Reference Matrix

We now find the null reference matrix elements of the circuit in Fig. 12.9 as functions of the *ABCD* parameters of the individual pair elements. The

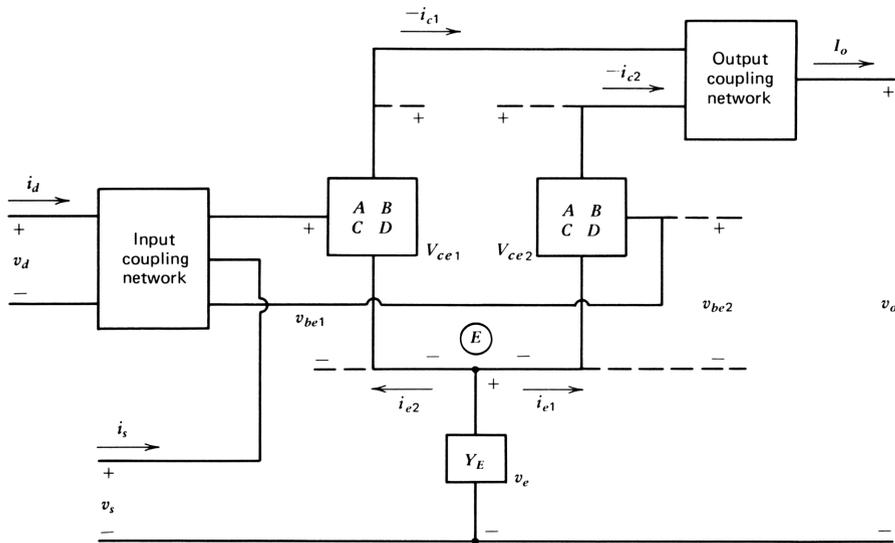


Figure 12.9. Separation of differential amplifier into three sections for general analysis.

differential and common mode characteristics of the pair can then be evaluated over the frequency range of interest.

The procedure to be followed in finding the null reference matrix involves three steps. First, we find the input voltage and current of the two pair elements:

$$\begin{bmatrix} v_{be1} \\ i_{b1} \\ v_{be2} \\ i_{b2} \end{bmatrix} = \begin{bmatrix} A_1 & B_1 & 0 & 0 \\ C_1 & D_1 & 0 & 0 \\ 0 & 0 & A_2 & B_2 \\ 0 & 0 & C_2 & D_2 \end{bmatrix} \begin{bmatrix} v_{ce1} \\ -i_{c1} \\ v_{ce2} \\ i_{c2} \end{bmatrix} \quad (12.4-1)$$

Next, we obtain the input transformation matrix by rewriting eqs. (12.1-1), (12.1-3), and (12.1-4) in matrix form:

$$\begin{bmatrix} v_d \\ i_d \\ i_s \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_{be1} \\ i_{b1} \\ v_{be2} \\ i_{b2} \end{bmatrix} \quad (12.4-2)$$

This gives us the desired input variables in terms of the input variables of the individual transistors; we can obtain them in terms of the output variables of the transistors by postmultiplying this matrix by the transistor matrix in (12.4-1):

$$\begin{bmatrix} v_d \\ i_d \\ i_s \end{bmatrix} = \begin{bmatrix} A_1 & B_1 & -A_2 & -B_2 \\ \frac{C_1}{2} & \frac{D_1}{2} & -\frac{C_2}{2} & -\frac{D_2}{2} \\ C_1 & D_1 & C_2 & D_2 \end{bmatrix} \begin{bmatrix} v_{ce1} \\ -i_{c1} \\ v_{ce2} \\ -i_{c2} \end{bmatrix} \quad (12.4-3)$$

This is as far as we can take the analysis without considering the specific output load arrangement and the basic differential amplifier constraint of (12.2-1), Kirchoff's current law at the emitter node. The third step is to express the column vector on the right in terms of the variables v_o , i_o , and v_s . At this point we make the simplifying assumption that the common mode input voltage is equal to the signal voltage v_e at node E in Fig. 12.9:

$$v_s \simeq v_e \quad (12.4-4)$$

To investigate the effect of current mirror unbalance, we employ a current mirror active load on the pair. We assume that the output current to input current ratio of the mirror is m (rather than unity), as shown in Fig. 12.10. Furthermore, by setting $m=0$, we can find the characteristics of a single-ended output load arrangement.

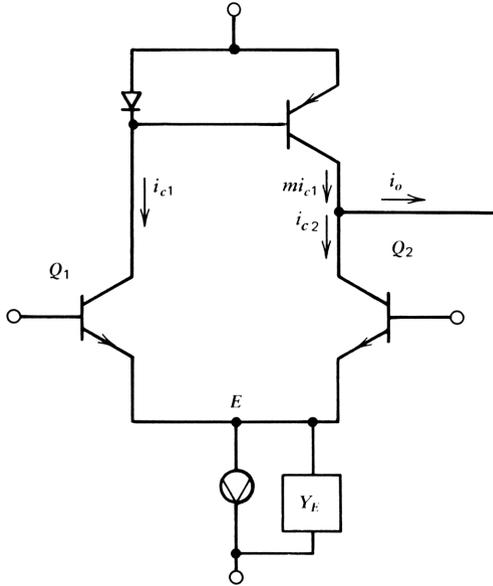


Figure 12.10. Differential amplifier with current mirror load that has a current mirror ratio of m .

The current mirror may be considered a feedforward spanning network characterized by its h parameters. In Chapter 9 we showed that h_{12} may usually be neglected, as may h_{11} . Output admittance h_{22} of the mirror may be incorporated into the load; it is generally small, and at high frequencies it is capacitive. The capacitance can be added to the parasitic capacitance of the output node to ground. For our purposes, only $h_{21}=m$ is of importance. Furthermore, we have seen that the current mirror is a broadband stage (of the order of the unity loss frequency) so that the frequency dependence of h_{21} usually can be ignored.

With the foregoing in mind, we can write the following relationship between the output currents of the two pair elements:

$$-i_{c2} = -mi_{c1} + i_o \tag{12.4-5}$$

We now invoke the basic differential amplifier constraint at the emitter node in Fig. 12.10:

$$i_{e1} + i_{e2} + v_e Y_E = 0 \tag{12.4-6}$$

The emitter currents can be related to v_{ce} and i_c of the pair elements by eq. (12.4-7), from the sixth row of eq. (9.1-1) in Chapter 9 with appropriate attention to the signs:

$$i_e = -Cv_{ce} - (1-D)i_c \tag{12.4-7}$$

in which i_e is considered positive flowing into the emitter. The basic constraint becomes

$$-C_1 v_{ce1} - (1 - D_1) i_{c1} - C_2 v_{ce2} - (1 - D_2) i_{c2} + Y_E v_e = 0 \quad (12.4-8)$$

Substituting (12.4-5) in this equation, we can solve for i_{c1} :

$$-i_{c1} = \frac{C_1 v_{ce1} + C_2 v_{ce2} + (1 - D_2) i_o - Y_E v_e}{1 - D_1 + m(1 - D_2)} \quad (12.4-9)$$

But, since $v_{c1} = 0$

$$v_{ce1} = -v_e \quad (12.4-10)$$

and

$$v_{ce2} = v_o - v_e \quad (12.4-11)$$

so that

$$-i_{c1} = \frac{C_2 v_o + (1 - D_2) i_o - (C_1 + C_2 + Y_E) v_e}{d} \quad (12.4-12)$$

where

$$d = 1 - D_1 + m(1 - D_2) \quad (12.4-13)$$

Using (12.4-5), we can also solve for $-i_{c2}$:

$$-i_{c2} = \frac{mC_1 v_o - (1 - D_1) i_o - m(C_1 + C_2 + Y_E) v_e}{d} \quad (12.4-14)$$

We can write these equations in matrix form:

$$\begin{bmatrix} v_{ce1} \\ -i_{c1} \\ v_{ce2} \\ -i_{c2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 \\ \frac{C_1}{d} & -\frac{1 - D_2}{d} & -\frac{C_1 + C_2 + Y_E}{d} \\ 1 & 0 & -1 \\ \frac{mC_1}{d} & \frac{1 - D_1}{d} & -\frac{m(C_1 + C_2 + Y_E)}{d} \end{bmatrix} \begin{bmatrix} v_o \\ i_o \\ v_e \end{bmatrix} \quad (12.4-15)$$

The final step in finding the null reference matrix for the stage is to premultiply the matrix in (12.4-15) by that in (12.4-3), using the approximation in (12.4-4), whereupon we obtain the matrix given in Table 12.1. This null

Table 12.1 Null Reference Matrix^a for a Differential Stage^b

v_d	$-A_2 + \frac{(B_1 - mB_2)C_2}{d}$	$-\frac{B_1(1 - D_2) + B_2(1 - D_1)}{d}$	$-\Delta A - (B_1 - mB_2) \frac{C_1 + C_2 + Y_E}{d}$	v_o
i_d	$-\frac{C_2}{2} \left(\frac{1 + m - 2D_1}{d} \right)$	$-\frac{D_1(1 - D_2) + D_2(1 - D_1)}{2d}$	$-\frac{\Delta C}{2} - (D_1 - mD_2) \frac{C_1 + C_2 + Y_E}{2d}$	i_o
i_s	$C_2 \left(\frac{1 + m}{d} \right)$	$\frac{D_2 - D_1}{d}$	$-(C_1 + C_2) - (D_1 - mD_2) \frac{C_1 + C_2 + Y_E}{d}$	v_s

^aIn this matrix

$m = h_{21}$ of the current mirror

$d = 1 - D_1 + m(1 - D_2)$

$\Delta A = A_1 - A_2$

$\Delta C = C_1 - C_2$

^bNotes: (1) when $m = 1$, load is ideal current mirror; (2) when $m = 0$, the circuit is a single-ended stage (noninverting); and (3) to find the matrix for single-ended inverting case, change signs of v_d, i_d .

reference matrix gives us the small-signal performance of the differential amplifier within the approximation that the common emitter voltage is equal to the common mode voltage. This matrix may be partitioned as follows;

$$T = \begin{bmatrix} T_{do} & T_{ds} \\ T_{so} & Y_{ss} \end{bmatrix}$$

The expressions for the elements of the null reference matrix appear complicated. Nevertheless, they can give us considerable insight into the operation of the stage. To illustrate this, we take $m=1$ (ideal current mirror load) and observe each element in turn.

Differential Two-Port Submatrix

The differential submatrix T_{do} is given by

$$T_{do} = - \begin{bmatrix} -A_2 & \bar{B} \\ \frac{C_2}{2} & \bar{D} \end{bmatrix} \quad (12.4-16)$$

in which we have ignored $B_1 - B_2$ and have assumed $1 - D_1 = 1 - D_2$; \bar{B} and \bar{D} are the average values of these parameters for the two transistors. This matrix is that of a transistor similar to each of the transistors of the pair, except that the input current is halved as a result of the active load. This equation for T_{do} gives the differential performance of the pair over the complete frequency range in familiar terms.

The two-port matrix in eq. (12.4-16) is a two-port description of the differential amplifier. It is a complete description for applications in which the common mode rejection and common mode input current are unimportant. The 2×2 matrix may be used with the two-port analysis of Chapters 8 and 9 in these applications.

Common Mode Rejection Submatrix

The common mode rejection submatrix in Table 12.1 is nominally zero when the $ABCD$ parameters of the two transistors are equal. Therefore, we can obtain good common mode rejection over a wide frequency range by keeping the parameters closely matched. We have seen that A is minimized at low frequencies by making $V_{ce1} = V_{ce2}$. Hence the dc collector voltages should be equal, and the signal voltage at the output should be minimized to minimize ΔA and ΔC in Table 12.1.

To minimize the second term of A_{ds} in the Table 12.1, both factors can be minimized separately. The first factor, $B_1 - mB_2$, tends to be small even where m departs from unity. Ignoring base resistance, $B_1 = kT/qI_1$ and $B_2 = kT/qI_2$. But the current mirror controls the split of the current between the transistors of the pair, making $I_2 = mI_1$ so that $mB_2 = kT/qI_1$, equal to B_1 . This is true

only for zero differential signal current (or zero signal intensity), so that when signals are applied, the common mode rejection suffers.

To see the effect of signal-induced current unbalance, we assume that $m=1$ and write $B_1 - B_2$ as a function of the signal current intensity as discussed in Section 12.2:

$$\begin{aligned} B_1 - B_2 &= \frac{kT}{qI_c} \left(\frac{1}{1-\gamma} - \frac{1}{1+\gamma} \right) \\ &= -\frac{kT}{qI_c} \frac{2\gamma}{1-\gamma^2} \end{aligned} \quad (12.4-17)$$

Under signal conditions, therefore, the common mode rejection rapidly worsens as the signal intensity departs from zero. Under slewing conditions, where γ approaches unity, $|B_1 - B_2|$ becomes very large, and common mode signals easily pass through the amplifier unless the second factor of the right-hand term in A_{ds} in Table 12.1 is made small. The variation of $B_1 - B_2$ with signal intensity is shown in Fig. 12.11, as given by (12.4-17).

To minimize the second factor, $(C_1 + C_2 + Y_E)/d$, we note that C_1 and C_2 are negative admittances whereas Y_E is usually a positive admittance. We can

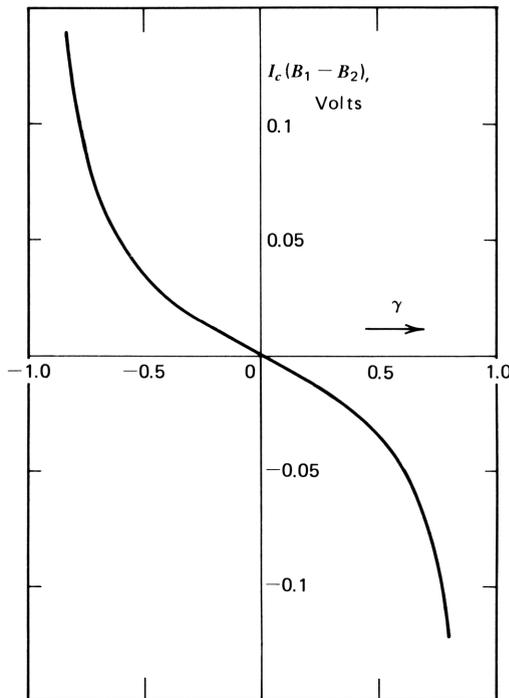


Figure 12.11. Difference between B of the two transistors as a function of signal intensity.

make these admittances roughly cancel to improve the CMRR. At low frequencies $C_1 = -\delta_1 g_{ce1}$ and $C_2 = -\delta_2 g_{ce2}$. By use of a Wilson or cascode current source, Y_E can be brought to the same vicinity as $-(C_1 + C_2)$. As seen from eq. (9.2-12) and the associated discussion, the output conductance of the Wilson current source is $2\delta g_{ce}$; g_{ce} for the current source is twice that for the individual transistors of the pair; thus the current source conductance is roughly twice too large to cancel the collector conductances of the pair.

Common Mode Rejection at High Frequencies

Common mode rejection tends to deteriorate at high frequencies because of collector capacitances of the two transistors and the output capacitance of the current source. Although we can find the common mode rejection parameters A_{ds} and C_{ds} from the expressions in Table 12.1, it is intuitively useful to see how these three capacitances affect common mode performance individually and in combination. Their individual effects are analyzed with the help of the three diagrams in Fig. 12.12.

When a common mode voltage v_s appears at the bases of the pair, it drives currents through the collector capacitances to the collectors. In Fig. 10.12a, only C_{jc1} is considered, with C_{jc2} and C_E (the emitter current source capacitance) ignored. The current through C_{jc1} is $C_{jc1}sv_s$, and it divides at the collector node. Ignoring base currents, we see that one part, $i_{c1}/(1+m)$, flows through the diode of the current source. The other part, $mi_{c1}/(1+m)$, flows down through the collector and into the emitter of the second transistor. This latter current supplies the current into the current source transistor; thus no output current change occurs (in accordance with the definitions of A_{ds} and C_{ds}). This current causes a differential input voltage to appear, given by

$$v_d \Big|_{i_{c1}} = \frac{mi_{c1}}{1+m} (r_{e1} + r_{e2}) \quad (12.4-18)$$

When C_{jc2} is added to the second transistor, the current also splits at the collector, except that in this case the current mirror current is into the transistor and is $mi_{c2}/(1+m)$. The rest of the current, $i_{c2}/(1+m)$, flows down through the second transistor and up through the first, causing a differential input voltage to appear, given by

$$v_d \Big|_{i_{c2}} = -\frac{i_{c2}}{1+m} (r_{e1} + r_{e2}) \quad (12.4-19)$$

Finally, we add C_E to the circuit, causing current $C_E sv_s$ to flow into the emitter source output admittance. This current divides according to the current mirror division, so the differential input voltage caused by C_E is given by

$$v_d \Big|_{i_{CE}} = \frac{i_{CE}}{1+m} (r_{e1} - mr_{e2}) \quad (12.4-20)$$

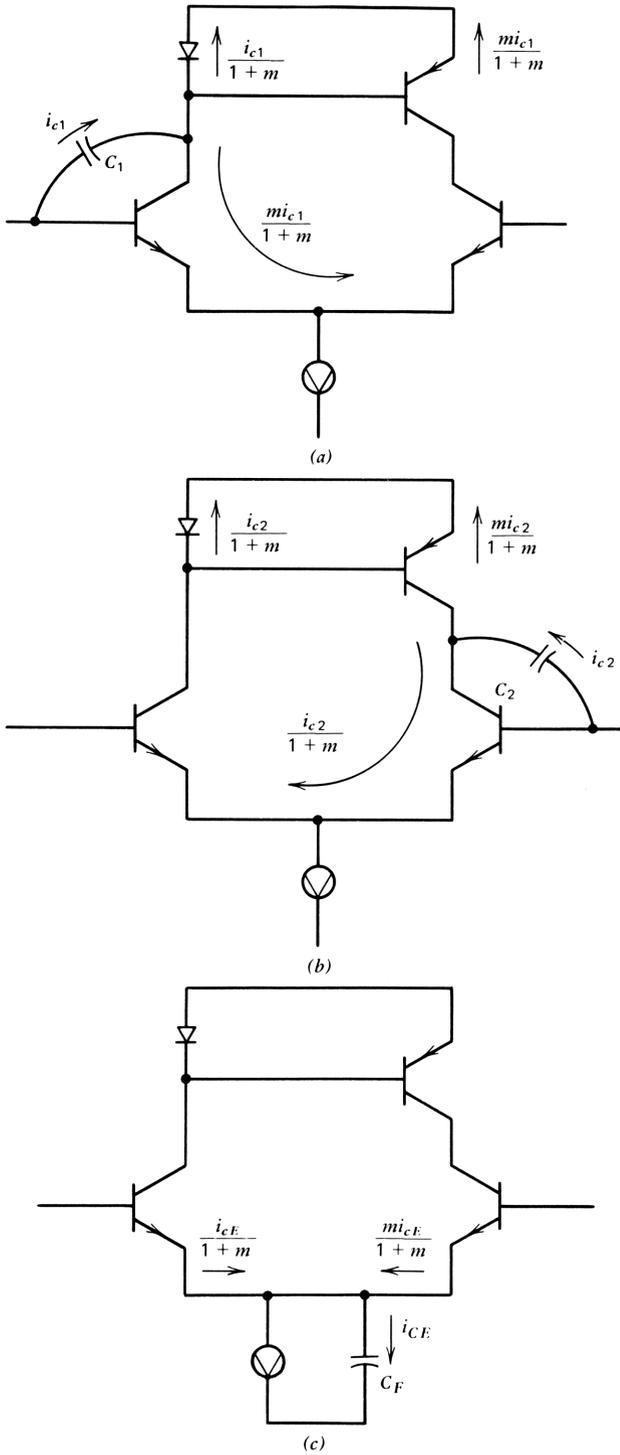


Figure 12.12. Analysis of the three contributors to CMRR at high frequencies in a bipolar differential amplifier.

We find the CMRR of (12.1.10) at high frequencies by adding these three components of input differential voltage and dividing by v_s . Thus we can write

$$A_{ds} = \frac{r_{e1} + r_{e2}}{1 + m} (mC_{jc1} - C_{jc2})s + \frac{C_{ES}}{1 + m} (r_{e1} - mr_{e2}) \quad (12.4-21)$$

The first term is minimized by making $mC_{jc1} = C_{jc2}$ as closely as possible and by minimizing the values of r_e ; at high frequencies, common mode rejection is aided by operating at reasonably high currents. (Note, however, that we have ignored r_b in this analysis.) The second term is minimized by operating at low signal intensity; thus the difference term is minimized. If we set $B = -r_e$ and $C = -C_{jc}s$ in eq. (12.4-21), the equation can be manipulated to give the value of A_{ds} in Table 12.1. A similar analysis gives C_{ds} by considering the differential input current caused by the three components of capacitive currents, and this value also corresponds to C_{ds} in Table 12.1.

Common Mode Input Current Parameters

The common mode current submatrix T_{so} , is given by

$$T_{so} = \begin{bmatrix} \frac{C_2}{1 - \bar{D}} & 0 \end{bmatrix} \quad (12.4-22)$$

and the common mode input admittance is

$$Y_{ss} = -2\bar{C} - \bar{D} \left(\frac{C_1 + C_2 + Y_E}{1 - \bar{D}} \right) \quad (12.4-23)$$

$$= -2\bar{C} - \frac{\bar{D} Y_E}{1 - \bar{D}} \quad (12.4-24)$$

$$= -2\bar{C} - \bar{D} Y_E \quad (12.4-25)$$

This is the parallel combination of the collector-to-base admittance of the two transistors and the current source admittance reduced by the factor \bar{D} . This completes the analysis of the null reference matrix. The exact values of the elements given in Table 12.1 are easily programmed, so that the effect of the approximations made here may be found.

12.5 OPERATIONAL AMPLIFIERS

An operational amplifier consists of the tandem combination of a differential stage or section and a two-port amplifier. It is characterized by the same set of independent and dependent signal variables as the differential amplifier; the output voltage and current and the common mode input voltage are taken as the independent signal variables. We can write the null reference matrix as the

product of the null reference matrix of the differential stage and the two-port matrix of the amplifier: the latter matrix is “built out” to a 3×3 matrix as follows:

$$\begin{aligned} \begin{bmatrix} v_d \\ i_d \\ i_s \end{bmatrix} &= \begin{bmatrix} A_{do} & B_{do} & | & A_{ds} \\ C_{do} & D_{do} & | & C_{ds} \\ \hline C_{so} & D_{so} & | & Y_{ss} \end{bmatrix} \begin{bmatrix} A_2 & B_2 & | & 0 \\ C_2 & D_2 & | & 0 \\ \hline 0 & 0 & | & 1 \end{bmatrix} \begin{bmatrix} v_o \\ i_o \\ v_s \end{bmatrix} \\ & \\ \begin{bmatrix} u_d \\ i_s \end{bmatrix} &= \begin{bmatrix} T_{do} & T_{ds} \\ T_{so} & Y_{ss} \end{bmatrix} \begin{bmatrix} T_2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_o \\ i_o \end{bmatrix} \\ &= \begin{bmatrix} T_d T_2 & T_{ds} \\ T_{so} T_2 & Y_{ss} \end{bmatrix} \begin{bmatrix} u_o \\ i_o \end{bmatrix} \end{aligned} \quad (12.5-1)$$

where T_2 is the $ABCD$ matrix of the two-port amplifier. The most important difference between the small-signal characteristics of the operational and differential amplifiers is that the differential submatrix of the differential amplifier is postmultiplied by the $ABCD$ matrix of the two-port. The common mode rejection properties and the common mode input admittance of the differential stage are retained in the operational amplifier.

In the 741 operational amplifier, for example, the $ABCD$ parameters of the output section are entirely dominated by C_2 , which is the negative of the feedback admittance of the frequency compensation capacitor. The voltage loss of the 741 is given by $B_{do}C_2$. As discussed in Chapter 5, there is considerable signal delay in the lateral pnp level shifter transistors; thus

$$B_{do(741)} \approx 2r_e e^{\tau_d s} \quad (12.5-4)$$

Therefore, the voltage loss of the 741 (the dominant loss) is given by $2r_e C_F S e^{\tau_d s}$. The term C_F is chosen so that unity is added to (12.5-4), the loss of the resulting unity gain follower is quadratic Butterworth. The common mode properties are those of the input differential stage.

The two-port used to convert a differential stage to an operational amplifier usually has other specific large-signal requirements. These are illustrated by the two circuits in Fig. 12.13. In Fig. 12.13a the output two-port is an emitter follower circuit. The amplifier output signal voltage therefore appears at the output of the differential stage, which means that if the output voltage becomes more negative than the input common mode voltage, the input transistors will saturate. This seriously restricts the allowable output voltage range of the amplifier. Put the other way around, for a given output voltage range, the input common mode voltage is restricted.

The limitation is removed in the circuit in Fig. 12.13b, in which a (pnp) common emitter stage replaces the emitter follower. This arrangement enables

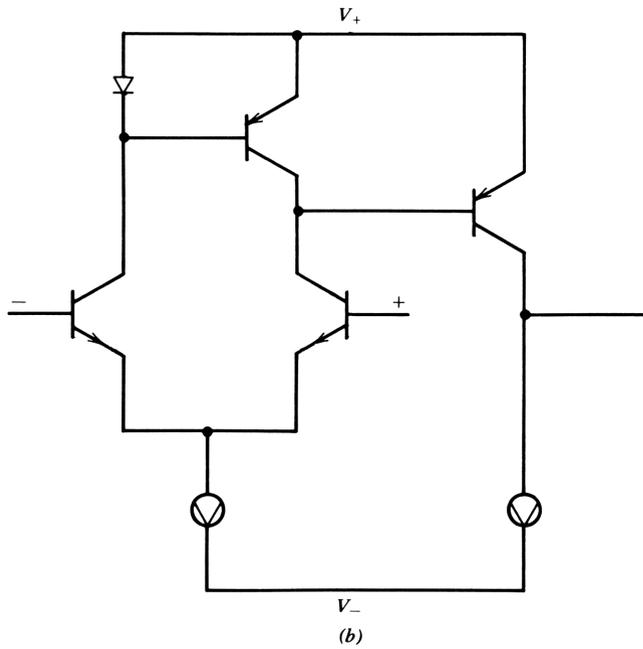
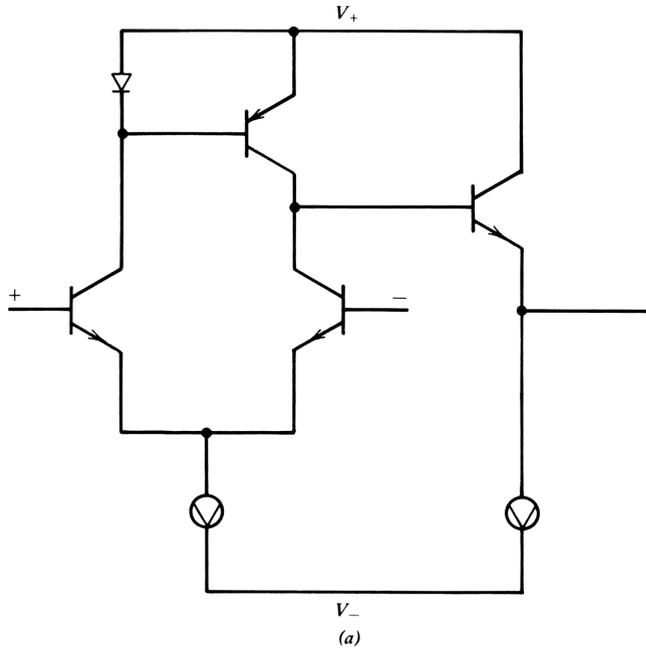


Figure 12.13. Connection of a two-port amplifier (a transistor) to the output of a differential stage to form an operational amplifier: two simple alternatives.

the output voltage to range over almost the entire span of the power supply voltages from the negative to the positive power supply voltage limited only by the small drops required by the transistor and the current source.

The key to this improved capability is that a collector junction is interposed between the differential stage output and the output of the amplifier, in this case the collector junction of the common emitter *pnp* transistor. This can be done only because of the availability of transistors of both conductivity types. It also affords isolation of the signal path from the power supplies, giving good power supply rejection.

The arrangement of Fig. 12.13*b* is that used in the 741 amplifier, except that the simple common emitter stage is replaced by a Darlington pair to give higher current gain. Also, in the 741 amplifier a Class B emitter follower output stage is added to give both current gain and higher output signal current capability. The extremely high current gain (or extremely small value of D) ensures that the input current of the two port is dominated by current through C_F .

A further consequence of use of the circuit in Fig. 12.13*b* rather than Fig. 12.13*a* is that the collector voltages of the differential pair remain virtually equal (0.7 V less than the positive supply) over the entire signal range of the amplifier. As noted previously, this is the condition required for good common mode rejection properties under conditions of high differential signal excitation.*

An All-*nnp* Operational Amplifier

Why even mention the circuit in Fig. 12.13*a* if the performance of that in Fig. 12.13*b* is superior in so many ways? The circuit in Fig. 12.13*a* is used where transistors of both conductivity types are not available in the same integrated circuit, or where the performance of one conductivity type is unequal to the performance requirements. In the 741, for example, the poor performance of the lateral *pnp* transistor restricts the amplifier to the audio range, and not even to the top of that range in critical applications, as we have seen in Section 4.3.

A broadband operational amplifier using only *nnp* transistors is shown in Fig. 12.14.³ The differential pair is conventional, and an emitter follower Q_G is used as the first stage of the two-port amplifier. A level shifter Q_F couples the emitter follower output to a Class A output stage such as that shown in Fig. 10.7 and as discussed in Section 10.1.

The output stage is arranged to have a voltage gain of 3, so that one-third of the output voltage appears at its input. With transistors that have an f_T of 4 GHz, the gain is constant with frequency to about 1 GHz. The level shifter (described in the following paragraphs) has roughly unity voltage loss as does

*Strictly speaking, common mode rejection is defined for a differential output signal of zero. It is desirable, however, for the amplifier to retain this rejection over the entire range of differential signal excitation.

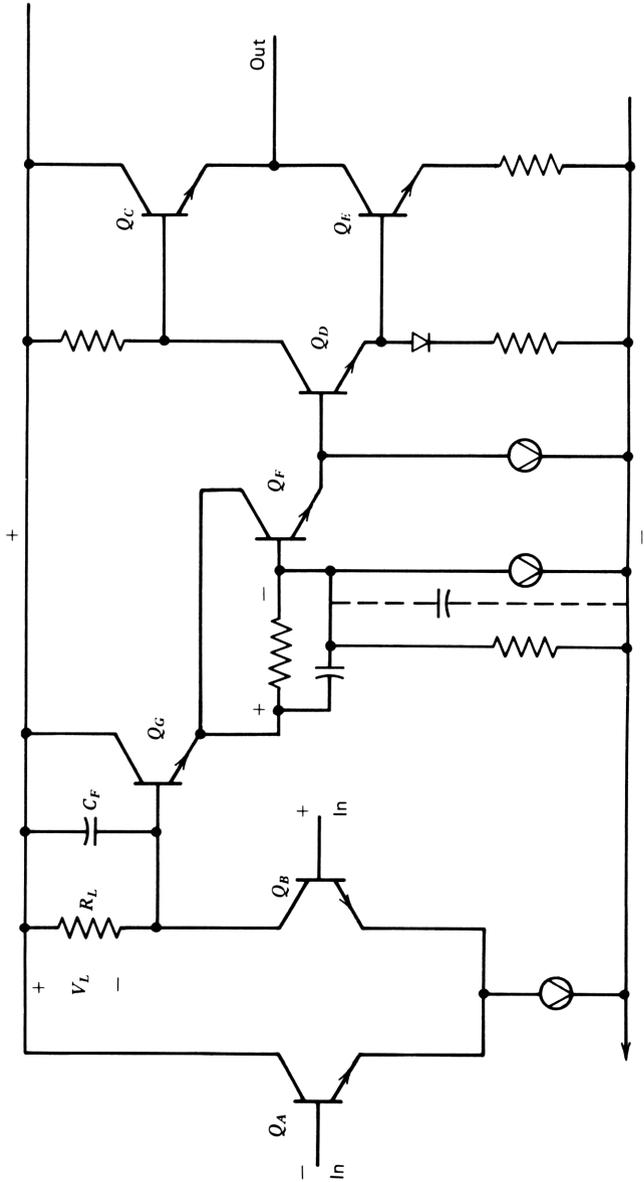


Figure 12.14. Broadband all-npn operational amplifier.

the emitter follower driving it. Thus the output voltage of the differential pair is also one-third of the output voltage of the amplifier. The linear frequency term is obtained by connecting a capacitor C_F from the base of the emitter follower to signal ground.

All subcircuits of the amplifier have been discussed previously except the level shifter. Figure 12.15 shows the level shifter as providing a resistive dc voltage drop connected to the base of Q_F . The collector of Q_F is connected to the input, giving unity D feedback. The feedback causes the response to be broadband. The capacitor around the series dc drop resistor compensates for shunt capacitance of the current source; the shunt conductance in parallel with the current source maintains flat response over the entire frequency range. The current source itself has a positive temperature coefficient to cancel the negative temperature coefficient of the base-emitter drops from the differential stage output through the base-emitter drops of Q_G , Q_F , Q_D , and Q_E to the

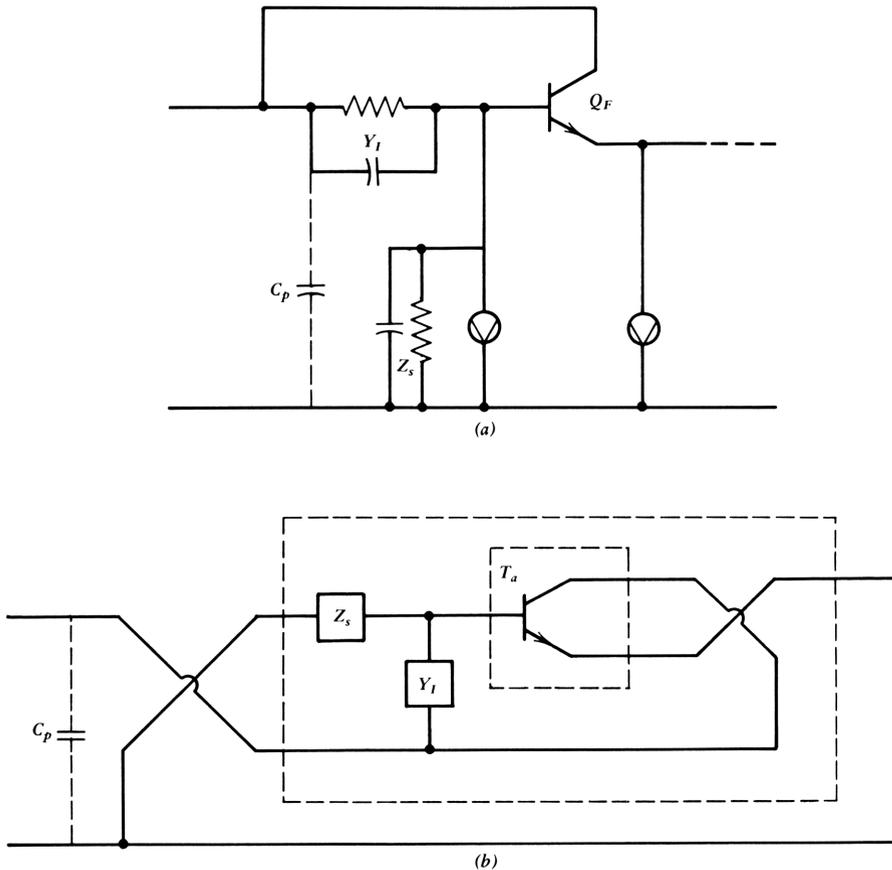


Figure 12.15. Level shifter analysis for all-npn operational amplifier.

negative power supply rail. The collector voltage of the differential pair output transistor is thereby held constant with temperature.

Level shifter response is readily analyzed by using the permutative feedback concept of Section 9.1. When drawn as in Fig. 12.15*b*, the *ABCD* parameters of the level shifter can be written by inspection as

$$T_{LS} = P_I [T_N * P_o [T_a]] \quad (12.5-5)$$

where T_N is the *ABCD* matrix of the voltage divider network (Z_s, Y_I) and T_a is that of the common emitter transistor. For a nominal voltage loss of 1.10, Z_s should be made one-ninth of $1/Y_I$. The parasitic capacitance C_p is the capacitance between the bottom electrode of C_1 and the substrate of the integrated circuit, typically one-fourth of C_1 . The *ABCD* matrix of its admittance premultiplies T_{LS} of eq. (12.5-5). The voltage loss of the level shifter is flat with frequency to the vicinity of one-third of f_T of the transistor. If the compensating resistance in parallel with the current source is omitted, a doublet of response is introduced with a magnitude of 1.1.

The feedback from the collector to the input of the level shifter effectively reduces the series impedance of the level shifter to the vicinity of 30 Ω .

The dc gain of the amplifier is limited by the differential-stage load resistor. Ignoring the loading of the emitter follower, the dc loss of the differential pair is $2r_e/R_L$. Since the collector current is V_L/R_L , the loss is $2kT/qV_L$. With $V_L = 3$ V, for example, the dc loss of the differential pair is $0.052/3 = 0.017$, or -35 dB; with the output stage loss of -9.5 dB (one-third) and a level shifter loss of $+1$ dB, the nominal loss of the amplifier is -43.5 dB. This loss is high for an operational amplifier and is a consequence of using transistors of only one conductivity type.

Operational Amplifier with Complementary Transistors

Availability of transistors of both conductivity types enhances design freedom. Where both types are similarly broadband, the bandwidth restriction of the lateral *pn*p transistor of the 741 amplifier is avoided. More care is required in the design, however, since all signal path transistors contribute to the rise of loss at high frequencies. Hence the loss polynomial requires more attention than the narrow-band 741.

Figure 12.16 shows a classic design for an operational amplifier of this type.⁴ All circuits have been discussed except for the two transistors, Q_6 and Q_8 , used to bias the output stage. The bias provided is equal to the base-emitter drops of the two transistors less the drop across the resistor connected between the collectors. The resistive drop reduces the output quiescent current relative to that of the driver. Although this circuit is bistable (i.e., a flip-flop) when connected to a low-impedance source, it is stable here, where it is fed from current source Q_9 .

The loss polynomial for this circuit is effectively a cubic plus delay if a capacitive load is to be accommodated. The loss of the amplifier (without

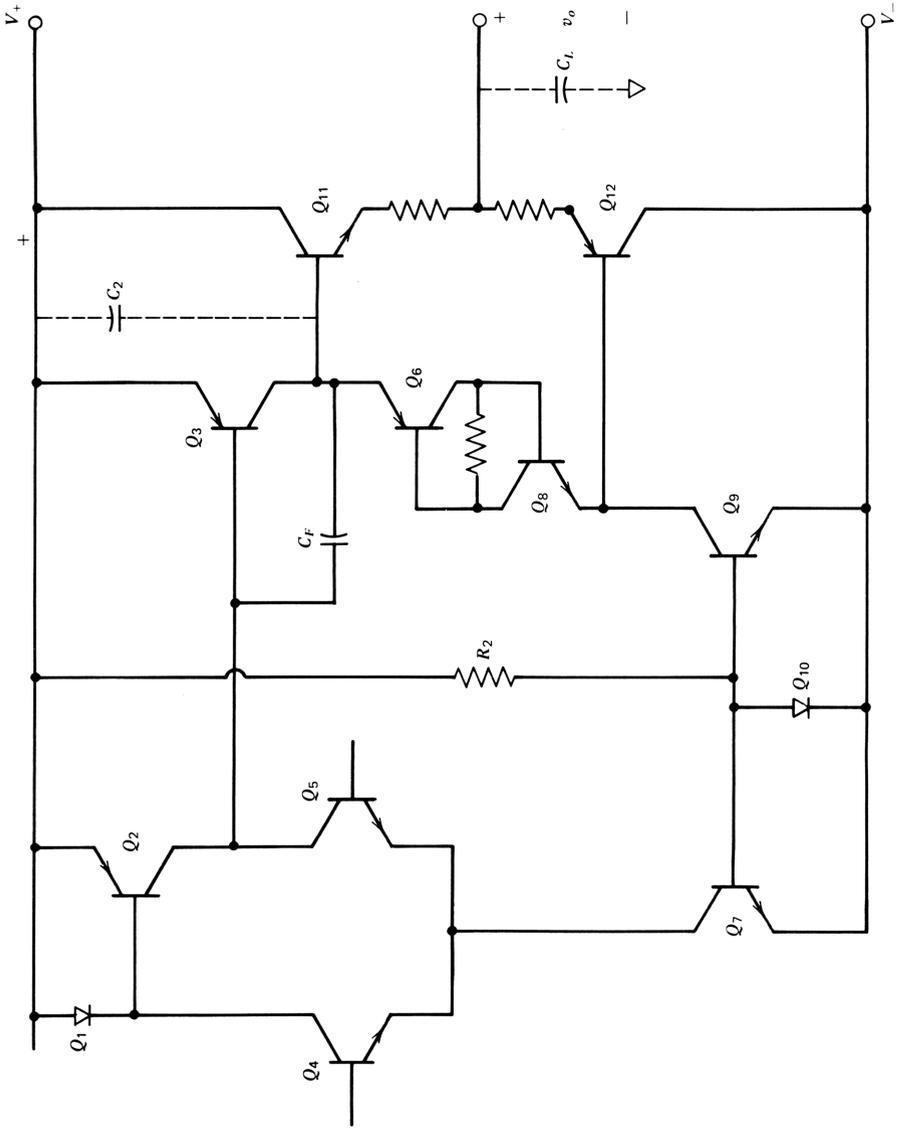


Figure 12.16. Operational amplifier using nnp and $pnnp$ transistors of similar design.

external feedback) is given by

$$L_a(s) = (a_0 + a_1s + a_2s^2 + a_3s^3) e^{\tau_d s} \quad (12.5-6)$$

In this equation the delay comprises the excess phase delay in one transistor of the input pair in Q_3 and effective delay introduced by base resistance of the input differential pair (r_{b4}/r_{e4}) τ_4 . (Because of the current mirror, which is broadband, this term is only half the total for r_{b4} and r_{b5} .) The dominant cubic term is $C_L\tau_{11}\tau_3r_{e4}$ and controls the bandwidth that can be achieved with capacitive loading. The a_0 term is $G_L\delta_{11}\delta_3r_{e4}$, usually negligible compared with the externally applied feedback. The middle two coefficients are the sums of terms; a_1 is controlled by the term $C_F r_{e3}$ as for the 741 amplifier. Quadratic term a_2 may be controlled by adding shunt capacitor C_2 to ground from the bases of the output stage. There is already considerable parasitic capacitance to substrate at this node, so that C_2 may not be needed. In this case the quadratic term (plus delay) controls the bandwidth that can be provided. Addition of C_2 makes the amplifier less sensitive to a variety of loads at the cost of bandwidth.

A specific loss polynomial (e.g., Butterworth) can be obtained for only one value of external feedback and overall loss. With external feedback, the polynomial is

$$L(s) = \beta_0 + (a_0 + a_1s + a_2s^2 + a_3s^3) e^{\tau_d s}$$

Each coefficient of the active path contains r_{e4} (or r_{e5}) as a factor, however, so that by varying the quiescent current of the differential pair, we can make the coefficients track the externally applied feedback. At least approximately, therefore, the desired polynomial can be realized over a wide range of feedback simply by adjusting the current supplied by the differential pair emitter current sink. The slew rate, given as $2I_{C4}/C_F$, will rise as the loss is reduced; the unity loss condition gives the most sluggish slew rate.

Where (as here) the delay is not excessive, the loss can be represented as a quartic with no delay as discussed in Chapter 5. The quartic can be selected to give sensitivity performance equivalent to polynomials of lower degree in band and to give a specified out-of-band margin against oscillation, as discussed in Chapter 5.

Operational Transconductance Amplifier⁵

Most operational amplifiers are characterized by a nonzero value of A in the null reference matrix, at least in the cutoff region. The operational transconductance amplifier (OTA), shown in Fig. 12.17, is characterized by its nonzero value of B , the reciprocal of its transconductance. The circuit is formed from a differential pair whose output is connected to a system of current mirrors that provide for combining current signals from the two sides of the pair. The current mirrors also provide downward level shift to the circuit output.

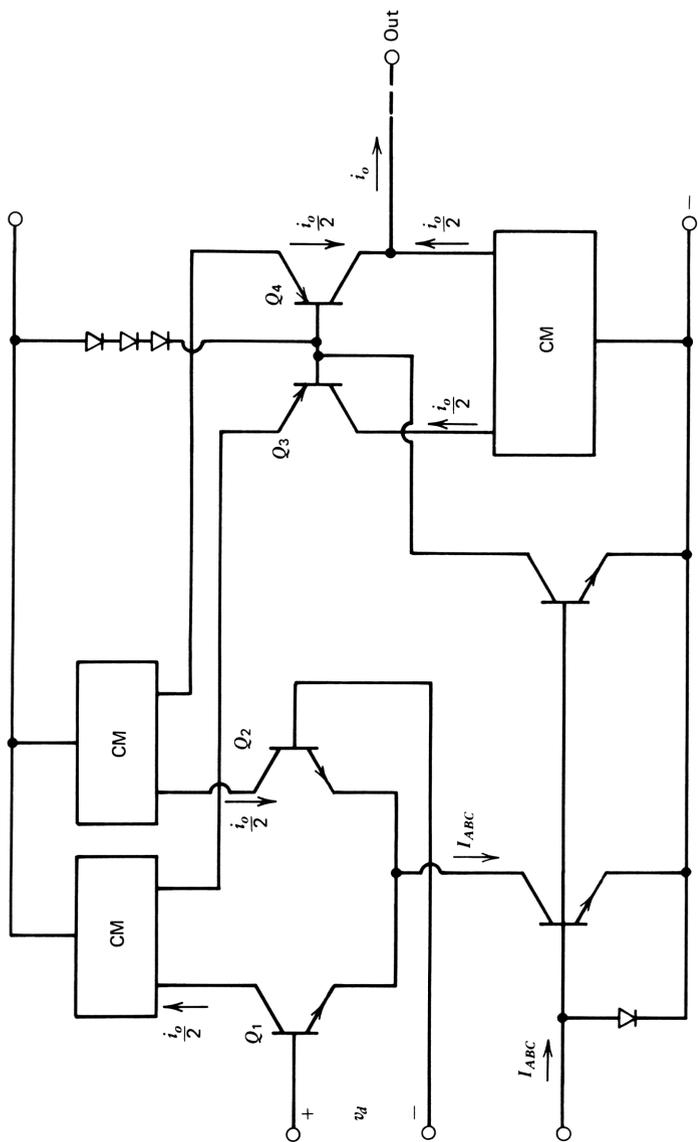


Figure 12.17. Operational transconductance (or B-type) amplifier.

Assuming all current mirrors to have a unity current ratio, the signal currents can be obtained in terms of the output current as shown in Fig. 12.17. The low-frequency null reference matrix for the amplifier can be derived as follows. The signal current in each transistor of the differential pair is $i_o/2$; the collector bias current is $I_{ABC}/2$, so we have

$$M_{NR} = \begin{bmatrix} 0 & \frac{2kT}{qI_{ABC}} \frac{1}{1-\gamma^2} & 0 \\ 0 & \frac{\delta}{2} & 0 \\ 0 & \delta & 0 \end{bmatrix} \quad (12.5-8)$$

High-frequency behavior is dominated by the output capacitance of the circuit, consisting of the collector capacitances of the output transistors and the collector-to-substrate capacitance of the lower current mirror. This adds the term $C_o B_{do}$ to A_{do} of the null reference matrix. Output capacitance serves the function of C_F of the 741 amplifier in cutting off the gain at 6 dB per octave. Where lateral *pnp* transistors are used in the upper current mirrors and the common base stages, the delay is large, as in the 741 amplifier. Because two *pnp* transistors are used effectively in common base tandem connection, the delay is twice that of the 741, so where the delay controls stability, the bandwidth is half that of the 741. With full-bandwidth *pnp* transistors, the circuit can be used for broadband applications.

Since B_{do} is inversely proportional to the differential stage bias current, the output current of the circuit is proportional to the product of the input voltage and the bias current. Thus the circuit can be used as a *signal multiplier* or a *varioloesser* in an AGC system. One signal is applied conventionally to the differential input, and the other is applied to the I_{ABC} bias lead.

The slew rate of this amplifier is limited by output capacitance connected to ensure stability. For a large differential input voltage, the amplifier output current is limited by I_{ABC} , the total bias current of the input pair. Therefore, the maximum rate at which the output capacitance can be charged sets the slew rate at I_{ABC}/C_o ; with $I_{ABC} = 20 \mu\text{A}$, for example, and $C_o = 30 \text{ pF}$. The slew rate is $0.02/30 = 0.67 \text{ V}/\mu\text{s}$, as in the 741 amplifier. Since the compensation capacitance is external to the amplifier circuit, it can be omitted from the chip and added to the external circuit. Therefore, it can be tailored to the application more easily than in the 741 amplifier. The maximum slew rate is then limited by parasitic capacitances, including collector-to-substrate and collector junction capacitances of the output transistors.

Application of an Operational Amplifier with Low Output Admittance

It is often said that an ideal operational amplifier should have zero output impedance. Here, we have an operational amplifier whose output *admittance* is ideally zero. Are there applications where this property are desirable? One

nonlinear application where low output admittance is desirable is in the signal rectifier shown in Fig. 12.18, which is useful in analog-to-digital conversion.⁶

In this circuit, when the upper diode is conducting (i.e., for positive output), the upper output voltage is clearly $i_{in} - R_{F1}$ and the lower output is zero, assuming the differential input voltage to be zero. When the lower diode is conducting, the upper output voltage is zero and the lower is $i_{in} - R_{F2}$. Since the conducting diode has a voltage drop of roughly 0.7 V, the amplifier output waveform in response to a ramp input must be as shown in Fig. 12.18. This voltage must undergo a rapid change at the transition where conduction shifts from one diode to the other. This transition is limited by the slew rate of the amplifier, which seriously limits the accuracy of the rectifier for high frequencies. If a 741 amplifier were to be used, for example, the change transition would take $1.4 \text{ V}/0.67 \text{ V}/\mu\text{s} = 2.1 \mu\text{s}$, the time taken for the compensation capacitance C_F to charge through the two diode drops.

However, if the OTA configuration is used, the compensation capacitance need not be connected across the output terminals directly since the amplifier operates without significant feedback in the transition region. The capacitance can be connected to ground at the circuit output terminals on the other side of the diodes, where the voltage does not change rapidly. Since the amplifier need not supply the large charging current, the voltage at the amplifier output can change much more rapidly than in the case of the 741 amplifier.

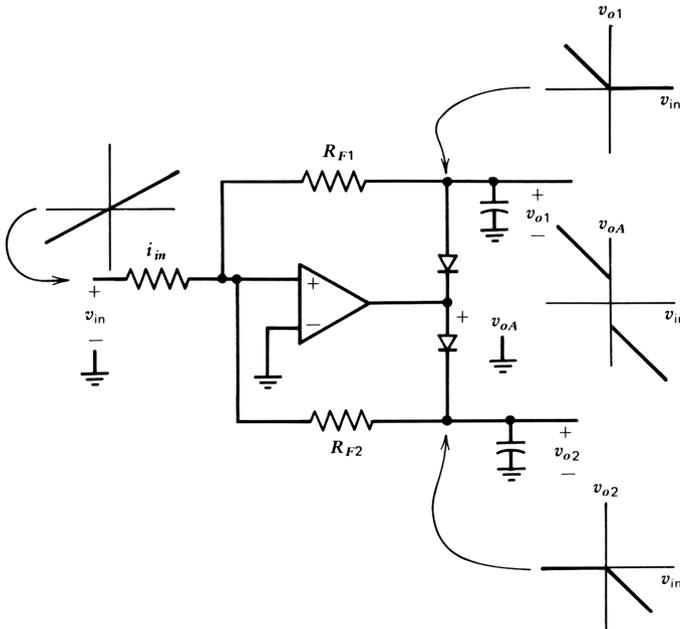


Figure 12.18. Signal rectifier circuit illustrating use of circuit shown in Fig. 12.17.

12.6 POWER SUPPLY ISOLATION

We have used a 3×3 matrix to represent the operational amplifier even though it has (usually) five external leads. The two extra leads, the positive and negative supply leads, carry the output signal current and can be considered part of the active circuit of the amplifier. This is illustrated in Fig. 12.19, which shows the output and power supply signal currents for an operational amplifier with a Class B output stage, such as the 741 under sine-wave excitation. Since the input lead currents are negligible, the sum of the two power supply signal currents must equal the output current, as shown. With both power supply leads connected firmly to signal ground, the power supply currents have no effect and can be ignored. In this usual case, the 3×3 matrix description is complete.

The power supply signal currents can perform a useful function, as shown in Fig. 12.20, in which the currents are used to drive a pair of output power transistors to obtain high slew rate.⁷ The output of the operational amplifier is connected to ground through a low-resistance R_0 to obtain drive current for

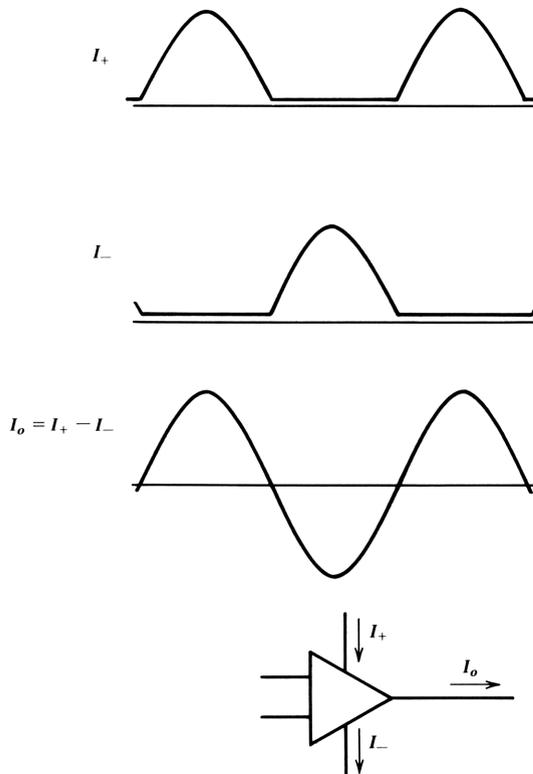


Figure 12.19. Power supply current waveforms for a 741 operational amplifier with a sine-wave output current.

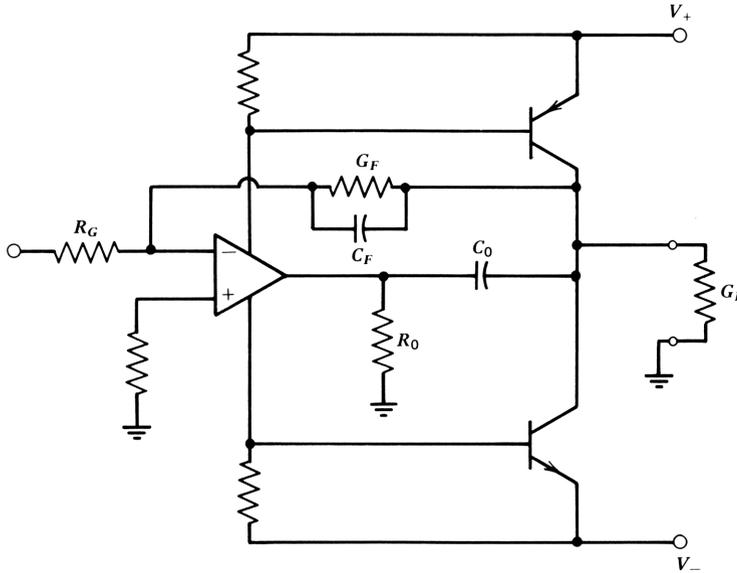


Figure 12.20. A circuit that uses power supply lead currents to drive a booster amplifier.⁷

the output transistors. Note that to improve the slew rate, the output voltage of the operational amplifier must be smaller than that of the output transistors. For the internal frequency compensation of the operational amplifier to work, the operational amplifier output is connected to the circuit output through a capacitor C_0 . Since this causes the operational amplifier output voltage to rise with frequency, the compensation capacitor now provides a *quadratic* rather than linear coefficient to the loss polynomial. The needed linear coefficient is provided by the external 10 pF capacitor connected from output to negative input. The constant coefficient of the loss polynomial is provided by the 51 k Ω feedback resistor; $R_G = 5.1$ k Ω , giving a loss of 0.1 for the current.

The approximate loss response of this circuit can be written by inspection, ignoring the loss of the output transistors, which are of broader bandwidth than the operational amplifier. Thus, ignoring $Y_F R_G$ relative to unity, we obtain

$$L(s) = R_G G_F + R_G C_F s + A_{do}(s) \frac{R_o C_o s}{1 + R_o C_o s} \tag{12.6-1}$$

in which $A_{do} \approx 0.15se^{0.08s}$ for the 741 amplifier. The equation can be used to proportion the coefficients of the loss polynomial.

Power Supply Rejection

Power supply rejection measures the ability of the amplifier to ignore signal voltages that (inadvertently) appear on the power supply leads. Such signals

may arise from other circuits fed from the same power supply, for example. As in noise analysis, the disturbance on the power supply is referred to the differential input. Therefore, it is similar in concept to common mode rejection and is ideally zero.*

We can expand the null reference matrix to include power supply rejection simply by adding power supply signal (noise) voltages as a pair of independent variables and power supply currents as the dependent variables, whereupon the null reference matrix becomes

$$\begin{bmatrix} v_d \\ i_d \\ i_s \\ i_p \\ i_n \end{bmatrix} = \begin{bmatrix} A_{do} & B_{do} & A_{ds} & A_{dp} & A_{dn} \\ C_{do} & D_{do} & C_{ds} & C_{dp} & C_{dn} \\ C_{so} & D_{so} & Y_{ss} & Y_{sp} & Y_{sn} \\ \hline C_{po} & D_{po} & Y_{ps} & Y_{pp} & Y_{pn} \\ C_{no} & D_{no} & Y_{ns} & Y_{np} & Y_{nn} \end{bmatrix} \begin{bmatrix} v_o \\ i_o \\ v_s \\ v_p \\ v_n \end{bmatrix} \quad (12.6-2)$$

in which subscripts p and n refer to the positive and negative supplies, respectively.

In this matrix the 3×3 submatrix in the upper left is the same as that in eq. (12.1-5). The power supply rejection is given for the positive and negative supplies by the column matrices at the upper right, $\{A_{dp}, C_{dp}\}$ and $\{A_{dn}, C_{dn}\}$, respectively. These submatrices give the differential input voltage and current that must be applied to cancel the effects of power supply ripple. The row matrix $|Y_{sp} Y_{sn}|$ gives the effect of power supply ripple on the input common mode current. The bottom two rows give the power supply signal currents in terms of the five independent variables and are seldom of interest, except in a circuit such as that in Fig. 12.20, where this current drives an output booster amplifier. For this case, the column matrix $\{D_{po}, D_{no}\}$ gives the signal current; the remaining terms represent spurious effects, usually small. The addition of the four extra port variables is seldom necessary, but the addition of the power supply voltages may be helpful where power supply rejection is to be investigated. Where this is the case, the upper 3×5 matrix is appropriate to present all the relevant information.

Where transistors of both conductivity types can be employed, power supply voltage rejection can be very high. The reason for this is that all internally used dc currents can be supplied through collector junctions of transistors, whose impedance is high. Such collector junctions effectively separate the circuit into isolated sections; the operational amplifier is separated into three such sections: the input section, the output section, and the power supply section. This is illustrated in the diagram in Fig. 12.21, in which dotted lines give the sectioning. The dotted lines cut the circuit only through collector junctions and

*Power supply rejection is usually expressed as a (large) number, the higher the better. As in the case of common mode rejection, we use the reciprocal formulation.

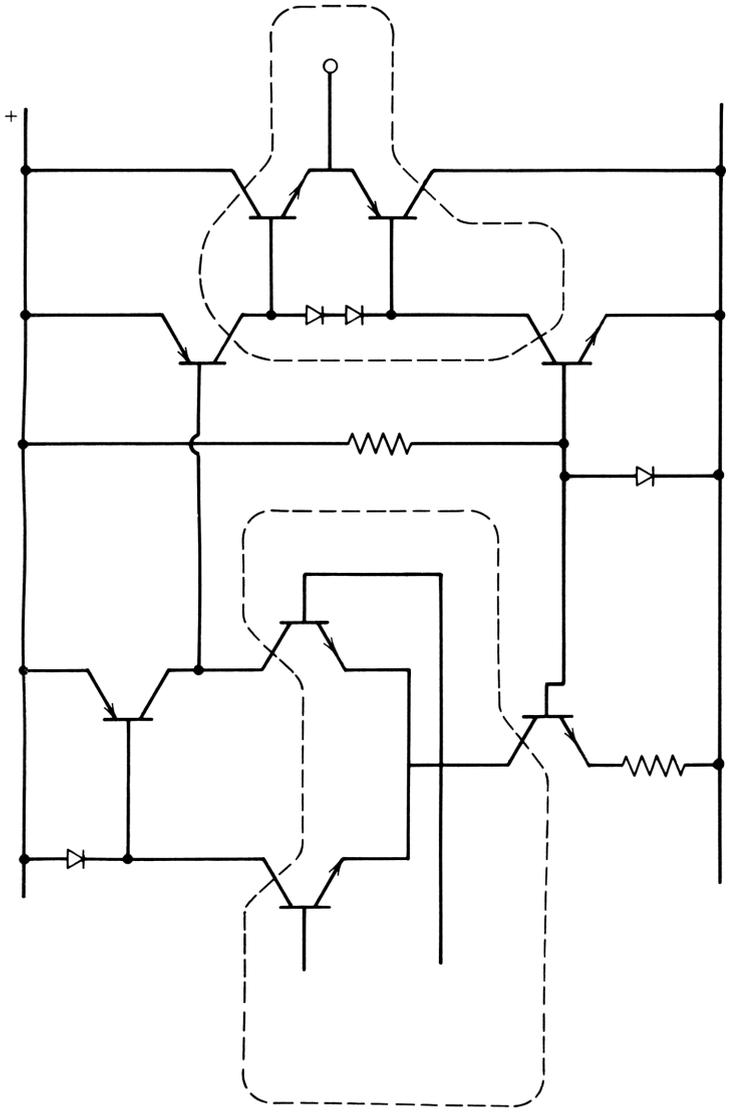


Figure 12.21. Operational amplifier showing separate sections isolated by collector junctions.

can be considered a collector junction cut-set. It is this double isolation that gives operational amplifiers such as the 741 their special properties of common mode rejection (input isolation) and power supply rejection (output isolation).

Conclusion

The concept of the null reference matrix can be extended to various multiport circuits. The key to the extension is the definition of the ideal case. The elements of the matrix then depict the various types of departure from the ideal case. The extension to multiple input–multiple output nonlinear systems, which is beyond our scope here, is treated by Desoer⁸ and requires more algebraic dexterity than needed here.

REFERENCES

- 1 R. D. Middlebrook, *Differential Amplifiers*, Wiley, New York, 1963.
- 2 L. J. Giacolletto, *Differential Amplifiers*, Wiley-Interscience, New York, 1970.
- 3 W. Kruppa and F. D. Waldhauer, "A UHF Monolithic Operational Amplifier," *ISSCC Digest* 21, IEEE Cat. No. 78CH 1298-9 SSC 1978, p. 74.
- 4 P. C. Davis, J. F. Graczyk, and W. A. Griffin, "Design of an Integrated Circuit for the TIC Low-Power Line Repeater," *IEEE Trans. COM-27* (2), 360–378 (February 1979).
- 5 C. F. Wheatley and H. A. Wittlinger, "OTA Obsoletes Op Amp," *Proceedings of the National Electronics Conference*, 1969.
- 6 F. D. Waldhauer, "Analog-to-Digital Converter," U.S. Patent 3,187,325, issued June 1965.
- 7 Fairchild Semiconductor, *Linear Integrated Circuits Data Book*, FAIRCHILD Semiconductor, 1976, p. 12–93.
- 8 C. A. Desoer and Y.-T. Wang, "Foundations of Feedback Theory for Nonlinear Dynamical Systems," *IEEE Trans. CAS-27* (2), 104 (February 1980).

Appendix A

Programs for Manipulating Polynomials

This first appendix contains 11 programs for investigating and manipulating polynomials. It allows us to represent polynomials in the frequency variable s in any of several forms: as polynomial coefficients, scaled in frequency and amplitude in any way we choose; as root positions on the complex s plane; or as Bode or Nyquist diagrams. If the information we have is expressed in any one of these three forms, it can be translated to either of the other forms. One program, "POLYTBL", generates polynomials suitable for various system applications—Butterworth, Bessel, Chebyshev, and others. Another, "SLC", finds the sensitivity of loss to each polynomial coefficient as a function of frequency.

Perhaps the biggest hurdle in using prepackaged programs is in getting started. The Hewlett-Packard HP 41C or HP 41CV calculator facilitates this process because of its alpha or word-use capability. To make repeated consultation of the appendixes unnecessary, the programs have been made interactive. The machine prompts the user for input data by name and tells the user how the data should be entered. Output data are identified. The next few paragraphs indicate what must be done to get started and how the input data are to be placed where needed.

HP 41C CONFIGURATION

For convenience, the same configuration of the calculator is used in all programs in this appendix. For the older HP 41C, two memory modules are required for the larger programs, and these are assumed to be present. For the newer HP 41CV, maximum user memory is contained in the machine. Many of the programs can be used together, and the larger capacity of the HP 41CV is convenient here. The HP 41C can be updated to the same capacity as the newer machine by obtaining a “quad memory module”; this is recommended for owners of the older machine. Total program memory required for all 11 programs in Appendix A is under 4.6 kbytes, or less than 37 kbits. Maximum data memory SIZE required is 36 data registers.

Several peripherals are available for this calculator, including a card reader (strongly recommended), a printer (needed in later appendixes), and a wand for reading bar code, an alternative to the card reader. In the later appendixes we find it convenient to record data, and the user may want to record programs of his or her own. Since the card reader is necessary for this and the wand functions only to read, not to write, the latter is not needed here.

The program is either keyed in from the program listing or is read into the machine from magnetic cards. It is run by pressing “XEQ filename”, where “filename” is the name of the program. The program must contain a label with a filename, such as LBL “N”.

For input data, the programs prompt the user by way of the display or the printer where available. The user keys in the requested data and presses the R/S key.

Data output will be displayed or will appear on the printed tape identified by suitable alpha characters in a form that should make its meaning clear. Data output can also be directed to magnetic cards by using the function WDTAX (write data under control of the number in the X register). If we key in 20.023 XEQ WDTAX, for example, the machine will prompt for a magnetic card. When this card is passed through the machine, the data in registers R20–R23 will be written into the first four data locations on the card. (There are 16 data locations on each side of a card.) Data obtained in this way can be used as input data to other programs. For further information, the operating manual should be consulted.

APPENDIX A PROGRAMS

Polynomial Scaling and Roots

A-1	“N”	Polynomial scaling in frequency and amplitude.
A-2	“PF”	Polynomial coefficients from linear and quadratic factors.
A-3	“Q”	Roots of quadratic equation.
A-4	“CU”	Roots of cubic equation.
A-5	“ROOTS”	Polynomial root extraction by Newton’s method.

Evaluation in Time and Frequency

A-6	“TIME”	Step response of systems with cubic loss ratios.
A-7	“BODE”	Evaluation of rational functions in s for $s = j\omega$; Bode and Nyquist diagrams.
A-8	“RCU”	Evaluation of cubic polynomial coefficients from loss and phase (reverse cubic evaluation).
A-9	“RQU”	Evaluation of quintic polynomial coefficients from loss and phase.

Polynomial Specifications

A-10	“POLYTBL”	Standard performance specification polynomials: Butterworth, Chebyshev, maximally flat delay, transitional.
A-11	“SLC”	Sensitivity magnitude of a rational function to its coefficients.

A-1 PROGRAM “N”: POLYNOMIAL SCALING (MAY USE “X-Y”)

This program scales a given polynomial in its dc value and its frequency as discussed in Section 2.2. In effect, it moves the Bode plot of the polynomial up or down and to the right or left. It can be used, for example, to scale a standard performance polynomial such as those given in Table 2.1 to a desired loss and cutoff frequency. (Such standard performance polynomials can be generated by using program “POLYTBL” in this appendix.)

To use the program, load it into the calculator from the keyboard or from a magnetic card. Then execute (XEQ) “N”, as shown in the first example (see Fig. A1.1). The program prompts for the degree of the polynomial (call it *polynomial A*). In the example the number 3 was pressed, followed by R/S. The polynomial coefficients stored in registers R00–R03 are then displayed in sequence or are printed if the printer is connected. If these are not the correct numbers, store the correct ones in ascending order in registers R00–R03 (for a quintic polynomial, this would be from R00 to R05). Press R/S, and the coefficients of polynomial A_A are displayed or printed. If they are correct, press R/S, and the machine prompts for b_0 , the dc coefficient of the desired, scaled polynomial B . In the example 0.1 was keyed in (a loss of 0.1, or -20 dB). The ratio of the cutoff frequencies of the two polynomials is then called for. Here, the user has a choice of responding by keying in the frequency ratio if it is known or by ignoring the prompt by pressing R/S, as was done in the example. In this latter case the frequency ratio is calculated from coefficient data to be supplied by the user, namely, the value of b_M , the value of the M th b coefficient. In the example we chose to make the quadratic coefficient b_2 equal to 0.1. The calculator prompts for M , to which we respond by pressing 2 and R/S. It then asks for b_M , to which we respond by pressing 0.1, R/S. The frequency ratio is then calculated and displayed or printed. Finally the coefficients of polynomial B are displayed or printed. As seen in the example, polynomial B is stored in registers R10–R13. In the second example this

		CLRG XEQ "N"			XEQ "N"	F0B/F0A?		
SCALE POLY			SCALE POLY			bM: M?		RUN
POLY A DEG?			POLY A DEG?				3	RUN
	3	RUN		3	RUN	bM?	1	RUN
R0= 0.000E0			R0= 1.000E0			F0B/F0A=6.844E-1		
R1= 0.000E0			R1= 2.000E0			R10= 1.000E0		
R2= 0.000E0			R2= 2.000E0			R11= 2.190E0		
R3= 0.000E0			R3= 1.000E0			R12= 2.056E0		
OK?			OK?			R13= 1.000E0		
	1	STO 00			RUN	R14= 2.193E-1		
		STO 03	POLY B			XEQ "X-Y"		
	2	STO 01	b0?		.1	XEQ "N"		
		STO 02	F0B/F0A?		.7	SCALE POLY		
		RUN				POLY A		
R0= 1.000E0			R10= 1.000E-1			DEG?	4	RUN
R1= 2.000E0			R11= 1.400E-1			R0= 1.000E0		
R2= 2.000E0			R12= 9.800E-2			R1= 2.190E0		
R3= 1.000E0			R13= 3.430E-2			R2= 2.056E0		
OK?		RUN				R3= 1.000E0		
POLY B					XEQ "N"	R4= 2.193E-1		
b0?	.1	RUN	SCALE POLY			OK?		
F0B/F0A?		RUN	POLY A			RUN		
bM:			DEG?		4	POLY B		
M?	2	RUN	R0= 1.000E0			b0?	.01	RUN
bM?	.1	RUN	R1= 3.200E0			F0B/F0A?	.4	RUN
			R2= 4.390E0			R10= 1.000E-2		
F0B/F0A=7.071E-1			R3= 3.120E0			R11= 8.760E-3		
			R4= 1.000E0			R12= 3.290E-3		
R10= 1.000E-1			OK?			R13= 6.400E-4		
R11= 1.414E-1					RUN	R14= 5.615E-5		
R12= 1.000E-1			POLY B					
R13= 3.536E-2			b0?		1			
					RUN			

Figure A1.1. Program "N": examples.

procedure is repeated except that the frequency ratio is given as 0.7. The resulting polynomial is then printed out directly.

In the third example (starting at the bottom of the second column) a quartic MFD polynomial from Table 2.2 is stored in registers R00–R04. It is then normalized to its cubic coefficient, to reduce the sensitivity of loss to its higher-degree coefficients; the cubic coefficient is reduced to 1.0 from 3.12, and the quartic coefficient is further reduced. We pay for this reduced sensitivity in bandwidth: as seen in the printout, $F_{0B}/F_{0A}=0.68$, so that 32% of the bandwidth is sacrificed. Next, we wish to scale the polynomial in amplitude and frequency to make it a loss polynomial for (say) a preamplifier with

```

01*LBL "N"
02 SF 21
03 SF 12
04 *SCALE POLY*
05 AVIEW
06 CF 12
07 ADV
08 *POLY A*
09 AVIEW
10 *DEG?*
11 PROMPT
12 | E-3
13 *
14 STO 09

15*LBL 03
16 RCL 09
17 XEQ 04
18 ADV
19 *OK?*
20 PROMPT
21 FS? 22
22 GTO 03
23 *POLY B*
24 AVIEW
25 *b0?*
26 PROMPT
27 STO 10
28 CF 22
29 *F0B/F0A?*
30 PROMPT
31 FS? 22
32 GTO 01
33 *bM:*
34 AVIEW
35 *M?*
36 PROMPT
37 STO 19
38 RCL 00
39 RCL IND 19
40 /
41 RCL 10
42 /
43 *bM?*
44 PROMPT
45 *
46 RCL 19
47 1/X

48 Y↑X
49 ADV
50 *F0B/F0A=*
51 ARCL X
52 AVIEW

53*LBL 01
54 STO 20
55 RCL 09
56 FRC
57 1
58 +
59 STO 09
60 11
61 STO 19

62*LBL 05
63 RCL 10
64 RCL 00
65 /
66 RCL IND 09
67 *
68 RCL 20
69 RCL 09
70 INT
71 Y↑X
72 *
73 STO IND 19
74 ISG 19
75 DEG
76 ISG 09
77 GTO 05
78 RCL 09
79 FRC
80 10.01
81 +
82 ADV
83 XEQ 04
84 TONE 9
85 RTN

86*LBL 04
87 FIX 0
88 CF 29
89 *R*
90 ENTER↑
91 X<>Y
92 RND

93 ARCL X
94 *t= "
95 X<>Y
96 SCI 3
97 SF 29
98 ARCL IND X
99 CF 22
100 AVIEW
101 FS? 22
102 STO IND Y
103 FS?C 22
104 X<>Y
105 ISG X
106 GTO 04
107 END

CAT 1

LBL"X-Y"
END
218 BYTES

LBL"X-Y"
.END.
34 BYTES

01*LBL "X-Y"
02 10
03 STO 19
04 .008
05 STO 09

06*LBL 07
07 RCL IND 09
08 X<> IND 19
09 STO IND 09
10 ISG 19
11 DEG
12 ISG 09
13 GTO 07
14 RTN
15 .END.

```

Figure A1.2. Programs "N" and "X-Y": listing.

—40 dB of loss and a bandwidth of 0.4 GHz. To do this, we copy the results of the first calculation into registers R00–R04 to begin a second scaling. The copying is done by use of a program "X–Y" that simply exchanges the contents of R00–R08 with those of R10–R18. (This data manipulation is useful in later programs as well; the program listing is shown after program "N"; see Fig. A1.2.)

The second scaling begins by setting $b_0=0.01$ (-40 dB loss) and $F_{0B}/F_{0A}=0.4$; the polynomial coefficients for the preamplifier are shown at the bottom of the printout.

Detailed Program Description

Since this is the first program, we describe it in more detail than later programs. Referring to the program listing, the first step is the program label. The second sets flag 21, the printer enable flag. When the printer is plugged in, flag 21 is automatically set, so the step is redundant; but without the printer, setting this flag stops the running program each time an AVIEW step is encountered, allowing the user to see the display at leisure. To start the program running again, poke the R/S key. Steps 03–07 title the program. Steps 08 and 09 indicate that polynomial A is to be displayed.

Steps 10 and 11 prompt the user to supply the degree of the polynomial to be scaled; steps 12–14 calculate and store (in R09) an index number used to control the display. If the degree is 3, for example, the index number is 0.003, indicating that registers R00–R03 are to be displayed. After label 03, the index number is recalled and a display subroutine is executed. (Where a printer is available, this subroutine can be replaced by the single command “PRREGX”, print the contents of the registers indicated by the number in the x register. This command is contained in a read-only memory (ROM) in the printer and does not exist in the calculator.) Without a printer, the register contents are displayed one at a time and are advanced in sequence by pressing R/S. When used without the printer, the subroutine also allows the contents of the displayed register to be changed or corrected one at a time. It is described later. Step 18 advances the printer tape if it is connected.

Steps 19 and 20 prompt the user with “OK?”. If the displayed polynomial is correct, press R/S. If not, store the correct number(s) in the indicated registers and press R/S. Flag 22 (at step 21) is set whenever keyboard input is encountered. (Flag 22 is cleared in subroutine 04.) If a correction was made, step 22 sends us back to label 03, and the display is repeated. When the numbers are correct, press R/S in response to “OK?” and arrive at steps 23 and 24, thus indicating that the characteristics of polynomial B are now to be dealt with. The AVIEW command stops the program when no printer is connected; press R/S to continue. The program then prompts for b_0 , the dc coefficient of the scaled polynomial, and stores it in R10.

Steps 29 and 30 prompt for the frequency scaling ratio F_{0A}/F_{0B} . Keying in a number in response to this prompt automatically sets flag 22 and sends us to label 01 (at step 53). If the ratio is not known, press R/S; steps 33–52 calculate the frequency scaling ratio from user-supplied values of two coefficients of the scaled polynomial b_0 and b_M , where M is the degree of the second coefficient to be used. The term M is prompted for as is b_M ; the frequency scaling ratio is obtained from the equation

$$\frac{F_{0B}}{F_{0A}} = \left(\frac{a_0 b_M}{a_M b_0} \right)^{1/M}$$

where M is stored in R19 (at step 37) and is used as an *index* to retrieve a_M (at step 39): a_M is divided into a_0 at step 40, and the result is divided by b_0 . Finally, b_M is prompted for: when supplied, it multiplies the previous result, giving the value in the brackets of the equation. The term M is then recalled, inverted, and raises the quantity in the brackets to $1/M$. The resulting frequency scaling ratio is stored in R20 at step 54.

Indirect storage indices are calculated for polynomials A and B in steps 55–60. The fractional part of the number in R09 is 0.003 for a third-degree polynomial; since we wish to calculate the b_i starting with b_1 , we add 1 and store 1.003 in R09. We want to store b_1 in R11, so that 11 is stored in R19, the index for polynomial B .

Steps 62–73 calculate the b_i according to the equation

$$b_i = \frac{a_i b_0}{a_0} \left(\frac{F_{0B}}{F_{0A}} \right)^i$$

and store the result in the register indicated by the number in R19. The next three steps increment the indices in R09 and R19 by one; "DEG" at step 75 is a "no operation" step that keeps the "ISG" (increment and skip if greater) from skipping step 76. When this step raises the number in R09 to 4.003 (for a cubic polynomial), it skips step 77. The next four steps develop the index for displaying the final resulting polynomial by subroutine 04. The tone signals that the computation is complete.

Subroutine 04

This subroutine can be used with or without a printer. Where a printer is available, it is shorter and takes less time to use "PRREGX", as noted previously. Without a printer, however, this subroutine has an interesting feature that allows the user to change the value of the number in the register being displayed.

With the number 10.013 in the x register, for example, this subroutine displays (or prints) the contents of R10–R13, as shown in the examples. By use of FIX 0 and CF 29 (clear flag 29), the number in x after rounding will appear as "10" without a decimal point. We use ARCL X to append the contents of X (10) to "R" and then append "=", giving "R10=". Before rounding, however, we copy the number in X into Y so as to have it available as a complete index (10.013). We then convert to SCI 3 format, set flag 29, and use ARCL IND X to append the number in register R10 into the alpha register for viewing by AVIEW. Before AVIEW, we clear flag 22, so that when AVIEW stops the program, we can change the contents of R10 should they need correction.

With no keyboard input, flag 22 remains clear and the following four steps are skipped; the number in X is incremented by one, and we go back to label 04 for the next register display. If there was keyboard input, however, the number in X is stored in the register indicated by the number in Y (because of the keyboard input, the number formerly in X is now in Y). The number in Y is then brought back to X, and operation continues as before.

Registers

Registers for program “N” are as follows:

R00–R08	Polynomial A
R09	Loop index for A
R10–R18	Polynomial B
R19	Loop index for B ; also M
R20	F_{0B}/F_{0A}

Registers for program “X–Y” are as follows:

R00–R08	Register set X
R09	Loop index for X
R10–R18	Register set Y
R19	Loop index for Y

A-2 PROGRAM “PF”: POLYNOMIAL COEFFICIENTS FROM LINEAR AND QUADRATIC FACTORS

This program multiplies linear or quadratic factors together to obtain the coefficients of the product polynomial. It can build a polynomial (of up to ninth degree) from individual factors. This is useful in feedback synthesis procedures.

The beginning polynomial is stored in data registers starting at R00; if it is desired to start from scratch to build a polynomial, store 1 in R00. The program multiplies the starting polynomial by a linear or quadratic factor. By a prompt, the program asks for the degree of the starting polynomial (to start from scratch, respond with 0). The program then asks for b_0 , b_1 , and b_2 . These values should be keyed in successively, each followed by pressing R/S. If the multiplier is to be a linear factor, zero should be keyed in for b_2 . The multiplication is then carried out, and the resulting polynomial coefficients are displayed or printed. To multiply by more factors, execute “B”, and key in the b_0 , b_1 , and b_2 coefficients as before. The resulting polynomial coefficients are in registers starting at R00 and are displayed or printed.

To continue building the polynomial, there is no need to go to the beginning of the program. Merely execute “B” to start the process at the point at which the new factors are keyed in.

In the example in Fig. A2.1, polynomials having roots of increasing multiplicity at $s = -1$ are generated.

Program Description

This program (see the listing in Fig. A2.2) is similar in several ways to program “N”, particularly the way in which the input and output data are handled. The same display subroutine is used (at label 04), and the initial polynomial coefficients are displayed and corrected in the same way. Following label “B”, the six registers above the highest-degree term register are cleared to make

```

        1 STO 00
        STO 01
        XEQ "PF"
BUILD POLY

POLY a
DEG?
        1 RUN
R00= 1.000E0
R01= 1.000E0

OK?
        RUN

POLY b:
b0, b1, b2
        1 RUN
        1 RUN
        0 RUN

R00= 1.000E0
R01= 2.000E0
R02= 1.000E0

        XEQ "P"
b0, b1, b2
        1 RUN
        2 RUN
        1 RUN

R00= 1.000E0
R01= 4.000E0
R02= 6.000E0
R03= 4.000E0
R04= 1.000E0

        XEQ "P"
b0, b1, b2
        1 RUN
        1 RUN
        0 RUN

R00= 1.000E0
R01= 5.000E0
R02= 1.000E1
R03= 1.000E1
R04= 5.000E0
R05= 1.000E0
    
```

Figure A2.1. Program "PF": examples.

room for the polynomial. This is done by locating the six statistics registers at the desired location with "ΣREG IND X" and clearing them. The program prompts for quadratic coefficients b_0 , b_1 , and b_2 . To multiply by a linear term, zero must be keyed in for b_2 . If it is zero, flag 09 is set. Steps 40–45 provide register R18 with an index equal to one or two greater than the degree of the initial polynomial, depending on whether the multiplier is a linear or a quadratic term. The actual multiplication is carried out at label 03 for both linear and quadratic terms; where the multiplier is a quadratic, subroutine 06 is called on.

```

01+LBL "PF"
02 SF 21
03 SF 12
04 "BUILD POLY"
05 AVIEW
06 CF 12
07 ADV
08 "POLY A"
09 AVIEW
10 "DEG?"
11 PROMPT
12 STO 17

13+LBL 01
14 RCL 17
15 1 E-3
16 *
17 XEQ 04
18 "OK?"
19 PROMPT
20 FS? 22
21 GTO 01
22 "POLY B"
23 AVIEW

24+LBL 0
25 RCL 17
26 1
27 +
28 ZREG IND X
29 CLΣ
30 "b0, b1, b2"
31 PROMPT
32 STO 10
33 STOP
34 STO 11
35 STOP
36 STO 12
37 CF 09
38 X=0?
39 SF 09
40 RCL 17
41 STO 18
42 1
43 ST+ 18
44 FC? 09
45 ST+ 18

46+LBL 03
47 RCL IND 18
48 RCL 10
49 *
50 DSE 18
51 DEG
52 RCL IND 18
53 RCL 11
54 *
55 +
56 FC? 09
57 XEQ 06
58 ISG 18
59 DEG
60 STO IND 18
61 DSE 18
62 GTO 03
63 2
64 FS? 09
65 1
66 ST+ 17
67 RCL 17
68 1 E-3
69 *
70 XEQ 04
71 TONE 9
72 RTN

73+LBL 06
74 DSE 18
75 DEG
76 RCL 18
77 X<0?
78 GTO 05
79 RDN
80 RCL IND 18
81 RCL 12
82 *
83 +
84 X<>Y

85+LBL 05
86 RDN
87 ISG 18
88 DEG
89 RTN

90+LBL 04
91 FIX 0
92 CF 29
93 "R0"
94 ENTER↑
95 X<>Y
96 RND
97 ARCL X
98 "I= "
99 X<>Y
100 SCI 3
101 SF 29
102 ARCL IND X
103 CF 22
104 AVIEW
105 FS? 22
106 STO IND Y
107 FS?C 22
108 X<>Y
109 ISG X
110 GTO 04
111 END

CAT 1

LBL*PF
LBL*P
END
223 BYTES

```

Figure A2.2. Program "PF": listing.

Program "PF": Registers

R00–R09	Polynomial
R10	b_0
R11	b_1
R12	b_2
R13–R16	Not used
R17	Degree of initial polynomial
R18	Loop index

A-3 PROGRAM "Q": QUADRATIC EQUATION

This program finds the roots of a quadratic equation. Load the program and execute "Q". The program will prompt for the coefficients in ascending order: a_0, a_1, a_2 . Enter these and press R/S. The roots are then printed. Listing and examples of program "Q" are given in Fig. A3.1.

```

01*LBL "Q"
02 "a0, a1, a2:"
03 PROMPT
04 STO 00
05 STOP
06 STO 01
07 STOP
08 STO 02

09*LBL 11
10 RCL 01
11 2
12 /
13 RCL 02
14 /
15 CHS
16 STO 03
17 X↑2
18 RCL 00
19 RCL 02
20 /
21 -
22 "ROOTS:"
23 AVIEW
24 CLA
25 X>0?
26 GTO 12
27 CHS
28 SQRT
29 ARCL 03
30 "+ -J"
31 ARCL X
32 AVIEW
33 ADY
34 RTN

35*LBL 12
36 SQRT
37 STO 04
38 RCL 03
39 +
40 ARCL X
41 "+, "
42 RCL 03
43 RCL 04
44 -
45 ARCL X
46 AVIEW
47 ADY
48 END

CAT 1
84 BYTES

XEQ "Q"
a0, a1, a2:
1.0000 RUN
2.0000 RUN
1.0000 RUN
ROOTS:
-1.0000 +-J0.0000

XEQ "Q"
a0, a1, a2:
1.0000 RUN
3.0000 RUN
1.0000 RUN
ROOTS:
-0.3820, -2.6180

XEQ "Q"
a0, a1, a2:
.0100 RUN
.1000 RUN
1.0000 RUN
ROOTS:
-0.0500 +-J0.0866
    
```

Figure A3.1. Program "Q": listing and examples.

The equation for which the roots are found is

$$a_0 + a_1x + a_2x^2 = 0$$

Program Description

The quadratic coefficients are stored in ascending order in response to a prompt in steps 02–08. The discriminant is then calculated in steps 10–21. At step 25 its sign is tested; if positive, we go to label 12 to form the two real roots. If negative, we continue to find the complex roots. In either case the roots are displayed, and if a printer is connected, are printed.

Program “Q”: Registers

R00	a_0
R01	a_1
R02	a_2
R03	$a_1/(2a_0)$
R04	Discriminant

A-4 PROGRAM “CU”: ROOTS OF CUBIC EQUATION

This program finds the roots of a cubic polynomial. On execution of “CU”, the program prompts for the cubic polynomial coefficients in ascending order: a_0 , a_1 , a_2 , and a_3 . After these are given, the program prints out the roots. Where there are complex roots, the locations are given in either rectangular form (clear flag 01) or in polar form (set flag 01). The equation programmed is

$$a_0 + a_1s + a_2s^2 + a_3s^3 = 0$$

Examples and listing of program “CU” are given in Figs. A4.1 and A4.2.

Program Description

This program is adapted from one written for the HP 45 (nonprogrammable) calculator by F. O. Simons, Jr. and R. C. Harden. Since the HP 45 had but one storage register, much use is made of the stack memory. While newer machines make this stack manipulation unnecessary, it is retained here as rather a curiosity.* The program is fast: finding the three roots takes about 15 seconds.

The first section of the program (up to step 54) calculates the discriminant of the cubic equation. If this is zero, two roots are real and equal, and these roots are calculated in subroutine B. If it is negative, the roots are real and separate

*For a complete description and explanation of such stack manipulations, see J. A. Ball's *Algorithms for RPN Calculators*, Wiley, 1978.

```

                                XEQ "CU"
a0, a1, a2, a3:
    1.0000  RUN
    3.0000  RUN
    3.0000  RUN
    1.0000  RUN
ROOTS:
-1.0000
-1.0000
-1.0000
                                SF 01
                                XEQ "CU"
a0, a1, a2, a3:
    .0100  RUN
    .2000  RUN
    2.0000  RUN
    10.0000  RUN
ROOTS:
-0.1000
CMPLX ROOTS:
0.1000+120.0000i

                                XEQ "CU"
a0, a1, a2, a3:
    1.0000  RUN
    2.4700  RUN
    2.4300  RUN
    1.0000  RUN
ROOTS:
-0.9305
CMPLX ROOTS:
1.0367+136.3219i

                                XEQ "CU"
a0, a1, a2, a3:
    1.0000  RUN
    1.0000  RUN
    1.0000  RUN
    1.0000  RUN
ROOTS:
-0.3820
-2.6180
-1.0000
                                SF 01
                                XEQ "CU"
a0, a1, a2, a3:
    1.0000  RUN
    1.0000  RUN
    1.0000  RUN
    1.0000  RUN
ROOTS:
-1.0000
CMPLX ROOTS:
1.0000+90.0000i

                                CF 01
                                XEQ "CU"
a0, a1, a2, a3:
    1.0000  RUN
    1.0000  RUN
    1.0000  RUN
    1.0000  RUN
ROOTS:
-1.0000
CMPLX ROOTS:
0.0000+11.0000i
    
```

Figure A4.1. Program "CU": examples.

and are calculated in subroutine C. If it is positive, two of the roots are complex, and the roots are calculated in subroutine A.

When there are two equal roots, round-off error may produce a small but nonzero discriminant. For this reason, the discriminant is rounded (at step 47) to direct the operation to subroutine B.

Program "CU": Registers

R01	Used
R02	Real root
R03, R04	Two other roots
R05	a_0/a_3
R06	a_1/a_3
R07	a_2/a_3

```

01*LBL "CU"
02 "a0, a1, a2, a3:"
03 PROMPT
04 STO 05
05 STOP
06 STO 06
07 STOP
08 STO 07
09 STOP
10 ST/ 05
11 ST/ 06
12 ST/ 07
13 RCL 07
14 3
15 /
16 STO 01
17 RCL 06
18 6
19 /
20 ENTER↑
21 ENTER↑
22 2
23 *
24 RCL 01
25 ENTER↑
26 *
27 -
28 ENTER↑
29 ENTER↑
30 RDN
31 RDN
32 +
33 RCL 01
34 *
35 RCL 05
36 2
37 /
38 -
39 RDN
40 *
41 *
42 RDN
43 *
44 X<>Y
45 RDN
46 +
47 "ROOTS:"
48 AVIEW
49 CLA
50 X=0?
51 GTO B
52 X<0?
53 GTO C
54 GTO A
55*LBL 30
56 SF 08
57 ARCL 02
58 XEQ 10
59 ARCL 03
60 XEQ 10
61 ARCL 04
62*LBL 10
63 AVIEW
64 PSE
65 CLA
66 RTN
67*LBL 31
68 ARCL 02
69 XEQ 10
70 RCL 04
71 RCL 03
72 FS? 01
73 R-P
74 ADV
75 "CMLX ROOTS:"
76 AVIEW
77 PSE
78 CLA
79 ARCL X
80 FC? 01
81 "I+J"
82 FS? 01
83 "IΔ"
84 ARCL Y
85 XEQ 10
86 RTN
87*LBL A
88 X<>Y
89 RDN
90 SQRT
91 ENTER↑
92 RDN
93 -
94 ENTER↑
95 RDN
96 RDN
97 +
98 RDN
99 *
100 X=0?
101 GTO 01
102 3
103 1/X
104 CHS
105 Y↑X
106 *
107 RDN
108 *
109 X=0?
110 GTO 02
111 3
112 1/X
113 CHS
114 Y↑X
115*LBL 03
116 *
117 ENTER↑
118 RDN

```

Figure A4.2. Program "CU": listing.

A-5 PROGRAM "ROOTS": POLYNOMIAL ROOT EXTRACTION BY NEWTON'S METHOD

This program finds the roots of polynomials up to twelfth degree, using Newton's method.* For accuracy, it finds the roots of the polynomial roughly in order of increasing size. A subroutine chooses a starting point that usually causes the smallest root to be found first. When the root is found, the

*The core of this program is due to R. K. Brush, in *High Level Math*, HP67/HP97 User's Library Solutions, Hewlett Packard, Corvallis, OR, May 1978.

```

119 +
120 RCL 01
121 -
122 STO 02
123 RCL 01
124 +
125 -2
126 /
127 RCL 01
128 -
129 STO 03
130 RDN
131 -
132 2
133 /
134 3
135 SQRT
136 *
137 STO 04
138 GTO 31

139*LBL 01
140 3
141 1/X
142 CHS
143 *
144 *
145 RDN
146 *
147 X=0?
148 GTO 02
149 3
150 1/X
151 CHS
152 Y↑X
153 GTO 03

154*LBL 02
155 3
156 1/X
157 CHS

158 *
159 GTO 03

160*LBL B
161 RDN
162 RDN
163 ENTER↑
164 *
165 3
166 1/X
167 SF 25
168 CHS
169 Y↑X
170 *
171 ENTER↑
172 ENTER↑
173 +
174 RCL 01
175 -
176 STO 02
177 RDN
178 RCL 01
179 +
180 CHS
181 STO 03
182 STO 04
183 GTO 30

184*LBL C
185 RDN
186 CHS
187 SQRT
188 ENTER↑
189 ENTER↑
190 3
191 1/X
192 Y↑X
193 2
194 *
195 RDN
196 /

197 ACOS
198 3
199 /
200 X<>Y
201 RDN
202 ENTER↑
203 COS
204 X<>Y
205 120
206 +
207 COS
208 RDN
209 *
210 RCL 01
211 -
212 STO 02
213 RDN
214 ENTER↑
215 RDN
216 *
217 RCL 01
218 -
219 ENTER↑
220 STO 03
221 RDN
222 RDN
223 ACOS
224 120
225 +
226 COS
227 *
228 RCL 01
229 -
230 STO 04
231 GTO 30
232 END

CAT 1
LBL↑CU
END
326 BYTES

```

Figure A4.2. Continued.

polynomial is deflated (the root is removed by synthetic division), and the process is repeated to find the next root. Where the root is complex, both roots are removed and the polynomial is reduced in degree by two. The process is usually automatic, continuing until all roots are found.

To use the program (with or without printer), load it into the machine, and store the polynomial whose roots are to be found in ascending order starting with a_0 in R20, a_1 in R21, and so forth up to any desired degree, limited only by the number of storage registers available. Clear flag 00 to suppress printout of the convergence process. Execute "ROOTS", and the machine will display "POLY ROOTS" and prompt for the degree of the polynomial. A high-pitched

tone accompanies any demand for data from the user. Key in the degree, and press R/S. The polynomial is then displayed and printed. Without a printer connected, the machine stops after displaying each coefficient; if a change is to be made, key in the new value and press R/S; if the displayed value is correct, just press R/S. At the end of the coefficient display, the user is prompted by "OK?". If so, press R/S; if not, make necessary changes and press R/S.

The rest of the process is automatic, and with printer connected, continues until all roots have been found and printed. Without printer, the machine stops after each root or root pair has been found to allow recording. In either case the finding of a root or a root pair is accompanied by a small fan-fare (BEEP). The coefficients of the deflated polynomial are then displayed. When the degree of the deflated polynomial reaches zero, the root finding is terminated.

After each iteration, two tones are sounded and the magnitude of $L(s)$ is displayed while the next iteration is being performed. If there is a problem with convergence, these $L(s)$ values can be observed and corrective action taken.

A tolerance of 10^{-8} on $|L(s)|$ is automatically stored in register R09 by the program. This may be changed when the program prompts for changes, at which time the polynomial can be checked and corrected if necessary. Flag 00 may be set to observe convergence in the event of difficulty. If flag 05 is set, the starting point for roots after the first will take the previously found root as the starting point: the first printed value of $L(s)$ (with flag 00 set) will be the residue for a previously calculated simple root. It will be related to the residues for complex roots: the magnitude of $L(s)$ is twice the magnitude of each residue, and the phase of $L(s)$ is the phase of the upper residue less 90° .

Program Description

The overall plan of the program is given in the flow chart in Fig. A5.1, containing five main sections labeled A through E. Section A initializes the calculator and allows a check of the input data. It also calls on Sections B, C, D, and E as subroutines and is the control program. It includes a display routine, subroutine 01, which was described in Appendix A-1 (as subroutine 04).

Section B calculates the starting point, described in the following paragraphs. Section C calculates a root or pair of complex roots of the polynomial and displays or prints the result. Section D deflates the original polynomial by synthetic division, reducing its degree by one or two. Roots of the deflated polynomial are then subjected to the same treatment until all roots are found. Section E ends the computation.

Root finding by Newton's method is discussed in Section 2.5 and in most texts on numerical analysis; for more detail, see References 1 and 2 in Chapter 2. As discussed in that chapter, we first estimate the location of the smallest root. Subroutine B does this. To find the starting point automatically, each coefficient a_i of the polynomial is divided into a_0 and the result raised to the $1/i$ power. The smallest of the resulting numbers establishes the starting

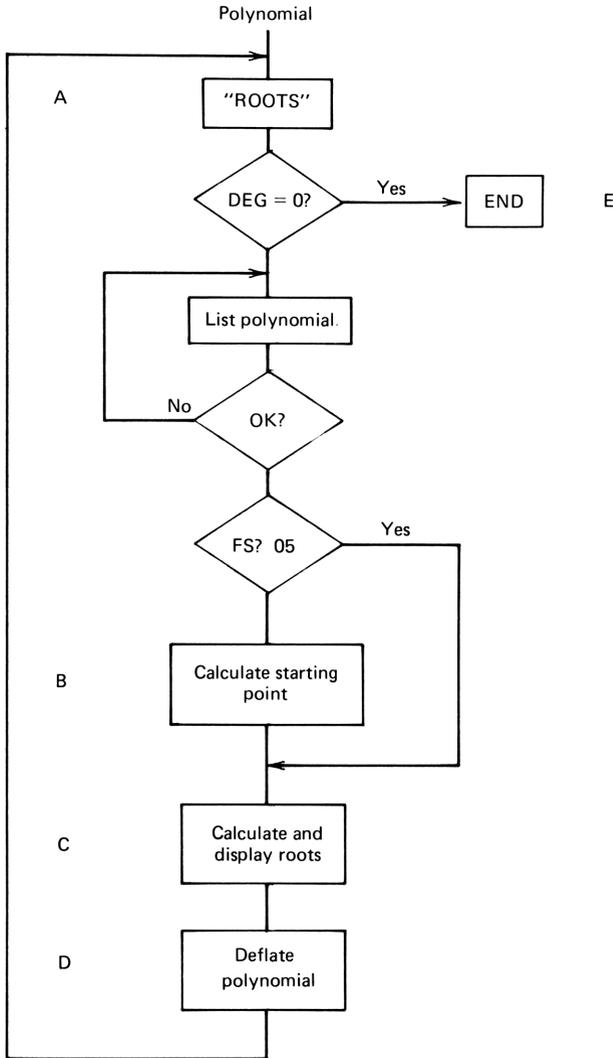


Figure A5.1. Flow chart for program "ROOTS".

point—the initial value of s —whose magnitude is taken as

$$|s| = \min_i \left(\frac{a_0}{a_i} \right)^{1/i}$$

The phase is taken as $180/i + \phi$ if a_0 and a_1 are of the same sign, and ϕ if they are of opposite sign, where ϕ is set (arbitrarily) at 7° in the program (at step 120) as discussed in the following paragraphs. The value of $L(s)$ is calculated

in subroutine C. If it is zero, the estimate was correct and the root is printed out. If the value of $L(s)$ is less than the allowable tolerance, the result is likewise printed out, except that an additional iteration is performed first. If the value of $L(s)$ is larger than the tolerance, a new value of s is calculated from

$$\Delta s = \frac{L(s)}{L'(s)}$$

which is subtracted from the previous value of s to obtain the new value.

The flow chart in Fig. A5.2 diagrams the operation of section C of the program. After setting the first iteration flag 02 (after label C), $|L(s)|$ is calculated (labels 03, 13, and 23) and several tests are made. If $|L(s)| = 0$, the C

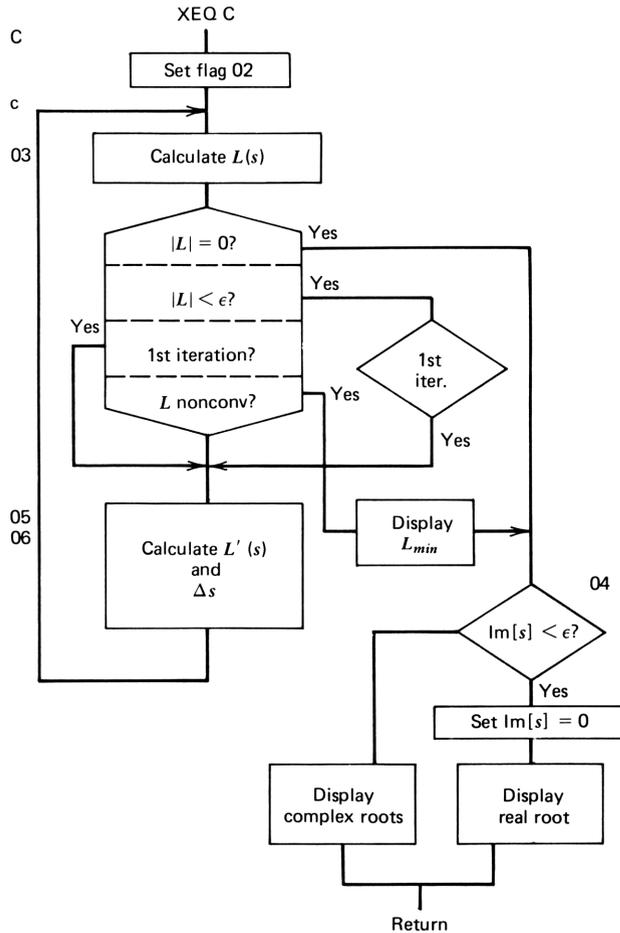


Figure A5.2. Flow chart for section C of program "ROOTS".

subroutine is exited and the root or roots are displayed by subroutine 04, which detects whether the root is real or complex. If $L(s)$ is less than tolerance ϵ , the subroutine is also exited, except that if this occurs on the first iteration, a second iteration is done. If $|L(s)|$ is larger than the previously calculated value, another iteration is performed with a smaller change in s (the change is halved). Sometimes the change in $L(s)$ is nil, and the iteration continues without improvement. To exit in this case, the current and previous values of $|L(s)|$ are compared. If they are the same, the value of $|L(s)| = L_{\min}$ is printed, and the value(s) of the root(s) is(are) printed. Often, L_{\min} will be satisfactorily small and the calculation can proceed to find the remaining roots. If it is not, as indicated by the display of L_{\min} , the calculation can be aborted and a new starting point can be chosen by changing the angle at step 120.

In most cases convergence will proceed normally below the decision box, where $L'(s)$ is calculated and Δs is found from $L(s)$ and $L'(s)$ (labels 05, 06,

<pre style="margin: 0;"> XEQ "ROOTS" ROOTS DEG? 12.0000 RUN POLY, ASC. ORDER R20= 1.000E0 R21= 0.000E0 R22= 0.000E0 R23= 0.000E0 R24= 0.000E0 R25= 0.000E0 R26= 0.000E0 R27= 0.000E0 R28= 0.000E0 R29= 0.000E0 R30= 0.000E0 R31= 0.000E0 R32= 1.000E0 OK? RUN S=1.0000+-415.00 S=1.0000+-4165.00 S=1.0000+-4105.00 S=1.0000+-445.00 S=1.0000+-475.00 S=1.0000+-4135.00 END </pre>	<pre style="margin: 0;"> XEQ "ROOTS" ROOTS DEG? 5.0000 RUN POLY, ASC. ORDER R20= 1.000E0 R21= 5.000E0 R22= 1.000E1 R23= 1.000E1 R24= 5.000E0 R25= 1.000E0 OK? RUN S=0.9772+-4-179.96 S=1.0376+-4180.00 S=-0.9704 END </pre>	<pre style="margin: 0;"> XEQ "ROOTS" ROOTS DEG? 3.0000 RUN POLY, ASC. ORDER R20= 1.000E0 R21= 2.000E0 R22= 2.000E0 R23= 1.000E0 OK? RUN L=3.79E-1 L=4.23E-2 L=1.79E-3 L=3.20E-6 L=7.14E-12 L=0.00E0 S=-1.0000 POLY, ASC. ORDER R20= 1.000E0 R21= 1.000E0 R22= 1.000E0 L=7.56E-1 L=1.27E-1 L=5.85E-3 L=1.14E-5 L=0.00E0 S=1.0000+-4120.00 END </pre>
---	---	---

Figure A5.3. Program "ROOTS": examples.

16, and 23). The iteration proceeds until exited by one of the means described in the previous paragraph.

Examples

Use of program "ROOTS" is illustrated in Figs. A5.3 and A5.4. The program listing is given in Fig. A5.5. For the first example, we find the roots of the equation

$$s^{12} + 1 = 0$$

which has 12 roots equally spaced about the origin on the unit circle. The root positions are symmetrical not only about the real axis, but about the imaginary axis as well. Because of these symmetries, if the first guess were to be either on the real or imaginary axes, convergence would be impossible since succeeding

```

                                XEQ "ROOTS"

ROOTS

DEG?                               7   RUN
POLY, ASC. ORDER
R20= 5.040E3
R21= 1.307E4
R22= 1.313E4
R23= 6.769E3
R24= 1.960E3
R25= 3.220E2
R26= 2.000E1
R27= 1.000E0

OK?                                  RUN

S=-1.0000

S=-2.0000

LMIN=1.0000E-6
LMIN=1.0000E-6
S=-3.0000

LMIN=8.0000E-7
LMIN=8.0000E-7
S=-4.0000

S=-5.0000

S=-6.0000

S=-7.0000

END

```

Figure A5.4. Program "ROOTS": examples (continued).

```

01*LBL "ROOTS"
02 SF 21
03 SF 10
04 ADV
05 SF 12
06 "ROOTS"
07 TONE 1
08 AVIEW
09 ADV
10 CF 12
11 CF 03
12 CF 02
13 CF 01
14 CF 04
15 TONE 9
16 "DEG?"
17 PROMPT
18 STO 00
19 1 E-8
20 STO 09

21*LBL A
22 RCL 00
23 X=0?
24 GTO E
25 20.02
26 STO 05
27 STO 06
28 FC? 00
29 CF 21
30 FS? 10
31 SF 21

32*LBL a
33 TONE 1
34 "POLY, ASC. "
35 "+ORDER"
36 AVIEW
37 RCL 00
38 1000
39 /
40 RCL 06
41 +
42 XEQ 01
43 ADV
44 TONE 9
45 "OK?"
46 CF 22
47 FS?C 10
48 PROMPT
49 FS? 22

50 GTO a
51 FC?C 05
52 XEQ B
53 XEQ C
54 XEQ D
55 GTO A
56 RTN

57*LBL 01
58 FIX 0
59 CF 29
60 "R"
61 ENTER↑
62 X<>Y
63 RND
64 ARCL X
65 "I= "
66 X<>Y
67 SCI 3
68 SF 29
69 ARCL IND X
70 CF 22
71 TONE 1
72 AVIEW
73 FS? 22
74 STO IND Y
75 FS?C 22
76 X<>Y
77 ISG X
78 GTO 01
79 RTN

80*LBL B
81 RCL 00
82 1 E3
83 /
84 RCL 06
85 +
86 1
87 +
88 STO 01
89 1
90 STO 02

91*LBL b
92 RCL IND 01
93 RCL 20
94 /
95 SIGN
96 STO 04
97 LASTX

98 ABS
99 RCL 01
100 RCL 06
101 -
102 INT
103 1/X
104 Y↑X
105 RCL 02
106 X<=Y?
107 XEQ 02
108 ISG 01
109 GTO b
110 RCL 02
111 1/X
112 STO 10
113 RCL 04
114 1
115 +
116 90
117 *
118 RCL 03
119 /
120 7
121 +
122 STO 11
123 RCL 10
124 P-R
125 STO 10
126 X<>Y
127 STO 11
128 RTN

129*LBL 02
130 X<>Y
131 STO 02
132 RCL 01
133 RCL 06
134 -
135 INT
136 STO 03
137 RTN

138*LBL C
139 SF 02

140*LBL c
141 CF 21
142 FS? 00
143 SF 21
144 XEQ 03
145 R-P

```

Figure A5.5. Program "ROOTS": listing.

146 CLA	196*LBL 03	243*LBL 24
147 TONE 5	197 RCL 00	244 BEEP
148 TONE 3	198 RCL 06	245 *S="
149 "L="	199 +	246 RCL 11
150 SCI 2	200 STO 05	247 RCL 10
151 ARCL X	201 0	248 R-P
152 AVIEW	202 ENTER↑	249 ARCL X
153 CLA	203 0	250 FIX 2
154 FIX 4		251 "++-Δ"
155 X=0?	204*LBL 13	252 ARCL Y
156 GTO 14	205 RCL IND 05	253 AVIEW
157 RCL 09	206 +	254 FIX 4
158 X>Y?	207 XEQ 23	255 ADV
159 GTO 04	208 DSE 05	256 RTN
160 RDN	209 GTO 13	
161 RCL 17	210 RCL IND 05	257*LBL 05
162 FS?C 02	211 +	258 RDN
163 GTO 05	212 RTN	259 STO 17
164 X>Y?		260 X<Y
165 GTO 05	213*LBL 04	261 STO 18
166 RDN	214 FS?C 02	262 XEQ 06
167 STO 03	215 GTO 05	263 R-P
168 X<Y	216 RCL 11	264 X=0?
169 STO 04	217 X=0?	265 GTO 15
170 RCL 16	218 GTO 14	266 1/X
171 RCL 15	219 ABS	267 RCL 17
172 R-P	220 RCL 09	268 *
173 X<Y	221 X<=Y?	269 CHS
174 RDN	222 GTO 14	270 X<Y
175 RCL 11	223 0	271 CHS
176 RCL 10	224 STO 11	272 RCL 18
177 R-P	225 GTO c	273 +
178 X<Y		274 X<Y
179 RDN	226*LBL 14	275 P-R
180 /	227 SF 21	276 STO 15
181 RCL 09	228 RCL 06	277 RCL 10
182 X>Y?	229 STO 05	278 +
183 GTO 15	230 RCL 11	279 STO 10
184 2	231 X=0?	280 X<Y
185 ST/ 15	232 DSE 05	281 STO 16
186 ST/ 16	233 X=0?	282 RCL 11
187 RCL 10	234 GTO 24	283 +
188 RCL 15	235 RCL 10	284 STO 11
189 -	236 BEEP	285 GTO c
190 STO 10	237 CLA	
191 RCL 11	238 *S="	286*LBL 15
192 RCL 16	239 ARCL X	287 TONE 8
193 -	240 AVIEW	288 TONE 9
194 STO 11	241 ADV	289 SF 21
195 GTO c	242 RTN	290 *LMIN="
		291 ARCL 03

Figure A5.5. Continued.

292 AVIEW	339 CF 01	389 X+2
293 ADV	340 RCL 11	390 RCL 11
294 FC? 00	341 X*0?	391 X+2
295 CF 21	342 SF 01	392 +
296 GTO 04	343 RCL 00	393 CHS
297 RTN	344 RCL 06	394 STO 14
	345 +	395 FS? 01
298*LBL 23	346 STO 05	396 GTO 07
299 STO 12	347 0	397 RCL 10
300 RCL 11	348 ISG 05	398 STO 12
301 *	349 DEG	399 0
302 X<>Y	350 STO IND 05	400 STO 14
303 STO 13	351 ISG 05	
304 RCL 10	352 DEG	401*LBL 07
305 *	353 STO IND 05	402 RCL 00
306 +	354 RCL 06	403 RCL 06
307 RCL 12	355 STO 05	404 +
308 RCL 10		405 STO 05
309 *	356*LBL d	
310 RCL 11	357 ISG 05	406*LBL 08
311 RCL 13	358 DEG	407 ISG 05
312 *	359 FS? 01	408 DEG
313 -	360 ISG 05	409 RCL IND 05
314 RTN	361 DEG	410 RCL 14
	362 RCL IND 05	411 *
315*LBL 06	363 DSE 05	412 DSE 05
316 RCL 00	364 DEG	413 RCL IND 05
317 RCL 06	365 FS? 01	414 RCL 12
318 +	366 DSE 05	415 *
319 STO 05	367 DEG	416 +
320 0	368 STO IND 05	417 DSE 05
321 ENTER↑	369 ISG 05	418 DEG
322 0	370 DEG	419 ST+ IND 05
	371 RCL 00	420 RCL 05
323*LBL 16	372 RCL 05	421 RCL 06
324 RCL 05	373 RCL 06	422 -
325 RCL 06	374 -	423 X)0?
326 -	375 X<=Y?	424 GTO 08
327 RCL IND 05	376 GTO d	425 RCL 00
328 *	377 1	426 RTN
329 +	378 FS? 01	
330 DSE 05	379 2	427*LBL E
331 XEQ 23	380 RCL 00	428 "END"
332 ISG 05	381 -	429 AVIEW
333 DEG	382 CHS	430 ADV
334 DSE 05	383 STO 00	431 END
335 GTO 16	384 RCL 10	
336 RTN	385 2	CAT 1
	386 *	
337*LBL D	387 STO 12	LBL*ROOTS
338 CF 04	388 RCL 10	END
		672 BYTES

Figure A5.5. Continued.

guesses could never leave the axis; roots on either side of the axis (real or imaginary) would pull equally; therefore, the initial phase offset (of 7°) is used. As shown, all 12 roots are found without difficulty. Running times tend to be long because many iterations are performed. It takes about 20 min to find all 12 roots and 5 min to find the first pair.

The second example finds five multiple roots at $s = -1$, a characteristically difficult case for accuracy with this method. The difficulty arises because all five roots contribute to driving $L(s)$ to zero, so that $L(s)$ becomes smaller than the tolerance even for values of s a small distance away from the actual root position. Accuracy can be improved somewhat by tightening the tolerance on $L(s)$. Running time for all five roots is about 10 minutes.

The third example illustrates the convergence process. With flag 00 set, the real root of a cubic MFA function is found. The process is then repeated to find the other two roots. The starting point is found as before, and $L(s)$ is found and displayed. Throughout the process $L(s)$ decreases in magnitude until it drops below the tolerance, after which one further iteration is performed. The root is then displayed.

A fourth example shows the solution for a seventh-degree polynomial:

$$\begin{aligned} L(s) &= (s+1)(s+2) \cdots (s+7) \\ &= s^7 + 28s^6 + 322s^5 + 1960s^4 + 6769s^3 \\ &\quad + 13,132s^2 + 13,068s + 5040 \end{aligned}$$

For a discussion of the factoring of this polynomial and the extreme sensitivity of the root positions to the exact values of the coefficients, see References 1 and 2 in Chapter 2. A change of the quadratic coefficient from 13,132 to 13,133, for example, changes six of the seven real roots to complex pairs!

Factoring of this polynomial shows how the program behaves when $L(s)$ becomes small but fails to converge. When finding the root at $s = -3.000$, $L(s)$ fails to go to zero and finds a minimum at 10^{-6} . (Note that compared with the dc loss, this is more than 200 dB down.) The value of s at this value of loss is negligibly different from the true root position. With the value of L_{\min} displayed, judgment can be exercised as to whether the value of s is a good estimate of the root position. In this case it clearly is. A check is available: by using program "PF", the root factors derived can be multiplied together and should give the original polynomial.

A-5 Program "ROOTS": Registers

R00	Degree
R01	Starting point loop index
R02	$(a_0/a_i)^{1/i}$
R03	$ L(s) $ from previous iteration
R04	$\angle L(s)$ from previous iteration
R05	Loop index

R06	Loop index beginning value
R07	Not used
R08	Not used
R09	Tolerance ϵ
R10	$\text{Re}[s]$
R11	$\text{Im}[s]$
R12	Used
R13	Used
R14	Used
R15	$\text{Re}[\Delta s]$
R16	$\text{Im}[\Delta s]$
R17	$ L(s) $
R18	$\angle L(s)$
R19	Used
R20-	Polynomial

A-6 PROGRAM "TIME": STEP RESPONSE OF CUBIC WITH COMPLEX ROOTS (USE WITH "CU")

This program calculates the step response for structures exhibiting cubic loss polynomials having a pair of complex roots. Its purpose is to illustrate the calculation of time response as discussed in Section 2.3. It is intended to be used with program "CU" which finds the roots of a cubic equation; alternatively, the real root can be stored in register R02, and the real and imaginary parts of either complex root can be stored in R03 and R04, respectively. To exercise the latter option, flag 00 should be set. Examples and listing of program "TIME" are given in Figs. A6.1 and A6.2.

The program calculates the residues by use of eq. (2.3-19); it normalizes the residues to a final value of unity. It then calculates the time response by use of eq. (2.3-13) for the real root and a root at $s=0$ for the step excitation; the complex root response is calculated from (2.3-18) and added. The program prompts for a time step; it then starts at $t=0$ and computes the response in equal increments of time. When a printer is used, the program must be stopped manually by pressing R/S. At each time point the time and the response are printed as shown in the examples.

The first example gives the (normalized) time response of the amplifier described in Section 2.1, stabilized with a 3 pF capacitor connected between input and output. The second example is the time response for a system with cubic Butterworth response. The third example is for a maximally flat delay response.

Without a printer, the program stops at every point where printing would occur as shown in the examples. To continue, press R/S. When it comes to the response list, the time is shown on the left side of the display, and the response is on the right side. To advance to the next time point, press R/S.

XEQ "TIME"			XEQ "TIME"			XEQ "TIME"		
STEP	RESP.		STEP	RESP.		STEP	RESP.	
a0, a1, a2, a3:			a0, a1, a2, a3:			a0, a1, a2, a3:		
	1.0000	RUN		1.0000	RUN		1.0000	RUN
	3.0000	RUN		2.0000	RUN		2.4700	RUN
	1.0000	RUN		2.0000	RUN		2.4300	RUN
	1.0000	RUN		1.0000	RUN		1.0000	RUN
ROOTS:			ROOTS:			ROOTS:		
-0.3611			-1.0000			-0.9305		
CMLX ROOTS:			CMLX ROOTS:			CMLX ROOTS:		
1.6641i101.0673			1.0000i120.0000			1.0367i136.3219		
K0			K0			K0		
1.0000			1.0000			1.0000		
K1			K1			K1		
-1.0376			-1.0000			-1.9712		
2KP \angle +90			2KP \angle +90			2KP \angle +90		
0.2252			1.1547			1.8246		
170.3937			-100.0000			147.8424		
TIME STEP?			TIME STEP?			TIME STEP?		
	1.0000	RUN		1.0000	RUN		1.0000	RUN
0.00 0.00000			0.00 0.00000			0.00 0.00000		
1.00 0.11417			1.00 0.09861			1.00 0.08980		
2.00 0.49097			2.00 0.44539			2.00 0.30193		
3.00 0.73517			3.00 0.81697			3.00 0.68667		
4.00 0.75012			4.00 1.03121			4.00 0.88467		
5.00 0.78431			5.00 1.00120			5.00 0.97592		
6.00 0.80792			6.00 1.04841			6.00 1.00392		
7.00 0.94036			7.00 1.00673			7.00 1.00634		
8.00 0.93658			8.00 0.98695			8.00 1.00292		
9.00 0.94003			9.00 0.98707			9.00 1.00038		
10.00 0.97603			10.00 0.99457			10.00 0.99951		
11.00 0.98628			11.00 1.00046			11.00 0.99952		
12.00 0.98371			12.00 1.00235			12.00 0.99975		
13.00 0.98768			13.00 1.00167			13.00 0.99992		
14.00 0.99505			14.00 1.00045			14.00 1.00000		
15.00 0.99673			15.00 0.99974			15.00 1.00001		

Figure A6.1. Program "TIME": examples.

Program "TIME": Registers

R00	—
R01	Used
R02	Real root
R03	Real part, complex roots
R04	Imaginary part, complex roots
R05	Time
R06	Time step
R07	k_0

```

01*LBL "TIME"
02 CF 08
03 SF 21
04 SF 12
05 "STEP RESP."
06 AVIEW
07 CF 12
08 ADV
09 FIX 4
10 FC? 00
11 XEQ "CU"
12 FS?C 08
13 GTO 11
14 XEQ "RES"

15*LBL "T"
16 ADV
17 TONE 9
18 "TIME STEP?"
19 PROMPT
20 STO 06
21 CLA
22 ADV
23 0
24 STO 05

25*LBL 10
26 RCL 05
27 RCL 04
28 *
29 R-D
30 RCL 10
31 +
32 SIN
33 RCL 05
34 RCL 03
35 *
36 E↑X
37 *
38 RCL 09
39 *
40 RCL 02
41 RCL 05
42 *
43 E↑X
44 RCL 08
45 *
46 +
47 RCL 07
48 +

49 STO 12
50 RCL 05
51 FIX 2
52 ARCL X
53 "+ "
54 RCL 12
55 FIX 5
56 RND
57 ARCL X
58 AVIEW
59 CLA
60 FIX 4
61 RCL 06
62 ST+ 05
63 GTO 10
64 RTN

65*LBL 11
66 CLX
67 1/X
68 RTN

69*LBL "RES"
70 "K0"
71 AVIEW
72 CLA
73 1
74 STO 07
75 ARCL 07
76 AVIEW
77 RCL 03
78 X↑2
79 RCL 04
80 X↑2
81 +
82 RCL 02
83 CHS
84 *
85 1/X
86 STO 11
87 ADV
88 "K1"
89 AVIEW
90 CLA
91 RCL 02
92 RCL 03
93 -
94 X↑2
95 RCL 04
96 X↑2

97 +
98 RCL 02
99 *
100 1/X
101 RCL 11
102 /
103 STO 08
104 ARCL 08
105 AVIEW
106 ADV
107 "2KP 4+90"
108 AVIEW
109 CLA
110 RCL 03
111 2
112 *
113 RCL 02
114 -
115 RCL 04
116 *
117 RCL 03
118 X↑2
119 RCL 04
120 X↑2
121 -
122 RCL 03
123 RCL 02
124 *
125 -
126 R-P
127 RCL 04
128 *
129 1/X
130 RCL 11
131 /
132 STO 09
133 X<>Y
134 CHS
135 STO 10
136 ARCL 09
137 AVIEW
138 CLA
139 ARCL 10
140 AVIEW
141 CLA
142 END

CAT 1

LBL"TIME
LBL"T
LBL"RES
END
223 BYTES

```

Figure A6.2. Program "TIME": listing.

R08	k_1
R09, 10	$2k_p$
R11	Used
R12	$f(t)$

A-7 PROGRAMS FOR EVALUATION OF RATIONAL FUNCTIONS

A series of three alternative programs and a subroutine used in common by each is described here. All three programs evaluate rational functions having numerator and denominator polynomials of up to seventh degree. They can be used for plotting Bode or Nyquist diagrams. The first program, "BODE", is intended primarily for use without a printer, whereas the second and third require a printer. The first program is called by two names, "BODER" and "BODEH", depending on whether the frequency is expressed in radians per second or in hertz, respectively. The second and third programs are identical except that one, "BORAD", works with the frequency expressed in radians per second (or gigradians per second), and the other, "BHZ" uses frequency in hertz (or gigahertz). These latter programs work the same as the first, adding the feature that the frequency is automatically incremented, either linearly or logarithmically.

Programs "BODER" and "BODEH" (Use with "FN")

Use of this program is illustrated in the first example, which shows the printout when the printer is connected (see Fig. A7.1). The left-justified printing is what one sees in the display, with or without printer. The right-justified printing is what the user keys in from the keyboard. To demonstrate how the program is used, we begin by clearing all registers (XEQ CLRG) and executing "BODER". The display prompts for the degree of the numerator. Suppose that we wish to find the response of a cubic maximally flat delay polynomial;

$$L(s) = 1 + 2.47s + 2.43s^2 + s^3$$

We then key in 3 for the degree and press R/S. The display shows R0=0.000E0, since we cleared all registers. The numerator polynomial is stored in registers R00-R07, so we key in 1 and press R/S. The display then shows R1=0.000E0; we key in 2.47 and R/S, and so forth. After keying in 1 for the cubic coefficient, the display prompts for the degree of the denominator polynomial; we key in 0, and the display prompts with R10=0.000E0. We key in 1, R/S, and the display prompts with "OK?". At this point, if we wish to correct a mistake, we can store the revised number in the indicated register, and the process of keying in the data will be repeated when R/S is pressed. In the absence of changes, the display will prompt with "2PIF= " * to which we

*We use "2PIF" or ω for $2\pi F$ to indicate angular frequency ω in these programs.

		CLRG		SF 01		XEQ "BODEH"
NUM DEG?	XEQ "BODER"		2PIF = 0.1000		NUM DEG?	
	3.0000	RUN	0.976 +J0.246		3.0000	RUN
R0= 0.000E0					R0= 1.000E0	RUN
R1= 0.000E0	1	RUN	2PIF = 0.2000		R1= 2.470E0	RUN
	2.47	RUN	0.903 +J0.486		R2= 2.430E0	RUN
R2= 0.000E0			2PIF = 0.3000		R3= 1.000E0	RUN
R3= 0.000E0	2.43	RUN	0.781 +J0.714			
	1	RUN	2PIF = 0.4000		DEN DEG?	
			0.611 +J0.924			0 RUN
DEN DEG?					R10= 1.000E0	RUN
R10= 0.000E0	0	RUN				
	1	RUN			OK?	
OK?					F = 0.0200	
2PIF =					0.09dB, 417.78	
	.2000	RUN			F = 0.0400	
2PIF = 0.2000					0.34dB, 435.55	
0.22dB, 428.29					F = 0.0600	
					0.79dB, 453.28	
2PIF =					F = 0.0800	
2PIF = 0.3000					1.43dB, 470.90	
0.49dB, 442.42					F = 0.1000	
					2.31dB, 488.21	
2PIF = 0.4000						
0.89dB, 456.52						
2PIF = 0.5000						
1.42dB, 470.53						
2PIF = 0.6000					F = 0.0200	
2.09dB, 484.35					0.962 +J0.308	SF 01
2PIF = 0.7000					F = 0.0400	
2.92dB, 497.83					0.847 +J0.605	

Figure A7.1. Programs "BODEH" and "BODER": examples.

respond by pressing .2. The display reads "2PIF=0.2000" until the response in dB and phase is displayed, in about 5 seconds.

With the printer connected, the prompt "2PIF= " appears in the display immediately following the printing of the response, prompting for the next frequency. Without a printer, the response remains in the display until R/S is pressed, at which point the frequency prompt appears in the display.

At this point in the example, the printer was switched to "Manual" operation to suppress the printing of prompts and user input in order to remove redundant printing. To plot a Nyquist diagram it is helpful to express the response in rectangular form; this can be done by setting flag 01, as shown in the second column. When the rational function has been stored, the input steps may be skipped by using the command XEQ "E".

Programs "BORAD" and "BHZ" (Use with "FN")

To try this program for the first time (see Figs. A7.2 and A7.3), clear all registers (XEQ CLRG) and XEQ BORAD. You will see the printed output shown in the first example. Suppose that we wish to find the loss of a cubic Butterworth polynomial. When the program prompts for NUM DEG?, press 3 and then R/S. The coefficients of the numerator (currently all zero, since we cleared the registers) are printed, with the registers in which they are contained

```

                                CLRG                SF 01
                                XEQ "BORAD"          XEQ "BORAD"
NUM DEG?                        OK?                NUM DEG?
                                .1000 STO 24
                                3.0000 STO 25
                                3.0000 STO 26
                                RUN                  3.0000 RUN

R00= 0.0000                      ΔFMIN,MAX        R00= 1.0000
R01= 0.0000                      R24= 0.1000    R01= 2.0000
R02= 0.0000                      R25= 3.0000    R02= 2.0000
R03= 0.0000                      R26= 3.0000    R03= 1.0000

DEN DEG?                          PTS/DEC:        DEN DEG?
                                0.0000 RUN          0.0000 RUN

R10= 0.0000                      R26= 3.0000
OK?                               OK?
                                1.0000 STO 00
                                STO 03
                                STO 10
                                2.0000 STO 01
                                STO 02
                                RUN
NUM DEG?                          ΔF           dB           PH           RUN
                                3.0000 RUN          0.100  0.00  11.48
                                R24= 0.1000
                                R25= 3.0000
                                0.215  0.00  24.89
                                0.464  0.04  55.51
                                1.000  3.01  135.00
                                2.154  20.04 -145.51

R00= 1.0000                      PTS/DEC:        R26= 3.0000
R01= 2.0000                      OK?
R02= 2.0000                      ΔF           RE           IM
R03= 1.0000                      0.100  0.980  0.199
                                0.215  0.907  0.421
                                0.464  0.569  0.828
                                1.000  -1.000  1.000
                                2.154  -0.283  -5.691

DEN DEG?                          RUN
                                0.0000 RUN

R10= 1.0000
OK?
                                ΔFMIN,MAX
                                R24= 0.0000
                                R25= 0.0000

PTS/DEC:
R26= 0.0000
    
```

Figure A7.2. Program "BORAD": Example 1.

(on the left). The program then prompts for denominator degree. Since we want the loss of a polynomial, with no denominator, press 0 and then R/S. The single denominator coefficient is in R10 and is also zero.

The next prompt is "OK?". Well, it certainly isn't. Not only is everything zero, but we're asked to divide by zero! So store 1 in R00 and R03 and 2 in R01 and R02 to place the cubic polynomial in the proper registers. Next, store 1 in R10, so that we don't divide by zero. Press R/S, and repeat the preceding steps, except that this time the numbers are in the registers.

CF 04			SF 04		
XEQ "BORAD"			XEQ "BORAD"		
NUM DEG?			NUM DEG?		
3.0000		RUN	3.0000		RUN
R00= 1.0000			R00= 1.0000		
R01= 2.4700			R01= 2.4700		
R02= 2.4300			R02= 2.4300		
R03= 1.0000			R03= 1.0000		
DEN DEG?			DEN DEG?		
3.0000		RUN	3.0000		RUN
R10= 1.0000			R10= 1.0000		
R11= -2.4700			R11= -2.4700		
R12= 2.4300			R12= 2.4300		
R13= -1.0000			R13= -1.0000		
OK?			OK?		
		RUN			RUN
ΔFMIN,MAX			ΔFMIN,MAX		
R24= 0.0250			R24= 0.0250		
R25= 1.5000			R25= 1.5000		
PTS/DEC:			PTS/DEC:		
R26= 6.6439			R26= 6.6439		
OK?			OK?		
		RUN			RUN
ΔF	dB	PH	ΔF	dB	PH
0.025	0.00	7.00	0.025	0.00	0.00
0.035	0.00	10.01	0.035	0.00	0.00
0.050	0.00	14.15	0.050	0.00	0.00
0.071	0.00	20.01	0.071	0.00	0.00
0.100	0.00	28.30	0.100	0.00	0.00
0.141	0.00	40.02	0.141	0.00	-0.01
0.200	0.00	56.59	0.200	0.00	-0.02
0.283	0.00	80.00	0.283	0.00	-0.05
0.400	0.00	113.03	0.400	0.00	-0.18
0.566	0.00	159.27	0.566	0.00	-0.84
0.800	0.00	221.54	0.800	0.00	-4.90
1.131	0.00	294.93	1.131	0.00	-25.30

Figure A7.3. Program "BORAD": Example 2.

When prompted by "OK?", press R/S and the program prompts for the minimum and maximum frequencies and the number of points per decade, all of which are currently zero. We then store an appropriate set of numbers in the indicated registers, and these are then listed as in the case of the coefficients. Finally, when the numbers are as we wish them, we press R/S and a list is printed giving the frequency (angular) and the magnitude (in decibels) and phase (in degrees). By setting flag 01, the response is printed in rectangular form, as shown in the example.

In this way any rational function having numerator and denominator up to seventh degree can be evaluated. Program "BHZ" (see Fig. A7.4) operates identically with the frequency expressed in hertz (or gigahertz). Only one of these two programs should be in the machine at one time; since they are virtually identical and contain the same global labels, there is the possibility of returning to the wrong program from subroutine "FN".

In a second example an all-pass function consisting of maximally flat delay polynomials for numerator and denominator is evaluated. Here, we use two

```

                                XEQ "BHZ"                                <F>
NUM DEG?                        3.0000  RUN                            R24= 0.0040
                                                                R25= 0.1751
R00= 1.0000
R01= 2.4700
R02= 2.4300
R03= 1.0000
                                                                P/D
                                                                R26= 6.6439
DEN DEG?                        3.0000  RUN                            OK?
                                                                F      dB      PH      RUN
R10= 1.0000
R11= -2.4700
R12= 2.4300
R13= -1.0000
OK?
                                                                0.004  0.00  7.00
                                                                0.006  0.00  10.01
                                                                0.008  0.00  14.15
                                                                0.011  0.00  20.01
                                                                0.016  0.00  28.30
                                                                0.023  0.00  40.02
                                                                0.032  0.00  56.59
                                                                0.045  0.00  80.00
                                                                0.064  0.00  113.03
                                                                0.090  0.00  159.27
                                                                0.127  0.00  221.54
                                                                RUN
<F>
R24= 0.0250
R25= 1.1000
P/D
R26= 6.6439
OK?
                                PI
                                2.0000  *
                                ST/ 24
                                ST/ 25
                                RUN

```

Figure A7.4. Program "BHZ": example.

points per octave (6.64 point per decade) to observe the linearity of the phase with frequency. Note that doubling the frequency doubles the phase with good precision up to an angular frequency of about 0.5.

This example is useful in demonstrating one other capability of the program — to remove linear phase. By setting flag 04, the linear phase contributed by the linear coefficients of both numerator and denominator is removed, with the results shown.

Program Description: “BODE”

Program “BODE” (see listings in Figs. A7.5) itself is a control program that calls on two subroutines, “FN” and “K” (see Fig. A7.6) to do most of the work. It also contains the display subroutine, “DX”, which is used when no printer is present. A flow chart for the program is shown in Fig. A7.7. It first calls on “FN” to get the desired function coefficients into the correct registers,

```

01*LBL "BODER"          32 FC? 02              65 FS? 03
   02 CF 02             33 GT0 11              66 XEQ "S"
   03 GT0 10            34 2                  67 ADV
                                35 *                  68 GT0 E
04*LBL "BODEH"          36 PI                  69 RTN
   05 SF 02             37 *
                                70*LBL "DX"
                                71 FIX 0
                                72 CF 29
                                73 "R"
                                74 ENTER↑
                                75 X<>Y
                                76 RND
                                77 ARCL X
                                78 "F= "
                                79 X<>Y
                                80 SCI 3
                                81 SF 29
                                82 ARCL IND X
                                83 CF 22
                                84 TONE 6
                                85 PROMPT
                                86 FS? 22
                                87 STO IND Y
                                88 FS?C 22
                                89 X<>Y
                                90 ISG X
                                91 GT0 "DX"
                                92 END
                                                CAT 1

06*LBL 10              38*LBL 11              70*LBL "DX"
   07 SF 21             39 STO 20              71 FIX 0
   08 CF 03             40 XEQ "K"            72 CF 29
   09 SF 09             41 FIX 2              73 "R"
                                42 FS? 01              74 ENTER↑
                                43 FIX 3                  75 X<>Y
                                44 RCL 23                  76 RND
                                45 RCL 22                  77 ARCL X
                                46 FS? 01                  78 "F= "
                                47 P-R                      79 X<>Y
                                48 FS? 01                  80 SCI 3
                                49 GT0 05                  81 SF 29
                                50 LOG                      82 ARCL IND X
                                51 20                      83 CF 22
                                52 *                        84 TONE 6
                                                85 PROMPT
                                                86 FS? 22
                                                87 STO IND Y
                                                88 FS?C 22
                                                89 X<>Y
                                                90 ISG X
                                                91 GT0 "DX"
                                                92 END
                                                CAT 1

10*LBL "V"            53*LBL 05              LBL "BODER"
   11 XEQ "FN"          54 RND                LBL "BODEH"
   12 "OK?"            55 ARCL X             LBL "V"
   13 TONE 6           56 FC? 01             LBL "DX"
   14 PROMPT           57 "FdB, Δ"          END
   15 FS? 22           58 FS? 01
   16 GT0 "V"          59 "F +J"
                                60 X<>Y
                                61 RND
                                62 ARCL X
                                63 TONE 0
                                64 AVIEW
                                END
                                205 BYTES

17*LBL E              60 X<>Y
   18 FIX 4             61 RND
   19 SF 21             62 ARCL X
   20 CLA               63 TONE 0
   21 FC? 02            64 AVIEW
   22 "2PI"            END
   23 "F= "            205 BYTES
   24 TONE 9
   25 PROMPT
   26 ARCL X
   27 FC? 55
   28 CF 21
   29 AVIEW
   30 SF 21
   31 CLA

```

Figure A7.5. Programs “BODEH” and “BODER”: listing.

01*LBL "FN"	43 CHS	88 RCL 11
02 0	44 STO 21	89 +
03 "NUM "	45 RCL 07	90 RCL 20
04 XEQ 10	46 *	91 *
05 RCL 18	47 RCL 05	92 RCL 16
06 STO 08	48 +	93 RCL 21
07 ADV	49 RCL 21	94 *
08 10.01	50 *	95 RCL 14
09 SF 08	51 RCL 03	96 +
10 "DEN "	52 +	97 RCL 21
11 XEQ 10	53 RCL 21	98 *
12 CF 22	54 *	99 RCL 12
13 RTN	55 RCL 01	100 +
	56 +	101 RCL 21
14*LBL 10	57 RCL 20	102 *
15 "+DEG?"	58 *	103 RCL 10
16 TONE 9	59 RCL 06	104 +
17 PROMPT	60 RCL 21	105 R-P
18 STO 18	61 *	106 ST/ 22
19 1 E3	62 RCL 04	107 STO 30
20 /	63 +	108 X<>Y
21 +	64 RCL 21	109 ST- 23
22 FC? 55	65 *	110 FS? 04
23 XEQ "DX"	66 RCL 02	111 XEQ 04
24 FS? 55	67 +	112 RTN
25 PRREGX	68 RCL 21	
26 RCL 18	69 *	113*LBL 04
27 1.007	70 RCL 00	114 RCL 23
28 +	71 +	115 RCL 01
29 10.01	72 R-P	116 RCL 00
30 X<>Y	73 STO 22	117 /
31 FS?C 08	74 STO 31	118 RCL 11
32 +	75 X<>Y	119 RCL 10
33 STO 09	76 STO 23	120 /
	77 RCL 17	121 -
34*LBL 03	78 RCL 21	122 RCL 20
35 CLX	79 *	123 *
36 X<> IND 09	80 RCL 15	124 R-D
37 ISG 09	81 +	125 -
38 GTO 03	82 RCL 21	126 RCL 22
39 RTN	83 *	127 P-R
	84 RCL 13	128 R-P
40*LBL "K"	85 +	129 X<>Y
41 RCL 20	86 RCL 21	130 STO 23
42 X12	87 *	131 END
		CAT 1
		219 BYTES

Figure A7.6. Subroutines "FN" and "K": listing.

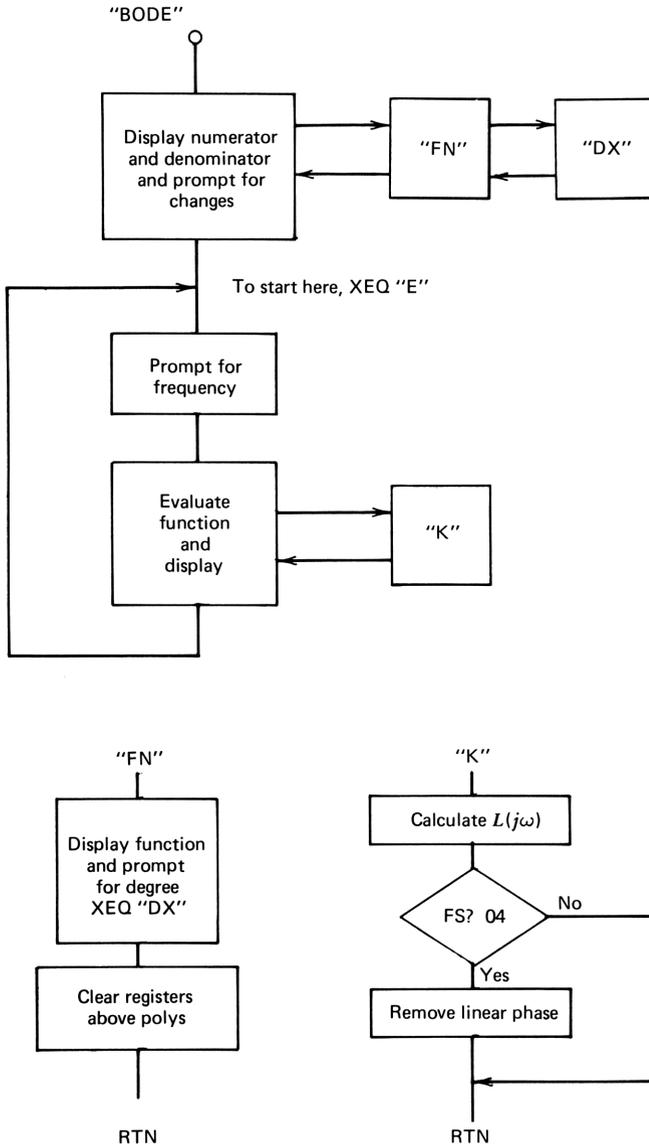


Figure A7.7. Flow charts for program "BODE" and subroutines "FN" and "K".

and it prompts for changes. At label "E" it prompts for the frequency at which the function is to be evaluated and puts the frequency in radians per second into R20. It then calls on subroutine "K" to evaluate the function. Finally, at label 05, it displays the results of the calculation in the desired form as controlled by flag 01. (Flag 03 is used for program "SLC", to be described in Section A-11.)

Subroutine “FN”

As noted previously, this subroutine prompts for the degree of numerator and denominator and displays the coefficients of each, allowing the user to change them, one at a time. It also clears the registers (label 03) such that the coefficients above the highest degree are zero (up to the seventh degree).

Subroutine “K”

This subroutine evaluates the numerator and the denominator separately and divides the latter into the former. It uses the polynomial in nested form, as described in Chapter 2. This subroutine could be much shorter (about 40 steps) by use of indirect addressing. We use the longer form because it cuts execution time roughly in half—from 8 to 4 s. The magnitudes of the numerator and the denominator are stored in R30 and R31 for use in program “SLC”. Linear phase is removed in subroutine 04 when flag 04 is found to be set.

01*LBL "BORAD"	33 FC? 01	65 GTO 05
02 CF 03	34 *dB PH*	66 LOG
	35 FS? 01	67 20
03*LBL *V*	36 *RE IM*	68 *
04 XEQ *FN*	37 ACA	
05 *OK?*	38 PRBUF	69*LBL 05
06 PROMPT	39 ADV	70 RND
07 FS? 22	40 RCL 26	71 ACX
08 GTO *V*	41 FS? 00	72 " "
	42 GTO 03	73 ACA
09*LBL 02	43 1/X	74 X<>Y
10 FIX 4	44 10↑X	75 RND
11 *ΔFMIN,MAX*		76 ACX
12 AVIEW	45*LBL 03	77 PRBUF
13 24.025	46 STO 27	78 FS? 03
14 PRREGX	47 RCL 24	79 XEQ *S*
15 ADV	48 STO 20	80 FIX 4
16 FS? 00		81 RCL 27
17 *ΔF INC*	49*LBL 11	82 FC? 00
18 FC? 00	50 XEQ *K*	83 ST* 20
19 *PTS/DEC:*	51 CLA	84 FS? 00
20 AVIEW	52 RCL 20	85 ST+ 20
21 26	53 FIX 3	86 RCL 25
22 PRREGX	54 RND	87 RCL 20
23 ADV	55 ACX	88 X<=Y?
24 CF 22	56 " "	89 GTO 11
25 *OK?*	57 ACA	90 END
26 PROMPT	58 FC? 01	
27 FS? 22	59 FIX 2	CAT 1
28 GTO 02	60 RCL 23	
29 6	61 RCL 22	LBL*FN
30 " ΔF"	62 FS? 01	LBL*K
31 ACA	63 P-R	END
32 SKPCHR	64 FS? 01	LBL*BORAD
		LBL*V
		END
		216 BYTES
		221 BYTES

Figure A7.8. Program “BORAD”: listing.

```

01*LBL "BHZ"          38 PRBUF          74 LOG
02 CF 03              39 ADV             75 20
                     40 RCL 26          76 *
03*LBL "Y"           41 FS? 00
04 XEQ "FN"          42 GTO 03
05 "OK?"             43 1/X
06 PROMPT            44 10↑X
07 FS? 22
08 GTO "Y"           45*LBL 03
                     46 STO 27
09*LBL 02            47 PI
10 FIX 4              48 2
11 "<F>"             49 *
12 AVIEW             50 RCL 24
13 24.025            51 *
14 PRREGX            52 STO 20
15 ADV
16 FS? 00            53*LBL 11
17 "F INC"           54 XEQ "K"
18 FC? 00            55 CLA
19 "P/D"             56 RCL 20
20 AVIEW             57 PI
21 26                58 2
22 PRREGX            59 *
23 ADV               60 /
24 CF 22             61 FIX 3
25 "OK?"             62 RND
26 PROMPT            63 ACX
27 FS? 22            64 " "
28 GTO 02            65 ACA
29 6                 66 FC? 01
30 " F"              67 FIX 2
31 ACA               68 RCL 23
32 SKPCHR            69 RCL 22
33 FC? 01            70 FS? 01
34 "dB PH"          71 P-R
35 FS? 01            LBL"BHZ
36 "RE IM"          LBL"Y
37 ACA              END
                     .END.
                     224 BYTES
                     08 BYTES
77*LBL 05
78 RND
79 ACX
80 " "
81 ACA
82 X<Y
83 RND
84 ACX
85 PRBUF
86 FS? 03
87 XEQ "S"
88 FIX 4
89 PI
90 2
91 *
92 RCL 27
93 FC? 00
94 ST* 20
95 *
96 FS? 00
97 ST+ 20
98 PI
99 2
100 *
101 RCL 25
102 *
103 RCL 20
104 X<=Y?
105 GTO 11
106 END
CAT 1

```

Figure A7.9. Program "BHZ": listing.

Programs "BORAD" and "BHZ"

These programs operate in the same way as the "BODE" program except for automatic frequency stepping and data format. In the program listings (see Figs. A7.8 and A7.9) frequency incrementing is provided by the steps following label 02, which also provides the heading for the columns of figures. The program terminates when the frequency exceeds the maximum value chosen (in R25) and is contained in the last six steps of "BORAD" and the last nine steps of "BHZ".

Programs for Evaluating Rational Functions: Registers

R00–R07	Numerator polynomial
R08	Degree of numerator
R09	Loop index for numerator
R10–R17	Denominator polynomial
R18	Degree of denominator
R19	Loop index for denominator
R20	Angular frequency ω
R21	$-\omega^2$
R22	Magnitude of $L(j\omega)$
R23	Phase of $L(j\omega)$
R24	F_{\min}
R24	F_{\max}
R26	F increment or points/decade
R27	Frequency incrementing constant
R28*	Coefficient index for sensitivity
R29*	Loop index for coefficient
R30*	Magnitude of numerator polynomial
R31	Phase of numerator polynomial
R32*	Magnitude of denominator polynomial
R33	Phase of denominator polynomial

A-8 PROGRAM “RCU”: REVERSE CUBIC EVALUATION

This and the following program “RQU” (reverse quintic evaluation) are useful in modeling structures having polynomial loss characteristics. Program “RCU” evaluates the coefficients of $L(j\omega)$ at two frequencies. It implements eqs. (2.5-6) to (2.5-9), allowing the user to model measured or calculated structures whose loss response is known to be cubic polynomial in character and whose frequency response is known in magnitude and phase at two frequencies.

Flags are used to accommodate various cases. If the magnitude is expressed in dB, clear flag 01; if a ratio, set flag 01. If the frequency is expressed in gigaradians per second, clear flag 02; if in gigahertz set flag 02. The frequency and the loss and phase data are stored in R10–R15; the polynomial is stored in R00–R03.

To use the program after it is loaded into the machine, “RCU” is executed. (See Fig. A8.1.) All flag status and data required are asked for by prompts.

With no printer, execution halts at each prompt for input data, allowing the user to modify the data, as in program “BODE”. With the printer connected, the input data are printed without halting the program. This is shown in the first example. If halting the program with printer connected is desired, set flag 00, as shown in the second column of the examples, in which the data are

*Used for Program “SLC”

```

      XEQ "RCU"
EVAL CU COEF
dB? CF 01
GHZ? SF 02

ΔF1: 0.5000
dB: 0.0700
PH: 60.2000

ΔF2: 1.0000
dB: 3.0100
PH: 135.0000

COEF:
a0 = 1.0013
a1 = 1.9994
a2 = 2.0013
a3 = 0.9995

      SF 00
      XEQ "RCU"
EVAL CU COEF
dB? CF 01
GHZ? SF 02

ΔF1: 0.5000
dB: 0.0700
PH: 60.2000

ΔF2: 1.0000
dB: 3.0100
PH: 135.0000

COEF:
a0 = 1.0000
a1 = 2.0006
a2 = 2.0000
a3 = 1.0006

      XEQ "RCU"
EVAL CU COEF
dB? CF 01
GHZ? SF 02

ΔF1: 0.5000
dB: 0.0690
PH: 60.2600

ΔF2: 1.0000
dB: 3.0100
PH: 135.0000

COEF:
a0 = 1.0004
a1 = 2.4704
a2 = 2.4307
a3 = 0.9993

      CF 00
      XEQ "RCU"
EVAL CU COEF
dB? CF 01
GHZ? SF 02

ΔF1: 0.2000
dB: 0.2200
PH: 28.2900

ΔF2: 0.7000
dB: 2.9200
PH: 97.8300

COEF:
a0 = 1.0004
a1 = 2.4704
a2 = 2.4307
a3 = 0.9993

```

Figure A8.1. Program "RCU": examples.

modified slightly. Both examples start with the data for a cubic Butterworth polynomial and find the corresponding coefficients. In the third column of the examples, loss and phase are changed to those of the first example of the "BODE" program, and the coefficients for the maximally flat delay polynomial are found. We then clear flag 00 to show the printed output for this condition. Round-off error in the loss and phase introduces slight errors in the coefficients.

Program Description

This program (Fig. A8.2) implements eqs. (2.5-6) through (2.5-9). The data are placed in the appropriate registers in the first 38 steps; subroutine 04 is essentially the same data display subroutine used in "BODE". The calculation is done in steps 39–121, and the output display is done by subroutine 06. The calculation takes less than 5 seconds, and somewhat more with the printer connected.

```

01*LBL "RCU"
02 SF 21
03 SF 12
04 "EVAL CU COEF"
05 AVIEW
06 CF 12
07 ADV
08 "dB? CF 01"
09 AVIEW
10 "GHZ? SF 02"
11 AVIEW
12 ADV
13 10
14 STO 04
15 CLA
16 FC? 02
17 "z"
18 "HF1"
19 XEQ 04
20 FS? 01
21 "MG"
22 FC? 01
23 "dB"
24 XEQ 04
25 "PH"
26 XEQ 04
27 ADV
28 FC? 02
29 "z"
30 "HF2"
31 XEQ 04
32 FS? 01
33 "MG"
34 FC? 01
35 "dB"
36 XEQ 04
37 "PH"
38 XEQ 04

39*LBL 02
40 FC? 02
41 GTO 03
42 2

43 PI
44 *
45 ST* 10
46 ST* 17

47*LBL 03
48 RCL 13
49 X12
50 RCL 10
51 X12
52 -
53 1/X
54 STO 07
55 CHS
56 STO 09
57 RCL 10
58 X12
59 *
60 STO 08
61 RCL 13
62 X12
63 RCL 07
64 *
65 STO 06
66 ADV
67 "COEF:"
68 AVIEW
69 RCL 12
70 RCL 11
71 FC? 01
72 XEQ 00
73 P-R
74 STO 16
75 X<>Y
76 RCL 10
77 /
78 STO 17
79 RCL 15
80 RCL 14
81 FC? 01
82 XEQ 00
83 P-R
84 STO 18

```

Figure A8.2. Program "RCU": listing.

85 X<Y	128 CF 29
86 RCL 13	129 ARCL X
87 /	130 "F = "
88 STO 19	131 FIX 4
89 RCL 06	132 ARCL IND X
90 RCL 16	133 AVIEW
91 *	134 X<Y
92 RCL 08	135 CLA
93 RCL 18	136 ISG X
94 *	137 GTO 06
95 +	138 RTN
96 STO 00	
97 RCL 06	139*LBL 00
98 RCL 17	140 20
99 *	141 /
100 RCL 08	142 10↑X
101 RCL 19	143 RTN
102 *	
103 +	144*LBL 04
104 STO 01	145 RCL IND 04
105 RCL 07	146 RND
106 RCL 16	147 "F: "
107 *	148 ARCL X
108 RCL 09	149 CF 22
109 RCL 18	150 FS? 00
110 *	151 PROMPT
111 +	152 FC? 00
112 STO 02	153 AVIEW
113 RCL 07	154 FC? 22
114 RCL 17	155 GTO 05
115 *	156 X<Y
116 RCL 09	157 RDN
117 RCL 19	158 STO IND 04
118 *	
119 +	159*LBL 05
120 STO 03	160 CLA
121 .003	161 ISG 04
	162 DEG
122*LBL 06	163 END
123 "a"	CAT 1
124 ENTER↑	LBL"RCU
125 X<Y	END
126 FIX 0	299 BYTES
127 RND	

Figure A8.2. Continued.

Program "RCU": Registers

R00–R03	Cubic coefficients (results)
R04	Input data index
R05	
R06–R09	Used
R10–R15	Input data: $F_1, L_1; F_2, L_2$.
R16–R19	Used

A-9 PROGRAM "RQU": POLYNOMIAL COEFFICIENTS FROM LOSS AND PHASE FOR QUINTIC POLYNOMIALS (USE WITH "MTR")

This program extends the "RCU" program for up to quintic polynomials. Operation is the same as "RCU" except that the program requires loss and phase at three frequencies as noted in Chapter 2.

The program implements eqs. (2.5-12) and (2.5-13). The input data are the three frequencies at which $L(j\omega)$ is evaluated and the three values of loss magnitude and phase at these frequencies. A subroutine, "MTR", which must be loaded in with "RQU", inverts the frequency matrix and performs the matrix multiplication, thereby finding the six coefficients of the polynomial.

Use of the program is self-explanatory in that the flag options are displayed or printed when the program is executed (XEQ "RQU"), as seen in the first example (see Fig. A9.1). After the flag options are listed, the input data is

XEQ "RQU" QUINTIC COEF FROM LOSS	SF 03 XEQ "RQU" QUINTIC COEF FROM LOSS	XEQ "RQU" QUINTIC COEF FROM LOSS
dB? CF 01	dB? CF 01	dB? CF 01
GHZ?, SF 02	GHZ?, SF 02	GHZ?, SF 02
DATA ON CARDS?	DATA ON CARDS?	DATA ON CARDS?
SF 03	SF 03	SF 03
READ DATA	READ DATA	READ DATA
$\angle F1 = 0.5000$	$\angle F1 = 0.5000$	$\angle F1 = 0.5000$
dB = 0.7540	dB = 0.7500	dB = 0.7970
.7970	PH = 68.5100	PH = 68.5000
PH = 68.5200		
68.5000		
$\angle F2 = 1.0000$	$\angle F2 = 1.0000$	$\angle F2 = 1.0000$
dB = 3.2670	dB = 3.2700	dB = 3.4370
3.4370	PH = 136.6800	PH = 136.5300
PH = 136.6800		
136.5300		
$\angle F3 = 2.0000$	$\angle F3 = 2.0000$	$\angle F3 = 2.0000$
dB = 14.3040	dB = 14.3000	dB = 14.9500
14.9500	PH = -112.8600	PH = -113.9800
PH = -112.8620		
-113.9800		
a0 = 1.0000	a0 = 0.9944	a0 = 1.0000
a1 = 2.3899	a1 = 2.3880	a1 = 2.3899
a2 = 2.4980	a2 = 2.4885	a2 = 2.4980
a3 = 1.4119	a3 = 1.4528	a3 = 1.4119
a4 = 0.4200	a4 = 0.4340	a4 = 0.4200
a5 = 0.0440	a5 = 0.0646	a5 = 0.0440
10.0180 WDTAX		

Figure A9.1. Program "RQU": examples.

displayed or printed, giving the frequencies and the loss and phase values at the three frequencies. As in "RCU", when no printer is connected, program execution halts at each prompt for input data; if this is desired when the printer is connected, set flag 00. This was done in the first column of the examples; flag 00 was cleared in the remaining columns. To evaluate the quintic polynomial coefficients, clear flag 01 if the loss magnitude is in decibels, set flag 02 if the frequencies are in hertz (or gigahertz), and press R/S. The resulting printout will give the coefficients.

The program can be used for polynomials of any degree up to quintic. For polynomials up to cubic, however, program "RCU" is shorter and faster. This program takes about five times as long as "RCU".

In the examples a quartic polynomial actually encountered in the design of a fiber optic preamplifier is given by

$$L(p) = 1 + 2.19p + 2.06p^2 + p^3 + 0.22p^4$$

For purposes of the examples, this polynomial will be inflated in two ways. Both have the effect of adding approximately 0.2 ns of delay to the polynomial. In the first, we add the delay by multiplying the polynomial by $e^{0.2p}$:

$$L_1(p) = (1 + 2.19p + 2.06p^2 + p^3 + 0.22p^4)e^{0.2p}$$

At frequencies of 0.5, 1.0, and 2.0 for ω/ω_0 , the magnitude and the phase of this polynomial with delay are

$$L_1(j.5) = 0.754 \text{ dB} \quad \underline{\angle 68.518^\circ}$$

$$L_1(j1.0) = 3.267 \text{ dB} \quad \underline{\angle 136.677^\circ}$$

$$L_1(j2.0) = 14.304 \text{ dB} \quad \underline{\angle -112.862^\circ}$$

This polynomial is evaluated for its coefficients in the first example.

In the second example the original quartic is inflated by a linear factor $(1 + 0.2p)$ approximating the delay:

$$\begin{aligned} L_2(p) &= (1 + 2.19p + 2.06p^2 + p^3 + 0.22p^4)(1 + 0.2p) \\ &= 1 + 2.39p + 2.498p^2 + 1.14p^3 + 0.42p^4 + 0.044p^5 \end{aligned}$$

and the evaluation of the polynomial by "BODE" gives

$$L_2(j0.5) = 0.797 \text{ dB} \quad \underline{\angle 68.499^\circ}$$

$$L_2(j1.0) = 3.437 \text{ dB} \quad \underline{\angle 136.528^\circ}$$

$$L_2(j2.0) = 14.949 \text{ dB} \quad \underline{\angle -113.979^\circ}$$

In the second example the coefficients of this polynomial are evaluated.

Program Description

Input data are treated similarly to those in program "RCU", with the frequencies and losses placed in registers R10–R18 in the first part of the program. (See Figs. A9.2 and A9.3.) Note that this portion calls on subroutine A, which in turn calls on 04, the display subroutine for the input data.

At label B, subroutine "M" in "MTR" is called. This subroutine places the frequency matrix described in Section 2.6 in R00–R08; this matrix is then

01*LBL "RQU"	43 *	83 RDN
02 SF 21	44 ST* 10	84 STO IND 09
03 SF 12	45 ST* 13	
04 "QUINTIC COEF"	46 ST* 16	85*LBL 05
05 AVIEW		86 TSC 09
06 "FROM LOSS"	47*LBL B	87 DEG
07 AVIEW	48 CLD	88 CLA
08 CF 12	49 XEQ "M"	89 RTN
09 ADV	50 XEQ C	
10 "dB? CF 01"	51 ADV	90*LBL 01
11 AVIEW	52 FC? 02	91 20
12 " GHZ?, SF 02"	53 RTN	92 /
13 AVIEW	54 PI	93 10↑X
14 10	55 2	94 RTN
15 STO 09	56 *	
16 "DATA ON CARDS?"	57 ST/ 10	95*LBL C
17 AVIEW	58 ST/ 13	96 RCL 12
18 " SF 03"	59 ST/ 16	97 RCL 11
19 AVIEW	60 RTN	98 FC? 01
20 "READ DATA"		99 XEQ 01
21 FS? 03	61*LBL A	100 P-R
22 AVIEW	62 ADV	101 STO 30
23 10.018	63 XEQ 04	102 X<>Y
24 FS? 03	64 FC? 01	103 RCL 10
25 RDTAX	65 "dB"	104 /
26 CLA	66 FS? 01	105 STO 31
27 FC? 02	67 "MAG"	106 RCL 15
28 "z"	68 XEQ 04	107 RCL 14
29 "HF1"	69 "PH"	108 FC? 01
30 XEQ A		109 XEQ 01
31 FC? 02	70*LBL 04	110 P-R
32 "z"	71 RCL IND 09	111 STO 32
33 "HF2"	72 RND	112 X<>Y
34 XEQ A	73 "f = "	113 RCL 13
35 FC? 02	74 ARCL X	114 /
36 "z"	75 CF 22	115 STO 33
37 "HF3"	76 FC? 00	116 RCL 18
38 XEQ A	77 AVIEW	117 RCL 17
39 FC? 02	78 FS? 00	118 FC? 01
40 GTO B	79 PROMPT	119 XEQ 01
41 2	80 FC? 22	120 P-R
42 PI	81 GTO 05	121 STO 34
	82 X<>Y	122 X<>Y

Figure A9.2. Program "RQU": listing.

```

123 RCL 16
124 /
125 STO 35
126 XEQ 02
127 RCL 06
128 STO 01
129 RCL 07
130 STO 03
131 RCL 08
132 STO 05
133 RCL 30
134 STO 31
135 RCL 32
136 STO 33
137 RCL 34
138 STO 35
139 XEQ 02
140 RCL 06
141 STO 00
142 RCL 07
143 STO 02
144 RCL 08
145 STO 04
146 ADV
147 .005
148 XEQ 10
149 RTN

150+LBL 02
151 20
152 STO 19
153 XEQ 06
154 STO 06
155 21
156 STO 19
157 XEQ 06
158 STO 07
159 22
160 STO 19
161 XEQ 06
162 STO 08
163 RTN

164+LBL 06
165 0
166 RCL 31
167 XEQ 07
168 RCL 33
169 XEQ 07
170 RCL 35
171 XEQ 07
172 RTN

173+LBL 07
174 RCL IND 19
175 *
176 +
177 3
178 ST+ 19
179 RDN
180 RTN

181+LBL 10
182 "a"
183 ENTER↑
184 X<>Y
185 FIX 0
186 CF 29
187 RND
188 ARCL X
189 "+ = "
190 FIX 4
191 ARCL IND X
192 RVIEW
193 CLA
194 X<>Y
195 ISG X
196 GTO 10
197 END

CAT 1
LBL'RQU
END
423 BYTES
    
```

Figure A9.3. Subroutine "MAT": listing.

inverted in "MTR" and the inverse is stored in R20–R28. In subroutine C, which is then called on, the column vector consisting of the real part of $L(j\omega)$ and the imaginary part divided by the frequency is set up in R30–R35. The matrix multiplication is carried out in subroutine 02, first for the imaginaries and then for the reals. This completes the calculation. Output data are displayed in subroutine 10.

Program "RQU": Registers

- R00–R08 Frequency matrix
- R00–R05 Quintic polynomial coefficients

01*LBL "MAT"	46 X↑2	91 RCL 02
02*LBL 08	47 STO 08	92 RCL 09
03 0	48 RCL 16	93 XEQ 03
04 STO 19	49 X↑2	94 STO 21
05 RCL 06	50 CHS	95 CLX
06 RCL 08	51 STO 06	96 RCL 02
07 XEQ 09	52 X↑2	97 RCL 06
08 RCL 04	53 STO 09	98 *
09 RCL 09	54 XEQ 08	99 RCL 03
10 XEQ 09	55 1/X	100 RCL 05
11 RCL 05	56 RCL 01	101 XEQ 03
12 RCL 07	57 RCL 09	102 STO 22
13 XEQ 09	58 *	103 CLX
14 CHS	59 RCL 03	104 RCL 05
15 RCL 03	60 RCL 07	105 RCL 09
16 RCL 08	61 XEQ 03	106 *
17 XEQ 09	62 STO 24	107 RCL 06
18 RCL 01	63 CLX	108 RCL 08
19 RCL 09	64 RCL 03	109 XEQ 03
20 XEQ 09	65 RCL 04	110 STO 20
21 RCL 02	66 *	111 CLX
22 RCL 07	67 RCL 01	112 RCL 06
	68 RCL 06	113 RCL 07
	69 XEQ 03	114 *
23*LBL 09	70 STO 25	115 RCL 04
24 ISG 19	71 CLX	116 RCL 09
25 DEG	72 RCL 02	117 XEQ 03
26 RCL IND 19	73 RCL 07	118 STO 23
27 *	74 *	119 CLX
28 *	75 RCL 01	120 RCL 04
29 +	76 RCL 08	121 RCL 08
30 RTN	77 XEQ 03	122 *
	78 STO 27	123 RCL 05
31*LBL "M"	79 CLX	124 RCL 07
32 1	80 RCL 01	125 XEQ 03
33 STO 01	81 RCL 05	126 STO 26
34 STO 02	82 *	127 RTN
35 STO 03	83 RCL 02	
36 RCL 10	84 RCL 04	128*LBL 03
37 X↑2	85 XEQ 03	129 *
38 CHS	86 STO 28	130 -
39 STO 04	87 CLX	131 *
40 X↑2	88 RCL 03	132 END
41 STO 07	89 RCL 08	CAT 1
42 RCL 13	90 *	
43 X↑2		LBL"MAT
44 CHS		LBL"M
45 STO 05		END
		187 BYTES

Figure A9.4. Continued.

R06–R08	Temporary coefficient storage
R09	Input data loop index
R10–R18	$F_1, L_1; F_2, L_2; F_3, L_3$
R19	Loop index
R20–R28	Inverse frequency matrix
R30–R35	Real and imaginary loss column vector

A-10 PROGRAM "POLYTBL": STANDARD PERFORMANCE SPECIFICATION POLYNOMIALS (CAN BE USED WITH "MFD" AND "CH")

This program provides a library or table of standard loss functions: maximally flat amplitude (Butterworth), maximally flat delay (Bessel), transitional between the two (Peless and Murikami), and Chebyshev. Equations used by the program are discussed in Section 2.5.

To use the program, load it into the machine and execute "POLYTBL" (see Fig. A10.1). The program will prompt for the degree desired; after this has been supplied, the program will ask for one of six options. There are four main functions and two modified functions available. The main four can be assigned keys (in user mode), for example: "B" for Butterworth, "C" for Chebyshev, "D" for maximally flat delay, and "TAN" for transitional. Each program will ask for the degree of the polynomial desired. In addition, the Chebyshev program will ask for the ripple width in decibels, and the transitional polynomial program will ask for the interpolation constant m , $0 \leq m \leq 1$, where $m=0$ is for MFD and $m=1$ is for Butterworth. The desired polynomial is then printed out. Polynomials up to sixteenth degree can be found.

The two modified functions, "DL" and "MFA", are obtained by executing "DL" or "MFA". Function "DL" is a renormalized MFD function to obtain Bessel polynomials normalized to unit delay (hence b_1 is made unity). It is useful for delay line analysis and design. Function "MFA" is the same as Butterworth but normalized in frequency to give a maximum in-band error (in dB) as specified by the user. If 3.0 dB is specified, the ordinary Butterworth polynomial is obtained, but any other error such as 0.1 dB may be specified.

The dc values of all polynomials except even-degree Chebyshev are normalized to unity; even-degree Chebyshev polynomials have a dc value equal to unity times the ripple width expressed as a ratio. Various frequency normalizations have been used in the literature. We choose an asymptotic cutoff frequency of unity for the Butterworth, the maximally flat delay and the transitional polynomials. The Chebyshev is normalized to the upper edge of the ripple width channel (it remains within the ripple specification up to unity frequency).

For convenience, "POLYTBL" has been broken into three parts, so that the user need not load in more of the program than is needed. The control program carries the label "POLYTBL" and is combined with "BUT" and "MFA", the Butterworth polynomial generator programs (see Fig. A10.2). The

```

XEQ "POLYTBL"
POLY TABLE
DEG?
      3.0000  RUN
B, C, D, TR, DL, MFA?
XEQ "BUT"
BUTTERWORTH
a20= 1.0000
a21= 2.0000
a22= 2.0000
a23= 1.0000
      ASN "MFA" 11
      XEQ "MFA"
MFA
ERROR TOL., dB?
      .2000  RUN
a20= 1.0000
a21= 1.2020
a22= 0.7224
a23= 0.2171
      XEQ "MFD"
MFD
a20= 1.0000
a21= 2.4662
a22= 2.4329
a23= 1.0000
      XEQ "DL"
DLY
a20= 1.0000
a21= 1.0000
a22= 0.4000
a23= 0.0667

XEQ "TR"
TRANSITIONAL
M?
      .5000  RUN
a20= 1.0000
a21= 2.2009
a22= 2.2050
a23= 1.0000
      XEQ "CH"
CHEBYSHEV
RIPPLE, dB?
      .2000  RUN
a20= 1.0000
a21= 1.0030
a22= 1.4148
a23= 0.8684
      20.0000
      XEQ "X-"
      XEQ "N-"
SCALE POLY
POLY A
DEG?
      3.0000  RUN
R0= 1.000E0
R1= 1.804E0
R2= 1.415E0
R3= 0.868E-1
OK?
      RUN
POLY B
b0?
      1  RUN
F0B/F0A?
      RUN
bM:
M?
      3  RUN
bM?
      1  RUN
F0B/F0A=1.048E0
R10= 1.000E0
R11= 1.891E0
R12= 1.554E0
R13= 1.000E0

```

Figure A10.1. Program "POLYTBL": examples.

```

01*LBL "POLYTBL"          40 ISG X                79 Y*X
   02 SF 21                41 GTO 01                80 *
   03 CLRG                 42 RTN                81 STO IND 14
   04 CF 06                43*LBL "MFA"           82 ISG 16
   05 SF 12                44 SF 12                83 DEG
06 "POLY TABLE"         45 "MFA"                84 ISG 14
   07 AVIEW                46 AVIEW                85 GTO 16
   08 CF 12                47 CF 12                86 GTO "P"
   09 ADV                  48 ADV                  87*LBL "BUT"
  10 20.02                 49 SF 05                88 CF 04
  11 STO 04                50 TONE 9               89 SF 12
  12 "DEG?"                51 "ERROR TOL., dB?"   90 "BUTTERWORTH"
  13 TONE 9                 52 PROMPT                91 AVIEW
  14 PROMPT                 53 10                    92 CF 12
  15 STO 00                 54 /                     93 ADV
  16 ADV                    55 10*X                 94*LBL "FA"
17 "B, C, D, TR, "       56 1                      95 RCL 00
  18 "FDL, MFA?"           57 -                      96 2
  19 PROMPT                 58 RCL 00                97 -
                             59 2                      98 2
  20*LBL "P"                60 *                      99 /
  21 RCL 00                 61 1/X                  100 STO 06
  22 1 E3                   62 Y*X                  101 0
  23 /                       63 STO 10                102 STO 01
  24 RCL 04                 64 XEQ "FA"             103 STO 02
  25 +                       65 CF 05                104 1
                             66 RCL 04                105 STO IND 04
  26*LBL 01                 67 RCL 00                106 RCL 04
  27 "a"                     68 1 E3                  107 RCL 00
  28 ENTER↑                 69 /                      108 +
  29 X<>Y                    70 +                      109 STO 05
  30 FIX 0                   71 1                      110 2
  31 CF 29                   72 STO 16                111 ST+ 05
  32 RND                     73 +                      112 0
  33 ARCL X                  74 STO 14                113 STO 19
  34 "f= "                   75*LBL 16                114*LBL 11
  35 FIX 4                   76 RCL IND 14            115 STO IND 05
36 ARCL IND X              77 RCL 10                116 DSE 05
  37 AVIEW                   78 RCL 16
  38 CLA
  39 X<>Y

```

Figure A10.2. Programs "POLYTBL", "BUT", and "MFA": listing.

Bessel polynomials generated by "MFD" and "DL" as well as the transitional polynomial generator, "TR" are contained in a second program (see Fig. A10.3) that must be run with "POLYTBL". Finally, "CH" (see Fig. A10.4), the Chebyshev polynomial generator, is in a separate program that also must be run with "POLYTBL".

All six options are illustrated in the accompanying examples of Fig. A10.1, which generate six cubic polynomials, one of each type possible with the program, and one twelfth-degree Chebyshev polynomial with 0.05 dB ripple. Once generated, polynomials up to eighth degree can be scaled to any desired dc loss and cutoff frequency by using program "N".

117 GTO 11	156 90	195 STO 21
118 RCL 00	157 *	196 STO 02
119 2	158 COS	197 FS? 04
120 /	159 2	198 XEQ H
121 FRC	160 *	199 RTN
122 X=0?	161 STO 12	
123 SF 06	162 1	200*LBL H
124 X#0?	163 STO 13	201 RCL 09
125 XEQ 10	164 STO 11	202 ST/ 21
	165 FS? 04	203 RTN
126*LBL 12	166 XEQ "TC"	
127 RCL 02	167 RTN	204*LBL "X-"
128 RCL 04		205 STO 19
129 +	168*LBL "ML"	206 .008
130 STO 03	169 RCL IND 03	207 STO 09
131 2	170 RCL 11	
132 ST+ 03	171 *	208*LBL 07
133 XEQ "TH"	172 DSE 03	209 RCL IND 09
134 XEQ "ML"	173 DEG	210 X(<) 19
135 2	174 RCL IND 03	211 STO IND 09
136 ST+ 02	175 RCL 12	212 ISG 19
137 1	176 *	213 DEG
138 ST+ 01	177 +	214 ISG 09
139 RCL 06	178 DSE 03	215 GTO 07
140 RCL 01	179 DEG	216 END
141 X<=Y?	180 RCL IND 03	CAT 1
142 GTO 12	181 RCL 13	LBL*POLYTBL
143 FC? 05	182 *	LBL*P
144 GTO "P"	183 +	LBL*MFA
145 RTN	184 2	LBL*BT
	185 ST+ 03	LBL*FA
146*LBL "TH"	186 RDN	LBL*TH
147 RCL 00	187 STO IND 03	LBL*ML
148 1	188 DSE 03	LBL*X-
149 -	189 GTO "ML"	END
150 RCL 01	190 RTN	432 BYTES
151 2		
152 *	191*LBL 10	
153 -	192 .5	
154 RCL 00	193 ST- 06	
155 /	194 1	

Figure A10.2. Continued.

An example of scaling is given in the third column of the examples. Since the scaling program, "N", operates on a polynomial in registers R00–R08 and "POLYTBL" leaves the generated polynomial in registers starting with R20, the latter polynomial (up to eighth degree) must be moved to the lower-numbered registers. This could be done manually, but it is easier and more accurate to have a program do this for us, as did program "X-Y" in Appendix A-1. Program "X-" exchanges the locations of two sets of nine registers; the first set is in R00–R08; the second set begins with a number that is keyed in before executing "X-". Thus the steps 10, XEQ "X-" would exchange the

```

01*LBL "DL"           43 ST+ 01           87 FS? 02
02 SF 12              44 20              88 RTN
03 "DLY"             45 STO 03           89 FC? 05
04 AVIEW              46*LBL 00           90 GTO "P"
05 ADV                47 DSE 01           91 RTN
06 CF 12              48 DEG              92*LBL 13
07 SF 05              49 RCL 00           93 I
08 XEQ "FD"          50 RCL 01           94 RCL 17
09 CF 05              51 +                95 -
10 RCL 21             52 FACT             96 Y+X
11 STO 10             53 RCL 01           97 RCL IND 02
12 RCL 04             54 FACT             98 RCL 17
13 RCL 00             55 /                99 Y+X
14 I E3              56 RCL 00           100 *
15 /                 57 RCL 01           101 RTN
16 +                 58 -
17 I                 59 FACT             102*LBL "TR"
18 STO 16            60 /                103 SF 12
19 +                 61 RCL 00           104 "TRANSITIONAL"
20 STO 14            62 FACT             105 AVIEW
                    63 RCL 00           106 CF 12
                    64 2                107 SF 02
                    65 *                108 "M?"
                    66 FACT             109 PROMPT
                    67 /                110 STO 17
                    68 RCL 01           111 RCL 00
                    69 RCL 00           112 STO 01
                    70 /                113 STO 16
                    71 Y+X           114 SF 05
                    72 *                115 XEQ "FA"
                    73 RCL 00           116 CF 05
                    74 RCL 01           117 RCL 16
                    75 -                118 STO 00
                    76 RCL 03           119 XEQ "FD"
                    77 +                120 CF 02
                    78 STO 02           121 GTO "P"
                    79 RDN              122 END
                    80 FS? 02           CAT 1
                    81 XEQ 13           LBL"DL
                    82 STO IND 02       LBL"MFD
                    83 ISG 01         LBL"FD
                    84 DEG           LBL"TR
                    85 DSE 01         END
                    86 GTO 00
                    -- -- --

```

Figure A10.3. Programs "MFD", "TR", and "DL": listing.

numbers in registers R00–R08 with those in R10–R18. This short program is incorporated in "POLYTBL" for convenience.

In the example, we key in 20, XEQ "X-" to move the cubic Chebyshev contained in R20–R23 down to R00–R03. (The remaining five of the eight registers moved play no part in the ensuing calculation.) The Chebyshev cubic, scaled to unity cutoff frequency, is shown in the example.

```

01*LBL "CH"          35 2          67 GT0 15
02 SF 12             36 /          68 RTN
03 *CHEBYSHEV*      37 STO 09
04 AVIEW             38 RCL 08          69*LBL "TC"
05 CF 12             39 RCL 07          70 RCL 12
06 *RIPPLE, dB?*    40 +          71 2
07 PROMPT           41 2          72 /
08 ENTER↑           42 /          73 ACOS
09 10               43 STO 08          74 1
10 /                44 SF 04          75 P-R
11 10↑X             45 SF 05          76 RCL 09
12 STO 10           46 RCL 00          77 *
13 1                47 XEQ "FA"        78 X<>Y
14 -                48 CF 05          79 RCL 08
15 1/X              49 CF 04          80 *
16 STO 08           50 FS? 06          81 X<>Y
17 1                51 XEQ 14          82 R-P
18 +                52 GT0 "P"        83 1/X
19 SQRT             53*LBL 14          84 X↑2
20 RCL 08           54 RCL 04          85 STO 13
21 SQRT             55 RCL 00          86 X<>Y
22 +                56 1 E3          87 COS
23 LN               57 /          88 2
24 STO 09           58 +          89 *
25 RCL 00           59 STO 19          90 RCL 13
26 /                60 RCL 18          91 SQRT
27 STO 09           61 SQRT          92 *
28 E↑X              62 STO 17          93 STO 12
29 STO 08           63*LBL 15          94 END
30 RCL 09           64 RCL 17          CAT 1
31 CHS              65 ST* IND 19
32 E↑X              66 ISG 19
33 STO 07
34 -

```

Figure A10.4. Program "CH": listing.

Program "POLYTBL" and Associated Programs: Registers

R00 Degree of desired polynomial
R01–R19 Used
R20– Polynomial (result)

A-11 PROGRAM "SLC": SENSITIVITY MAGNITUDE OF A RATIONAL FUNCTION TO ITS COEFFICIENTS (USE WITH "BODE" AND "FN")

This program (see Figs. A11.1 and A11.2), used in conjunction with any one of the three evaluation programs, "BODE", "BORAD", or "BHZ", evaluates the sensitivities of the function $L(j\omega)$ to its numerator and denominator coefficients. It is (of course) also applicable to polynomials, by setting the denominator to unity.

To use the program, several cards must be loaded into the machine: "SLC" itself, one of the three evaluation programs, and their associated subroutine "FN". By executing "SLC", these other programs are automatically invoked where needed. The user should be concerned only with the prompts for input data from the various programs. When program "BODE" is used, no printer is needed, although it is helpful; the other two evaluation programs require a printer.

To try the program, load "SLC", "BODE", and "FN" into the machine; the process is illustrated in the first example of Fig. A11.1. As in "BODE", the program prompts for the numerator and denominator degrees and coefficients. For the example, we used the Chebyshev polynomial normalized to unity asymptotic cutoff (it was already in the machine from the discussion of "POLYTBL"). After setting the input data, the prompt "2PIF=" appears, to which the response is .2. (*Remember:* When the printer is used, what the calculator does is left-justified and what you key in is right-justified.) The loss and phase are then calculated and displayed, followed by a display of each of the numerator coefficient sensitivities and the same for the denominator coefficients. In this case, the denominator is unity as is the sensitivity of loss to d_0 . The display then prompts for the next frequency, and the process is repeated, as shown.

For users with a printer, the same example is repeated by use of program "BORAD" instead of "BODE". Flag 00 is set to obtain a linear frequency increment. The results are the same, but the frequency is incremented automatically.

To show the evaluation of denominator coefficient sensitivities, we use an all-pass function having the numerator of the preceding example and an identical denominator except that the sign of the odd coefficients is changed. Such a function will have a magnitude of unity at all frequencies, and the sensitivity magnitudes of the numerator and denominator coefficients should be identical for each degree. The results are shown in the third column of the examples.

Program Description

This program extends the sensitivity analysis in Chapter 2 to include the sensitivities of the numerator and denominator coefficients of a rational function. Invoking the sum rule for sensitivities (chapter 1), we obtain

$$S_{a_i}^L = \frac{a_i s^i / D(s)}{N(s) / D(s)} = \frac{a_i s^i}{N(s)}$$

where

$$L(s) = \frac{N(s)}{D(s)}$$

XEQ "SLC"			XEQ "SLC"			XEQ "SLC"		
L			L			L		
S			S			S		
coef i			coef i			coef i		
NUM DEG?	3.0000	RUN	NUM DEG?	3.0000	RUN	NUM DEG?	3.0000	RUN
R00=	1.0000		R00=	1.0000		R00=	1.0000	
R01=	1.8907		R01=	1.8907		R01=	1.8907	
R02=	1.5544		R02=	1.5544		R02=	1.5544	
R03=	1.0000		R03=	1.0000		R03=	1.0000	
DEN DEG?	0.0000	RUN	DEN DEG?	0.0000	RUN	DEN DEG?	3.0000	RUN
R10=	1.0000		R10=	1.0000		R10=	1.0000	
OK?			OK?			R11=	-1.8907	
2PIF =		RUN	ΔFMIN,MAX		RUN	R12=	1.5544	
	.2000	RUN				R13=	-1.0000	
2PIF =	0.2000		R24=	0.2000		OK?		RUN
0.07dB, Δ21.54			R25=	2.0000		2PIF =		RUN
S, L TO COEF:			ΔF INC				.2000	RUN
NUM			R26=	0.2000		2PIF =	0.2000	
i3: -42.01 db			OK?			0.00dB, Δ43.00		
i2: -24.20 db						S, L TO COEF:		
i1: -8.52 db						NUM		
i0: -0.07 db						i3: -42.01 db		
DEN			ΔF dB PH		RUN	i2: -24.20 db		
i10: 0.00 db			0.200 0.07 21.54			i1: -8.52 db		
2PIF =						i0: -0.07 db		
	1.0000	RUN	S, L TO COEF:			DEN		
2PIF =	1.0000		NUM			i13: -42.01 db		
0.42dB, Δ121.90			i3: -42.01 db			i12: -24.20 db		
S, L TO COEF:			i2: -24.20 db			i11: -8.52 db		
NUM			i1: -8.52 db			i10: -0.07 db		
i3: -0.42 db			i0: -0.07 db			2PIF =		
i2: 3.41 db			DEN				1.0000	RUN
i1: 5.12 db			i10: 0.00 db			2PIF =	1.0000	
i0: -0.42 db			0.400 0.19 42.66			0.00dB, Δ243.00		
DEN			S, L TO COEF:			S, L TO COEF:		
i10: 0.00 db			NUM			NUM		
2PIF =			i3: -24.06 db					
			i2: -12.27 db					

Figure A11.1. Program "SLC": examples.

Sensitivities of $L(s)$ to the denominator coefficients is found similarly, since from Table 1.1 (Chapter 1),

$$S_{d_i}^{1/L} = -S_{d_i}^L$$

Therefore, by the sum rule

$$S_{d_i}^L = -\frac{d_i s^i / N(s)}{D(s) / N(s)} = -\frac{d_i s^i}{D(s)}$$

Since we are concerned here with the *magnitudes* of sensitivities, the minus sign is of no significance. This program finds and lists the sensitivities of both numerator and denominator coefficients.

The program, listed in Fig. A11-2, is in two sections, labeled "SLC" and "S". The first section titles the program, sets flag 03, and executes "V"; the latter is a global label contained in each of the three polynomial evaluation programs.

```

01*LBL "SLC"
02 ADV
03 SF 12
04 FC? 55
05 CF 21
06 " L"
07 AVIEW
08 " S"
09 AVIEW
10 SF 13
11 " COEF I"
12 AVIEW
13 CF 13
14 CF 12
15 ADV
16 SF 03
17 CF 01
18 SF 21
19 CF 04
20 XEQ "V"
21 RTN

22*LBL "S"
23 RCL 08
24 STO 28
25 1
26 +
27 STO 29
28 ADV
29 "S, L TO COEF:"
30 AVIEW
31 CF 10

32 "NUM"
33 AVIEW
34 XEQ 02
35 SF 10
36 "DEN"
37 AVIEW
38 XEQ 03
39 RTN

40*LBL 02
41 SF 13
42 " I"
43 CF 13
44 10
45 RCL 28
46 FS? 10
47 +
48 FIX 0
49 CF 29
50 ARCL X
51 "I: "
52 RCL 20
53 RCL 28
54 Y↑X
55 DSE 29
56 DEG
57 RCL IND 29
58 *
59 FC? 10
60 RCL 31
61 FS? 10

62 RCL 30
63 /
64 ABS
65 LOG
66 20
67 *
68 FIX 2
69 RND
70 ARCL X
71 "I dB"
72 AVIEW
73 FIX 4
74 ISG 29
75 DEG
76 DSE 28
77 DEG
78 DSE 29
79 GTO 02
80 RTN

81*LBL 03
82 RCL 18
83 STO 28
84 11.010
85 +
86 STO 29
87 XEQ 02
88 END
CAT 1

LBL'SLC
LBL'S
END
205 BYTES
    
```

Figure A11.2. Program "SLC": listing.

In these programs, after the function is evaluated in magnitude and phase, program execution is directed to the sensitivity evaluation section “S” of this program. In program “BODE”, for example, at step 65 flag 03 is tested. If set, “S” is executed.

Section “S” evaluates and prints the numerator sensitivities using subroutine 02, and the denominator sensitivities using subroutine 03 (which, in turn, calls on subroutine 02).

Program “SLC”: Registers

R00–R27	Used by associated programs
R28	Coefficient index for sensitivity
R29	Loop index for coefficient
R30	Magnitude of numerator polynomial
R31*	Phase of numerator polynomial
R32	Magnitude of denominator polynomial
R33*	Phase of denominator polynomial

*Used in later programs.

Appendix B

Feedback Analysis and Synthesis Programs

The 12 programs in Appendix B can be divided into three groups. The first five programs deal with the analysis of circuits and an evaluation of sensitivities of the circuit loss to its polynomial coefficients and to its component and device parameters. The statistical variation of loss is also considered, given the statistical variation of device and component parameters.

The second group gives three synthesis programs. The first is a general synthesis program for systems whose loss polynomial (cubic or quartic) is controlled by dominant feedback elements. These are varied in such a way that the loss polynomial converges on a desired one in a series of iterations. The second illustrates the design of an active filter (a resonator) as discussed in Section 4.3. The third illustrates the design of an equalizer whose loss is a ratio of polynomials—a biquadratic function.

The third group shows how delay is incorporated into system and circuit design. The first of the programs, “CFE”, finds the locus of roots of the classic feedback equation when the delay is varied. The second consists of two similar programs, each of which incorporates delay into the Design B case study amplifier. Both can be used with the general synthesis program. The third of these delay programs develops the characteristics of a transistor as a binomial loss function with delay in a way that is suitable for use with analysis

programs. Finally, program “PCM” gives the design of a quantized feedback system for removing dc wander from an ac coupled digital signal. In a quantized feedback system, the delay of the feedback signal governs the performance.

As we begin to use the calculator for more complex jobs involving a considerable amount of input data, a printer becomes almost essential to be able to scan the data. The calculator display is simply not large enough to do the job. The programs described here and in Appendix C are written assuming the presence of a printer. It is possible to convert the programs for use without the printer by the use of the prompting system used, for example, in program “RQU” in Appendix A. Similarly, the printer command, “PRREGX” can be replaced by “XEQ 04” where label 04 is the subroutine given in program “N”.

APPENDIX B PROGRAMS

Analysis of Circuits, Including Sensitivities

- B-1 Program “CM”: Lists components and device parameters for Design B. No computations are performed.
- B-2 Program “AN1”: Simple analysis of Design B; delay is not taken into account.
- B-3 Program “SCX”: Calculation of coefficient-to-component sensitivities for a feedback system. Used in conjunction with an analysis program (e.g., “AN1”).
- B-4 Program “SLX”: loss-to-component sensitivities as a function of frequency.
- B-5 Program “STAT”: mean and standard deviation of loss and phase for a system.

Synthesis

- B-6 Program “SJ”: general feedback system synthesis.
- B-7 Program “OPTRES”: finds optimum design of a single-amplifier biquad resonator (discussed in Chapter 4).
- B-8 Program “LED”: design of an equalizer for the bass response of a loudspeaker, illustrating rational function synthesis.

The Effects of Delay

- B-9 Program “CFE”: finds the root locations and root locus of the classic feedback equation. Loci are found for variation of either the control time constant or the delay.
- B-10 Programs “AN2” and “AN3”: analysis by binomial factors of the Design B amplifier using a two-step hierarchy of equations. Includes delay, load capacitance, and stabilizing capacitance.
- B-11 Program “DEV”: finds the individual stage dc loss, time constant and delay from the transistor characteristics and parasitics, and dc collector currents for the transistors of the Design B amplifier.
- B-12 Program “PCM”: quantized feedback design for a PCM system.

B-1 PROGRAM "CM": DESIGN B PARAMETERS

This program lists the components and device parameters of the Design B amplifier described in Chapter 3. It serves the function of giving the user the register locations and values of the parameters in registers R41–R61 and also serves as a checklist to ensure that all parameters are included. (see Fig. B1.1).

This program also lists the system requirements expressed as a quartic polynomial (the quartic coefficient may be zero) and a cutoff frequency. These are stored in registers R36–R39, with the cutoff frequency in R40. To use the program, execute "CM". No calculations are made in this program. This program was written to be used in conjunction with analysis programs such as "AN1", "AN2", or "AN3" as well as the synthesis program "SJ", all given in this appendix.

A listing like this one is intended for use in the analysis and synthesis of other amplifiers or feedback systems of arbitrary configurations in which three dominant control elements can be identified and used to realize a given set of

```

01*LBL "CM"
02 *5/16/80"
03 SF 12
04 "DES. B PAR."
05 AVIEW
06 CF 12
07 ADV
08 "GF, CF, G2:"
09 AVIEW
10 " = X0, X1, X2:"
11 AVIEW
12 41.043
13 PRREGX
14 ADV
15 "RG, GL:"
16 AVIEW
17 45.046
18 PRREGX
19 ADV
20 SF 13
21 "R1-R3"
22 AVIEW
23 CF 13
24 47.049
25 PRREGX
26 ADV
27 8
28 ACCHR
29 "1-3:"
30 ACA
31 PRBUF
32 50.052
33 PRREGX
34 ADV
35 "T1-T3:"
36 AVIEW
37 53.055
38 PRREGX
39 ADV
40 "Td1-3"
41 AVIEW
42 56.058
43 PRREGX
44 ADV
45 "CL, CB, G3:"
46 AVIEW
47 59.061
48 PRREGX
49 ADV
50 "b1-b4"
51 AVIEW
52 36.039
53 PRREGX
54 ADV
55 "2PI F0"
56 AVIEW
57 40
58 PRREGX
59 ADV
60 "OK?"
61 PROMPT
62 END
CAT 1
LBL"CM
END
221 BYTES

```

Figure B1.1. Program "CM": listing.

polynomial coefficient requirements. For each such design, a program carrying the label "CM" should be written or adapted from this program. In all such programs the three dominant elements are stored in R41–R43. Registers R45–R61 (this may be extended if the capacity of the machine allows) are available for storing circuit parameters—dc losses, stage time constants, delays, and the like. This program was written to be used in conjunction with analysis programs such as "AN1", "AN2", or "AN3" as well as the synthesis program "SJ", all given in this appendix.

An example of the output of this program is given in Section B-2, where the program is used to identify the values of the components of the amplifier analyzed by program "AN1". In the example note that the components are identified by both name and register location in which they are to be found. To change any component, store the new value in the indicated register.

1 Program "CM": Registers

R35	Index
R36–R39	Normalized polynomial
R40	F_0
R41–R43	G_F, C_F, G_2
R45–R61	Components

B-2 PROGRAM "AN1": ANALYSIS OF DESIGN B

This program implements the analysis given in eqs. (3.3-3) to (3.3-6) (in Chapter 3), calculating a_0 through a_3 . Those equations include the assumption that $r_3G_2 \ll 1$, and the program makes this assumption with flag 04 cleared. If flag 04 is set, it includes the effect of the denominator $1 - r_3G_2$; the equations programmed in the latter case are

$$a_0 = -R_G G_F$$

$$a_1 = -R_G C_F - \frac{r_1 \tau_2 G_2 (1 + r_3 G_L)}{1 - r_3 G_2}$$

$$a_2 = -\frac{R_G \tau_1 \tau_2 G_2 (1 + r_3 G_L) + r_1 \tau_2 \tau_3 (G_2 + G_L)}{1 - r_3 G_2}$$

$$a_3 = -\frac{R_G \tau_1 \tau_2 \tau_3 (G_2 + G_L)}{1 - r_3 G_2}$$

Thus the program calculates the polynomial coefficients in the simple case discussed in Chapter 3 and does not consider delay, input defect current of the devices, and similar factors. It is intended to demonstrate the use of programs "SCX", "SLX", and "SJ" in a simple case.

For a listing of the component and device parameters needed for this program, execute program "CM" and store the desired values in the indicated

DES. B

INCL. FEEDTHRU?
SF 04

R20= -0.0110
R21= -0.2200
R22= -2.2000
R23= -11.0000

RCL 45
XEQ "CM"

DES. B PAR.

GF, CF, G2:
= X0, X1, X2:

R41= 0.0110
R42= 0.1100
R43= 1.0000

RG, GL:

R45= 1.0000
R46= 10.0000

r1-r3

R47= 0.1000
R48= 0.1000
R49= 0.0100

Δ1-3:

R50= 0.0000
R51= 0.0000
R52= 0.0000

T1-T3:

R53= 1.0000
R54= 1.0000
R55= 1.0000

Td1-3

R56= 0.0000
R57= 0.0000
R58= 0.0000

CL, CB, G3:

R59= 0.0000
R60= 0.0000
R61= 0.0000

b1-b4

R36= 2.0000
R37= 2.0000
R38= 1.0000
R39= 0.0000

2PI F0

R40= 0.1000

OK?

XEQ "AN1"

DES. B

INCL. FEEDTHRU?
SF 04

R20= -0.0110
R21= -0.2200
R22= -2.2000
R23= -11.0000

SF 04
XEQ "AN1"

DES. B

INCL. FEEDTHRU?
SF 04

R20= -0.0110
R21= -0.2211
R22= -2.2222
R23= -11.1111

Figure B2.1. Programs "CM" and "AN1": examples.

```

01*LBL "AN1"          41 FS? 04          81 ST- IND 18
02 ADV              42 RCL 02          82 RCL 02
03 SF 12            43 FS? 04          83 FS? 04
04 "DES. B"        44 /              84 ST/ IND 18
05 AVIEW           45 CHS          85 RCL 45
06 CF 12           46 RCL 45          86 RCL 53
07 ADV             47 RCL 42          87 *
08 "INCL. FEEDTHRU?" 48 *              88 RCL 54
09 AVIEW           49 -              89 *
10 "SF 04"         50 ISG 18          90 RCL 55
11 AVIEW           51 DEG          91 *
12 20.019          52 STO IND 18     92 CHS
13 STO 18          53 RCL 46          93 ENTER↑
14 XEQ "AN"        54 RCL 49          94 RCL 46
15 20.023          55 *              95 RCL 43
16 PRREGX          56 1              96 +
17 RTN             57 +              97 *
                  58 RCL 53          98 ISG 18
                  59 *              99 DEG
18*LBL "AN"        60 RCL 45          100 STO IND 18
19 1               61 *              101 RCL 02
20 RCL 43          62 CHS          102 FS? 04
21 RCL 49          63 RCL 47          103 ST/ IND 18
22 *              64 RCL 55          104 RTN
23 -              65 *              105 .END.
24 STO 02         66 -
25 RCL 45         67 RCL 54
26 RCL 41         68 *
27 *              69 RCL 43
28 CHS           70 *
29 STO IND 18    71 ISG 18          LBL"CM
30 RCL 46         72 DEG          END          221 BYTES
31 RCL 49         73 STO IND 18    LBL"AN1
32 *              74 RCL 46          LBL"AN
33 1              75 RCL 47          END          211 BYTES
34 +              76 *              .END.       07 BYTES
35 RCL 43         77 RCL 54
36 *              78 *
37 RCL 54         79 RCL 55
38 *              80 *
39 RCL 47
40 *

```

Figure B2.2. Program "AN1": listing.

registers. To use the program, execute "AN1". The polynomial coefficients are stored in R20–R23 and are printed.

Example

The example in Fig. B2.1 shows the analysis of the Design B amplifier, both ignoring and including the effect of feedforward, or direct feedthrough in the third stage. Preceding the analyses, execute "CM" to display the component values and device parameters, then execute "AN1". The program listing is shown in Fig. B2.2.

component value is stored in register X, where X can be any register not used by "SCX". (Program "CM" can be used to locate the parameters.) It is used here to calculate the sensitivities of the components of Design B in Chapter 3, as given in Table 3.3. When executed, the program prompts for the X register desired. If the sensitivity of the a_i to G_F of Design B is required, for example, key in 41, the register containing G_F . If sensitivities of several or all components are desired, key in 41.0nn, where nn is the register of the last component desired. The program will calculate sensitivities for all components in registers R41-Rnn.

```

01*LBL "SCX"          43 1
02 ADV              44 -
03 SF 12            45 STO 33      LBL*CM
04 SF 13            46 RCL 15      END          221 BYTES
05 " AI"           47 RCL 22      LBL*ANI
06 AVIEW           48 /        LBL*AN
07 CF 13           49 1        END          204 BYTES
08 " S"            50 -        LBL*SCX
09 AVIEW           51 STO 32      .END.        198 BYTES
10 " X"            52 RCL 14
11 AVIEW           53 RCL 21
12 ADV             54 /
13 1.01            55 1
14 STO 19          56 -
15 SF 13           57 STO 31
16 "I=0,1,2,3"    58 RCL 13
17 AVIEW           59 RCL 20
18 ADV             60 /
19 CF 12           61 1
20 CF 13           62 -
21 ADV             63 STO 30
22 "X REGISTERS?" 64 RCL 19
23 PROMPT          65 1
24 STO 04          66 -
                  67 ST/ 33
25*LBL 01          68 ST/ 32
26 ΣREG 20         69 ST/ 31
27 CLΣ            70 ST/ 30
28 ΣREG 13        71 *S, X = "
29 CLΣ            72 ACA
30 20.019         73 RCL 04
31 STO 18         74 FIX 0
32 XEQ "AN"       75 ACX
33 RCL 19         76 PRBUF
34 ST* IND 04     77 FIX 3
35 13.012         78 30.033
36 STO 18         79 PRREGX
37 XEQ "AN"       80 ADV
38 RCL 19         81 FIX 4
39 ST/ IND 04     82 ISG 04
40 RCL 16         83 GTO 01
41 RCL 23         84 .END.
42 /

```

CAT 1

PRKEYS

PRKEYS:

11 ΣREG

-11 SIZE

14 "AN"

15 "SCX"

-21 PRKEYS

-22 RND

-23 PRFLAGS

24 "CM"

25 "AN1"

-42 DSE

-51 FS?C

-62 PRREGX

-63 PRBUF

-72 DEL

-73 WDTAX

-74 RDTAX

-83 R-D

PRFLAGS

STATUS:

SIZE= 070

Σ= 01

DEG

FIX 4

FLAGS:

F 00 CLEAR

F 01 CLEAR

F 02 CLEAR

F 03 CLEAR

F 04 CLEAR

Figure B3.2. Program "SCX": listing.

In operation, the nominal coefficient values are calculated and stored in R20–R23. Next, the component whose sensitivities are sought is changed by 1%, and the polynomial coefficients are recalculated, this time storing them in R13–R16. The differences are calculated, divided by the nominal value, and multiplied by 100. The result is then printed out for the four cubic coefficients.

Example

The example shows the calculation of the coefficient to component sensitivities for each circuit element of Design B in Chapter 3. It is run with programs "CM" and "AN1" in the machine. Program "CM" is executed first to give the component values. Program "SCX" is then run (see Fig. B3.1.) In response to the prompt for the X registers, the number 41.055 is keyed in order to calculate the sensitivities to all components in registers R41–R55. The resulting listing appears in Table 3.3. The program listing is given in Fig. B3.2.

B-4 PROGRAM "SLX": SENSITIVITY OF LOSS AND PHASE TO COMPONENT X

This program develops the data for plots such as those in Figs. 3.6 and 3.9. It is similar to program "SLC" in Appendix A, except that the separate sensitivities of the real and imaginary parts for each coefficient are added together to give the sensitivity of loss to the given component. Input data required are the polynomial coefficients (entered here as positive numbers), to be stored in registers R00–R0n, and the coefficient-to-component sensitivities for the given component (e.g., as obtained from program "SCX") to be stored in R10–R1n. The program prompts for the minimum and maximum frequencies and the frequency increment (flag 00 set) or the number of points per decade (flag 00 cleared). The program prints the sum of the real and imaginary parts of the component sensitivities expressed in dB and degrees.

Examples

The sensitivity of loss to r_3 of Design B as a function of frequency is shown in the example (see Fig. B4.1.) The coefficient-to-component sensitivities of the example in Section B-3 (the bottom four numbers of column 3) are stored in R10–R13, and the polynomial (from program "AN") is stored in R00–R03. The program is then run. A linear frequency increment is obtained (as in "BORAD") by setting flag 00. The sensitivity of loss to r_3 is then found as a function of frequency, as plotted in Fig. 3.6.

This process is repeated in columns 2 and 3 of Fig. B4.1 to find the sensitivities of loss to G_L and to τ_2 .

Note that unlike program "SCX", which must be run with an analysis program such as "AN1" in program memory, this program does not use other programs in the machine. Program "SLX" is listed in Fig. B4.2.

XEQ "SLX"				XEQ "SLX"						XEQ "SLX"					
SENS, LOSS TO CMPNT X				SENS, LOSS TO CMPNT X						SENS, LOSS TO CMPNT X					
DEG?				DEG?						DEG?					
3.0000 RUN				3.0000 RUN						3.0000 RUN					
S, COEF TO X:				S, COEF TO X:						S, COEF TO X:					
R10= 0.0000				R10= 0.0000						R10= 0.0000					
R11= 0.0450				R11= 0.0450						R11= 0.5000					
R12= 0.0450				R12= 0.5000						R12= 1.0000					
R13= 0.0000				R13= 0.9090						R13= 1.0000					
OK?				OK?						OK?					
RUN				RUN						RUN					
L POLY:				L POLY:						L POLY:					
R00= 0.0110				R00= 0.0110						R00= 0.0110					
R01= 0.2200				R01= 0.2200						R01= 0.2200					
R02= 2.2000				R02= 2.2000						R02= 2.2000					
R03= 11.0000				R03= 11.0000						R03= 11.0000					
OK?				OK?						OK?					
RUN				RUN						RUN					
ΔFMIN,MAX				ΔFMIN,MAX						ΔFMIN,MAX					
R24= 0.0200				R24= 0.0200						R24= 0.0200					
R25= 0.2000				R25= 0.2000						R25= 0.2000					
ΔF INC				ΔF INC						ΔF INC					
R26= 0.0200				R26= 0.0200						R26= 0.0200					
OK?				OK?						OK?					
RUN				RUN						RUN					
L				L						L					
S X :				S X :						S X :					
ΔF dB DEG				ΔF dB DEG						ΔF dB DEG					
0.0200 -34.7 78.2				0.0200 -27.7 141.9						0.0200 -13.6 89.5					
0.0400 -28.2 64.5				0.0400 -15.9 140.6						0.0400 -6.7 86.3					
0.0600 -24.2 46.8				0.0600 -8.4 127.5						0.0600 -2.0 77.8					
0.0800 -21.7 24.2				0.0800 -3.5 107.1						0.0800 1.3 62.9					
0.1000 -20.9 0.0				0.1000 -0.8 84.3						0.1000 3.0 45.0					
0.1200 -21.5 -20.1				0.1200 0.2 65.1						0.1200 3.3 30.1					
0.1400 -22.6 -34.4				0.1400 0.4 51.5						0.1400 3.0 20.0					
0.1600 -23.8 -44.3				0.1600 0.4 42.2						0.1600 2.6 13.7					
0.1800 -25.0 -51.2				0.1800 0.2 35.6						0.1800 2.2 9.7					
0.2000 -26.0 -56.3				0.2000 0.1 30.8						0.2000 1.9 7.1					

Figure B4.1. Program "SLX": examples (Design B). Left column: r_3 ; middle column: G_L ; right column: τ_2 .

Program "SLX": Registers

R01–R07	Loss polynomial coefficients
R08	Degree of loss polynomial
R09	Used
R10–R17	Sensitivities of polynomial coefficient to component

```

01*LBL "SLX"          36 STO 09          71 " X"
02 FIX 4              37 CF 22          72 AVIEW
03 SF 12              38*LBL 03          73 CF 12
04 "SENS. LOSS"      39 CLX             74 " ΔF"
05 AVIEW              40 X(> IND 09     75 ACA
06 "TO CMPNT X"      41 ISG 09         76 5
07 AVIEW              42 GT0 03         77 SKPCHR
08 CF 12              43 CF 22         78 "dB"
09 ADV                44 "OK?"         79 ACA
                    45 PROMPT        80 SKPCHR
                    46 FS? 22        81 "DEG"
                    47 GT0 02        82 ACA
                    48*LBL 04        83 PRBUF
14 "S, COEF TO X:"  49 FIX 4          84 ADV
15 AVIEW              50 "ΔFMIN,MAX"    85 RCL 26
16 1 E3               51 AVIEW          86 FS? 00
17 /                  52 24.025        87 GT0 05
18 10.01              53 PRREGX        88 1/X
19 +                  54 FS? 00        89 10↑X
20 PRREGX             55 "ΔF INC"      90*LBL 05
21 "OK?"              56 FC? 00        91 STO 27
22 CF 22              57 "PTS/DEC:"    92 RCL 24
23 PROMPT             58 AVIEW          93 STO 20
24 FS? 22             59 26             94*LBL C
25 GT0 01             60 PRREGX        95 XEQ E
                    61 CF 22          96 XEQ A
                    62 "OK?"         97 FIX 4
                    63 PROMPT        98 RCL 27
                    64 FS? 22        99 FC? 00
                    65 GT0 04       100 ST+ 20
                    66 SF 12       101 FS? 00
                    67 " L"       102 ST+ 20
                    68 AVIEW       103 RCL 25
                    69 "S ::"    104 RCL 20
                    70 AVIEW       105 X<=Y?
26*LBL 02
27 "L POLY:"
28 AVIEW
29 RCL 08
30 1 E3
31 /
32 PRREGX
33 RCL 08
34 1.007
35 +

```

Figure B4.2. Program "SLX": listing.

- R18, R19 —
- R20 Angular frequency
- R21 Used
- R22,R23 Loss magnitude and phase
- R24, R25 F_{\min} , F_{\max}
- R26 Points/decade (CF00) or ΔF increment
- R27 Used
- R28 Degree index
- R29 Loop index
- R30–R33 —
- R34, R35 Real and imaginary parts of sensitivities
- R36 Sensitivity loop index

106 GTO C	141 STO 22	175 X<>Y
107 RTN	142 X<>Y	176 P-R
	143 STO 23	177 ST+ 34
108*LBL E	144 RTN	178 X<>Y
109 RCL 20		179 ST+ 35
110 X†2	145*LBL A	180 FIX 4
111 CHS	146 0	181 ISG 29
112 STO 21	147 STO 34	182 DEG
113 RCL 07	148 STO 35	183 DSE 28
114 *	149 RCL 08	184 DEG
115 RCL 05	150 STO 28	185 DSE 36
116 +	151 1	186 DEG
117 RCL 21	152 +	187 DSE 29
118 *	153 STO 29	188 GTO a
119 RCL 03	154 9	189 CLA
120 +	155 +	190 ARCL 20
121 RCL 21	156 STO 36	191 RCL 35
122 *		192 RCL 34
123 RCL 01	157*LBL a	193 R-P
124 +	158 RCL 20	194 X=0?
125 RCL 20	159 RCL 28	195 1 E-8
126 *	160 Y†X	196 LOG
127 RCL 06	161 DSE 29	197 20
128 RCL 21	162 DEG	198 *
129 *	163 RCL IND 29	199 "+ "
130 RCL 04	164 *	200 FIX 1
131 +	165 RCL 22	201 ARCL X
132 RCL 21	166 /	202 "+ "
133 *	167 STO 37	203 ARCL Y
134 RCL 02	168 RCL IND 36	204 AVIEW
135 +	169 *	205 FIX 4
136 RCL 21	170 90	206 END
137 *	171 RCL 28	CAT 1
138 RCL 00	172 *	
139 +	173 RCL 23	LBL'SLX
140 R-P	174 -	END
		431 BYTES

Figure B4.2. Continued.

B-5 PROGRAM "STAT": MEAN AND STANDARD DEVIATION OF LOSS AND PHASE

This program calculates the mean and standard deviation of the loss in db and phase in degrees at a given frequency (see Figs. B5.1 to B5.4). The mean and standard deviation of the variation of each of the components of a system must be known or estimated. Required input data are the polynomial coefficients, to be stored in R20-R2n. Other input data are called for as the program runs, including the number of components.

Equations are developed in Section 3.5. Coefficient-to-component sensitivities such as those given in Table 3.3 are keyed in manually as called for by prompts in the program. The mean and standard deviation for the variation of each component are also keyed in manually. Since there are many key entries to be made, an identifying tone system has been included: a different set of

two tones is sounded for each sensitivity value, and a different set of tones is sounded for component mean and standard deviation entry.

A list of coefficient-to-component sensitivities is required, as previously calculated for the system using program "SCX" in conjunction with an analysis program for the system such as program "ANI" for the Design B amplifier. Such a program produces a list of sensitivities: as shown in the example for "SCX", for a cubic system there are four sensitivity values for each component, and these are printed out in groups of four.

These are the sensitivity values to be *entered manually* in running this program. The program then calculates the sensitivity of loss to each coefficient and multiplies it by the keyed in sensitivity to obtain the sensitivity of loss to the component, a complex quantity. To calculate the mean of the loss variation, this latter sensitivity is multiplied by the mean of the component variation and added to a running sum of such loss variations. To calculate the standard deviation, this sensitivity is multiplied by the standard deviation of the component value, squared, and added to a running sum of squared deviation values. Real and imaginary components of the preceding summed values are added separately. Thus there are two running sums for the mean and two more for the standard deviation.

When the specified number of components have been processed in this fashion, the total of the real part of the mean gives the magnitude change in nepers; this is converted to dB. The imaginary part of the mean sum gives the phase change in radians, and this is converted to degrees. The standard deviation is handled in exactly the same way, except that the square root of the real and imaginary parts is taken before conversion to db and degrees.

In using the program, wait for the tone before keying in the component sensitivities or means and standard deviations. If data are keyed in incorrectly for any component, it can be reentered when the prompt "OK?" appears. If the data are correct, press R/S. If the data are no good ("NG"), press N and R/S. In the latter case program execution returns to the beginning of data entry for the component. This is illustrated in Fig. B5.1.

Example 1 carries out the calculation for the circuit in Fig. 3.10 by using the numbers cited in Section 3.6. The coefficient to component sensitivities are given in eq. (3.6-6); we take R as X_0 , G as X_1 , and C as X_2 . In Example 1 there are three components, the degree is unity, and we analyze the statistical variation at an angular frequency of 0.5. The polynomial is $L = 1 + s$, and the two coefficients are printed. The loss is then calculated, and the program prompts for the coefficient-to-component sensitivities. The response is 1 for the sensitivity of a_0 to X_0 and 1 for that of a_1 to X_0 ; we mistakenly keyed in 37. The program then prompts for the mean of the variation of X_0 , 0.1, and the standard deviation of X_0 , 0.02. The error prompt comes next. Since 37 is incorrect, we press N and R/S and reenter the data, as shown. The program then steps to the next component. When data for all three have been keyed in correctly, the mean of the loss variation in decibels and degrees is calculated and printed, followed by the standard deviation.

```

                                XEQ "STAT"
MEAN AND                                RUN
STD. DEV.                                S, COEF. TO X1?
LOSS + PH.                                1.0000 RUN
                                           0.0000 RUN
NO. OF CMPNTS?                             $\mu$ 1?
    3.0000 RUN                            .1 RUN
DEG?                                        $\sigma$ 1?
    1.0000 RUN                            .02 RUN
4F?                                       OK? R/S
    .5000 RUN                            NG? N, R/S
                                           RUN
POLY                                       S, COEF. TO X2?
                                           0.0000 RUN
R00= 1.0000                                1.0000 RUN
R01= 1.0000                                 $\mu$ 2?
OK?                                         0 RUN
                                            $\sigma$ 2?
                                RUN                            .02 RUN
LOSS:                                       OK? R/S
1.0 dB, 426.6 DEG                            NG? N, R/S
S, COEF. TO X0?                                RUN
    1.0000 RUN                                 $\mu$ L =
    37.0000 RUN                                -0.17 dB, 4-2.29 DEG
 $\mu$ 0?                                          $\sigma$ L =
    -.1 RUN                                    0.23 dB, 40.65 DEG
 $\sigma$ 0?
    .02 RUN
OK? R/S
NG? N, R/S
N RUN
S, COEF. TO X0?
    1.0000 RUN
    1.0000 RUN
 $\mu$ 0?
    -.1 RUN
 $\sigma$ 0?
    .02 RUN
OK? R/S
NG? N, R/S

```

Figure B5.1. Program "STAT": Example 1.

In the Example 2 (Fig. B5.2) the component sensitivities are taken from the example given for program "SCX" in this appendix. Only components with nonzero sensitivities (of which there are 10) are included. Total running time for all 10 components is under 10 minutes including data entry.

Note that in the program listing (see Fig. B5.3) four steps are not programmable from the keyboard but were added to the program by using the bar code wand. The steps are numbers 148, 155, 181, and 199 and include the lowercase Greek characters μ and σ . If a wand is not available, substitute keyboard-accessible symbols for the mean and standard deviation. If a wand is available, μ and σ can be placed in the program line by reading the bar code given under

```

      XEQ "STAT"
MEAN AND          .5450  RUN       $\mu$ 6?
STD. DEV.        .0910  RUN
LOSS + PH.       $\sigma$ 2?           $\sigma$ 6?          -.1  RUN
NO. OF CMPNTS?          .1  RUN          .05  RUN
      10.0000  RUN      OK? R/S          OK? R/S
DEG?              3.0000  RUN      NG? N, R/S          NG? N, R/S
      .1000  RUN          S, COEF. TO X7?
POLY              1.0000  RUN          0.0000  RUN
      .5000  RUN          0.0000  RUN
      1.0000  RUN          1.0000  RUN
R00= 0.0110       $\mu$ 3?           $\sigma$ 7?          -.1  RUN
R01= 0.2200       $\sigma$ 3?          OK? R/S          .05  RUN
R02= 2.2000      .01  RUN      NG? N, R/S
R03= 11.0000     OK? R/S          S, COEF. TO X8?
OK?              NG? N, R/S          0.0000  RUN
LOSS:            S, COEF. TO X4?          .5000  RUN
-36.2 dB,  $\Delta$ 135.0 DEG      0.0000  RUN          1.0000  RUN
S, COEF. TO X0?      .0450  RUN          1.0000  RUN
      1.0000  RUN          .4500  RUN
      0.0000  RUN          .9090  RUN
      0.0000  RUN       $\mu$ 4?           $\sigma$ 8?          -.1  RUN
      0.0000  RUN          .1  RUN          .05  RUN
      .1  RUN       $\sigma$ 4?          OK? R/S          .05  RUN
      .01  RUN      NG? N, R/S          NG? N, R/S
OK? R/S          S, COEF. TO X5?          S, COEF. TO X9?
NG? N, R/S      0.0000  RUN          0.0000  RUN
      RUN          .5000  RUN          .5000  RUN
S, COEF. TO X1?      .5000  RUN          1.0000  RUN
      0.0000  RUN          0.0000  RUN
      .5000  RUN           $\mu$ 9?          -.1  RUN
      0.0000  RUN          0.0000  RUN
      0.0000  RUN       $\mu$ 5?           $\sigma$ 9?          .05  RUN
      0.0000  RUN          -.1  RUN          OK? R/S
      0  RUN       $\sigma$ 5?          NG? N, R/S
      .01  RUN      OK? R/S          RUN
      .01  RUN      NG? N, R/S           $\mu$ L =
OK? R/S          S, COEF. TO X6?          -1.35 dB,  $\Delta$ -14.61 DEG
NG? N, R/S      0.0000  RUN           $\sigma$ L =
      RUN          .0450  RUN          0.62 dB,  $\Delta$ 5.00 DEG
S, COEF. TO X2?      .0450  RUN
      0.0000  RUN          .0450  RUN
      .5000  RUN          0.0000  RUN

```

Figure B5.2. Program "STAT": Example 2.

```

01*LBL "STAT"
02 SF 12
03 "MEAN AND"
04 AVIEW
05 "STD. DEV."
06 AVIEW
07 "LOSS + PH."
08 AVIEW
09 CF 12
10 ADV
11 SF 04
12 "NO. OF CMPNTS?"
13 PROMPT
14 STO 09
15 "DEG?"
16 PROMPT
17 STO 08
18 "ΔF?"
19 PROMPT
20 STO 20
21 10.01
22 STO 36
23 ADV

24*LBL 01
25 "POLY"
26 AVIEW
27 10.01
28 STO 36
29 RCL 08
30 1 E3
31 /
32 STO 18
33 PRREGX
34 RCL 08
35 1.007
36 +
37 STO 30

38*LBL 02
39 CLX
40 X<> IND 30
41 ISG 30
42 GTO 02
43 CF 22

44 "OK?"
45 PROMPT
46 FS? 22
47 GTO 01
48 "LOSS:"
49 AVIEW
50 CLA
51 XE0 E
52 FIX 1
53 RCL 22
54 LOG
55 20
56 *
57 RND
58 ARCL X
59 "F dB, Δ"
60 ARCL 23
61 "F DEG"
62 AVIEW
63 ADV

64*LBL B
65 0
66 STO 28
67 STO 29
68 ΣREG 30
69 CLΣ

70*LBL 07
71 CLA
72 "S, COEF. TO X"
73 RCL 35
74 FIX 0
75 CF 29
76 ARCL X
77 "F?"
78 AVIEW
79 RCL 18

80*LBL 03
81 TONE IND 35
82 TONE IND X
83 FIX 4
84 RCL 36
85 +

86 STOP
87 STO IND Y
88 RDN
89 RCL 36
90 -
91 ISG X
92 GTO 03
93 R↑
94 R↑
95 "N"
96 ASTO Y
97 AON
98 "OK? R/S"
99 AVIEW
100 "NG? N, R/S"
101 PROMPT
102 ASTO X
103 AOFF
104 X*Y?
105 GTO 08
106 RDN
107 RDN
108 GTO 07

109*LBL 08
110 0
111 STO 24
112 STO 25
113 RCL 18
114 STO 19
115 RCL 36
116 +
117 STO 37

118*LBL 14
119 RCL IND 37
120 FIX 4
121 RCL IND 19
122 *
123 RCL 20
124 RCL 19
125 RCL 18
126 -
127 Y↑X
128 *

```

Figure B5.3. Program "STAT": listing.

the program listing.* Note that while μ is a displayable character, σ appears in the display as a "wagon wheel" (a square wheel at that). Both print nicely, however.

*These bar codes are taken from a bar code character table in W. C. Wickes, *Synthetic Programming on the HP-41C*, Larken Publications, College Park, MA, 1980.

129 RCL 22	173 RCL 24	217 FIX 4
130 /	174 RCL 26	218 RTN
131 RCL 19	175 *	
132 RCL 18	176 ST+ 28	219*LBL E
133 -	177 RCL 35	220 RCL 20
134 90	178 RCL 09	221 X+Y
135 *	179 X+Y?	222 CHS
136 RCL 23	180 GTO 07	223 STO 21
137 -	181 "uL ="	224 RCL 07
138 X<>Y	182 AVIEW	225 *
139 P-R	183 CLA	226 RCL 05
140 ST+ 24	184 RCL 28	227 +
141 X<>Y	185 E+X	228 RCL 21
142 ST+ 25	186 LOG	229 *
143 ISG 37	187 20	230 RCL 03
144 DEG	188 *	231 +
145 ISG 19	189 FIX 2	232 RCL 21
146 GTO 14	190 ARCL X	233 *
147 FIX 0	191 "t dB, Δ"	234 RCL 01
148 "v"	192 RCL 29	235 +
149 ARCL 35	193 R-D	236 RCL 20
150 "t?"	194 ARCL X	237 *
151 AVIEW	195 "t DEG"	238 RCL 06
152 TONE 6	196 AVIEW	239 RCL 21
153 STOP	197 CLA	240 *
154 STO 26	198 ADV	241 RCL 04
155 "σ"	199 "σL ="	242 +
156 ARCL 35	200 AVIEW	243 RCL 21
157 "t?"	201 CLA	244 *
158 TONE 8	202 RCL 31	245 RCL 02
159 TONE 9	203 SORT	246 +
160 PROMPT	204 E+X	247 RCL 21
161 STO 27	205 LOG	248 *
162 FIX 4	206 20	249 RCL 00
163 RCL 25	207 *	250 +
164 *	208 ARCL X	251 R-P
165 RCL 24	209 "t dB, Δ"	252 STO 22
166 RCL 27	210 RCL 33	253 X<>Y
167 *	211 SORT	254 STO 23
168 E+	212 R-D	255 END
169 RCL 25	213 ARCL X	CAT 1
170 RCL 26	214 "t DEG"	CAT 1
171 *	215 AVIEW	LBL*STAT
172 ST+ 29	216 CLA	END
		529 BYTES



Figure B5.3. Continued.

Program "STAT": Registers

R00–R07	Loss polynomial coefficients
R08	Degree
R09	Number of components
R10–R17	Buffer store for $S_{X_i}^a$ values
R18	Sensitivity index
R19	Coefficient index
R20	Angular frequency ω
R21	$-\omega^2$
R22, R23	Loss magnitude and phase
R24	Sum of real part $S_{a_i}^L$
R25	Sum of imaginary part $S_{a_i}^L$
R26	Mean of component variation
R27	Sigma of component variation
R28	Magnitude variation of the mean in nepers
R29	Phase variation of the mean in radians
R30–R35	Statistics registers
R31	Square of standard deviation of loss (nepers) ²
R33	Square of standard deviation of phase (radians) ²
R35	Component index
R36	10.010
R37	Index for coefficient-to-component sensitivities

B-6 PROGRAM "SJ": GENERAL FEEDBACK SYSTEM SYNTHESIS (USE WITH "MAT" AND "AN")

This program implements the synthesis method discussed in Section 3.7 and is generally applicable to the synthesis of linear and quasilinear single-input, single-output systems, or systems that can be cast into this form. This applies to systems whose loss can be expressed by a polynomial of up to fourth degree. Within these limitations the synthesis program is not tied to any specific system. The characteristics of the specific system are contained in an analysis program, written according to guidelines given in conjunction with the description of programs "AN2" and "AN3" in this appendix. Program "SJ" calls on such an analysis program as well as a component identification program, "CM", also tied to a specific system. In addition, program "MAT" a general matrix inversion and multiplication program described later in this section, is required.

To make sufficient room in program memory of the HP 41C for these programs, three memory modules must be used; this means that the printer and the card reader must be used alternately in the fourth plug-in slot. When this is done, 544 bytes (77.7 registers) are available in the program memory for the "AN" and "CM" programs. The HP 41CV contains all user memory internally, so that printer and card reader can be used together. In this

machine 990 bytes are available for the other programs. The HP 41C with a quad memory module also has the larger capacity.

Program "SJ" operates by finding numerical values of the nine partial derivatives of eq. (3.7-8), three at a time, relating three dominant feedback elements to three polynomial coefficients under at least partial control of these elements. It does this by calling on the analysis program (which must carry the label "AN") to analyze the system in the nominal case—that is, with the three dominant elements that have their initially selected values. It then perturbs the value of each of these elements by 1% one at a time and recalculates the coefficients, again using the analysis program. The 1% change in each coefficient is then divided by the change in the element to obtain the partial derivative. In this process the analysis program is invoked four times, once for the nominal case and once when each dominant element is perturbed. The nine partial derivatives thus found yield a Jacobian matrix that is then inverted by use of program "MAT".

The inverted matrix is then postmultiplied by a column vector whose three components are the functions whose values are to approach zero. These functions are the difference between the desired values of the polynomial coefficients and the values actually obtained in the nominal case. The three vector components resulting from this multiplication give the *change* in the values of the three dominant elements needed for the next iteration.

The feedback system is then reanalyzed by using the revised values of the dominant elements to obtain revised polynomial coefficient values. These values are normalized and compared with the design specification. If differences still exist, the process is repeated.

This procedure will usually converge in a few iterations even where the coefficient values are not as strongly tied to the dominant elements as they are in the Design B examples shown.

Provision is made in program "SJ" for monitoring a quartic coefficient in case the analysis program calculates it. If the quartic coefficient is found to be greater than the value stored in register R39, the iteration halts, allowing the user to modify the circuit to reduce the quartic. An example of this is given in Chapter 5, where C_B is introduced into Design B for this purpose. Alternatively, the design specification can be modified at this point. In either case, the iteration can proceed from this point.

Program "MAT"

This auxiliary program inverts a 3×3 matrix and multiplies it by a column vector. It is adapted with minimal changes from the HP-67 *Standard Pac*, pages 10-01–10-07 and L10-01.

Memory locations of the data have been coordinated among the several programs involved in analysis, synthesis, and sensitivity analysis. The organization of the data is shown in Fig. B6.1.

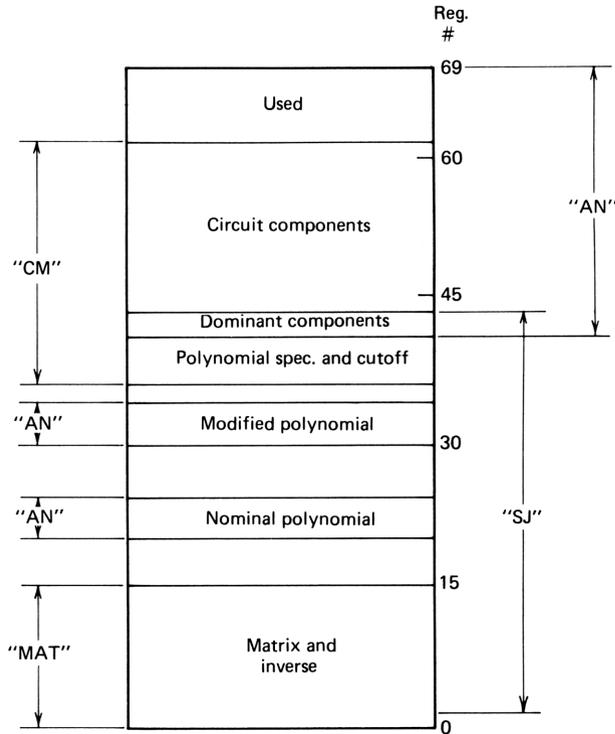


Figure B6.1. Diagram showing data register locations for the various programs and subroutines used with program "SJ".

Examples

Two examples serve to show how "SJ" operates. Both use analysis program "AN1" for the Design B amplifier; the first program ignores direct feedthrough (see Fig. B6.2), whereas the second (see Fig. B6.3) takes it into account. As noted in the text, convergence in the first example is immediate since the relation between the dominant components and the coefficients is linear. In the second example three iterations are required to complete the convergence from a remote starting point (compare the initial and final values of G_F , C_F , and G_2).

The first two examples show the synthesis of a Design B amplifier having a Butterworth cutoff response. The third example (see Fig. B6.4) compares the synthesis for Chebyshev (0.2 dB ripple width) and MFD responses *at the same asymptotic bandwidth*. This is the appropriate comparison where the transistors control the bandwidth. The component values (except for the 3 dominant components) were the same for all cases; printing of the components was suppressed (by a trivial program change) in the latter two examples. Note that

```

          CF 04
          XEQ "SJ"
JAC. SYN.
DES. B PAR.

GF, CF, G2:
= X0, X1, X2:

R41= 10.000
R42= 10.000
R43= 10.000

RG, GL:

R45= 1.000
R46= 10.000

r1-r3

R47= 0.100
R48= 0.100
R49= 0.010

Δ1-3

R50= 0.000
R51= 0.000
R52= 0.000

T1-T3:

R53= 1.000
R54= 1.000
R55= 1.000

Td1-3

R56= 0.000
R57= 0.000
R58= 0.000

CL, CB, G3:

R59= 0.000
R60= 0.000
R61= 0.000

b1-b4

R36= 2.000
R37= 2.000
R38= 1.000
R39= 0.000

2PI F0

R40= 0.100

OK?

RUN

X0, X1, X2:

R41= 0.0110
R42= 0.1100
R43= 1.0000

a0, b1-b4:

R20= -0.0110
R21= 2.0000
R22= 2.0000
R23= 1.0000
R24= 0.0000

DONE
    
```

Figure B6.2. Program "SJ": example using "AN1" ignoring direct feedthrough (clear flag 04).

G_2 (X2) for the MFD case is about 3 times that for the Chebyshev and $1\frac{1}{2}$ times that for the Butterworth design. To compare the performance as to magnitude and phase response and coefficient sensitivities, see Fig. 2.26.

It should be emphasized that these designs ignore delay and thus are not practical. Designs developed in Chapter 5 and later in this appendix include delay and are practical. Listings for programs "SJ" and "MAT" are given in Figs. B6.5 and B6.6.

```

                SF 04
                CF 00
                XEQ "SJ"
JAC. SYN.      CL, CB, G3:
                a0, b1-b4:
DES. B PAR.    R59= 0.000
                R60= 0.000
                R61= 0.000
                R20= -0.0111
                R21= 2.0071
                R22= 2.0079
                R23= 1.0007
                R24= 0.0000
GF, CF, G2:    b1-b4
= X0, X1, X2:
R41= 10.000
R42= 10.000
R43= 10.000
                R36= 2.000
                R37= 2.000
                R38= 1.000
                R39= 0.000
                X0, X1, X2:
                R41= 0.0111
                R42= 0.1111
                R43= 1.0000
RG, GL:        2PI F0
R45= 1.000
R46= 10.000
                R40= 0.100
                a0, b1-b4:
r1-r3          OK?
                R20= -0.0111
                R21= 2.0000
                R22= 2.0000
                R23= 1.0000
                R24= 0.0000
R47= 0.100
R48= 0.100
R49= 0.010
                X0, X1, X2:
                R41= 0.0111
                R42= 0.1111
                R43= 1.8273
                DONE
Δ1-3:
R50= 0.000
R51= 0.000
R52= 0.000
                a0, b1-b4:
                R20= -0.0111
                R21= 2.8427
                R22= 2.9269
                R23= 1.0043
                R24= 0.0000
T1-T3:
R53= 1.000
R54= 1.000
R55= 1.000
                X0, X1, X2:
                R41= 0.0111
                R42= 0.1111
                R43= 1.0071
Td1-3
R56= 0.000
R57= 0.000
R58= 0.000

```

Figure B6.3. Program "SJ": example using "AN1" with direct feedthrough (clear flag 04).

B-7 PROGRAM "OPTRES": OPTIMUM ACTIVE RESONATOR DESIGN

This program uses the development described in Section 4.3 to design an active resonator that has equal contributions to loss variation from the passive elements and from the amplifier. The program requires estimates of the standard deviations of the amplifier time constant and the passive resistors; the deviations of the capacitors are assumed to be sufficiently small not to affect the outcome significantly. Other input data required are the resonant frequency (in R16) the required Q (in R15), the amplifier time constant τ_1 (in R06), and

XEQ "SJ" JAC. SYN.	XEQ "SJ" JAC. SYN.
b1-b4	b1-b4
R36= 1.891	R36= 2.470
R37= 1.554	R37= 2.430
R38= 1.000	R38= 1.000
R39= 0.000	R39= 0.000
2PI F0	2PI F0
R40= 0.100	R40= 0.100
OK?	OK?
RUN	RUN
X0, X1 ,X2:	X0, X1 ,X2:
R41= 0.0106	R41= 0.0117
R42= 0.1415	R42= 0.1214
R43= 0.5330	R43= 1.5036
a0, b1-b4:	a0, b1-b4:
R20= -0.0106	R20= -0.0117
R21= 1.8931	R21= 2.4791
R22= 1.5570	R22= 2.4400
R23= 1.0002	R23= 1.0009
R24= 0.0000	R24= 0.0000
X0, X1 ,X2:	X0, X1 ,X2:
R41= 0.0106	R41= 0.0117
R42= 0.1415	R42= 0.1214
R43= 0.5307	R43= 1.4943
a0, b1-b4:	a0, b1-b4:
R20= -0.0106	R20= -0.0117
R21= 1.8907	R21= 2.4700
R22= 1.5544	R22= 2.4300
R23= 1.0000	R23= 1.0000
R24= 0.0000	R24= 0.0000
DONE	DONE

Figure B6.4. Program "SJ": examples comparing Chebyshev (0.2 dB) (left column) and MFD (right column) synthesis (same asymptotic cutoff).

01*LBL "SJ"	41 30.029	81 DSE 16
02 "5/16/80"	42 STO 18	82 RCL 17
03 SF 12	43 XEQ "AN"	83 RCL 40
04 "JAC. SYN."	44 RCL 19	84 *
05 AVIEW	45 ST/ IND 44	85 ST- IND 16
06 CF 12	46 RCL 32	86 RCL 19
07 ADV	47 RCL 22	87 1
08 FC? 00	48 -	88 -
09 XEQ "CM"	49 STO IND 16	89 RCL IND 44
10 SF 00	50 RCL 31	90 *
	51 RCL 21	91 2
11*LBL 05	52 -	92 ST+ 16
12 FC? 02	53 DSE 16	93 RDN
13 CF 21	54 STO IND 16	94 ST/ IND 16
14 ΣREG 20	55 RCL 30	95 DSE 16
15 CLΣ	56 RCL 20	96 ST/ IND 16
16 CF 05	57 -	97 DSE 16
17 1.01	58 DSE 16	98 ST/ IND 16
18 STO 19	59 STO IND 16	99 RTN
19 20.019	60 RCL 33	
20 STO 18	61 RCL 23	100*LBL 22
21 XEQ "AN"	62 -	101 RCL 23
22 20.024	63 RCL 40	102 RCL 40
23 PRREGX	64 *	103 *
24 43	65 RCL 38	104 RCL 38
25 STO 44	66 /	105 /
26 9	67 STO 17	106 STO 10
27 STO 16	68 RCL 37	107 STO 11
28 XEQ D	69 *	108 STO 12
29 DSE 44	70 2	109 RCL 37
30 DSE 16	71 ST+ 16	110 ST* 12
31 XEQ D	72 RDN	111 RCL 22
32 DSE 44	73 ST- IND 16	112 ST- 12
33 DSE 16	74 DSE 16	113 RCL 40
34 XEQ D	75 RCL 40	114 ST* 10
35 GTO 22	76 ST* 17	115 ST* 11
	77 RCL 17	116 RCL 36
36*LBL D	78 RCL 36	117 ST* 11
37 ΣREG 30	79 *	118 RCL 21
38 CLΣ	80 ST- IND 16	119 ST- 11
39 RCL 19		120 RCL 40
40 ST* IND 44		

Figure B6.5. Program "SJ": listing.

the frequency compensation time constant τ_2 (in R07). If double-slope frequency compensation is absent, a large time constant is to be stored in R07 (e.g., $10^6\tau_1$). In addition, the impedance level (we chose 10 kΩ) is stored in R24. The program calculates the optimum value of R_A/R_B from (4.3-31):

$$\frac{R_A}{R_B} = \frac{\sigma_A L_a}{4\sigma_R} \quad \text{where} \quad L_a = \frac{\tau_1 \tau_2 s^2}{1 + \tau_2 s} \quad (\text{B7-1})$$

```

121 ST* 10
122 RCL 20
123 ST- 10
124 1.009
125 PRREGX
126 ADV
127 10.012
128 PRREGX
129 ADV
130 XEQ "INV"
131 XEQ "MUL"
132 ST+ 41
133 VIEW X
134 RDN
135 ST+ 42
136 VIEW X
137 RDN
138 ST+ 43
139 VIEW X
140 ADV
141 FIX 4
142 SF 21
143 "X0, X1 ,X2:"
144 AVIEW
145 41.043
146 PRREGX
147 ADV

148*LBL 30
149 ΣREG 20
150 CLS
151 20.019
152 STO 18
153 XEQ "AN"

154*LBL 87
155 "a0, b1-b4:"
156 AVIEW
157 RCL 40
158 RCL 20
159 /
160 ST* 21

161 ST* 22
162 ST* 23
163 ST* 24
164 RCL 40
165 ST* 22
166 ST* 24
167 X↑2
168 ST* 23
169 ST* 24
170 20.024
171 PRREGX
172 ADV
173 FIX 3
174 RCL 22
175 RCL 37
176 -
177 RND
178 X#0?
179 SF 01
180 RCL 21
181 RCL 36
182 -
183 RND
184 X#0?
185 SF 01
186 "DONE"
187 FC? 01
188 AVIEW
189 FC?C 01
190 RTN

191*LBL 06
192 SF 00
193 RCL 39
194 RCL 24
195 ">b4 LIM."
196 X>Y?
197 PROMPT
198 GTO 05
199 END

LBL'SJ
END 446 BYTES
LBL'MAT
LBL'INV
LBL'MUL
END 263 BYTES
LBL'CM
END 222 BYTES
LBL'ANB
LBL'AN
END 319 BYTES
.END. 03 BYTES
CAT 1

```

Figure B6.5. Continued.

The values of R_A and R_B are calculated individually from the impedance level:

$$R_B = R_0 \sqrt{R_B/R_A}, \quad R_A = \frac{R_0}{\sqrt{R_B/R_A}} \quad (B7-2)$$

Then $1/Q_0$ is obtained from (4.3-15)

$$\frac{1}{Q_0} = \frac{2}{(R_B/R_A)^{1/2}} \quad (B7-3)$$

01*LBL "MAT"	41 XEQ 07	81 RCL 01
02*LBL 10	42 RCL 02	82 RCL 05
03 0	43 RCL 07	83 *
04 GTO 05		84 RCL 02
	44*LBL 07	85 RCL 04
05*LBL 11	45 ISG 15	86 XEQ 03
06 3	46 DEG	87 STO 00
07 GTO 05	47 RCL IND 15	88 CLX
	48 *	89 RCL 03
08*LBL 12	49 *	90 RCL 00
09 6	50 +	91 *
10 GTO 05	51 RTN	92 RCL 02
		93 RCL 09
11*LBL 13	52*LBL "INV"	94 XEQ 03
12 9	53 XEQ 15	95 STO 01
	54 1/X	96 CLX
13*LBL 05	55 ADV	97 RCL 02
14 STO 15	56 PRX	98 RCL 06
15 XEQ 06	57 RCL 01	99 *
16 XEQ 06	58 RCL 09	100 RCL 03
	59 *	101 RCL 05
17*LBL 06	60 RCL 03	102 XEQ 03
18 R↑	61 RCL 07	103 STO 03
19 ISG 15	62 XEQ 03	104 CLX
20 DEG	63 STO 13	105 RCL 05
21 STO IND 15	64 CLX	106 RCL 09
22 RTN	65 RCL 03	107 *
	66 RCL 04	108 RCL 06
23*LBL 15	67 *	109 RCL 08
24 0	68 RCL 01	110 XEQ 03
25 STO 15	69 RCL 06	111 STO 02
26 RCL 06	70 XEQ 03	112 CLX
27 RCL 08	71 STO 14	113 RCL 06
28 XEQ 07	72 CLX	114 RCL 07
29 RCL 04	73 RCL 02	115 *
30 RCL 09	74 RCL 07	116 RCL 04
31 XEQ 07	75 *	117 RCL 09
32 RCL 05	76 RCL 01	118 XEQ 03
33 RCL 07	77 RCL 08	119 STO 06
34 XEQ 07	78 XEQ 03	120 CLX
35 CHS	79 STO 15	
36 RCL 03	80 CLX	
37 RCL 08		
38 XEQ 07		
39 RCL 01		
40 RCL 09		

Figure B6.6. Program "MAT": listing.

We then obtain the positive feedback ratio n from

$$\frac{n}{1-n} = \frac{1}{2Q_0} \left(\frac{1}{Q_0} - \frac{1}{Q} \right) = b \quad (\text{B7-4})$$

whereupon

$$n = \frac{b}{1+b} \quad (\text{B7-5})$$

```

121 RCL 04
122 RCL 08
123 *
124 RCL 05
125 RCL 07
126 XEQ 03
127 RCL 15
128 RCL 00
129 XEQ 12
130 RCL 02
131 RCL 01
132 RCL 03
133 XEQ 10
134 RCL 06
135 RCL 13
136 RCL 14
137 XEQ 11
138 CLX
139 RTN

140*LBL 03
141 *
142 -
143 *
144 RTN

145*LBL "MUL"
146 1
147 STO 15
148 XEQ 01
149 STO 13
150 2
151 STO 15
152 XEQ 01
153 STO 14
154 3
155 STO 15
156 XEQ 01
157 STO 00
158 0
159 RCL 00
160 RCL 14

161 RCL 13
162 RTN

163*LBL 01
164 0
165 RCL 10
166 XEQ 04
167 RCL 11
168 XEQ 04
169 RCL 12
170 XEQ 04
171 RTN

172*LBL 04
173 RCL IND 15
174 *
175 +
176 3
177 ST+ 15
178 RDN
179 RTN
180 END

```

Figure B6.6. Continued.

We take $C_A = C_B$, so that

$$C_A = C_B = \frac{1}{R_0 \omega_0} \tag{B7-6}$$

The nominal loss at resonance is given by

$$L_{\text{res., nom.}} = \frac{1}{Q} \sqrt{\frac{R_B}{R_A}} \tag{B7-7}$$

The sensitivities of loss to the amplifier loss and the passive components at resonance are

$$S_{L_a}^L = \frac{|L_a|}{\frac{1}{Q}\sqrt{R_B/R_A}} = QL_a\sqrt{\frac{R_A}{R_B}} \quad (\text{B7-8})$$

and

$$S_R^L = \frac{Q}{Q_0} \quad (\text{B7-9})$$

The standard deviation of loss is computed from (4.3-30):

$$\sigma_L = 2Q \left[(\sigma_A |L_a| Q_0)^2 + \left(\frac{\sigma_R}{Q_0} \right)^2 \right]^{1/2} \quad (\text{B7-10})$$

Finally, the upper bound on the Q and resonant frequency deviations is calculated for the optimized case from (4.3-33) and (4.3-44)

$$\sigma_Q < Q\sqrt{2\sigma_R\sigma_A\tau_1\omega_0} \quad (\text{B7-11})$$

and

$$\sigma_{\omega_0} < \sqrt{2\sigma_R\sigma_A\tau_1\omega_0} \quad (\text{B7-12})$$

Examples

Two examples are given (see Fig. B7.1), as described in Section 4.3. In Example 1 a simple 741-type amplifier is used ($\tau_1 = 1.59 \times 10^{-4}$ ms) to realize a 2.0 kHz resonator with a Q of 20. Note that the units used are volts, milliamperes, and milliseconds, so that derived units are kilohms, microfarads, kilohertz, and kiloradians per second. Impedance level R_0 is taken as 10 k Ω . The standard deviation of the amplifier time constant is 15%, and that of the resistors is 0.12%. All results are printed, including the amplifier loss at resonance $1/Q_0$ and circuit loss at resonance. The sensitivities of loss to the amplifier (0.16) and the resistors (10) are also printed. Following this, the standard deviation of loss at resonance is given (3.4%) followed by its interpretation as an upper bound on the standard deviation of Q (1.7%) and the resonant frequency (0.085%). Finally, all circuit values are printed.

Example 2 (in Fig. B7.1) repeats this procedure for a 10 kHz resonator with the use of an amplifier having double slope (“T”) compensation. Sensitivity performance, even at 10 kHz, is twice as good as the simple 2 kHz resonator.

Program “OPTRES” listing is given in Fig. B7.2.

```

Example 1.
La = τ1s
f0 = 2 kHz
XEQ "OPTRES"
OPTIMUM
RESONATOR
DESIGN
INPUT:
σA, σR
R09= 0.1500
R10= 0.0012
R0
R00= 10.0000
La: T1, T2:
R06= 1.592-04
R07= 1.000+03
Q, 2PI F0:
R11= 20.0000
R12= 12.5664
RESULTS:
La AT RES.
-53.98dB < 90.00
1/Q0 = 0.50
L0 AT RES. = -38.06dB

SENSITIVITIES:
AT RES. OF L TO
La: 0.1600
R: 10.0014
σL = 33.9-03
OPTIMUM CASE:
UPPER BOUND ON
σQ = 17.0-03
σΔF/F = 849.-06
RA, RB, CA, CB, N:
R01= 40.0+00
R02= 2.50+00
R03= 7.96-03
R04= 7.96-03
R05= 101.-03
DONE.

R0
R00= 10.0000
La: T1, T2:
R06= 1.592-04
R07= 1.000+03
Q, 2PI F0:
R11= 20.0000
R12= 62.8319
RESULTS:
La AT RES.
-66.03dB < 177.14
1/Q0 = 0.25
L0 AT RES. = -44.09dB

Example 2.
La =  $\frac{\tau_1 \tau_2 s^2}{1 + \tau_2 s}$ 
XEQ "OPTRES"
OPTIMUM
RESONATOR
DESIGN
INPUT:
σA, σR
R09= 0.1500
R10= 0.0012
R0
R00= 10.0000
La: T1, T2:
R06= 1.592-04
R07= 1.000+03
Q, 2PI F0:
R11= 20.0000
R12= 12.5664
RESULTS:
La AT RES.
-53.98dB < 90.00
1/Q0 = 0.50
L0 AT RES. = -38.06dB

SENSITIVITIES:
AT RES. OF L TO
La: 0.0000
R: 4.9976
σL = 17.0-03
OPTIMUM CASE:
UPPER BOUND ON
σQ = 8.48-03
σΔF/F = 424.-06
RA, RB, CA, CB, N:
R01= 80.0+00
R02= 1.25+00
R03= 1.59-03
R04= 1.59-03
R05= 24.4-03
DONE.

```

Figure B7.1. Program "OPTRES": examples.

B-8 PROGRAM "LED": LOUDSPEAKER EQUALIZER DESIGN

This program (see Figs. B8.1 and B8.2) illustrates the design of an equalizer network having a biquadratic response requirement; the focus is on the feedforward aspects of the design since the feedback portion was already done in the Sallen-Key analysis in Chapter 4.

The program finds the values of the circuit elements for the low-frequency equalization of a closed-box loudspeaker for the circuit in Fig. 4.15, as

```

01*LBL "OPTRES"          41 11.012              81 RCL 09
02 SF 12                 42 PRREGX              82 *
03 ADV                   43 ADV                 83 RCL 10
04 "OPTIMUM"            44 -1                  84 /
05 AVIEW                 45 RCL 07              85 STO 15
06 "RESONATOR"          46 RCL 12              86 SQRT
07 AVIEW                 47 *                   87 RCL 00
08 "DESIGN"             48 1/X                 88 *
09 AVIEW                 49 CHS                 89 STO 02
10 CF 12                 50 R-P                 90 RCL 15
11 ADV                   51 1/X                 91 /
12 "INPUT:"             52 RCL 12              92 STO 01
13 AVIEW                 53 *                   93 RCL 15
14 ADV                   54 RCL 06              94 SQRT
15 9                     55 *                   95 2
16 ACCHR                 56 X<>Y               96 *
17 "R, "                57 CHS                 97 STO 16
18 ACA                   58 STO 14              98 RCL 11
19 9                     59 X<>Y               99 1/X
20 ACCHR                 60 STO 13              100 -
21 "R:"                 61 ADV                 101 RCL 16
22 ACA                   62 "RESULTS:"          102 "1/Q0 = "
23 PRBUF                 63 AVIEW               103 ACA
24 9.01                  64 ADV                 104 ACX
25 PRREGX                65 LOG                 105 PRBUF
26 ADV                   66 20                  106 *
27 "R0"                 67 *                   107 2
28 AVIEW                 68 "La AT RES."        108 /
29 0                     69 AVIEW               109 ENTER↑
30 PRREGX                70 FIX 2               110 ENTER↑
31 ADV                   71 ACX                 111 1
32 "La: T1, T2:"        72 "dB <"             112 +
33 AVIEW                 73 ACA                 113 /
34 SCI 3                 74 RCL 14              114 STO 05
35 6.007                 75 ACX                 115 RCL 12
36 PRREGX                76 ADV                 116 RCL 00
37 ADV                   77 ADV                 117 *
38 FIX 4                 78 RCL 13              118 1/X
39 "Q, 2PI F0:"         79 4                   119 STO 03
40 AVIEW                 80 /                   120 STO 04

```

Figure B7.2. Program "OPTRES": listing.

discussed in Section 4.4. Input data consists of the resonant frequency of the loudspeaker, its actual Q , the resonant frequency desired, and the Q desired. The values of R_1 , R_2 , and R_F are also specified initially. The program finds values of C_1 , C_2 , $1/G_0$, C_0 , and $1/G_{FF}$.

Example

The example shows the computation for the speaker and equalizer discussed in section 4.4. Registers R10–R12 contain the numerator coefficients, and R13–R15 contain the denominator coefficients.

```

121 RCL 15
122 SQRT
123 RCL 11
124 /
125 LOG
126 20
127 *
128 ADV
129 "L "
130 ACA
131 5
132 ACCHR
133 " AT RES. = "
134 ACA
135 RDN
136 ACX
137 "dB"
138 ACA
139 PRBUF
140 ADV
141 "SENSITIVITIES:"
142 AVIEW
143 "AT RES. OF L TO"
144 AVIEW
145 "La: "
146 ACA
147 RCL 15
148 SQRT
149 1/X
150 RCL 11
151 *
152 RCL 13
153 *
154 STO 00
155 FIX 4
156 ACX
157 ADV
158 "R: "
159 ACA
160 RCL 11
161 RCL 16
162 *
163 ACX
164 ADV
165 9
166 ACCHR
167 "L = "
168 ACA
169 RCL 09
170 RCL 13
171 *
172 RCL 16
173 /
174 X^2
175 RCL 10
176 RCL 16
177 *
178 X^2
179 +
180 SQRT
181 RCL 11
182 *
183 2
184 *
185 ENG 2
186 ACX
187 PRBUF
188 ADV
189 "OPTIMUM CASE:"
190 AVIEW
191 "UPPER BOUND ON"
192 AVIEW
193 9
194 ACCHR
195 "Q = "
196 ACA
197 RCL 09
198 RCL 10
199 *
200 RCL 13
201 *
202 2
203 *
204 SQRT
205 RCL 11
206 *
207 ACX
208 ADV
209 9
210 ACCHR
211 8
212 ACCHR
213 "F/F = "
214 ACA
215 RDN
216 RDN
217 RCL 11
218 /
219 ACX
220 ADV
221 ADV
222 "RA, RB, CA, CB,"
223 "F N:"
224 AVIEW
225 1.005
226 PRREGX
227 ADV
228 "DONE."
229 AVIEW
230 FIX 4
231 ADV
232 ADV
233 END
CAT 1
LBL*OPTRES
END
.END.
513 BYTES
05 BYTES

```

Figure B7.2. Continued.

B-9 PROGRAMS "CFE" AND "DE": CLASSIC FEEDBACK EQUATION

Program "CFE" calculates the root loci as the delay or the control time constant is varied in the equation

$$L(s) = 1 + \tau_1 s e^{\tau_d s} \tag{B9-1}$$

Two normalizations are used; the first, obtained by executing "CFE", normalizes the frequency variable to the control time constant τ_1 . Setting $p = \tau_1 s$,

$$L(p) = 1 + p e^{T_d p} \tag{B9-2}$$

where we have set $T_d = \tau_d / \tau_1$.

```

                                XEQ "LED"
SPKR EQ DES                                N:
QE, QS, FE, FS                                R10= 1.0000
                                                R11= 4.0000
                                                R12= 16.0000
R16= 1.0000
R17= 1.0000
R18= 0.0398
R19= 0.0995
                                                D:
R1, R2, RF:                                R13= 6.2500
                                                R14= 10.0000
                                                R15= 16.0000
R01= 100.0000
R02= 100.0000
R03= 100.0000
                                                C1, C2:
                                                R08= 0.0800
                                                R09= 0.0200
                                                1/G0, C0, 1/GFF
                                                R20= 19.0476
                                                R21= 0.0600
                                                R22= 100.0000

```

Figure B8.1. Program "LED": example.

For real roots, as shown in section 5.2, we have $\omega = 0$ and $p = \sigma$, so that

$$\ln(-\sigma) = -\sigma T_d$$

or

$$T_d = -\frac{\ln(-\sigma)}{\sigma} \quad (\text{B9-3})$$

For complex roots, $p = \sigma + j\omega$; as noted in the text, we choose as a dummy variable the angle $T_d\omega$. This is the angle from the negative real axis to the line passing through the origin and the root position. From eqs. (5.2-14) to (5.2-16), we find σ , ω , and T_d , and these are displayed or printed (no printer is required for this program).

Use of the program is illustrated in the example shown in Fig. B9.1. After executing "CFE", the program prompts the user to set flag 04 if the dummy variable angle $T_d\omega$ is to be expressed in degrees. The program then prompts for the dummy variable.

If the angle is zero, the roots are real, and the program prompts for sigma (σ). It then displays the position of the root on the real axis and calculates and displays T_d . If the data are other than zero, the program calculates the complex root position and displays it. It then calculates T_d and displays it. The program then repeats the prompt for the next angle. By responding with an angle greater than 360° , higher-order root locations can be found.

```

01*LBL "LED"          41 RCL 17          81 STO 22
02 SF 12              42 ST/ 14         82 RCL 13
03 *SPKR. EQ."       43 1             83 RCL 10
04 AVIEW             44 STO 10         84 RCL 05
05 CF 12             45 "N:"         85 *
06 ADV              46 AVIEW         86 -
07 *QE, QS, FE, FS" 47 10.012        87 RCL 03
08 AVIEW            48 PRREGX        88 /
09 16.019           49 ADV           89 1/X
10 PRREGX           50 "D:"         90 STO 20
11 ADV              51 AVIEW         91 RCL 14
12 *R1, R2, RF:"    52 13.015        92 RCL 11
13 AVIEW            53 PRREGX        93 RCL 05
14 1.003            54 ADV           94 *
15 PRREGX           55 RCL 11        95 -
16 ADV              56 RCL 01        96 RCL 03
17 RCL 18           57 RCL 02        97 /
18 PI               58 +            98 STO 21
19 2                59 /            99 *1/G0, C0, 1/GFF"
20 *               60 STO 09        100 AVIEW
21 *               61 RCL 12        101 20.022
22 1/X             62 RCL 01        102 PRREGX
23 STO 11          63 /            103 ADV
24 X+2            64 RCL 02        104 .END.
25 STO 12          65 /
26 STO 15          66 RCL 09
27 RCL 19          67 /
28 RCL 18          68 STO 08
29 /              69 *C1, C2:"
30 X+2            70 AVIEW
31 STO 13          71 8.009
32 STO 14          72 PRREGX
33 RCL 19          73 ADV
34 PI             74 RCL 12
35 2              75 RCL 15
36 *              76 /
37 *              77 STO 05
38 ST/ 14         78 RCL 03
39 RCL 16         79 /
40 ST/ 11         80 1/X

                                CAT 1
                                .END.      224 BYTES

                                PRKEYS

                                USER KEYS:
                                11 ZREG
                                -11 SIZE
                                -21 PRKEYS
                                -22 RND
                                -23 PRFLAGS
                                25 "LED"

                                PRFLAGS

                                STATUS:
                                SIZE= 024

```

Figure B8.2. Program "LED": listing.

The second normalization, corresponding to that used to plot the root locus in Fig. 5.8, is obtained by executing "DE" (see Fig. B9.2). Here, the frequency variable is normalized to the delay time, so that $p = \tau_d s$. Hence

$$L(p) = 1 + T_1 p e^p \tag{B9-4}$$

where we have set $T_1 = \tau_1 / \tau_d$. In this case we can write for the real roots

$$\ln(-\sigma T_1) = -\sigma$$

	XEQ "CFE"		Td*OMEGA?		Td*OMEGA?
Td*OMEGA:			15.000	RUN	390.000 RUN
IN DEG? SF 04			S= -2.566 +J0.688		114.216.932 +J65.943.177
			Td = 0.381		Td = 1.032E-4
Td*OMEGA?			Td*OMEGA?		Td*OMEGA?
	SF 04		30.000	RUN	420.000 RUN
SIGMA? 0.000	RUN		S= -2.145 +J1.238		S= -34.434 +J59.642
S= -1.000 +J0.000	RUN		Td = 0.423		Td = 0.123
Td = 0.000			Td*OMEGA?		Td*OMEGA?
Td*OMEGA?			45.000	RUN	435.000 RUN
SIGMA? 0.000	RUN		S= -1.551 +J1.551		S= -1.979 +J7.386
S= -1.500 +J0.000	RUN		Td = 0.506		Td = 1.028
Td = 0.270			Td*OMEGA?		Td*OMEGA?
Td*OMEGA?			60.000	RUN	420.000 RUN
SIGMA? 0.000	RUN		S= -0.915 +J1.585		S= -34.434 +J59.642
S= -2.000 +J0.000	RUN		Td = 0.661		Td = 0.123
Td = 0.347			Td*OMEGA?		Td*OMEGA?
Td*OMEGA?			75.000	RUN	425.000 RUN
SIGMA? 0.000	RUN		S= -0.368 +J1.372		S= -13.432 +J28.804
S= -2.000 +J0.000	RUN		Td = 0.954		Td = 0.258
Td = 0.347			Td*OMEGA?		Td*OMEGA?
Td*OMEGA?			90.000	RUN	430.000 RUN
SIGMA? 0.000	RUN		S= 1.745E-10 +J1.000		S= -5.252 +J14.431
S= -2.500 +J0.000	RUN		Td = 1.571		Td = 0.520
Td = 0.367			Td*OMEGA?		Td*OMEGA?
Td*OMEGA?			105.000	RUN	
SIGMA? 0.000	RUN		S= 0.158 +J0.591		
S= -4.000 +J0.000	RUN		Td = 3.100		
Td = 0.347			Td*OMEGA?		
Td*OMEGA?			120.000	RUN	
SIGMA? 0.000	RUN		S= 0.149 +J0.258		
S= -10.000 +J0.000	RUN		Td = 8.104		
Td = 0.230					

Figure B9.1. Program "CFE": examples.

or

$$T_1 = -\frac{e^{-\sigma}}{\sigma} \quad (\text{B9-5})$$

We can then specify σ and find the value of the control time constant at that value.

```

                                XEQ "DE"

    OMEGA?
        0.000  RUN
    SIGMA?
        -.010  RUN
    S= -0.010 +J0.000
    T1 = 101.005

    OMEGA?
        0.000  RUN
    SIGMA?
        -.100  RUN
    S= -0.100 +J0.000
    T1 = 11.052

    OMEGA?
        0.000  RUN
    SIGMA?
        -.500  RUN
    S= -0.500 +J0.000
    T1 = 3.297

    OMEGA?
        0.000  RUN
    SIGMA?
        -1.000  RUN
    S= -1.000 +J0.000
    T1 = 2.718

    OMEGA?
        0.000  RUN
    SIGMA?
        -2.000  RUN
    S= -2.000 +J0.000
    T1 = 3.695

    OMEGA?
        0.000  RUN
    SIGMA?
        -4.000  RUN
    S= -4.000 +J0.000
    T1 = 13.650

                                OMEGA?
                                        .010  RUN
                                S= -1.000 +J0.010
                                T1 = 2.718

                                OMEGA?
                                        .500  RUN
                                S= -0.915 +J0.500
                                T1 = 2.395

                                OMEGA?
                                        1.000  RUN
                                S= -0.642 +J1.000
                                T1 = 1.599

                                OMEGA?
                                        1.500  RUN
                                S= -0.106 +J1.500
                                T1 = 0.740

                                OMEGA?
                                        1.5708  RUN
                                S= 5.770E-6 +J1.571
                                T1 = 0.637

                                OMEGA?
                                        2.000  RUN
                                S= 0.915 +J2.000
                                T1 = 0.182

                                OMEGA?
                                        3.000  RUN
                                S= 21.046 +J3.000
                                T1 = 3.407E-11

                                OMEGA?
                                        5.000  *
                                        2.000  /
                                        7.854  ***
                                        RUN
                                S= 1.371E-8 +J7.854
                                T1 = 0.127
    
```

Figure B9.2. Program "CFE", subroutine "DE": examples.

For complex roots, we choose a value of ω as the dummy variable (since the normalization is to unit delay) and obtain the root positions by the equations

$$T_1\sigma = -\exp \frac{\omega}{\tan \omega} \cos \omega \tag{B9-6}$$

$$T_1\omega = \exp \frac{\omega}{\tan \omega} \sin \omega \tag{B9-7}$$

01*LBL "CFE"	37 *T1 = "	74 RCL 00
02 SF 21	38 ARCL 03	75 /
03 *TD*OMEGA: "	39 AVIEW	76 STO 03
04 AVIEW	40 CLA	77 1/X
05 *IN DEG? "	41 GTO "DE"	78 STO 04
06 *+SF 04"		79 RTN
07 AVIEW	42*LBL C	
08 CLA	43 ADV	80*LBL A
	44 *+OMEGA?"	81 0
09*LBL B	45 PROMPT	82 STO 02
10 CF 08	46 FC? 04	83 *SIGMA?"
11 CLA	47 R-D	84 PROMPT
12 *Td*"	48 ENTER↑	85 STO 01
13 XEQ C	49 TAN	86 CHS
14 *S= "	50 X=0?	87 FS? 08
15 ARCL 01	51 GTO A	88 GTO a
16 *+ +J*"	52 X<>Y	89 LN
17 ARCL 02	53 D-R	90 RCL 01
18 AVIEW	54 STO 00	91 /
19 *Td = "	55 R-D	92 CHS
20 ARCL 04	56 TAN	93 STO 04
21 AVIEW	57 RCL 00	94 1
22 GTO B	58 X<>Y	95 STO 03
	59 /	96 RTN
23*LBL "DE"	60 E↑X	
24 CF 04	61 STO 05	97*LBL a
25 SF 21	62 CHS	98 E↑X
26 SF 08	63 RCL 00	99 RCL 01
27 CLA	64 R-D	100 /
28 XEQ C	65 COS	101 CHS
29 RCL 04	66 *	102 STO 03
30 ST* 01	67 STO 01	103 1
31 ST* 02	68 RCL 05	104 STO 04
32 *S= "	69 RCL 00	105 END
33 ARCL 01	70 R-D	
34 *+ +J*"	71 SIN	CAT 1
35 ARCL 02	72 *	
36 AVIEW	73 STO 02	END
		213 BYTES

Figure B9.3.

Since we started with a given value of ω , we find T_1 from the second equation and divide it into the first to obtain σ .

Registers

Six registers are used. The dummy variable is stored in R00, σ and ω are stored in R01 and R02, T_1 is in R03, and T_d is in R04; R05 is also used. The program is listed in Fig. B9.3.

B-10 PROGRAMS "AN2" AND "AN3": ANALYSIS PROGRAMS FOR DESIGN B WITH DELAY

These two programs analyze the Design B amplifier with fewer simplifying approximations than used in the analysis given in chapter 3. Program "AN2"

is a simple program that establishes the framework on which the more elaborate program, "AN3" is based. Program "AN2" treats the amplifier polynomial as a cubic, with delay assigned to the feedback path. Program "AN3" treats the amplifier as a quartic and includes the effects of load capacitance C_L and third-stage input-output capacitance C_B . The latter can be augmented to exert separate control of the third-stage characteristics in the presence of arbitrary amounts of delay and load capacitance. The analysis also includes interstage shunt parasitic capacitance and series base resistance.

Each program includes a global label "AN" to allow its use with program "SJ" (that program calls on "AN" as a generic analysis program to be used in synthesis by inversion of a 3×3 Jacobian matrix). Each can be used for analysis only, simply by calling for "AN2" or "AN3". Register R18 contains the address of the registers to be used to store the polynomial coefficients of the analysis.

The analysis is carried out in an essentially two-step hierarchy of circuit equations: the characteristics of each stage are analyzed and stored (as a binomial in frequency plus delay). The stage binomials are then multiplied together to give the loss of the active path. Finally, the feedback polynomial coefficients (including the effect of delay) are added to the result. The final loss polynomial is stored in the registers determined by the address in R18, and the results are printed (when the program is used for analysis only).

To use the programs it is convenient to clear the machine and use a status card to reestablish the machine status. Programs "SJ", "MAT", "AN2" or "AN3", and "CM" are then read into the machine and all flags cleared.* Circuit data can be read into the memory registers from a previously recorded card; registers R36–R61 are filled by using the card and the command 36.061 RDTAX. Then "SJ" is executed.

Examples

The equations programmed for "AN2" are given in Section 5.4, and those for "AN3", in Section 5.5. The examples shown in Figs. B10.1 and B10.2 are for the 300 MHz amplifier design described in Section 5.6; in using "AN2", C_L is not taken into account, nor is C_B . The effect of these capacitances on the design is seen by comparing the two examples. Load capacitance can be important to the design.

Values of the device parameters were obtained from program "DEV", to be described later. In each case the initial values of the dominant elements were (arbitrarily) set to unity.

*When either of these programs is used with "SJ", the HP 41C without the quad memory module requires three memory modules, thus requiring the card reader and printer to share the fourth plug-in port of the calculator.

```

                                XEQ "S.J"
JAC. SYN.                        b1-b4
DES. B PAR.                       R36= 2.000
                                   R37= 1.850
                                   R38= 0.940
                                   R39= 0.230
GF, CF, G2:
= X0, X1, X2:

R41= 1.000                         2PI F0
R42= 1.000                         R40= 1.885
R43= 1.000

RG, GL:                            OK?
                                   RUN
R45= 0.150                         X0, X1 ,X2:
R46= 6.670

r1-r3                               R41= 0.0134
                                   R42= 0.0075
                                   R43= 0.6349

R47= 0.028                         a0, b1-b4:
R48= 0.009                         R20= -0.0024
R49= 0.005                         R21= 2.0040
                                   R22= 1.8534
                                   R23= 0.9394
                                   R24= 0.0000

Δ1-3:                               X0, X1 ,X2:
                                   R41= 0.0134
                                   R42= 0.0075
                                   R43= 0.6329

R50= 0.010                         a0, b1-b4:
R51= 0.019                         R20= -0.0024
R52= 0.010                         R21= 2.0000
                                   R22= 1.8500
                                   R23= 0.9400
                                   R24= 0.0000

T1-T3:

R53= 0.101
R54= 0.076
R55= 0.045

Td1-3

R56= 0.050
R57= 0.115
R58= 0.114

CL, CB, G3:                         DONE

R59= 0.000
R60= 0.000
R61= 0.000

```

Figure B10.1. Program "AN2": example.

Loss Variation

Once the design is completed, the sensitivity of the polynomial coefficients to the components may be evaluated by using program "SCX", as shown in Fig. B10.3. The sensitivities thus generated can then be used to find the mean and standard deviation of the *variation* of loss and phase of the amplifier by using program "STAT" as illustrated in Fig. B10.4. To evaluate accuracy, the means and standard deviations of the individual components must be found. The

```

                                XEQ "SJ"
JAC. SYN.                        b1-b4

DES. B PAR.                      R36= 2.0000
                                R37= 1.8500
                                R38= 0.9400
GF, CF, G2:                      R39= 0.2300
= X0, X1, X2:

R41= 1.0000                      2PI F0
R42= 1.0000                      R40= 1.8350
R43= 1.0000

RG, GL:                          OK?
                                RUN
R45= 0.1500                      X0, X1 ,X2:
R46= 6.6700                      R41= 0.0389
                                R42= 0.0196
r1-r3                             R43= 1.9511
                                a0, b1-b4:
R47= 0.0280                      R20= -0.0070
R48= 0.0085                      R21= 1.9205
R49= 0.0053                      R22= 1.7286
                                R23= 0.9908
                                R24= 0.2288
Δ1-3:                             X0, X1 ,X2:
R50= 0.0100                      R41= 0.0395
R51= 0.0185                      R42= 0.0183
R52= 0.0100                      R43= 2.2319
                                a0, b1-b4:
T1-T3:                             R20= -0.0073
                                R21= 2.0002
R53= 0.1000                      R22= 1.8501
R54= 0.0764                      R23= 0.9400
R55= 0.0446                      R24= 0.2251
                                a0, b1-b4:
Td1-3                             R20= -0.0073
                                R21= 2.0002
R56= 0.0501                      R22= 1.8501
R57= 0.1155                      R23= 0.9400
R58= 0.1140                      R24= 0.2251

CL, CB, G3:                       DONE
R59= 2.5000
R60= 0.3000
R61= 1.3300

```

Figure B10.2. Program "AN3": example.

values used here are merely estimates and are not necessarily typical. Care should be taken to obtain good estimates for the particular integrated circuit process being used.

Preparation of an Analysis Program "AN" for Other Systems

To take advantage of the synthesis program and the several programs for sensitivity analysis, the reader should be able to write an analysis program for

```

XEQ "SCX"
      a i
      S
      X

i=0,1,2,3

X REGISTERS?
      41.061 RUN
S, X = 41. GF
R30= 0.814
R31= -0.210
R32= 0.065
R33= 0.003

S, X = 42. CF
R30= 0.000
R31= 0.356
R32= -0.199
R33= 0.112

S, X = 43. G2
R30= 0.181
R31= 0.784
R32= 0.662
R33= -0.088

S, X = 44. --
R30= 0.000
R31= 0.000
R32= 0.000
R33= 0.000

S, X = 45. RG
R30= 0.823
R31= 0.275
R32= 0.679
R33= 0.766

S, X = 46. GL
R30= 0.015
R31= 0.086
R32= 0.268
R33= 0.354

S, X = 47. rE1
R30= 0.178
R31= 0.730
R32= 0.327
R33= 0.239

S, X = 48. rE2
R30= 0.000
R31= 0.000
R32= 0.000
R33= 0.000

S, X = 49. rE3
R30= 0.013
R31= 0.060
R32= 0.100
R33= 0.095

S, X = 50. Δ1
R30= 0.009
R31= 0.039
R32= 0.017
R33= 0.013

S, X = 51. Δ2
R30= 0.188
R31= 0.129
R32= 0.056
R33= 0.031

S, X = 52. Δ3
R30= 0.007
R31= 0.032
R32= 0.047
R33= 0.032

S, X = 53. τ1
R30= 0.000
R31= 0.090
R32= 0.798
R33= 0.643

S, X = 54. τ2
R30= 0.000
R31= 0.730
R32= 1.086
R33= 0.863

S, X = 55. τ3
R30= 0.000
R31= 0.029
R32= 0.289
R33= 0.774

S, X = 56. τd1
R30= 0.000
R31= -0.038
R32= -0.014
R33= 0.038

S, X = 57. τd2
R30= 0.000
R31= -0.088
R32= -0.033
R33= 0.087

S, X = 58. τd3
R30= 0.000
R31= -0.105
R32= -0.204
R33= -0.214

S, X = 59. CL
R30= 0.000
R31= 0.005
R32= 0.066
R33= 0.373

S, X = 60. CB
R30= 0.000
R31= 0.021
R32= 0.212
R33= 0.323

S, X = 61. G3
R30= 0.005
R31= 0.022
R32= 0.033
R33= 0.023

```

Figure B10.3. Sensitivity analysis of 300 MHz Design B amplifier.

XEQ "STAT"		SNSTVY, aI TO X 2.?	SNSTVY, aI TO X 7.?
3/30/80		.1810 RUN	.0090 RUN
		.7840 RUN	.0390 RUN
		.6620 RUN	.0170 RUN
		-.0880 RUN	.0130 RUN
MEAN AND		μ 2.?	μ 7.?
STD. DEV.		.01 RUN	.1 RUN
LOSS + PH.		σ 2.?	σ 7.?
		.01 RUN	.1 RUN
NO. OF CMPNTS?		SNSTVY, aI TO X 3.?	SNSTVY, aI TO X 8.?
19.0000	RUN	.0230 RUN	.1800 RUN
DEG?		.2750 RUN	.1290 RUN
3.0000	RUN	.6790 RUN	.0560 RUN
2PI F?		.7660 RUN	.0310 RUN
.1500	PI	μ 3.?	μ 8.?
2.0000	*	.01 RUN	.1 RUN
	*	σ 3.?	σ 8.?
0.9425	***	.01 RUN	.1 RUN
	RUN	SNSTVY, aI TO X 4.?	SNSTVY, aI TO X 9.?
POLY		.0150 RUN	.0070 RUN
		.0860 RUN	.0320 RUN
R20= -0.0073		.2680 RUN	.0470 RUN
R21= -0.0077		.3540 RUN	.0320 RUN
R22= -0.0038		μ 4.?	μ 9.?
R23= -0.0010		.01 RUN	.1 RUN
		σ 4.?	σ 9.?
L, dB+PH:		.01 RUN	.1 RUN
-42.3974	***	SNSTVY, aI TO X 5.?	SNSTVY, aI TO X 10.?
-122.0005	***	.1780 RUN	0.0000 RUN
		.7300 RUN	.0900 RUN
SNSTVY, aI TO X 0.?		.3270 RUN	.7980 RUN
.0140	RUN	.2390 RUN	.6430 RUN
-.2100	RUN	μ 5.?	μ 10.?
.0650	RUN	.05 RUN	.1 RUN
.0030	RUN	σ 5.?	σ 10.?
μ 0.?		.05 RUN	.1 RUN
.01	RUN	SNSTVY, aI TO X 6.?	SNSTVY, aI TO X 11.?
σ 0.?		.0130 RUN	0.0000 RUN
.01	RUN	.0600 RUN	.7300 RUN
SNSTVY, aI TO X 1.?		.1000 RUN	1.0060 RUN
0.0000	RUN	.0950 RUN	.8630 RUN
.3560	RUN	μ 6.?	μ 11.?
-.1990	RUN	.05 RUN	.1 RUN
.1120	RUN	σ 6.?	σ 11.?
μ 1.?		.05 RUN	.1 RUN
.01	RUN		
σ 1.?			
.01	RUN		

Figure B10.4. Example: statistical analysis of 300 MHz amplifier at 150 MHz.

						RESULTS :
SNSTVY, aI TO X 12.?			SNSTVY, aI TO X 16.?			
	0.0000	RUN		0.0000	RUN	μ L
	.0290	RUN		.0050	RUN	
	.2890	RUN		.0660	RUN	dB= 0.4127
	.7740	RUN		.3730	RUN	PH= 6.2064
μ 12.?			μ 16.?			
	.1	RUN		.05	RUN	σ L
σ 12.?			σ 16.?			
	.1	RUN		.05	RUN	dB= 0.4315
SNSTVY, aI TO X 13.?			SNSTVY, aI TO X 17.?			PH= 4.7820
	0.0000	RUN		0.0000	RUN	
	-.0300	RUN		.0210	RUN	
	-.0140	RUN		.2120	RUN	
	.0300	RUN		.3230	RUN	
μ 13.?			μ 17.?			
	.1	RUN		.05	RUN	
σ 13.?			σ 17.?			
	.1	RUN		.05	RUN	
SNSTVY, aI TO X 14.?			SNSTVY, aI TO X 18.?			
	0.0000	RUN		.0050	RUN	
	-.0000	RUN		.0220	RUN	
	-.0330	RUN		.0330	RUN	
	.0070	RUN		.0230	RUN	
μ 14.?			μ 18.?			
	.1	RUN		.01	RUN	
σ 14.?			σ 18.?			
	.1	RUN		.01	RUN	
SNSTVY, aI TO X 15.?						
	0.0000	RUN				
	-.1050	RUN				
	-.2040	RUN				
	-.2140	RUN				
μ 15.?						
	.1	RUN				
σ 15.?						
	.1	RUN				

Figure B10.4. Continued.

the circuit or system in question. The means for extending the type of analysis used in the case study amplifier in Chapters 3 and 5 is given in Chapters 8 and 9; here, we are interested in the formal requirements of an analysis program so that it will operate in conjunction with the synthesis program.

The key requirement is that the storage of the polynomial be under the control of the address in register R18, thus allowing the synthesis program to locate the nominal polynomial in one address and the modified polynomial in another. Therefore, the analysis program, in calculating the polynomial coefficients of the system in question, must use indirect storage of these coefficients. Each analysis program discussed so far provides examples of this procedure. Often, calculation of the polynomial coefficients will require multiplication of the polynomial in storage by a binomial and, occasionally, by a quadratic function. A subroutine for multiplying by a linear factor is given in "AN2" or "AN3" in label 40. A subroutine for multiplying by a quadratic (as well as a

```

01*LBL "AN2"          41 XEQ 40          81 RCL 56
02 20.019            42 RCL 47          82 +
03 STO 18            43 RCL 50          83 STO 65
04 XEQ "AN"         44 RCL 45          84 RCL 41
05 20.024            45 *              85 *
06 PRREGX           46 +              86 -
07 RTN              47 STO 63          87 RCL 45
                   48 RCL 53          88 *
                   49 RCL 45          89 ISG 18
                   50 *              90 DEG
08*LBL "AN"         51 STO 64          91 ST- IND 18
09 TONE 0           52 3              92 RCL 65
10 RCL 46            53 ST+ 18         93 2
11 RCL 43            54 RDN             94 /
12 +                55 XEQ 40         95 RCL 41
13 RCL 41            56 GTO 41         96 *
14 +                57*LBL 40          97 RCL 42
15 STO 65           58 RCL IND 18     98 -
16 RCL 49           59 *              99 RCL 65
17 RCL 43           60 ISG 18         100 *
18 *                61 DEG           101 RCL 45
19 RCL 52           62 ST+ IND 18   102 *
20 +                63 DSE 18         103 ISG 18
21 *                64 RCL 63         104 DEG
22 RCL 43           65 ST* IND 18   105 ST- IND 18
23 +                66 RCL 64         106 RCL 65
24 STO IND 18      67 DSE 18         107 X12
25 RCL 55           68 GTO 40         108 2
26 RCL 65           69 RTN             109 /
27 *                70*LBL 41          110 RCL 42
28 RCL 43           71 ISG 18         111 *
29 RCL 58           72 DEG           112 RCL 45
30 *                73 RCL 45         113 *
31 -                74 RCL 41         114 ISG 18
32 ISG 18           75 *              115 DEG
33 DEG             76 ST- IND 18   116 ST- IND 18
34 STO IND 18     77 RCL 42         117 END
35 RCL 51           78 RCL 58
36 CHS             79 RCL 57
37 STO 63         80 +
38 RCL 54
39 CHS
40 STO 64

```

Figure B10.5. Programs "AN2" and "AN3": listing.

linear term) is given in program "PF" in Appendix A. Listings for programs "AN2" and "AN3" are given in Fig. B10.5.

B-11 PROGRAM "DEV": DEVICE-LEVEL ANALYSIS OF THE STAGES OF THE DESIGN B AMPLIFIER

The purpose of this program is to derive the values of the dc loss, the unity loss time constant, and the delay of each of the three transistors of design B from the physical device parameters of the transistors and the dc collector currents.

	CAT 1	01*LBL "AN3"	41 RCL 69
LBL*CM		02 20.019	42 RCL 59
END	221 BYTES	03 STO 18	43 RCL 42
LBL*SJ		04 XEQ "AN"	44 +
END	446 BYTES	05 RTN	45 RCL 60
LBL*MAT			46 +
LBL*INV		06*LBL "AN"	47 STO 66
LBL*NUL		07 RCL 43	48 *
END	263 BYTES	08 RCL 49	49 +
LBL*AN2		09 *	50 RCL 68
LBL*AN		10 STO 68	51 /
END	216 BYTES	11 RCL 61	52 ISG 18
.END.	09 BYTES	12 RCL 49	53 DEG
		13 *	54 STO IND 18
	PRKEYS	14 +	55 RCL 60
		15 RCL 52	56 RCL 49
USER KEYS:		16 +	57 RCL 60
11 ZREG		17 STO 69	58 *
-11 SIZE		18 RCL 46	59 RCL 68
14 "AN"		19 RCL 43	60 /
15 "SJ"		20 +	61 CHS
-15 "AN2"		21 RCL 41	62 RCL 58
-21 PRKEYS		22 +	63 +
-22 RND		23 STO 65	64 STO 65
-23 PRFLAGS		24 *	65 RCL 43
24 "CM"		25 RCL 68	66 *
-42 DSE		26 I	67 -
-51 FS?C		27 -	68 ST- IND 18
-62 PRREGX		28 STO 68	69 ISG 18
-63 PRBUF		29 /	70 DEG
-72 DEL		30 STO IND 18	71 RCL 67
-73 XROM 30.00		31 RCL 43	72 RCL 66
-74 XROM 30.03		32 ST- IND 18	73 *
-83 R-I		33 RCL 55	74 RCL 68
		34 RCL 49	75 /
		35 RCL 60	76 RCL 60
	PRFLAGS	36 *	77 RCL 65
		37 +	78 *
STATUS:		38 STO 67	79 +
SIZE= 070		39 RCL 65	80 RCL 65
Σ= 20		40 *	
DEG			
FTY ?			

Figure B10.5. Continued.

It is intended to complement programs "AN2" or "AN3" and "SJ" for the design of the amplifier by using sets of the equations corresponding to the system ("SJ"), the circuit ("AN2" or "AN3"), and the devices ("DEV"). More generally, this program is intended to illustrate the breakup of the design process into three hierarchical levels to clarify and simplify the rationale underlying the design. For clarity, simplifications have been introduced, particularly in ignoring the effect of dc collector voltages on collector junction capacitance and device transit time. Early effect (the effect of collector voltage

```

81 X+2          121 ST+ IND 18          161 +
82 2            122 DSE 18          162 RCL 63
83 /            123 RCL 63          163 *
84 RCL 43       124 ST* IND 18       164 RCL 45
85 *            125 RCL 64          165 *
86 -            126 DSE 18          166 ST- IND 18
87 STO IND 18   127 GTO 40          167 ISG 18
88 RCL 51       128 RTN            168 DEG
89 STO 63       129*LBL 41          169 RCL 42
90 RCL 54       130 RCL 56          170 2
91 STO 64       131 RCL 57          171 /
92 XEQ 40       132 +            172 RCL 63
93 RCL 50       133 RCL 65          173 X+2
94 RCL 41       134 +            174 *
95 RCL 47       135 CHS            175 RCL 45
96 *            136 STO 63          176 *
97 +            137 RCL 45          177 ST- IND 18
98 RCL 45       138 RCL 41          178 END
99 *            139 *                CAT 1
100 RCL 47      140 ISG 18          LBL*AN3
101 +            141 DEG          LBL*AN
102 STO 63      142 ST- IND 18      END
103 RCL 53      143 ISG 18          306 BYTES
104 RCL 42      144 DEG
105 RCL 47      145 RCL 42
106 *            146 RCL 41
107 +            147 RCL 63
108 RCL 45      148 *
109 *            149 +
110 STO 64      150 RCL 45
111 4            151 *
112 ST+ 18      152 ST- IND 18
113 RDM         153 ISG 18
114 XEQ 40      154 DEG
115 GTO 41      155 RCL 41
                156 RCL 63
                157 *
                158 2
                159 /
                160 RCL 42
116*LBL 40
117 RCL IND 18
118 *
119 ISG 18
120 DEG

```

Figure B10.5. Continued.

on the junction spacing) is also ignored. These effects are included in the device programs in Appendix C.

Transistor data required by the program are given in the following list, which gives values for the parameters used in the design described in Section 5.6. These values were obtained from transistor measurements taken on actual microwave integrated circuit transistors over a frequency range from low frequencies to 2 GHz:

r'_e	Emitter contact resistance	0.002 k Ω
δ	Dc current defect ratio ($1/h_{fe}$)	0.01

r'_b	Base resistance	0.050 k Ω
τ_F	Base transit time	0.032 ns
C_{je}	Emitter junction capacitance	1.0 pF
C_{jc}	Collector junction capacitance	0.30 pF
C_{cs}	Collector-to-substrate capacitance	0.30 pf

In addition, certain parasitic elements are included as well as conductances used in supplying biases:

G_{1-3}	Base-to-ground shunt conductances
I_{C1-3}	DC collector currents, taken as 1, 4, and 8 mA, respectively
C_{p1-3}	Parasitic base-to-ground capacitances including that of the biasing diodes of fig. 5.19.

The equations programmed begin with the total emitter resistance of the i th stage:

$$r_{ei} = \frac{kT}{qI_{Ci}} + r'_{ei} \quad (\text{B11-1})$$

The dc current loss of the i th stage is

$$\Delta_i = \delta_i + G_i r_{Ei} \quad (\text{B11-2})$$

The unity loss time constant is

$$\tau_i = \tau_{Fi} + r_{Ei}(C_{jei} + C_{pTi}) + r_{B(i+1)}C_{jci} \quad (\text{B11-3})$$

in which r_{Bi} is the total series base resistance, including that of the series biasing diodes, and C_{pTi} is the total parasitic shunt capacitance from base to ground. For the three stages diagrammed in Fig. 5.19, these capacitances are given by

$$\begin{aligned} C_{pT1} &= C_{jc1} + C_{p1} \\ C_{pT2} &= C_{cs1} + C_{jc1} + C_{jc2} + C_{p2} \\ C_{pT3} &= C_{cs2} + C_{jc2} + C_{jc3} + C_{p3} \end{aligned} \quad (\text{B11-4})$$

These capacitances are used internally by "DEV" but are not used in the analysis program. Their effects are included in the time constants and delays. The stage delays are given by the sum of the transit time delay and the equivalent delays introduced by direct feedthrough (through the collector capacitance) and the combination of base resistance and parasitic capacitance:

$$\tau_{di} = \frac{\tau_{Fi}}{2} + r_{Ei}C_{jci} + r_{Bi} \frac{\tau_i C_{pTi}}{\tau_i + r_{Ei}C_{pTi}} \quad (\text{B11-5})$$

This completes the set of equations for the three stages of Design B.

The results of these equations are stored in the memory locations required by the analysis programs; by using this program in conjunction with "AN3", for example, the amplifier can be analyzed starting with the basic transistor parameters. Program "AN3", in conjunction with "SJ", can then be used to design the amplifier for a wide variety of bias conditions, polynomial specifications, and cutoff frequencies. The device-level equations do not interact directly with the synthesis program; only adjacent levels of the three-tiered hierarchy interact.

REQ "DEV"		RUN
DEV. LEVEL ANALYSIS		
PAR.		r1-3
$\delta, G1-3, Re'$	R47= 0.0280	
	R48= 0.0085	
	R49= 0.0053	
R00= 0.0100		$\Delta 1-3$
R01= 0.0000		
R02= 1.3300		R50= 0.0100
R03= 0.0000		R51= 0.0185
R04= 0.0020		R52= 0.0100
rb1-3		CPTOT 1-3
R05= 0.0500		
R06= 0.0800		R19= 0.6000
R07= 0.0800		R20= 1.4000
TF, CJE, CJC		R21= 1.4000
R08= 0.0320		T1-3
R09= 1.0000		
R10= 0.3000		R53= 0.1000
CP1-3		R54= 0.0764
		R55= 0.0446
R12= 0.3000		Td1-3:
R13= 0.5000		
R14= 0.5000		R56= 0.0501
IC1-3		R57= 0.1155
		R58= 0.1137
R15= 1.0000		
R16= 4.0000		
R17= 8.0000		
KT/Q		
R18= 0.0260		
OK?		

Figure B11.1. Program "DEV": example. Left column: Input data. Right column: Amplifier stage data.

Once the design is realized, its sensitivities and variations may be investigated directly by use of programs "SCX" and "STAT". The data are stored in memory locations appropriate for these programs.

Example

The physical device parameters are stored in registers R00–R18 in locations indicated in the example. The results from the program are stored in registers R47–R58, and the total interstage parasitic capacitances are stored in R19–R21. Program "DEV" is listed in Fig. B11.2.

01*LBL "DEV"	41 15.017	81 RCL 01
02 "5/29/80"	42 PRREGX	82 RCL 47
03 SF 12	43 ADV	83 *
04 "DEV. LEVEL"	44 "KT/Q"	84 RCL 00
05 AVIEW	45 AVIEW	85 +
06 "ANALYSIS"	46 18	86 STO 50
07 AVIEW	47 PRREGX	87 RCL 02
08 CF 12	48 ADV	88 RCL 48
09 ADV	49 "OK?"	89 RCL 00
10 "PAR."	50 PROMPT	90 +
11 AVIEW	51 SF 13	91 STO 51
12 18	52 "R1-3"	92 RCL 03
13 ACCHR	53 AVIEW	93 RCL 49
14 ", G1-3, Re"	54 CF 13	94 *
15 ACA	55 RCL 18	95 RCL 00
16 39	56 RCL 15	96 STO 52
17 ACCHR	57 /	97 50.052
18 PRBUF	58 RCL 04	98 PRREGX
19 .004	59 +	99 ADV
20 PRREGX	60 STO 47	100 RCL 10
21 ADV	61 RCL 18	101 RCL 12
22 SF 13	62 RCL 16	102 +
23 "RB1-3"	63 /	103 STO 19
24 AVIEW	64 RCL 04	104 RCL 09
25 CF 13	65 +	105 +
26 5.007	66 STO 48	106 RCL 47
27 PRREGX	67 RCL 18	107 *
28 ADV	68 RCL 17	108 RCL 06
29 "TF, CJE, CJC"	69 /	109 RCL 10
30 AVIEW	70 RCL 04	110 *
31 8.01	71 +	111 +
32 PRREGX	72 STO 49	112 RCL 08
33 ADV	73 47.049	113 +
34 "CP1-3"	74 PRREGX	114 STO 53
35 AVIEW	75 ADV	115 RCL 10
36 12.014	76 8	116 2
37 PRREGX	77 ACCHR	117 *
38 ADV	78 "1-3"	118 RCL 11
39 "IC1-3"	79 ACA	119 +
40 AVIEW	80 PRBUF	120 RCL 13

Figure B11.2. Program "DEV": listing.

```

121 +                161 RCL 53                201 RCL 55
122 STO 20           162 RCL 19                202 RCL 21
123 RCL 09           163 *                    203 *
124 +                164 RCL 05                204 RCL 07
125 RCL 48           165 *                    205 *
126 *                166 RCL 47                206 RCL 49
127 RCL 07           167 RCL 19                207 RCL 21
128 RCL 10           168 *                    208 *
129 *                169 RCL 53                209 RCL 55
130 +                170 +                    210 +
131 RCL 00           171 /                    211 /
132 +                172 RCL 10                212 RCL 49
133 STO 54           173 RCL 47                213 RCL 10
134 RCL 10           174 *                    214 *
135 2                175 +                    215 +
136 *                176 RCL 00                216 RCL 00
137 RCL 11           177 2                    217 2
138 +                178 /                    218 /
139 RCL 14           179 +                    219 +
140 +                180 STO 56                220 STO 58
141 STO 21           181 RCL 54                221 56.058
142 RCL 09           182 RCL 20                222 PRREGX
143 +                183 *                    223 ADV
144 RCL 49           184 RCL 06                224 RTN
145 *                185 *                    225 .END.
146 RCL 00
147 +
148 STO 55
149 *CPTOT 1-3*     189 RCL 54                CAT 1
150 AVIEW           190 +
151 19.021           191 /                LBL'DEV
152 PRREGX           192 RCL 48                END          441 BYTES
153 ADV             193 RCL 10                .END.       07 BYTES
154 *T1-3*         194 *
155 AVIEW           195 +                PRKEYS
156 53.055         196 RCL 00
157 PRREGX           197 2                USER KEYS:
158 ADV             198 /                11 SREG
159 *Td1-3:*       199 +                -11 SIZE
160 AVIEW           200 STO 57                15 "DEV"

```

Figure B11.2. Continued.

B-12 PROGRAM "PCM": QUANTIZED FEEDBACK DESIGN FOR A PCM SYSTEM

This program finds the quantized feedback and forward-path filter functions and circuit element values for a PCM regenerator or repeater, as discussed in Section 5.7. The circuits that eliminate dc wander are shown in Fig. B12.1 and include four sections. Three of these sections make up the quantized feedback system, including the quantized feedback filter and a forward-path filter including a doublet section and a low-frequency cutoff. The fourth section provides phase correction for all remaining low-frequency cutoffs in the system between the transmitter and the decision circuit input. While applied here to a PCM regenerator for digital transmission, the system is equally applicable to

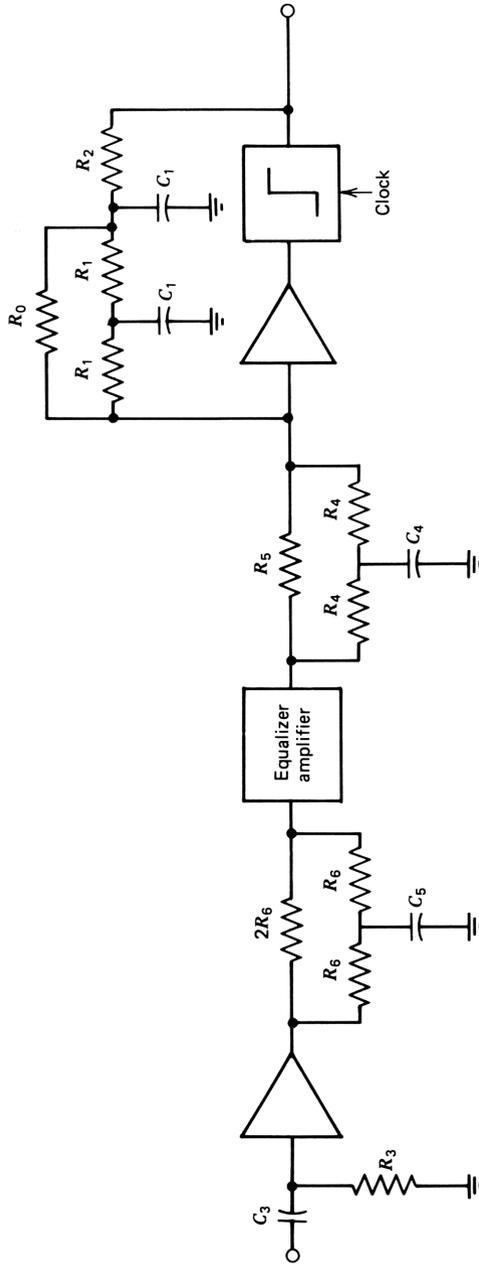


Figure B12.1. RCM regenerator showing the quantized feedback forward path and feedback filter circuits.

any PCM system as that found in digital audio or video recording, for example. Its function is to reduce the octave bandwidth of the system to about four to six octaves.

The following system parameters are known or assumed at the outset:

- f_B The signaling rate, or system clock rate
- f_1 The low-frequency cutoff of the forward path introduced by the quantized feedback system; typical values: $f_B/50$ – $f_B/20$
- τ_d The delay of the decision circuit, including sampling delay and any analog delay that may be present.

In addition, certain circuit parameters and elements are also assumed or chosen at the outset; these control the impedance levels of the circuits and also the relative signal levels at the decision circuit input and output. These are the following:

- R_1 The value of two resistors of the quantized feedback filter. (A third resistor, R_2 , to be found in this program, is approximately equal to R_1 .)
- R_3 The resistor controlling the low-frequency cutoff at f_1 .
- R_6 The resistors of the phase correction network, chosen to give a convenient impedance level.
- K_v Ratio of signal voltages at the output of the decision circuit to the output of the linear channel.

With these elements given, the program finds the remaining circuit elements of the networks to give zero error at the sampling instants. The design allows for decision circuit delays of up to 1.5 time slots. (The error may also be negligible for delays somewhat larger than this.)

The quantized feedback design is based on the equation

$$H(s) + Q(s)e^{-\tau_d s} = 1 \quad (\text{B12-1})$$

We assume the pulses to be 100 percent duty factor, so that a long string of "ones" will be a step waveform.

The quantized feedback system works as follows. A step waveform, representing a long string of ones is applied to $H(s)$, with the result shown in Fig. B12.2. The response of $H(s)$ in the time domain is arranged such that the amplitude at the first sampling instant is correct. Later, when the response begins to sag, the quantized feedback waveform becomes available (after delay τ_d) to correct the sagging waveform. The characteristic equations of both $H(s)$ and $Q(s)$ are quadratics.

Forward Path

The step response of $H(s)$ is given in the time domain by

$$u(t) * h(t) = k_1 e^{-t/\tau_1} + k_2 e^{-t/\tau_2} \quad (\text{B12-2})$$

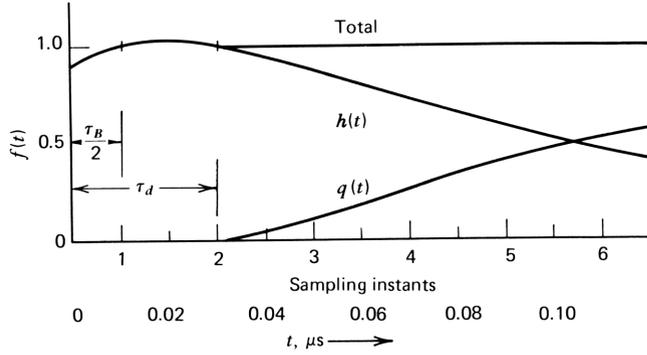


Figure B12.2. Step responses of the feedback and feedforward paths and their total: $f_B = 50$ MHz; $\tau_d = 0.03 \mu\text{s}$ (1.5 time slots); $f_1 = 2$ MHz.

where k_1 and k_2 are the residues in the roots of the characteristic equation and τ_1 and τ_2 are the time constants of the (real) roots of the characteristic equation. Our purpose is to make the response correct at the first sampling instant, $t = \tau_B/2$, and also at time τ_d , just before the quantized feedback signal becomes available (see Fig. B12.2). Thus we can write

$$k_1 e^{-\tau_B/2\tau_1} + k_2 e^{-\tau_B/2\tau_2} = 1 \quad (\text{B12-3})$$

and

$$k_1 e^{-\tau_d/\tau_1} + k_2 e^{-\tau_d/\tau_2} = 1 \quad (\text{B12-4})$$

solving these equations simultaneously, we obtain

$$k_1 = \frac{1 - \exp(\tau_d - \tau_B/2)}{\exp(-\tau_B/2\tau_2) - \exp\left(\frac{\tau_d - \tau_B/2}{\tau_2} - \frac{\tau_d}{\tau_1}\right)} \quad (\text{B12-5})$$

and, from (B12-3) directly,

$$k_2 = (1 - k_1 e^{-\tau_B/2\tau_1}) e^{\tau_B/2\tau_2} \quad (\text{B12-6})$$

The program uses these equations to calculate the residues. These residues and the time constants, τ_1 and τ_2 , completely define $H(s)$: the Laplace transform of (B12-2) is

$$\frac{1}{s} H(s) = \frac{k_1 \tau_1}{1 + \tau_1 s} + \frac{k_2 \tau_2}{1 + \tau_2 s} \quad (\text{B12-7})$$

so that

$$H(s) = \frac{(k_1\tau_1 + k_2\tau_2)s + (k_1 + k_2)\tau_1\tau_2s^2}{(1 + \tau_2s)(1 + \tau_2s)} \quad (\text{B12-8})$$

$$= K_H \cdot \frac{\tau_1s}{1 + \tau_1s} \cdot \frac{1 + \tau_ns}{1 + \tau_2s} \quad (\text{B12-9})$$

where we have broken $H(s)$ into a forward-path gain factor K_H , an ac cutoff at a frequency of $1/2\pi\tau_1$, and a doublet with numerator time constant τ_n and denominator time constant τ_2 . The forward-path gain factor is

$$K_H = k_1 + \frac{k_2\tau_2}{\tau_1} \quad (\text{B12-10})$$

In the example, K_H is 1.46. The numerator time constant is given by

$$\tau_n = \frac{(k_1 + k_2)\tau_1\tau_2}{k_1\tau_1 + k_2\tau_2} \quad (\text{B12-11})$$

Quantized Feedback Path

A step waveform from the decision circuit output is applied to the quantized feedback filter at time τ_d , after which the forward-path and feedback-path waveforms are to be made complementary. Thus

$$u(t - \tau_d) * q(t - \tau_d) = 1 - k_3e^{-(t - \tau_d)/\tau_1} - k_4e^{-(t - \tau_d)/\tau_2} \quad (\text{B12-12})$$

so that we can write

$$k_3 = k_1e^{-\tau_d/\tau_1} \quad (\text{B12-13})$$

$$k_4 = k_2e^{-\tau_d/\tau_2} \quad (\text{B12-14})$$

Note that since the quantized feedback signal is zero at $t = \tau_d$, $k_3 + k_4 = 1$ from eq. (B12-12). The program calculates these quantized feedback filter residues.

The quantized feedback filter is defined by these residues and the time constants. Thus, taking the Laplace transform of (B12-12), we obtain

$$\frac{1}{s}Q(s) = \frac{1}{s} - \frac{k_3\tau_1}{1 + \tau_1s} - \frac{k_4\tau_2}{1 + \tau_2s} \quad (\text{B12-15})$$

and thus

$$Q(s) = \frac{1 + [(1 - k_3)\tau_1 + (1 - k_4)\tau_2]s + \tau_1\tau_2(1 - k_3 - k_4)s^2}{(1 + \tau_1s)(1 + \tau_2s)} \quad (\text{B12-16})$$

Since $k_3 + k_4 = 1$, the quadratic term in the numerator drops out. The linear term in the numerator provides small slope on the quantized feedback waveform at $t = \tau_d$. This is sufficiently small that the numerator can usually be left out without affecting accuracy greatly; it only adds a single resistor to the circuit, however, so we retain it. The numerator time constant is $(1 - k_3)\tau_1 + (1 - k_4)\tau_2$.

The program calculates the numerator time constants of both $Q(s)$ and $H(s)$ as well as the forward-path gain factor K_H . To do this, we assume a value of τ_2 of one-third τ_1 . From the standpoint of quantized feedback, this ratio is relatively unimportant; the optimum is quite broad. For a given total series resistance, as the roots are moved closer together in frequency, the center resistor rises in value, making the circuit loading due to the outer resistors more serious. Conversely, as the roots move farther apart, the quantized feedback cutoff frequency must move downward for a given slope on the waveform in the first time slot. Making the resistors equal weighs these considerations roughly equally and facilitates the realization of the filter.

In Fig. B12.3a the ladder circuit is analyzed by assuming the current in the left resistor to be unity. Then the voltage v_1 is simply R_1 and the current through the left capacitor is $R_1 C_1 s$, giving I_A as $1 + R_1 C_1 s$. The voltage v_2 is then

$$v_2 = R_1(2 + R_1 C_1 s) \quad (\text{B12-17})$$

From this voltage we obtain the current through the right capacitor and I_B , from which, in turn, we obtain v_3 :

$$\begin{aligned} v_3 &= 3R_1 + 4R_1^2 C_1 s + R_1^3 C_1^2 s^2 \\ &= 3R_1(1 + R_1 C_1 s) \left(1 + \frac{R_1 C_1}{3} s \right) \end{aligned}$$

so that we take $R_1 C_1 = \tau_1$; τ_2 is then one-third of τ_1 .

Addition of R_0 to the circuit in Fig. B12.3 to obtain the numerator of $Q(s)$ is a small perturbation on the procedure just described. By continuing to take the current in the left resistor as unity for a starting point, v_2 remains unchanged, and we find the input current immediately as the sum of v_2/R_0 and unity

$$i_{\text{in}} = 1 + \frac{R_1}{R_0} (2 + R_1 C_1 s)$$

This function gives the required numerator time constant of $Q(s)$ in (B12-16). For this circuit, we obtain

$$\tau_{nQ} = \frac{\tau_1 R_1}{R_0 + 2R_1} \quad (\text{B12-18})$$

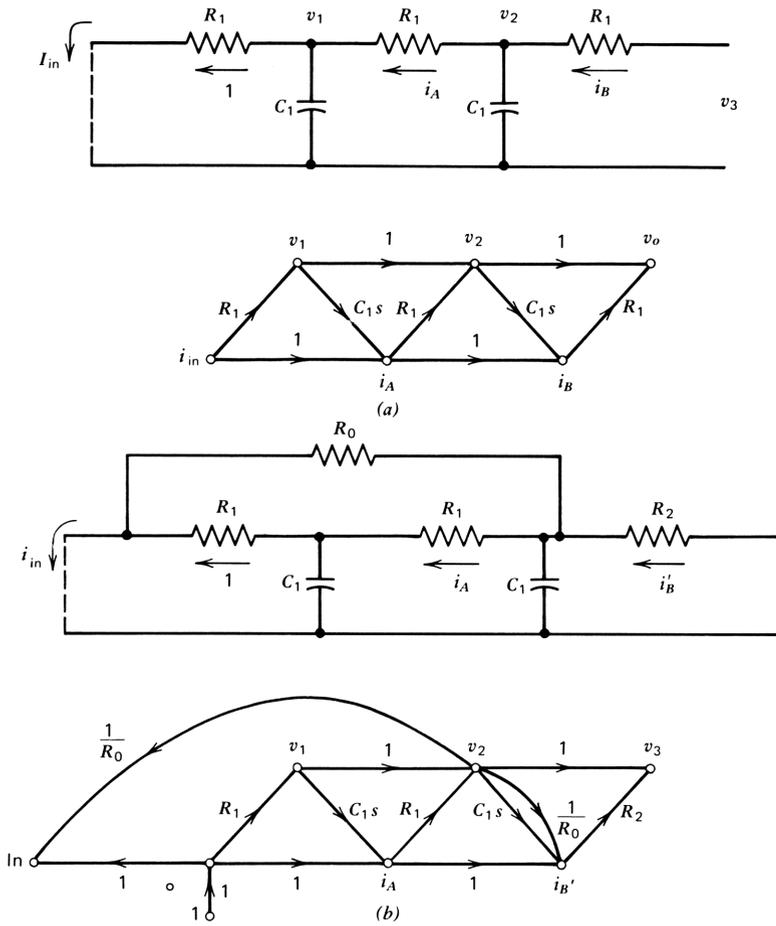


Figure B12.3. Feedback filter analysis.

from which

$$R_0 = R_1 \left(\frac{\tau_1}{\tau_{nQ}} - 2 \right) \tag{B12-19}$$

We can retain the denominator (the characteristic equation) by increasing the value of the right-hand resistor from R_1 to R_2 , where the parallel combination of R_0 and R_2 is equal to R_1 . Thus

$$R_2 = \frac{R_1 R_0}{R_0 - R_1} \tag{B12-20}$$

The Thevenin equivalent of R_2 and the loading of R_0 at the v_2 node is then R_1 , and the characteristic equation is retained; v_3 is increased by the constant term $1 + R_2/R_0$. The transmission of the quantized feedback filter at dc is conveniently expressed as a transresistance R_T and is obtained as

$$\frac{v_3(0)}{i_{in}(0)} = R_T = \frac{1}{Q(0)} = \frac{2R_1(R_0 + R_2) + R_2R_0}{R_0 + 2R_1} \quad (\text{B12-21})$$

This completes the design of the quantized feedback filter. We used the values of f_B , f_1 , and τ_d to obtain the residues and the transfer function of the filter; R_1 is selected on the basis of circuit loading of the decision circuit. The program calculates the values of R_0 , R_2 , and C_1 by use of preceding equations. A reasonable choice for f_1 might be $0.02f_B$. Since the program runs quickly, several such choices may be tried quite conveniently.

The forward-path filter is realized in two parts; the ac coupling is placed at the regenerator input and introduces a low-frequency cutoff at f_1 , using C_3 in combination with R_3 (the combination of the source and load resistances of the input network).

The forward-path gain factor K_H and the doublet defined in eqs. (B12-9) to (B12-11) are realized in the circuit to the left of the summing node in Fig. B12.1. At high frequencies the only transmission is through R_5 . From eq. (B12-8), this high-frequency transmission is just $k_1 + k_2$, so that

$$R_5 = \frac{(k_1 + k_2)}{K_v} R_T \quad (\text{B12-22})$$

where K_v is the voltage gain to be achieved from the output of the linear channel (at the left of the circuit under consideration) and the decision circuit output. Resistance R_T is the quantized feedback transresistance factor given in (B12-21). At low frequencies transmission is through both R_5 and the two resistors R_4 . By analogy with operational amplifier gain, in which R_T is the feedback resistor, and the parallel combination of $2R_4$ and R_5 is the source resistance, we can write

$$\frac{1}{R_5} + \frac{1}{2R_4} = \frac{1}{R_T} K_H K_v$$

where $K_H K_v$ is the desired gain and K_H is given by (B12-10). Solving for R_4 , we obtain

$$R_4 = \frac{0.5R_5 R_T \tau_1}{R_5 K_v (k_1 \tau_1 + k_2 \tau_2) - R_T \tau_1} \quad (\text{B12-23})$$

Finally, the capacitor C_4 is of the value required to give the denominator time

constant τ_2 . This time constant is easily shown to be $C_4(R_4/2)$, so that

$$C_4 = \frac{2\tau_2}{R_4} \quad (\text{B12-24})$$

thereby completing the design of the forward-path filter.

It is usually advisable to include a phase correction network to remove the phase put in by low-frequency cutoffs in the linear channel other than the cutoff of the quantized feedback forward path. Low-frequency cutoffs degrade the time domain performance almost entirely as a result of phase shift, and this is easily removed by adding a low-frequency one-octave doublet. The pole of loss of the doublet is placed at a frequency equal to the *sum of all cutoff frequencies* of the channel (excluding the R_3C_3 cutoff), and the zero of loss of the doublet is placed at twice this frequency. At frequencies considerably higher than this latter frequency, the phase of the loss is given by

$$\phi \simeq -\frac{f}{f_c} - \frac{f}{f_c} + \frac{f}{2f_c} = 0 \quad (\text{B12-25})$$

where f_c is the sum of the low-frequency cutoffs and f is the frequency at which the phase is evaluated. The first term is due to the low-frequency cutoffs, the second is due to the pole of loss, and the third is due to the zero of loss. This correction should be applied to any system regardless of whether quantized feedback is used. The circuit is shown in Fig. B-12.1; R_6 is selected on the basis of impedance level. Capacitor C_5 is then given by

$$C_5 = \frac{1}{\pi f_c R_6} \quad (\text{B12-26})$$

As shown in Fig. B12.4, the program starts with the input data f_B , f_1 , and τ_d . It then calculates the residues and uses them to calculate all circuit values, first for the quantized feedback filter and then the forward-path and (finally) low-frequency cutoff phase correction, using the equations developed in the preceding paragraphs. It then uses the residues in eqs. (B12-3) and (B12-13) to calculate the time response of the combination of the forward-path signal and the delay quantized feedback signal. It prints out the response at the sampling instants as well as at the transition between pulses.

As an example of a low-frequency channel design for a regenerator, we let $f_b = 50$ MHz, let the quantized feedback cutoff be 1 MHz, and assume a decision circuit delay of 1.5 time slots, or $0.03 \mu\text{s}$. The calculations are shown in the calculator printout, and the resulting design is given in Fig. B12.5. The performance is perfect at the sampling instants, as expected, and the maximum deviation in the time slot occurs initially, where the response is down 2%. Program "PCM" is listed in Fig. B12.6.

XEQ "PCM"
 PCM REGEN
 DESIGN WITH
 QUANTIZED
 FEEDBACK

UNITS: MHZ, US,
 KOHM, HF.

INPUT DATA:
 FB, F1

R17= 50.0000
 R18= 1.0000

Td

R00= 0.0300

DELAY = 1.5 TIME SLOTS
 OK?

RUN

Td, T1, T2=T1/3

R00= 0.0300
 R01= 0.1592
 R02= 0.0531

RESIDUES
 K1-K4:

R03= 1.7064
 R04= -0.7275
 R05= 1.4133
 R06= -0.4133

R1, R3, R6, KV

R30= 3.0000
 R31= 0.1000
 R32= 1.0000
 R33= 10.0000

SUM OF ALL LF
 C.O. FREQ.
 EXCEPT R3-C3

R34= 0.1000

OK?

RUN

QF FILTER

R1, R0, R2, C1:

R39= 3.0000
 R40= 45.8948
 R41= 3.2098
 R42= 0.0531

FORWARD
 PATH FIL.

R4, R5, C4.

R43= 0.8779
 R44= 0.8699
 R45= 0.1209

C3, C5

R46= 1.5915
 R47= 3.1831

PERFORMANCE

> = SAMPLE PT.

TIME SLOT	RESPONSE h(t)	sum
1.	0.979	0.979
1. >	1.000	1.000
2.	1.006	1.006
2. >	1.000	1.000
3.	0.985	1.000
3. >	0.963	1.000
4.	0.936	1.000
4. >	0.905	1.000
5.	0.871	1.000
5. >	0.836	1.000

THE END

Figure B12.4. Program "PCM": example.

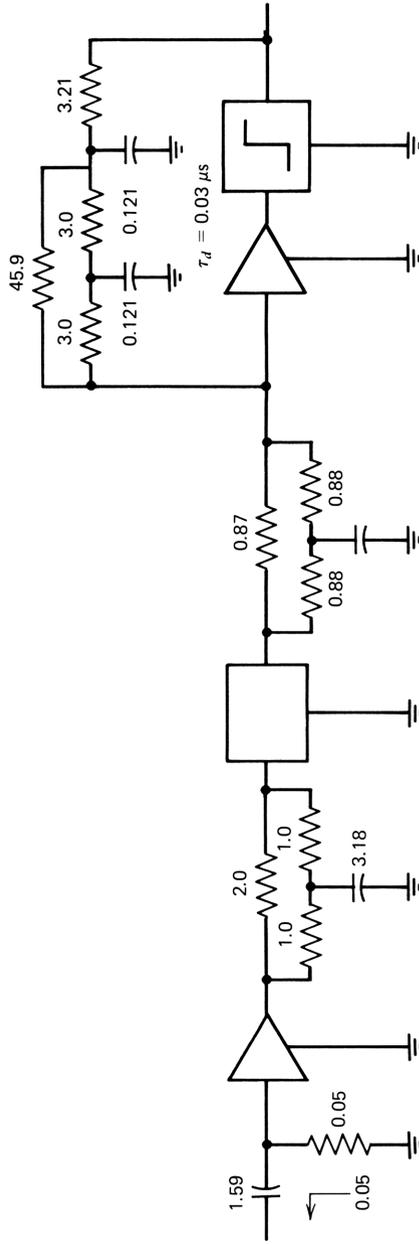


Figure B12.5. Circuit values calculated by program "PCM". $f_B = 50 \text{ MWz}$; all R in kilohms; all C in nanofarads.

01*LBL "PCM"	45 "OK?"	90 1
02 SF 12	46 PROMPT	91 RCL 00
03 "PCM REGEN"	47 FS? 22	92 RCL 24
04 AVIEW	48 GT0 00	93 -
05 "DESIGN WITH"	49 RCL 17	94 RCL 02
06 AVIEW	50 1/X	95 /
07 "QUANTIZED"	51 STO 08	96 E+X
08 AVIEW	52 RCL 18	97 -
09 "FEEDBACK"	53 2	98 RCL 23
10 AVIEW	54 PI	99 /
11 CF 12	55 *	100 STO 03
12 ADV	56 *	101 RCL 24
13 FIX 4	57 1/X	102 RCL 01
14 ADV	58 STO 01	103 /
15 "UNITS: MHZ, US,"	59 3	104 CHS
16 AVIEW	60 /	105 E+X
17 "KOHM, NF,"	61 STO 02	106 *
18 AVIEW	62 "Td, T1, T2=T1/3"	107 CHS
19 ADV	63 AVIEW	108 1
20 "INPUT DATA:"	64 .002	109 +
21*LBL 00	65 PRREGX	110 RCL 24
22 AVIEW	66 ADV	111 RCL 02
23 "FB, F1"	67 "RESIDUES:"	112 /
24 AVIEW	68 AVIEW	113 E+X
25 17.018	69 RCL 17	114 *
26 PRREGX	70 2	115 STO 04
27 ADV	71 *	116 RCL 00
28 "Td"	72 1/X	117 RCL 02
29 AVIEW	73 STO 24	118 /
30 0	74 RCL 01	119 CHS
31 PRREGX	75 /	120 E+X
32 ADV	76 CHS	121 *
33 "DELAY = "	77 E+X	122 STO 06
34 ACA	78 RCL 00	123 RCL 00
35 FIX 1	79 RCL 24	124 RCL 01
36 RCL 00	80 -	125 /
37 RCL 17	81 RCL 02	126 CHS
38 *	82 /	127 E+X
39 ACX	83 RCL 00	128 RCL 03
40 " TIME SLOTS"	84 RCL 01	129 *
41 ACA	85 /	130 STO 05
42 PRBUF	86 -	131 "K1-K4:"
43 FIX 4	87 E+X	132 AVIEW
44 CF 22	88 -	133 3.006
	89 STO 23	134 PRREGX

Figure B12.6. Program "PCM": listing.

135 ADV	179 STO 40	224 +
136*LBL "OF"	180 1/X	225 *
137 "R1, R3, R6, KV"	181 RCL 30	226 RCL 27
138 AVIEW	182 1/X	227 RCL 01
139 30.033	183 X<>Y	228 *
140 PRREGX	184 -	229 -
141 ADV	185 1/X	230 RCL 44
142 "SUM OF ALL LF"	186 STO 41	231 RCL 27
143 AVIEW	187 RCL 01	232 *
144 "C.O. FREQ."	188 RCL 30	233 RCL 01
145 AVIEW	189 /	234 *
146 "EXCEPT R3-C3"	190 STO 42	235 X<>Y
147 AVIEW	191 RCL 40	236 /
148 34	192 RCL 41	237 2
149 PRREGX	193 +	238 /
150 "OK?"	194 RCL 30	239 STO 43
151 CF 22	195 *	240 RCL 02
152 PROMPT	196 2	241 2
153 FS? 22	197 *	242 *
154 GTG "OF"	198 RCL 41	243 RCL 43
155 SF 12	199 RCL 40	244 /
156 "OF FILTER"	200 *	245 STO 45
157 AVIEW	201 +	246 RCL 01
158 CF 12	202 RCL 30	247 RCL 31
159 ADV	203 2	248 /
160 1	204 *	249 STO 46
161 RCL 05	205 RCL 40	250 RCL 34
162 -	206 +	251 RCL 32
163 RCL 01	207 /	252 *
164 *	208 STO 27	253 PI
165 1	209 RCL 33	254 *
166 RCL 06	210 /	255 1/X
167 -	211 RCL 03	256 STO 47
168 RCL 02	212 RCL 04	257 RCL 30
169 *	213 +	258 STO 39
170 +	214 /	259 "R1, R0, R2, C1:"
171 STO 25	215 STO 44	260 AVIEW
172 RCL 01	216 RCL 33	261 39.042
173 /	217 *	262 PRREGX
174 1/X	218 RCL 01	263 ADV
175 2	219 RCL 03	264 SF 12
176 -	220 *	265 "FORWARD"
177 RCL 30	221 RCL 02	266 AVIEW
178 *	222 RCL 04	267 "PATH FIL."
	223 *	268 AVIEW

Figure B12.6. Continued.

Appendix C

Two-Ports, Transistors, and ABCD Matrices

The programs in this appendix enable the user to calculate the characteristics of active or passive networks including transistors, passive immittances, diodes, and delays and parasitics. The first program converts between any two sets of two-port parameters. The second and third provide modeling of the bipolar transistor and the derivation of the $ABCD$ parameters from the equivalent circuit for any bias condition or emitter area. The next two programs derive the network characteristics of a network whose $ABCD$ parameters are known at two frequencies and are limited to networks whose loss is represented by a cubic polynomial.

Program “ABCD” provides a programmable calculator for two-port networks. It enables the user to calculate the $ABCD$ parameters of any two-port. Examples of both manual and programmable operation are given. Finally, a program that evaluates the effect of spanning networks in an equivalent ladder (as described in Section 8.4) is intended for use as a subroutine of program “ABCD”.

Two-Port Conversions

C-1 Program “CNV”: two-port parameter conversions.

Transistor Analysis and Modeling

- C-2 Program "E>A": transistor $ABCD$ parameters from equivalent circuit with modification for collector bias conditions and emitter area.
- C-3 Program "A>E": transistor modeling program. Equivalent circuit from $ABCD$ parameters.

Network Calculations from ABCD Matrices

- C-4 Programs "T>P3" and "TPOLY": Cubic polynomial coefficients of $ABCD$ parameters from $ABCD$ magnitude and phase at two frequencies.
- C-5 Programs "P3>LA", "FR", and "NMR" and "NWCALC": network characterization including loss and immittances and sensitivities from cubic $ABCD$ coefficients.

ABCD Matrices of Circuits

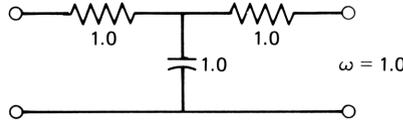
- C-6 Program "ABCD": a two-port programmable network calculator.
- C-7 Programs for the "ABCD" Calculator: six sample programs.
- C-8 Program "SP": effect of feedback and feedforward spanning networks on two-port networks. Used with "CNV" and "ABCD", this program completes the two-port programmable network calculator.

C-1 PROGRAMS "CNV" AND "CNV-S": TWO-PORT PARAMETER CONVERSIONS

Program "CNV" implements the two-port parameter conversions of Table 6.1 (in Chapter 6), allowing conversion from $ABCD$ parameters to h , z , y , or g parameters or the reverse in one step. Thus conversion is possible from any set to any other in at most two steps as can be seen in the examples of Fig. C1.1. Program "CNV-S" converts back and forth between $ABCD$ and S parameters, completing the implementation of Table 6.1. Examples are shown in Fig. C1.2.

The original matrix is stored in registers R01–R08, with the frequency at which the parameters are evaluated in R00. The matrix elements are stored as magnitude and phase. After conversion, the new matrix elements replace the old in R01–R08, and the original matrix is preserved (in R21–R28 for "CNV" and in R11–R18 for "CNV-S").

The conversions involving the h , z , y , and g parameters are based on Table C1.1, which is, in turn, based on Table 6.1. Table C1.1 contains the operations that are common to all the conversions; all require calculation of the determinant of the original matrix, and all use one of the original matrix elements as a denominator of the final matrix. Furthermore, the numerator of one of the elements of the final matrix is ± 1 . The determinant is calculated and placed in the appropriate register of the final matrix. The denominator element from the original matrix is placed in the denominator registers R21 and R22, and ± 1 is placed in the appropriate register of the final matrix. Subroutine "R" divides each element of the final matrix by the denominator. The result is then printed. Subroutine "X-" exchanges the locations of the original and final matrices.



ORIG:	XEQ "CNV"	XEQ "A-Z" XEQ "W"	XEQ "Y-A" XEQ "W"
R00= 1.0000		R00= 1.0000	R00= 1.0000
R01= 1.4142		R01= 1.4142	R01= 1.4142
R02= 45.0000		R02= -45.0000	R02= 45.0000
R03= 3.6056		R03= 1.0000	R03= 3.6056
R04= 33.6901		R04= -90.0000	R04= 33.6901
R05= 1.0000		R05= 1.0000	R05= 1.0000
R06= 90.0000		R06= -90.0000	R06= 90.0000
R07= 2.2361		R07= 2.2361	R07= 2.2361
R08= 63.4349		R08= -26.5651	R08= 63.4349
CONV: ?-?	XEQ "A-H"	XEQ "Z-A" XEQ "W"	XEQ "A-G" XEQ "W"
R00= 1.0000		R00= 1.0000	R00= 1.0000
R01= 1.6125		R01= 1.4142	R01= 0.7071
R02= -29.7449		R02= 45.0000	R02= 45.0000
R03= 0.4472		R03= 3.6056	R03= -0.7071
R04= -63.4349		R04= 33.6901	R04= -45.0000
R05= -0.4472		R05= 1.0000	R05= 0.7071
R06= -63.4349		R06= 90.0000	R06= -45.0000
R07= 0.4472		R07= 2.2361	R07= 2.5495
R08= 26.5651		R08= 63.4349	R08= -11.3099
	XEQ "H-A" XEQ "W"	XEQ "A-Y" XEQ "W"	XEQ "G-A" XEQ "W"
R00= 1.0000		R00= 1.0000	R00= 1.0000
R01= 1.4142		R01= 0.6202	R01= 1.4142
R02= 45.0000		R02= 29.7449	R02= 45.0000
R03= 3.6056		R03= -0.2774	R03= 3.6056
R04= 33.6901		R04= -33.6901	R04= 33.6901
R05= 1.0000		R05= -0.2774	R05= 1.0000
R06= 90.0000		R06= -33.6901	R06= 90.0000
R07= 2.2361		R07= 0.3922	R07= 2.2361
R08= 63.4349		R08= 11.3099	R08= 63.4349

Figure C1.1. Program "CNV": examples.

To shorten the program, advantage is taken of the relations

$$[y] = [z]^{-1}$$

and

$$[g] = [h]^{-1}$$

<pre> R0, KOHM: R25= 0.0500 ORIG: R00= 1.0000 R01= 1.4142 R02= 45.0000 R03= 3.6056 R04= 33.6901 R05= 1.0000 R06= 90.0000 R07= 2.2361 R08= 63.4349 CONV: ?-? </pre>	<pre> XEQ "CNV-S" </pre>	<pre> R0, KOHM: R25= 0.0500 ORIG: R00= 2.5133 R01= 0.0154 R02= -76.1781 R03= 0.0061 R04= -145.3320 R05= 0.5846 R06= -79.7403 R07= 0.0811 R08= -94.9995 CONV: ?-? </pre>	<pre> XEQ "CNV-S" </pre>
<pre> R00= 1.0000 R01= 0.9477 R02= -1.7841 R03= 0.0265 R04= -34.7744 R05= 0.0265 R06= -34.7744 R07= 0.9624 R08= -0.4608 </pre>	<pre> XEQ "A-S" </pre>	<pre> R00= 2.5133 R01= 0.4500 R02= 281.5873 R03= 0.0275 R04= 51.9365 R05= 9.2277 R06= 116.0259 R07= 0.7234 R08= -20.6728 </pre>	<pre> XEQ "A-S" </pre>
<pre> R00= 1.0000 R01= 1.4142 R02= 45.0000 R03= 3.6056 R04= 33.6901 R05= 1.0000 R06= 90.0000 R07= 2.2361 R08= 63.4349 </pre>	<pre> XEQ "S-A" XEQ "W" </pre>	<pre> R00= 2.5133 R01= 0.0154 R02= -76.1781 R03= 0.0061 R04= -145.3320 R05= 0.5846 R06= -79.7403 R07= 0.0811 R08= -94.9995 </pre>	<pre> XEQ "S-A" XEQ "W" </pre>

Figure C1.2. Program "CNV-S": examples.

The inversions are performed in subroutine "INV". The program listing for "CNV" is shown in Fig. C1.3.

The conversion between $ABCD$ and S parameters and the reverse is accomplished in "CNV-S", which implements the equations of the bottom entries in Table 6.1 directly. The program listing is given in Fig. C1.4.

Examples

There are two ways to use these programs. In the first, execute "CNV" (or "CNV-S") which serves as a supervisory program that calls on the appropriate

Table C1.1 Two-Port Matrix Conversions^a

	k_{11}	k_{12}	k_{21}	k_{22}	Denominator
$A-h$	B	Δ_A	-1	C	D
$A-z$	A	Δ_A	1	D	C
$A-y$	D	$-\Delta_A$	-1	A	B
$A-g$	C	$-\Delta_A$	1	B	A
$h-A$	Δ_k	k_{11}	k_{22}	1	$-k_{21}$
$z-A$	k_{11}	Δ_k	1	k_{22}	k_{21}
$y-A$	k_{22}	1	Δ_k	k_{11}	$-k_{21}$
$g-A$	1	k_{22}	k_{11}	Δ_k	k_{21}

^aNote: $k_{ij}/\text{Den.} = h_{ij}, z_{ij}, y_{ij}, \text{ or } g_{ij}$.

```

01*LBL "CNV"
02 SF 03
03 *ORIG:"
04 AVIEW
05 PSE
06 CLD
07 XEQ "RD"
08 "CONV: ?-?"
09 PROMPT

10*LBL 00
11 TONE 4
12 FS?C 03
13 XEQ "H"
14 RTN

15*LBL "A-H"
16 XEQ A
17 GTO 00

18*LBL "A-Z"
19 XEQ B
20 GTO 00

21*LBL "A-Y"
22 XEQ B

23*LBL 09
24 SF 10
25 XEQ "INV"
26 GTO 00

27*LBL "A-G"
28 XEQ A
29 GTO 09

30*LBL "H-A"
31 XEQ a
32 GTO 00

33*LBL "Z-A"
34 XEQ B
35 GTO 00

36*LBL "Y-A"
37 SF 10
38 XEQ "INV"
39 XEQ B
40 GTO 00

41*LBL "G-A"
42 SF 10
43 XEQ "INV"
44 XEQ a
45 GTO 00

46*LBL A
47 XEQ 00
48 CHS
49 STO 25
50 RCL 04
51 STO 22
52 RCL 03
53 STO 21
54 RCL 06
55 STO 28
56 RCL 05
57 STO 27
58 20
59 XEQ "X-"
60 RCL 28
61 RCL 27
62 XEQ "R"
63 RTN

64*LBL B
65 XEQ 08
66 STO 25

67 RCL 01
68 STO 21
69 RCL 02
70 STO 22
71 RCL 07
72 STO 27
73 RCL 08
74 STO 28
75 20
76 XEQ "X-"
77 RCL 26
78 RCL 25
79 XEQ "R"
80 RTN

81*LBL a
82 SF 09
83 XEQ 08
84 STO 21
85 X<>Y
86 STO 22
87 RCL 02
88 STO 24
89 RCL 01
90 STO 23
91 RCL 08
92 STO 26
93 RCL 07
94 STO 25
95 1
96 STO 27
97 0
98 STO 28
99 20
100 XEQ "X-"
101 RCL 26
102 RCL 25
103 CHS
104 XEQ "R"

```

Figure C1.3. Program "CNV": listing.

```

105 RTN
106*LBL 08
107 RCL 08
108 STO 20
109 XEQ "DT"
110 RCL 24
111 RCL 23
112 R-P
113 FS?C 09
114 RTN
115 STO 23
116 X<>Y
117 STO 24
118 0
119 STO 26
120 I
121 RTN

122*LBL "R"
123 ST/ 01
124 ST/ 03
125 ST/ 05
126 ST/ 07
127 X<>Y
128 ST- 02
129 ST- 04
130 ST- 06
131 ST- 08
132 RTN

133*LBL "DT"
134 RCL 02
135 RCL 08
136 +
137 RCL 01
138 RCL 07
139 *
140 P-R

141 STO 23
142 X<>Y
143 STO 24
144 RCL 04
145 RCL 06
146 +
147 RCL 03
148 RCL 05
149 *
150 P-R
151 ST- 23
152 X<>Y
153 ST- 24
154 RTN

155*LBL "INV"
156 XEQ "DT"
157 RCL 01
158 X<> 07
159 STO 01
160 RCL 02
161 X<> 08
162 STO 02
163 I
164 FS?C 10
165 CHS
166 ST* 03
167 ST* 05
168 RCL 24
169 RCL 23
170 R-P
171 XEQ "R"
172 RTN

173*LBL "%-"
174 STO 19
175 .008
176 STO 09

177*LBL 07
178 RCL IND 09
179 X<> IND 19
180 STO IND 09
181 ISG 19
182 DEG
183 ISG 09
184 GTO 07
185 RTN

186*LBL "RD"
187 SF 06

188*LBL "W"
189 .008
190 FS? 06
191 RDTAX
192 PRREGX
193 ADV
194 FC?C 06
195 WDTAX
196 END
CAT 1

LBL"CHV
LBL"A-H
LBL"A-Z
LBL"A-Y
LBL"A-G
LBL"H-A
LBL"Z-A
LBL"Y-A
LBL"G-A
LBL"R
LBL"DT
LBL"INV
LBL"X-
LBL"RD
LBL"W
END
447 BYTES

```

Figure C1.3. Continued.

subroutines. The program will prompt for the original matrix, asking for a CARD in the display, assuming that the *ABCD* matrix is stored on a magnetic card. It then reads the supplied card and prints the (angular) frequency and the matrix elements. If the matrix and its associated frequency are already in registers R00–R08, press R/S twice, and the program will proceed, ignoring the read instruction: the matrix is printed as before. The program then prompts for the conversion desired: convert from what matrix to what matrix? At this point, the user executes the appropriate conversion, “A-H”, “H-A”, “A-Y”, or any of the others. When the conversion is complete, the final matrix is printed, and the display prompts for recording a magnetic card with “RDY 01 OF 01”. If no card is needed, press R/S, thus completing the process. This

```

01*LBL "CNV-S"           42 RCL 01                85 R-P
   02 SF 03              43 P-R                86 STO 17
03 "R0, KOHM:"          44 ST+ 11               87 X<>Y
   04 AVIEW              45 ST+ 13               88 STO 18
   05 25                 46 ST- 15               89 RCL 25
   06 PRREGX            47 ST- 17               90 ST* 13
   07 ADV                48 X<>Y                91 ST/ 15
   08 "ORIG:"           49 ST+ 12               92 10
   09 AVIEW              50 ST+ 14               93 XEQ "X-"
   10 PSE                 51 ST- 16               94 RCL 16
   11 CLD                 52 ST- 18               95 RCL 15
   12 XEQ "RD"           53 RCL 08                96 2
13 "CONV: ?-?"          54 RCL 07                97 *
   14 PROMPT             55 P-R                  98 XEQ "R"
                               56 ST- 11               99 GTO 00
                               57 ST+ 13
                               58 ST- 15
                               59 ST+ 17
                               60 X<>Y
                               61 ST- 12
                               62 ST+ 14
                               63 ST- 16
                               64 ST+ 18
                               65 RCL 12
                               66 RCL 11
                               67 R-P
                               68 STO 11
                               69 X<>Y
                               70 STO 12
                               71 RCL 14
                               72 RCL 13
                               73 R-P
                               74 STO 13
                               75 X<>Y
                               76 STO 14
                               77 RCL 16
                               78 RCL 15
                               79 R-P
                               80 STO 15
                               81 X<>Y
                               82 STO 16
                               83 RCL 18
                               84 RCL 17
                               85 R-P
                               86 STO 17
                               87 X<>Y
                               88 STO 18
                               89 RCL 25
                               90 ST* 13
                               91 ST/ 15
                               92 10
                               93 XEQ "X-"
                               94 RCL 16
                               95 RCL 15
                               96 2
                               97 *
                               98 XEQ "R"
                               99 GTO 00

15*LBL 00
16 BEEP
17 FS?C 03
18 XEQ "W"
19 RTN

20*LBL "S-A"
21 RCL 00
22 STO 10
23 XEQ "DT"
24 RCL 23
25 STO 13
26 STO 15
27 CHS
28 STO 11
29 STO 17
30 RCL 24
31 STO 14
32 STO 16
33 CHS
34 STO 12
35 STO 18
36 1
37 ST+ 11
38 ST+ 13
39 ST+ 15
40 ST+ 17
41 RCL 02

100*LBL "A-S"
101 RCL 00
102 STO 10
103 XEQ "DT"
104 RCL 24
105 RCL 23
106 R-P
107 STO 23
108 2
109 *
110 STO 13
111 X<>Y
112 STO 14
113 2
114 STO 15
115 0
116 STO 16
117 RCL 02
118 RCL 01
119 P-R
120 STO 21
121 STO 11
122 CHS
123 STO 17
124 X<>Y
125 STO 22
126 STO 12

```

Figure C1.4. Program "CNV-S": listing.

mode of operation was used in the initial calculation shown in the examples of Fig. C1.1.

The second way to use these programs is to call on each of the preceding operations manually, as needed (or as an instruction in another program; more later). If a card is to be read, execute "RD", which will prompt for a card and print the matrix. If it is desired to print the matrix in R00-R08, execute "RD" and press R/S twice to ignore the read instruction. Next, execute the desired conversion. A tone will indicate completion. If the result is to be written on a

127 CHS	171 R-P	213 P-R
128 STO 18	172 STO 17	214 ST- 23
129 RCL 04	173 X<>Y	215 X<>Y
130 RCL 03	174 STO 18	216 ST- 24
131 RCL 25	175 RCL 22	217 RTN
132 /	176 RCL 21	
133 P-R	177 R-P	218*LBL "X-"
134 ST+ 21	178 10	219 STO 19
135 ST+ 11	179 XEQ "X-"	220 .008
136 ST+ 17	180 RCL 22	221 STO 09
137 X<>Y	181 RCL 21	
138 ST+ 22	182 R-P	222*LBL 07
139 ST+ 12	183 XEQ "R"	223 RCL IND 09
140 ST+ 18	184 GTO 00	224 X<> IND 19
141 RCL 06		225 STO IND 09
142 RCL 05	185*LBL "R"	226 ISG 19
143 RCL 25	186 ST/ 01	227 DEG
144 *	187 ST/ 03	228 ISG 09
145 P-R	188 ST/ 05	229 GTO 07
146 ST+ 21	189 ST/ 07	230 RTN
147 ST- 11	190 X<>Y	
148 ST- 17	191 ST- 02	231*LBL "RD"
149 X<>Y	192 ST- 04	232 SF 06
150 ST+ 22	193 ST- 06	
151 ST- 12	194 ST- 08	233*LBL "W"
152 ST- 18	195 RTN	234 .008
153 RCL 08		235 FS? 06
154 RCL 07	196*LBL "DT"	236 RDTAX
155 P-R	197 RCL 02	237 PRREGX
156 ST+ 21	198 RCL 08	238 ADV
157 ST- 11	199 +	239 FC?C 06
158 ST+ 17	200 RCL 01	240 WDTAX
159 X<>Y	201 RCL 07	241 END
160 ST+ 22	202 *	CAT 1
161 ST- 12	203 P-R	LBL*CNV-S
162 ST+ 18	204 STO 23	LBL*S-A
163 RCL 12	205 X<>Y	LBL*A-S
164 RCL 11	206 STO 24	LBL*R
165 R-P	207 RCL 04	LBL*DT
166 STO 11	208 RCL 06	LBL*X-
167 X<>Y	209 +	LBL*RD
168 STO 12	210 RCL 03	LBL*W
169 RCL 18	211 RCL 05	END
170 RCL 17	212 *	440 BYTES

Figure C1.4. Continued.

card, execute "W", and after the matrix is printed, it will prompt for writing a card. This may be bypassed by pressing R/S. The remaining examples of Fig. C1.1 were done in this way.

Thus in the examples we begin by executing "CNV". In response to the prompt for the original matrix, we read in a card containing the *ABCD* matrix of the network at the top of the page: the frequency and the matrix are shown in the first group of numbers at the top of the first column. In response to the prompt, we then execute "A-H", whereupon the *h* parameters of the network

are printed. Since these parameters are in position for further conversion, we illustrate the "H-A" conversion manually, by the second method. When complete, we execute "W" to print the results, the original $ABCD$ matrix. On the rest of the page of examples, we illustrate all possible conversions with "CNV" for the network, going back and forth from $ABCD$ parameters to k parameters.

The examples for "CNV-S" shown in Fig. C1.2 were carried out in the same way. In the first column the S parameters for the simple network were calculated with $R_0 = 0.05 \Omega$. In the second column the S parameters of a transistor at 400 MHz (2.513 Grad/s) were found from the $ABCD$ parameters, and the reverse conversion was also done.

Program "CNV": Registers

R00	Frequency at which matrix is evaluated
R01, R02	Magnitude and phase of A or k_{11}
R03, R04	Magnitude and phase of B or k_{12}
R05, R06	Magnitude and phase of C or k_{21}
R07, R08	Magnitude and phase of D or k_{22}
R09	Register index for subroutine "X-"
R10–R18	Not used
R19	Second register index for subroutine "X-".
R20	Frequency at which second matrix is evaluated (copied from R00)
R21–R28	Magnitude and phase of second matrix

Program "CNV-S": Registers

R00	Angular frequency at which matrix is evaluated
R01–R08	Original and final matrix
R09	Index for "X-"
R10–R18	Final and original matrix
R19	Index for "X-"
R20	—
R21, R22	Denominator for S parameters
R23, R24	Determinant of $ABCD$ parameters
R25	R_0

C-2 PROGRAM "E>A": TRANSISTOR EQUIVALENT CIRCUIT ANALYSIS PROGRAM

This program analyzes the equivalent circuit shown in Fig. 7.19*b*, finding the two-port $ABCD$ parameters. It incorporates a subroutine that modifies the equivalent circuit for different collector bias conditions and also modifies the equivalent circuit for changes in the size of the transistor. With this program, the equivalent circuit can be used to generate the $ABCD$ parameters

of transistors used in analog integrated circuits, for example, as discussed in Chapters 8 and 9. In the event that other two-port parameters are required, program “CNV” can be used for the conversion.

The starting point, the equivalent circuit, can be obtained from two-port measurements on the transistor. The translation from two-port parameters to equivalent circuit is done by using program “A>E”, to be described in Section C-3. Alternatively, where similar transistors have been previously manufactured, two-port measurements may already exist; these can be used to find an equivalent circuit whose elements can then be scaled. The scaling for changing the emitter area (actually emitter stripe length, assuming constant stripe width) is contained in subroutine “O”, discussed later.

The equations for the $ABCD$ parameters are developed in Chapter 7; with slight modifications discussed in the following paragraphs, eqs. (7.5-9) (in Chapter 7) become

$$\begin{aligned}
 A &= - \left\{ r_e g_{ce} + \left[r_b (k_b C_{jc} + \delta C'_{ce}) + r_E C'_{ce} \right] s + r_b \tau_T C'_{ce} k_a s^2 \right\} e^{r_E C_{jc} s} \\
 B &= - \left[r_E + \delta r_b + r_b \tau_T s \right] \exp(\mu \tau_T + r_E C_{jc}) s \\
 C &= - \left[\delta g_{ce} + g_{cb} + C_{jc} s + \tau_T C'_{ce} s^2 \right] e^{r_E C_{jc} s} \\
 D &= - \left[\delta + \tau_T s \right] \exp(\mu \tau_T + r_E C_{jc}) s
 \end{aligned} \tag{C2-1}$$

The main simplifying assumption is to ignore the distributed nature of the collector capacitance at the base in all equations except the first, where it is accounted for by the constant k_b . In the two-capacitor approximation, k_b is the proportion of C_{jc} connected to the external terminal. Furthermore, constant k_a has been introduced in the quadratic term of A to account for small effects that have been ignored.

Excess phase factors multiply B and D , and the denominator phase $r_E C_{jc}$ multiplies all four parameters. At high frequencies where excess phase is important, delay of the $r_b \tau_T$ term of B dominates; thus the delay of $\mu \tau_T$ is appropriate.

As in Appendix B, memory register usage is coordinated among the programs for dealing with two-port parameters. The angular frequency is stored in R00 and the $ABCD$ parameters are stored, magnitude and phase, in R01–R08 in alphabetical order. For the bipolar transistor programs, the equivalent circuit elements and modeling parameters are stored as shown in Table C2.1.

Subroutine “O”

When flag 00 is set, the equivalent circuit parameters are modified before calculating the $ABCD$ parameters. Changes in transistor size—emitter stripe length or area—as well as changes in collector voltage and current can be accommodated.

Table C2.1 Register Locations of Parameters for Transistor Programs

00	ω_0	23	r_E
01-08	<i>ABCD</i> magnitude and phase	24	C_{j_e}
09	V_{c_e}	25	τ_T
10	I_c	26	r_b
11	A_e	27	k_b
12	V_T	28	g_{c_e}
13	V_A	29	C'_{c_e}
14	I_{Kirk}	30	g_{c_b}
15	Γ_c	31	k_a
16	Γ_τ	32	Used
17	BV_{EBO}	33	Used
18	μ	34	Loop termination
19	τ_F	35	V_{c_e2}
20	C_{j_e}	36	I_{c2}
21	δ	37	A_{e2}
22	r'_e		

The subroutine begins by scaling r_b and r'_e inversely with emitter area, and C_{j_e} , C_{j_c} , and C_{c_e} directly with emitter area. (For very small transistors, this underestimates C_{j_c} somewhat.) The Kirk current is likewise scaled in proportion to emitter area.

Capacitance C'_{c_e} is assumed to be primarily junction capacitance of the isolating junction. It and C_{j_c} are assumed to vary with collector voltage as discussed in Chapter 7 and are as indicated in Table 7.1, where all bias dependencies are listed. Both δ and g_{c_e} vary inversely with $V_A + V_{c_e}$. The prime reason for the variation of the *ABCD* parameters with collector current is the inverse relation of r_e to I_c . Other equivalent circuit parameter variations with current are much less pronounced, as shown in Table 7.1.

The reason for separating τ_T into its components— τ_F , $r_e C_{j_e}$, and $r_E C_{j_c}$ —is that each component varies with collector bias in a different way. Capacitance C_{j_c} does not vary significantly with collector current over most of its operating range, but C_{j_e} , τ_F , and r_e do, as shown in Table 7.1. Similarly, C_{j_e} does not vary with collector voltage, but τ_F and C_{j_c} do. Their separate variations are taken into account, and τ_T is recalculated for the new set of bias conditions. This completes the transformation; the subroutine terminates and *ABCD* parameters are then calculated for the new equivalent circuit.

To monitor the calculation, the revised equivalent circuit values are printed after the subroutine calculations are made. This printing may be suppressed by setting flag 04, which also suppresses the printing of the initial values of the equivalent circuit parameters.

This program has the additional function of providing *ABCD* parameters on magnetic cards for use in later programs in this appendix. To this end, the

program is arranged to fit into the HP 41C with both the card reader and the printer connected, as well as two memory modules. At the conclusion of the computation, the program calls for "WDTAX"—write data card. The first 15 register contents are recorded on a magnetic card for later use. This step may be skipped by pressing R/S twice.

Example

This example begins with equivalent circuit data for an integrated circuit transistor previously obtained and recorded on two sides of a magnetic card (see Figs. C2.1 and C2.2). (The equivalent circuit was obtained from transistor measurements by using program "A>E", to be described in Section C-3.) The

	XEQ "E>A"	
EQ>ABCD	BIAS PAR.	rb, kb, qce, cce'
CHANGE BIAS, Ae? SF 00	VCE, IC, Ae:	r26= 0.1063
		r27= 0.3780
EQ. CCT.	R09= 3.0000	r28= 0.0964
	R10= 8.0000	r29= 0.3621
CORE:	R11= 250.0000	
δ, re pr, rE, CJC, TT:	IKIRK, GC, T, BVEBO:	qcb, ka:
R21= 0.0102	R14= 00.0000	R30= 0.0000
R22= 0.0015	R15= 0.4300	R31= 1.1100
R23= 0.0047	R16= 0.2300	μ:
R24= 0.2290	R17= 7.0000	R18= 0.4324
R25= 0.0320		
rb, kb, qce, cce'	NEW VCE, IC, Ae:	TF, CJE
r26= 0.0404	R35= 0.7500	R19= 0.0362
r27= 0.3780	R36= 2.0000	R20= 0.3214
r28= 0.3480	R37= 95.0000	
r29= 0.5250	OK?	FREQ., GHZ?
qcb, ka:		.0316 RUN
R30= 0.0000	NEW EQ. CCT.	NEW ABCD
R31= 1.1100	CORE:	
μ:	δ, re pr, rE, CJC, TT:	R00= 0.1985
R18= 0.4500	R21= 0.0113	R01= 0.0027
TF, CJE	R22= 0.0039	R02= -115.4484
R19= 0.0280	R23= 0.0169	R03= 0.0182
R20= 0.9360	R24= 0.1579	R04= -176.8920
	R25= 0.0430	R05= 0.0314
		R06= -90.8392
		R07= 0.0142
		R08= -142.6802
		WDTAX
(a)	(b)	(c)

Figure C2.1. Program "E>A": example.

```

                CF 00
                XEQ "E>A"
EQ>ABCD
CHANGE BIAS, Ae? SF 00
                RUN
EQ. CCT.
CORE:
δ, re nr, rE, CJC, TT:
R21= 0.0113
R22= 0.0039
R23= 0.0169
R24= 0.1579
R25= 0.0430
rb, kb, qce, cce:
r26= 0.1063
r27= 0.3780
r28= 0.0964
r29= 0.3621
qcb, ka:
R30= 0.0000
R31= 1.1100
μ:
R10= 0.4324
FREQ., GHZ?
                .1000    RUN
    
```

(d)

```

NEW ABCD
R00= 0.6283
R01= 0.0079
R02= -93.7525
R03= 0.0104
R04= -170.2341
R05= 0.0994
R06= -86.9842
R07= 0.0293
R08= -111.9268
WDTAX
                SF 04
                XEQ "E>A"
EQ>ABCD
CHANGE BIAS, Ae? SF 00
                RUN
FREQ., GHZ?
                .3160    RUN
NEW ABCD
WDTAX
                PRREGX
R00= 1.9855
R01= 0.0255
R02= -76.0986
R03= 0.0203
R04= -150.9920
R05= 0.3193
R06= -78.8051
R07= 0.0862
R08= -95.1172
    
```

(e)

```

XEQ "E>A"
EQ>ABCD
CHANGE BIAS, Ae? SF 00
                RUN
FREQ., GHZ?
                1.0000    RUN
NEW ABCD
WDTAX
                PRREGX
R00= 6.2832
R01= 0.1060
R02= -46.7506
R03= 0.0340
R04= -114.6067
R05= 1.1670
R06= -57.2892
R07= 0.2706
R08= -84.7331
    
```

(f)

Figure C2.1. Continued.

collector voltage is then changed from 3 to 0.75 V, and the collector current from 3.0 to 2.0 mA for use in a circuit computation. The emitter area is changed from 250 to 95 μm^2 . The program calculates a new equivalent circuit for the revised bias conditions and emitter area. The *ABCD* parameters of the revised equivalent circuit are then calculated in the example for four frequencies at half-decade intervals from 0.0316 to 1.0 GHz.

Columns *a* to *e* in the example (Figs. C2.1 and C2.2) give the complete printouts with flag 04 clear, to show the computation in detail. Columns *e* and *f* show the printout with flag 04 set.

In columns *a*, *b*, and part of *c* the equivalent circuit for the original bias condition is shown followed by the revised equivalent circuit for the new bias.

01*LBL "E>A"	41 ADV	81 X↑2
02 SF 12	42 RTN	82 *
03 "EQ>ABCD"		83 RCL 31
04 AVIEW	43*LBL 21	84 *
05 CF 12	44 XEQ "OL"	85 RCL 23
06 ADV	45 "OK?"	86 RCL 22
07 "CHANGE BIAS, .	46 BEEP	87 -
	47 PROMPT	88 RCL 28
08 "+Ae?"	48 XEQ "0"	89 *
09 ACA	49 RCL 35	90 -
10 " SF 00"	50 RCL 36	91 R-P
11 ACA	51 RCL 37	92 STO 01
12 PRBUF	52 STO 11	93 X<>Y
13 STOP	53 RDN	94 STO 02
14 FC? 04	54 STO 10	95 RCL 25
15 XEQ "EL"	55 RDN	96 RCL 26
16 FS? 00	56 STO 09	97 *
17 XEQ 21	57 SF 12	98 RCL 00
18 "FREQ., GHZ?"	58 "NEW "	99 *
19 TONE 6	59 ACA	100 CHS
20 PROMPT	60 XEQ "EL"	101 RCL 26
21 PI	61 RTN	102 RCL 21
22 2		103 *
23 *		104 RCL 23
24 *	62*LBL "EA"	105 +
25 STO 00	63 RCL 24	106 CHS
26 XEQ "EA"	64 RCL 27	107 R-P
27 SF 12	65 *	108 STO 03
28 "NEW ABCD"	66 RCL 26	109 X<>Y
29 AVIEW	67 *	110 STO 04
30 CF 12	68 RCL 23	111 RCL 24
31 ADV	69 RCL 29	112 RCL 00
32 .008	70 *	113 *
33 FC? 04	71 +	114 CHS
34 PRREGX	72 RCL 00	115 RCL 25
35 "WD TAX"	73 *	116 RCL 29
36 AVIEW	74 CHS	117 *
37 TONE 9	75 RCL 26	118 RCL 00
38 SF 14	76 RCL 25	119 X↑2
39 .015	77 *	120 *
40 WDTAX	78 RCL 29	
	79 *	
	80 RCL 00	

Figure C2.2. Program "E>A": listing.

After this computation is complete, the program prompts for the frequency (in gigahertz) at which the $ABCD$ parameters are to be evaluated. It then lists the angular frequency and the magnitude and phase of the $ABCD$ parameters and prompts for a data card on which to record this information. (The biases are also recorded for reference.)

To calculate the $ABCD$ parameters at the next frequency (0.1 GHz), flag 00 is cleared since we do not wish to change the bias. Columns d and e show this

```

121 RCL 28
122 RCL 21
123 *
124 -
125 RCL 30
126 -
127 R-P
128 STO 05
129 X<>Y
130 STO 06
131 RCL 25
132 RCL 00
133 *
134 CHS
135 RCL 21 D
136 CHS
137 R-P
138 STO 07
139 X<>Y
140 STO 08
141 RCL 23
142 RCL 24
143 *
144 RCL 00 ADD
145 *
146 R-D EXCESS
147 ST+ 02
148 ST+ 06 PHASE
149 RCL 18
150 RCL 25
151 *
152 RCL 00
153 *
154 R-D
155 +
156 ST+ 04
157 ST+ 08
158 RTN
159*LBL "BIAS" SCALING
160*LBL "0"

161 RCL 37
162 RCL 11
163 X<>Y TRANSISTOR
164 STO 11
165 / SIZE
166 ST/ 14
167 ST/ 20
168 ST* 22
169 ST/ 24
170 ST* 26
171 ST/ 29
172 RCL 09
173 RCL 35
174 / Cjc and
175 RCL 15
176 Y↑X Cce(Vce)
177 ST* 24
178 ST* 29
179 RCL 09
180 RCL 35
181 /
182 RCL 16 τF(IC)
183 Y↑X
184 RCL 19
185 *
186 STO 38
187 RCL 13
188 RCL 09 δ and
189 +
190 RCL 13 gce(Vce)
191 RCL 35
192 +
193 /
194 ST* 21
195 ST* 28
196 RCL 12
197 RCL 36 re and
198 /
199 RCL 22 rE(IC)
200 +

201 STO 23
202 RCL 36
203 RCL 10
204 /
205 ST* 28
206 40
207 RCL 11 Cje(IC)
208 *
209 RCL 10
210 /
211 LN
212 40
213 RCL 11
214 *
215 RCL 36
216 /
217 LN
218 /
219 SORT
220 ST* 20
221 RCL 36
222 X↑2
223 RCL 14
224 X↑2
225 +
226 RCL 10 τF(IKirk)
227 X↑2
228 RCL 14
229 X↑2
230 +
231 /
232 ST* 38
233 RCL 12
234 RCL 36
235 /
236 RCL 20 τT
237 *
238 RCL 23
239 RCL 24
240 *

```

Figure C2.2. Continued.

calculation. The equivalent circuit parameters are printed—these are identical to those shown for the “new equivalent circuit” as calculated previously.

Finally, in column *e* flag 04 is set to suppress this printing, and the *ABCD* parameters are calculated for the last two frequencies and recorded on magnetic cards.

In this way a “library” of *ABCD* parameters may be built up for later circuit calculations at different frequencies and bias conditions for various transistors. Program “E>A” is listed in Fig. C2.2.

in an integrated circuit. This latter function is accomplished through use of program "E>A" described in Section C-2.

The $ABCD$ parameters of the transistor may be found by a two-port measurement on the transistor, preferably in the cutoff region, between $f_T/20$ and $f_T/5$. Better accuracy can be obtained by repeating the procedure at more than one frequency and averaging the results.

The equivalent circuit element values are obtained in an iterative procedure on the basis of the following equations for the $ABCD$ parameters. The equations are derived from the $ABCD$ equations in program "E>A". The excess phase is subtracted from the phase of the $ABCD$ parameters, and the equivalent circuit elements are derived from the real and imaginary parts of the resulting parameters. Letting C' represent the phase-reduced value of C , and similarly for the other three parameters, we have

$$\begin{aligned}
 C_{jc} &= -\frac{\text{Im}[C']}{\omega} \\
 r_E &= -\text{Re}[B'] - \delta r_b \\
 r'_e &= r_E - \frac{kT}{qI_c} \\
 \tau_T &= -\frac{\text{Im}[D']}{\omega} \\
 r_b &= -\frac{\text{Im}[B']}{\tau_T \omega} \\
 \mu &\doteq \frac{\text{Re}[D] + \delta}{\tau_T^2 \omega^2} - \frac{r_E C_{jc}}{\tau_T} \\
 C'_{ce} &= \frac{\delta g_{ce} + g_{cb} + \text{Re}[C']}{\tau_T \omega^2} \\
 k_b &= -\frac{\text{Im}[A'] + r_E C'_{ce}}{r_b C_{jc} \omega} \\
 k_a &= \frac{\text{Re}[A'] + r_e g_{ce}}{r_b \tau_T C'_{ce} \omega^2}
 \end{aligned} \tag{C3-1}$$

No formal matrix inversion procedure is needed to find the equivalent circuit elements from the $ABCD$ parameters. This is because of the almost 1:1 relationships of r_E to B , C_{jc} to C , and τ_T to D . Base resistance at the input and C'_{ce} at the output tend to blur these relationships; thus iteration is

required. Calculation of the parameters begins with the dominant elements, thus establishing the main parameters at the beginning of the calculation. In the equation for r_E , for example, we do not initially know r_b , but in the real part of B' , r_b is multiplied by δ , and hence its effect is small. It is calculated later in the program, so its effect on r_E is accounted for in a succeeding iteration. The excess phase factor μ is calculated late in the program. It is sensitive to the exact value of D and is one of the more variable parameters from iteration to iteration. When it stabilizes to a constant value from one iteration to the next (to four decimal places), the iteration stops and the equivalent circuit parameters are printed. This normally takes only three or four iterations.

Note that τ_F and C_{je} do not appear explicitly in the preceding equations; their combined effect is incorporated in τ_T . At a given bias level they cannot be separated and need not be. To allow program "E>A" to find the $ABCD$ parameters over a broad range of bias conditions, it is necessary to determine the components of τ_T explicitly since their variations with bias are not the same.

The program finds C_{je} in a subroutine that is called on when flag 00 is set. The subroutine calculates C_{je} from eq. (7.4-22) and requires that the emitter area and the breakdown voltage be known or estimated. The value of τ_F is then calculated from the known values of τ_T , r_E , and C_{je} by using eq. (7.5-3).

Some of the equivalent circuit values are best obtained from separate measurements on the transistor. The Early conductance, for example, cannot be obtained very accurately from high-frequency measurements since $r_e g_{ce}$ is only a small part of A . The program assumes that V_A is known for the transistors of the integrated circuit process in question, so that

$$g_{ce} = \frac{I_C}{V_A + V_{CE}}$$

Note that the excess phase factor is sensitive to the exact value of the real part of D , whose main component at low frequencies is δ . At low frequencies the effect of excess phase is extremely small, so that the real part is insensitive to μ ; put the other way, μ is extremely sensitive to the exact value of the real part. Consequently, μ should not be evaluated from low-frequency measurements since small errors in the real part cause wild fluctuations in the value of μ . If operation is not to extend to high frequencies, μ is unimportant and may be neglected. In this case ($\leq f_T/20$) delay is ignored, and the circuit used becomes equivalent to the hybrid pi model.

If the measurement frequency is not too high, δ may be obtained from the real part of D ; better accuracy may be obtained by a separate, low-frequency measurement of D , and this is assumed in the equations.

To summarize, the complete equivalent circuit is obtained from (1) the $ABCD$ parameters at a high frequency, (2) a separate low-frequency measurement or calculation to obtain g_{ce} (and g_{cb}) and δ , and (3) a separate calculation

of C_{je} or τ_F . The latter evaluation is needed only if the size or bias conditions are to be changed.

Example

The *ABCD* parameters of a transistor at 0.4 GHz are given in this example (see Fig. C3.1), as well as the bias voltage V_A and the current and emitter area. The Early voltage is also known (or estimated), as are BV_{EB0} and δ . The program lists (see Fig. C3.2) these values and prompts for corrections, as shown, (i.e., $BV_{BE0} = 7.0$ V and $\delta = 0.0102$). After R/S is pressed, the value of μ on

```

XEQ "A>E"
ABCD>EQ.
FIND CJE, TF?
SF 00
RUN
ABCD:
F= 0.4000
ABCD
R01= 0.0154
R02= -76.1781
R03= 0.0061
R04= -145.3320
R05= 0.5846
R06= -79.7403
R07= 0.0811
R08= -94.9995
VCE, IC, Ae, VT, VA:
R09= 3.0000
R10= 8.0000
R11= 250.0000
R12= 0.0260
R13= 20.0000
BVEB0:
R17= 0.0000
delta
R21= 0.0000
OK?
7.0000 STO 17
.0102 STO 21

RUN
BIAS PAR.
IK, IC, T:
R14= 0.0000
R15= 0.0000
R16= 0.0000
OK?
80.0000 STO 14
.4300 STO 15
.2300 STO 16
RUN
WDTA

Iterate on {
0.0000
0.4454
0.4512
0.4511
}
mu
EQ. CCT.
delta, re, rF, rE, CJC, TT:
R21= 0.0102
R22= 0.0015
R23= 0.0047
R24= 0.2290
R25= 0.0320
rb, kb, gce, cce:
r26= 0.0404
r27= 0.3780
r28= 0.3478
r29= 0.5250
gcb, ka:
R30= 0.0000
R31= 1.1141
mu:
R18= 0.4511
TF, CJE
R19= 0.0279
R20= 0.9362

```

Figure C3.1. Program "A>E": example.

```

01*LBL "A>E"
02 SF 12
03 "ABCD>EQ."
04 AVIEW
05 CF 12
06 ADV
07 "FIND CJE, TF?"
08 ACA
09 PRBUF
10 " SF 00"
11 PROMPT
12 FC? 04
13 XEQ 00
14 "OK?"
15 FC? 04
16 PROMPT
17 XEQ "AE"
18 FS? 00 CONTROL
20 XEQ 01 PROGRAM
21 "BIAS PAR."
22 AVIEW
23 ADV
24 "IK,"
25 ACA
26 6
27 ACCHR
28 "C, T:"
29 ACA
30 PRBUF
31 14.016
32 PRREGX
33 ADV
34 "OK?"
35 PROMPT
36 "WDTA"
37 AVIEW
38 .031
39 WDTAX
40 ADV
41 RTN

42*LBL "AE"-----
43 XEQ 11
44 RCL 34
45 RCL 18 iter.
46 X<>Y
47 STO 18 loop
48 -
49 RND
50 X#0?
51 GTO "AE"-----
52 RTN

53*LBL 11 EQUIV.
54 VIEW 18
55 RCL 06 CKT.
56 0
57 XEQ 10 PARAM.
58 RCL 05
59 P-R
60 STO 39
61 X<>Y Cjc
62 RCL 00
63 /
64 CHS
65 STO 24 -----
66 RCL 04
67 XEQ 09
68 RCL 03
69 P-R
70 RCL 26 rE
71 RCL 21
72 *
73 +
74 CHS
75 STO 23 -----
76 RCL 12
77 RCL 10
78 / re'
79 -
80 STO 22 -----

81 X<>Y
82 STO 38
83 RCL 08
84 XEQ 09
85 RCL 07
86 P-R T_T
87 X<>Y
88 RCL 00
89 /
90 CHS
91 STO 25 -----
92 RCL 08
93 RCL 07
94 P-R
95 RCL 21
96 +
97 RCL 25 mu
98 /
99 RCL 00
100 X+2
101 /
102 RCL 23
103 RCL 24
104 *
105 -
106 RCL 25
107 /
108 STO 34 -----
109 RCL 10
110 RCL 13
111 RCL 09 gce
112 +
113 /
114 STO 28 -----
115 RCL 21
116 *
117 RCL 30
118 +
119 RCL 39
120 +

```

Figure C3.2. Program "A>E": listing.

successive iterations is viewed as shown. When no further change in μ occurs, the equivalent circuit values are printed. Finally, bias parameters I_{Kirk} , Γ_C , and Γ_r as stored are printed. These values are not needed in the equivalent circuit calculation but will be needed by program "E>A" in calculating the $ABCD$ parameters at various bias values. Measured values (or default values) should be stored in the appropriate registers at this point, as shown. The data are now in proper form for recording (on both sides of a magnetic card).

121 RCL 25	161 +	201 STO 20
122 /	162 CHS k_b	202 RCL 25
123 RCL 00 C_{ce}	163 RCL 24	203 RCL 24
124 X↑2	164 /	204 RCL 23
125 /	165 RCL 26	205 *
126 STO 29	166 /	206 -
127 RCL 38	167 STO 27	207 RCL 23 τ_F
128 CHS	168 RTN	208 RCL 22
129 RCL 00 r_b		209 -
130 /	169*LBL 09 SR	210 RCL 20
131 RCL 25	170 RCL 18	211 *
132 /	171 RCL 25 Subtract	212 -
133 STO 26	172 *	213 STO 19
134 RCL 02		214 RTN
135 0	173*LBL 10	
136 XEQ 10	174 RCL 23 Phase	215*LBL 01
137 RCL 01	175 RCL 24	216 SF 12
138 P-R	176 *	217 *EQ. CCT.*
139 RCL 28	177 +	218 ACA
140 RCL 12	178 RCL 00	219 PRBUF LIST
141 *	179 *	220 CF 12
142 RCL 10 k_a	180 R-D	221 ADV
143 /	181 -	222 18
144 +	182 RTN	223 ACCHR
145 RCL 25		224 SF 13
146 /	183*LBL "CJE"	225 ", RE PR, R"
147 RCL 29	184 RCL 11	226 ACA
148 /	185 .01	227 CF 13
149 RCL 00	186 * $C_{je}(A_e,$	228 *E, CJC, TT:*
150 X↑2	187 RCL 11	229 ACA
151 /	188 40	230 PRBUF
152 RCL 26	189 * $BV_{EB0},$	231 21.025
153 /	190 RCL 10	232 PRREGX
154 STO 31	191 / $I_c)$	233 ADV
155 RDN	192 LN	234 SF 13
156 RCL 00	193 SQRT	235 *RB, KB, GCE, *
157 /	194 /	236 *FCCE"
158 RCL 23	195 7	237 ACA
159 RCL 29	196 RCL 17	238 39
160 *	197 /	239 ACCHR
	198 1.25	240 PRBUF
	199 Y↑X	
	200 *	

Figure C3.2. Continued.

C-4 PROGRAMS "TPOLY", "T>P3", AND "T>N3/D3": POLYNOMIAL AND RATIONAL FUNCTION COEFFICIENTS FROM MAGNITUDE AND PHASE AT TWO FREQUENCIES (USE "T>N3/D3" WITH "TPOLY")

The basic program of this section, "TPOLY", takes the *ABCD* parameters of a network evaluated in magnitude and phase at two frequencies and converts this information into four sets of cubic polynomial coefficients—one set for each of the four *ABCD* parameters. By so doing, the *ABCD* parameters are

```

241 26.029
242 PRREGX
243 ADV
244 *GCB, KA:*
245 ACA
246 PRBUF
247 CF 13
248 30.031
249 PRREGX
250 ADV
251 12
252 ACCHR
253 ":"
254 ACA
255 PRBUF
256 18
257 PRREGX
258 ADV
259 FC? 00
260 RTN
261 *TF, CJE*
262 AVIEW
263 19.02
264 PRREGX
265 ADV
266 RTN

267*LBL 00
268 SF 12 LIST
269 *ABCD:*
270 ACA ABCD
271 PRBUF
272 CF 12
273 ADV
274 *F=*
275 ACA
276 RCL 00
277 PI
278 2
279 *
280 /

281 ACX
282 PRBUF
283 ADV
284 *ABCD*
285 AVIEW
286 1.000
287 PRREGX
288 ADV
289 *VCE, IC, Ae, VT*
290 ACA
291 ", VA:*
292 ACA
293 PRBUF
294 9.013
295 PRREGX
296 ADV
297 *BVEBO:*
298 AVIEW
299 17
300 PRREGX
301 ADV
302 18
303 ACCHR
304 PRBUF
305 21
306 PRREGX
307 ADV
308 END

USER KEYS:
11 ZREG
-11 SIZE
14 "A>E"
-21 PRKEYS
-22 RND
-23 PRFLAGS
-42 DSE
-51 FS?C
-62 PRREGX
-63 PRBUF
-72 DEL
-73 WDTAX
-74 RDTAX
-83 R-D

PRKEYS

PRFLAGS

STATUS:
SIZE= 040
Z= 01
DEG
FIX 4

FLAGS:
F 00 SET
F 01 CLEAR
F 02 CLEAR
F 03 CLEAR
F 04 CLEAR
F 05 CLEAR
F 06 CLEAR

CAT 1
LBL'A>E
LBL'AE
LBL'CJE
END
.END.
657 BYTES
00 BYTES
XEQ "A>E"
PRP ""

```

Figure C3.2. Continued.

described over the entire frequency range over which the cubic polynomial description is accurate. (For a cubic system, this is the entire frequency range.) This enables us to evaluate the loss and input and output immittances as a function of frequency by the methods discussed in Section 8.2. In addition, once the polynomial coefficients are known, the sensitivity analyses in Appendix A and B become available to us; the sensitivities of loss to the system components and device parameters can be calculated and statistical analysis carried out.

We have already seen—in program “RCU” in Appendix A—how the magnitude and phase of loss at two frequencies can be converted into cubic polynomial coefficients. This program uses exactly the same method, but it applies it to all four of the $ABCD$ parameters to obtain four cubic polynomials. The use of the polynomial coefficients in the design of feedback structures was discussed in Section 8.3 in connection with Fig. 8.13, where these coefficients were used to design an amplifier with desired loss and input and output immittances. Programs that make use of these polynomial coefficients are discussed in Section C-5.

Example

An example of the use of “T>P3” is given in the first column of the example in Fig. C5.1, Section C-5. We find the cubic polynomial coefficients of the $ABCD$ parameters of the bipolar transistor whose magnitude and phase were found in the example in program “E>A”. There, we recorded the magnitude and phase of the $ABCD$ parameters at four frequencies for a $95 \mu\text{m}^2$ transistor at $V_{CE}=0.75 \text{ V}$ and $I_C=2.0 \text{ mA}$. We now take (any) two of the four resulting matrices and calculate the cubic polynomial coefficients from them. The two frequencies used in the example are 0.0316 GHz (0.1985 Grad/s) and 0.10 GHz (0.623 Grad/s). (Use of any other pair of frequencies gives negligibly different results.)

In the example in the first and second columns of Fig. C5.1, four of the 16 polynomial coefficients are recognizable as the “core” coefficients of the transistor (with the sign changed because of the phase reversal). Equivalent circuit parameters r_E , C_{jc} , δ , and τ_T are found in registers R04, R09, R12, and R13, respectively, as indicated in eqs. (C3-1). Programs “TPOLY” and “T>P3” are listed in Fig. C4-1.

Program “T>N3/D3” (Use with “T>P3”)

This extension of Program “T>P3”, also listed in Fig. C4.1, likewise uses the magnitude and phase of a function at two frequencies, but it models the function as a rational function—a ratio of two cubic functions—in which the denominator coefficients are known. A complete development of the equations is given in Section 9.3. Memory register usage is compatible with “TPOLY” and “T>P3” as well as the programs described in the next section, Section C-5. The denominator coefficients d_0 , d_1 , d_2 , and d_3 are stored in R52–R55. They can be stored manually in response to a prompt, or can be read in from a card by executing “RDTAX” at the same prompt.

The program calls on “T>P3” as a subroutine but substitutes its own subroutine “M2” for subroutine “02” of “T>P3”. Subroutine “M2” performs the calculations described in Section 9.3 to obtain $L(j\omega)D(j\omega)$.

An example of the use of this program is given in Fig. 9.13 for the Wilson current source; the denominator is the (normalized) loss polynomial of the simple current source.

```

01*LBL "TPOLY"
  02 XEQ "T1"
  03 CF 03
  04 XEQ "T>P3"
  05 RTN

  06*LBL "T1"
07 "T AT F1: R-"
  08 AVIEW
  09 20.028
  10 RDTAX
  11 PRREGX
  12 ADV
  13 "F2: R-"
  14 AVIEW
  15 30.038
  16 RDTAX
  17 PRREGX
  18 ADV
  19 RTN

20*LBL "T>P3"
  21 RCL 21
  22 STO 12
  23 RCL 22
  24 STO 13
  25 RCL 31
  26 STO 14
  27 RCL 32
  28 STO 15
  29 .1
  30 XEQ 01
  31 RCL 23
  32 STO 12
  33 RCL 24
  34 STO 13
  35 RCL 33
  36 STO 14
  37 RCL 34
  38 STO 15
  39 4.1
  40 XEQ 01
  41 RCL 25
  42 STO 12
  43 RCL 26
  44 STO 13
  45 RCL 35
  46 STO 14
  47 RCL 36

  48 STO 15
  49 8.1
  50 XEQ 01
  51 RCL 27
  52 STO 12
  53 RCL 28
  54 STO 13
  55 RCL 37
  56 STO 14
  57 RCL 38
  58 STO 15
  59 12.1
  60 XEQ 01
  61 "A-POLY:"
  62 AVIEW
  63 .003
  64 PRREGX
  65 ADV
  66 "B"
  67 AVIEW
  68 4.007
  69 PRREGX
  70 ADV
  71 "C"
  72 AVIEW
  73 8.011
  74 PRREGX
  75 ADV
  76 "D"
  77 AVIEW
  78 12.015
  79 PRREGX
  80 ADV
  81 TONE 4
  82 "WR?"
  83 AVIEW
  84 PSE
  85 CLD
  86 .015
  87 WDTAX
  88 RTN

  89*LBL 01
  90 STO 18
  91 RCL 30
  92 X+2
  93 RCL 20
  94 X+2
  95 -

  96 1/X
  97 STO 29
  98 CHS
  99 STO 39
  100 RCL 20
  101 X+2
  102 *
  103 STO 17
  104 RCL 30
  105 X+2
  106 RCL 29
  107 *
  108 STO 16
  109 FC? 03
  110 XEQ 02
  111 FS? 03
  112 XEQ "M2"
  113 RCL 16
  114 RCL 40
  115 *
  116 RCL 17
  117 RCL 42
  118 +
  119 +
  120 STO IND 18
  121 ISG 18
  122 RCL 16
  123 RCL 41
  124 *
  125 RCL 17
  126 RCL 43
  127 *
  128 +
  129 STO IND 18
  130 ISG 18
  131 RCL 29
  132 RCL 40
  133 *
  134 RCL 39
  135 RCL 42
  136 *
  137 +
  138 STO IND 18
  139 ISG 18
  140 RCL 29
  141 RCL 41
  142 *
  143 RCL 39
  144 RCL 43

```

Figure C4.1. Programs "TPOLY" and "T>P3": listing.

Program “T>P3”: Registers

R00–R15	<i>ABCD</i> cubic polynomial coefficients
R16, R17	Used
R18	Coefficient index
R19	—
R20–R28	<i>ABCD</i> parameters at F_1
R29	Used
R30–R38	<i>ABCD</i> parameters at F_2
R39–R43	Used

Program “T>N3/D3”

In addition to the preceding registers for “T>P3”, we use

R52–R55	Denominator coefficients d_0 to d_3
R56, R57	Used

C-5 PROGRAMS FOR NETWORK CHARACTERIZATION FROM ABCD PARAMETERS: “P3>LA”, “FR”, “NMR”, AND “NWCALC”

These programs develop the voltage loss, input admittance, and output impedance of a network with given source and load resistances as discussed in Section 8.2. They begin with the (cubic) polynomial coefficients of the *ABCD* parameters (e.g., as found, by program “T>P3” or “TPOLY”). Program “P3>LA” implements eqs. (8.2-3), (8.2-4), and (8.2-9) for an *ABCD* matrix expressed by its cubic polynomial coefficients; “FR” calculates the loss in dB and phase as well as the real and imaginary parts of the input admittance and the output impedance as functions of frequency. Using program “NMR”, the user may normalize the constituent parts of the loss polynomial and immittances (i.e., A , B/R_L , CR_G , and DR_G/R_L) in three ways to aid the design process, as illustrated in Section 8.3. (the single stage hybrid feedback amplifier).

Program “NWCALC” is a control program that calls on “T>P3” as well as the programs of this section. “NWCALC” organizes the programs of this and the previous section to find a desired set of network properties from the *ABCD* parameters at two frequencies. It starts with the *ABCD* parameters of a network expressed in loss and phase at two frequencies (recorded on cards). It finds the polynomial coefficients and from these, using “P3>LA”, finds the polynomial coefficients of the loss and the rational function coefficients of Y_{in} and Z_o . It then lists the loss response and the input and output immittances as a function of frequency using “FR”. Finally, it normalizes the loss coefficients using “NMR”.

Program “P3>LA” requires that the polynomial coefficients of the *ABCD* parameters be stored initially in registers R00–R15. Source and load resistances are stored in registers R44 and R45; the program lists these and

prompts for corrections by asking “OK?”. The program then generates the sum represented by eq. (8.2-3) in registers R20–R23; the numerator of the input admittance, given by (8.2-4), is stored in registers R24–R27 and the denominator in R28–R31. The numerator of the output impedance is stored in R32–R35, and the denominator is stored in R36–R39. These locations are the ones addressed by program “FR” in calculating the response and immittances as a function of frequency. Program “P3>LA” also calculates the normalized or scaled polynomial of loss as well as the dc loss in dB and the cutoff frequency in gigahertz. The loss polynomial is scaled such that the dc and the M th coefficient are unity, as discussed in Section 2.2 and in program “N” in Appendix A. Coefficient M must be stored in register R46. If the polynomial is to be normalized to its quadratic coefficient, for example, 2 should be stored in R46.

Program “FR” is similar to program “BODE” in Appendix A, except that it calculates Bode plot information for the cubic polynomial in R20–R23 and Nyquist diagram information for the ratios of polynomials representing the input admittance and output impedance. The minimum and maximum (angular) frequencies for the calculations and the frequency increment are listed in the beginning of the calculation, and a prompt is provided to change any of them. Figure 8.3 (in Chapter 8) gives an example of the calculations of programs “P3>LA” and “FR”; another example for the hybrid feedback amplifier in Fig. 8.12 is given in Fig. 8.13.

Program “NMR” is intended as an aid in the design process. It allows the user to compare the relative contributions to the loss of the four $ABCD$ parameters on a coefficient-by-coefficient basis, enabling the user to correct a deficient loss polynomial or an undesirable port immittance. With flag 04 set, the original loss polynomial coefficients are printed, but with B divided by R_L and C multiplied by R_G ; thus the actual values of the four contributors of eq. (8.2-3) are printed, allowing an assessment of the importance of the various contributors to each coefficient. With flag 02 set, each contributor is divided by the coefficient of the loss polynomial of like degree, giving the sensitivity of each loss coefficient to A , B , C , or D . Examples of results using program “NMR” are given in Figs. 8.4 and 8.13c.

Example

In this example (see Fig. C5.1) we model the bipolar transistor as a set of cubic polynomials for A , B , C , and D . The program can be used for any two-port that can be modeled as a set of cubic polynomials; we develop the programs for calculating the $ABCD$ parameters of arbitrary two-ports in the remaining programs in this appendix. For purposes of showing how this series of programs works however, the modeling of the bipolar transistor—familiar ground by now—lends clarity to the process.

To use the program, we start with cleared program memory and set SIZE at 58 registers. Program “NWCALC” is read in from cards and executed, as shown in the example. The user is immediately prompted to read in program

```

XEQ "NWCALC"
R-TPOLY
T AT F1: R-
R20= 0.1985
R21= 0.0027
R22= -115.4484
R23= 0.0182
R24= -176.8928
R25= 0.0314
R26= -90.8392
R27= 0.0142
R28= -142.6802

F2: R-
R30= 0.6283
R31= 0.0079
R32= -93.7525
R33= 0.0184
R34= -170.2341
R35= 0.0994
R36= -86.9842
R37= 0.0293
R38= -111.9268

A-POLY:
R00= -0.0013
R01= -0.0125
R02= -0.0019
R03= -4.9667-06

B
R04= -0.0181
R05= -0.0050
R06= -0.0001
R07= -1.0652-06

C
R08= -0.0011
R09= -0.1579
R10= -0.0160
R11= -4.2274-05

D
R12= -0.0113
R13= -0.0433
R14= -0.0009
R15= -9.7628-06

WR?
R-P3>LA
RG, RL:
R44= 0.0000
R45= 0.0000
OK?
.0750 STO 44
STO 45
RUN

RG, RL:
R44= 0.0750
R45= 0.0750
OK?
LA, NUM
R20= -0.2546
R21= -0.1338
R22= -0.0053
R23= -3.2102-05

YI:
N
R24= -0.1518
R25= -0.7349
R26= -0.0262
R27= -0.0002

D
R28= -0.2432
R29= -0.0786
R30= -0.0032
R31= -1.9169-05

Z0:
N
R32= -0.0190
R33= -0.0082
R34= -0.0002
R35= -1.7974-06

D
R36= -0.0013
R37= -0.0243
R38= -0.0031
R39= -0.1372-06

NRM LA
bM=1, M:
R46= 0.0000
OK?
1.0000 STO 46
RUN

bM=1, M:

```

Figure C5.1. Program "NWCALC": example.

"TPOLY". When this is done, "TPOLY" prompts for cards containing the $ABCD$ parameters at two frequencies— F_1 and F_2 . It then prints the polynomial coefficients of the $ABCD$ parameters and asks if a card is to be written. If so, pass a card through the machine to save the polynomial coefficients; if not, press R/S. Operation returns to the control program, "NWCALC", which prompts for "P3>LA" to be read in. When this is done, "TPOLY" is automatically cleared from the machine, making room for succeeding programs. Program "P3>LA" prompts for source and load resistances R_G and R_L by printing the contents of R44 and R45. These are initially zero (assuming that memory was cleared), and the desired values are stored in these two registers (0.075 k Ω in the example). The corrected values are then printed; if

```

R46= 1.0000          LA
OK?                  F      L, dB  PH
                    0.010  -11.88  -178.11
                    0.022  -11.86  -175.93
                    0.046  -11.80  -171.28
                    0.100  -11.50  -161.59
                    0.215  -10.33  -143.53
                    0.464  -7.09   -118.25
                    1.000  -1.58   -93.00
                    2.154  5.47   -67.29
                    R-NMR
                    NORM. ABCD
                    ABCD COEF:
                    P.U. OF LA COEFS?
                    SF 02
                    SUM TO LA POLY?
                    SF 04
                    SUM TO NRM POLY?
                    CLFLGS
                    RUN

R40= 1.0000          YIN
R41= 1.0000          F      RE      IM      A
R42= 0.0761          0.010  0.627  0.177  R24= 0.0049
R43= 0.0009          0.022  0.639  0.381  R25= 0.0934
                    0.046  0.692  0.817  R26= 0.3504
                    0.100  0.932  1.718  R27= 0.1547
                    0.215  1.915  3.331
                    0.464  4.754  4.870
                    1.000  8.411  3.961
                    2.154  9.895  1.557
                    B/RL
                    R24= 0.9504
                    R25= 0.4946
                    R26= 0.2532
                    R27= 0.4424

R47= 0.0000          Z0
R48= 0.0000          F      RE      IM
R49= 0.0000          0.010  6.341  -6.937
OK?                  0.022  2.239  -4.897
                    0.046  0.726  -2.565
                    0.100  0.347  -1.236
                    0.215  0.259  -0.601
                    0.464  0.223  -0.326
                    1.000  0.160  -0.215
                    2.154  0.097  -0.137
                    DRG/RL
                    R24= 0.0003
                    R25= 0.08806
                    R26= 0.2246
                    R27= 0.0988

                    .0100 STO 47
                    1.0000 STO 48
                    3.0000 STO 49
                    RUN
                    FMIN, MAX,
                    PTS/DEC:
                    R47= 0.0100
                    R48= 1.0000
                    R49= 3.0000
                    OK?
                    RUN
                    R24= 0.0444
                    R25= 0.3235
                    R26= 0.1718
                    R27= 0.3041
    
```

Figure C5.1. Continued.

correct, press R/S. The loss polynomial and the numerator and denominator of the input admittance and output impedance are then printed.

The loss polynomial is then normalized as in program “N” in Appendix A. The program prompts for the coefficient whose value is to be equated to unity, either 1, 2, or 3. (Noninteger values may also be used.) The cutoff frequency and the dc loss (magnitude and dB) are then printed, followed by the coefficients of the normalized polynomial. (For a discussion of the value of this normalized polynomial description, see Section 2.2.)

The control program then prompts for cards containing program “FR”, which prints the loss and input and output immittances as a function of frequency. The program prompts for the minimum and maximum frequencies

```

TOTAL                                     CRG                                     CF 02
R28= 1.0000                               R24= -0.0001                               CF 04
R29= 1.0000                               R25= -0.0118                               RUN
R30= 1.0000                               R26= -0.0012
R31= 1.0000                               R27= -3.1705-06                            R24= 0.0049
                                           R25= 0.0934
                                           R26= 0.0266
                                           R27= 0.0001

DONE                                     DRG/RL
                                           R24= -0.0113
                                           R25= -0.0433
                                           R26= -0.0009
                                           R27= -9.7628-06                            B/RL
                                           R24= 0.9504
                                           R25= 0.4946
                                           R26= 0.0193
                                           R27= 0.0004

XEQ "NMR"
NORM. ABCD
ABCD COEF:
P.U. OF LA COEFS?
SF 02
SUM TO LA POLY?
SF 04
SUM TO NRM POLY?
CLFLGS
SF 04
RUN
A
R24= -0.0013
R25= -0.0125
R26= -0.0019
R27= -4.9667-06
B/RL
R24= -0.2420
R25= -0.0661
R26= -0.0014
R27= -1.4202-05

XEQ "NMR"
NORM. ABCD
ABCD COEF:
P.U. OF LA COEFS?
SF 02
SUM TO LA POLY?
SF 04
SUM TO NRM POLY?
CLFLGS
TOTAL
R28= -0.2546
R29= -0.1338
R30= -0.0053
R31= -3.2102-05
CRG
R24= 0.0003
R25= 0.0086
R26= 0.0171
R27= 0.0001
DRG/RL
R24= 0.0444
R25= 0.3235
R26= 0.0131
R27= 0.0003
TOTAL
R28= 1.0000
R29= 1.0000
R30= 0.0761
R31= 0.0009

```

Figure C5.1. Continued.

and the number of points per decade for this listing, followed by the listing itself.

The control program, "NWCALC", then prompts for program "NMR" to be read in. The latter program lists the flag options with a prompt. (The program sets flag 02 since this has been found to be most useful in design.) With flag 02 set, the coefficients of loss, the $ABCD$ parameters, are normalized to the coefficients of the loss polynomial of each degree individually. Thus we find the per unit (P.U.) contribution of the i th coefficient of A , for example, to the i th coefficient of L_A . Since A , B , C , and D are related directly to the circuit components, we can modify the circuit to give a desired performance. The sum of the four contributions is (of course) unity for this normalization, and the numbers printed are, by the sum rule, the sensitivities of the i th coefficient of loss to the i th coefficient of A , B , C , and D .

```

01*LBL "NWCALC"
02 CF 03
03 GTO 10

04*LBL "NW-D"
05 SF 03

06*LBL 10
07 FC? 03
08 "R-TPOLY"
09 FS? 03
10 "R-T>N3/D3"
11 XEQ 03
12 FC? 03
13 XEQ "TPOLY"
14 FS? 03
15 XEQ "T>N3/D3"
16 "R-P3>LA"
17 XEQ 03
18 XEQ "P3>LA"
19 "R-FR"
20 XEQ 03
21 XEQ "FR"
22 "R-NMR"
23 XEQ 03
24 XEQ "NMR"
25 BEEP
26 "DONE"
27 AVIEW
28 RTN

29*LBL 02
30 XEQ 02
31 RSUB
32 RTN

33*LBL 02
34 TONE 8
35 AVIEW
36 PSE
37 CLD
38 END
CAT 1

LBL"NWCALC
LBL"NW-D
END 130 BYTES

01*LBL "P3>LA"
02*LBL 00
03 "RC, RL:"
04 AVIEW
05 44.045
06 PRREGX
07 CF 22
08 "OK?"
09 PROMPT
10 FS? 22
11 GTO 00
12 4
13 STO 18

14*LBL 01
15 DSE 18
16 DEG
17 RCL IND 18
18 20
19 ST+ 18
20 RDN
21 STO IND 18
22 8
23 ST+ 18
24 RDN
25 STO IND 18
26 8
27 ST+ 18
28 RDN
29 STO IND 18
30 35
31 ST- 18
32 DSE 18
33 GTO 01
34 7.003
35 STO 18

36*LBL 02
37 RCL IND 18
38 RCL 45
39 /
40 16
41 ST+ 18
42 RDN
43 ST+ IND 18
44 8
45 ST+ 18
46 RDN

47 ST+ IND 18
48 4
49 ST+ 18
50 RDN
51 STO IND 18
52 28
53 ST- 18
54 DSE 18
55 GTO 02
56 11.007
57 STO 18

58*LBL 03
59 RCL IND 18
60 RCL 44
61 *
62 12
63 ST+ 18
64 RDN
65 ST+ IND 18
66 4
67 ST+ 18
68 RDN
69 STO IND 18
70 12
71 ST+ 18
72 RDN
73 ST+ IND 18
74 28
75 ST- 18
76 DSE 18
77 GTO 03
78 15.011
79 STO 18

80*LBL 04
81 RCL IND 18
82 RCL 45
83 /
84 RCL 44
85 *
86 8
87 ST+ 18
88 RDN
89 ST+ IND 18
90 4
91 ST+ 18
92 RDN

```

Figure C5.2. Programs "NWCALC" and "P3>LA": listing.

With flag 04 set (the status of flag 02 is immaterial in this case), the actual contributions of A , B , C , and D to the loss polynomial are printed, and the total is the loss polynomial itself. This option is convenient for determining the actual values of components to be added to the circuit to obtain desired performance.

With both flags 02 and 04 clear, the contributions of A , B , C , and D to the normalized loss polynomial are printed, and the total is the normalized polynomial. This option keeps the relative importance of the coefficients of loss in perspective; in the example, which models a bipolar transistor as a set of cubic polynomials, the higher-degree coefficients (the quadratic and cubic) are of lesser importance, and this option reflects this state of affairs.

```

93 ST+ IND 18
94 8
95 ST+ 18
96 RDN
97 ST+ IND 18
98 20
99 ST- 18
100 DSE 18
101 GTO 04
102 "LA, NUM"
103 AVIEW
104 20.023
105 PRREGX
106 ADV
107 "YI:"
108 AVIEW
109 RCL 44
110 ST/ 24
111 ST/ 25
112 ST/ 26
113 ST/ 27
114 "N"
115 AVIEW
116 24.027
117 PRREGX
118 ADV
119 "D"
120 AVIEW
121 28.031
122 PRREGX
123 ADV
124 "Z0:"
125 AVIEW
126 RCL 45
127 ST* 32
128 ST* 33
129 ST* 34
130 ST* 35
131 "N"
132 AVIEW
133 32.035

134 PRREGX
135 ADV
136 "D"
137 AVIEW
138 36.039
139 PRREGX
140 ADV
141 "NRN LA"
142 AVIEW

143*LBL 05
144 "bM=1, M:"
145 AVIEW
146 46
147 PRREGX
148 "OK?"
149 CF 22
150 TONE 6
151 PROMPT
152 FS? 22
153 GTO 05
154 RCL 46
155 20
156 +
157 STO 18
158 RCL 20
159 RCL IND 18
160 /
161 RCL 46
162 1/X
163 Y1X
164 STO 19
165 "FG="
166 PI
167 2
168 *
169 /
170 ARCL X
171 "F GHZ"
172 AVIEW
173 RCL 21

174 STO 41
175 RCL 22
176 STO 42
177 RCL 23
178 STO 43
179 RCL 19
180 ST* 41
181 ST* 43
182 X^2
183 ST* 42
184 ST* 43
185 RCL 20
186 ST/ 41
187 ST/ 42
188 ST/ 43
189 1
190 STO 40
191 ADV
192 "a0"
193 AVIEW
194 20
195 PRREGX
196 RCL 20
197 ABS
198 LOG
199 20
200 *
201 ACX
202 "dB"
203 ACA
204 PRBUF
205 ADV
206 "NP:"
207 AVIEW
208 40.043
209 PRREGX
210 ADV
211 END

CAT 1
LBL*P3>LA
END
443 BYTES

```

Figure C5.2. Continued.

The entire process represented in the example, including the three “NMR” flag options, takes about 10 min. During those 10 min., you may think of a better way to do (1) the circuit you are working on or (2) the program.

PROGRAM “NW-D”:

This program is identical to “NWCALC”, except that it allows the inclusion of a known denominator polynomial in the loss function. Instead of calling on program “TPOLY”, it calls on “T>N3/D3” (which, in turn, calls on

```

01+LBL "FR"                45 35                36 ST+ 18
02 "FMIN, MAX,"          46 XEQ 00            87 XEQ 05
03 AVIEW                 47 ADV              88 ST/ 50
04 "PTS/DEC:"           48 PI                89 X<>Y
05 AVIEW                 49 2                 90 ST- 51
06 47.049               50 *                 91 4
07 PRREGX               51 ST/ 47            92 ST- 18
08 CF 22                52 ST/ 48            93 RTN
09 "OK?"                53 RCL 49
10 PROMPT               54 LOG              94+LBL 02
11 FS? 22               55 1/X              95 FC? 03
12 GTO "FR"             56 STO 49            96 RTN
13 PI                   57 RTN              97 32
14 2                    58+LBL 00            98 ST+ 18
15 *                    59 STO 18            99 XEQ 05
16 ST* 47               60 RCL 47            100 ST/ 50
17 ST* 48               61 STO 16            101 X<>Y
18 RCL 49               62+LBL 03            102 ST- 51
19 1/X                  63 XEQ 10            103 32
20 10+X                 64 FS? 01            104 ST- 19
21 STO 49               65 XEQ 06            105 RTN
22 RCL 47               66 XEQ 11            106+LBL 05
23 STO 16               67 RCL 16            107 RCL IND 18
24 "LA"                 68 RCL 48            108 RCL 17
25 FS? 03               69 X(=Y?            109 *
26 "I, NUM"             70 RTN              110 DSE 18
27 AVIEW                 71 RCL 49            111 DSE 18
28 SF 01                72 ST* 16            112 RCL IND 18
29 CF 05                73 GTO 03            113 +
30 XEQ 08               74+LBL 10            114 RCL 16
31 23                   75 RCL 16            115 *
32 XEQ 00               76 X+2              116 ISG 18
33 ADV                  77 CHS              117 DEG
34 "YIN"                78 STO 17            118 RCL IND 18
35 AVIEW                 79 XEQ 05            119 RCL 17
36 SF 05                80 STO 50            120 *
37 CF 01                81 X<>Y            121 DSE 18
38 XEQ 08               82 STO 51            122 DSE 18
39 27                   83 FC? 05            123 RCL IND 18
40 XEQ 00               84 GTO 02            124 +
41 ADV                  85 4                 125 3
42 "Z0"                 126 ST+ 18
43 AVIEW                127 RDN
44 XEQ 00

```

Figure C5.3. Programs “FR” and “NMR”: listing.

“TPOLY”). In the course of the analysis, the user is prompted for a set of cubic denominator coefficients. The remainder of the program proceeds as in “NWCALC”: “P3>LA” finds the numerator of the loss polynomial. Since the denominator is common to all four *ABCD* parameters, the input and output immittances are unaffected by it. The response calculation of “FR” includes the denominator. Program “NMR” finds the normalized numerator coefficients.

```

128 R-P
129 RTN
130*LBL 06
131 RCL 50
132 LOG
133 20
134 *
135 STO 50
136 RTN
137*LBL 08
138 - F -
139 ACA
140 FS? 01
141 "L, dB PH"
142 FC? 01
143 - RE IM"
144 ACA
145 PRBUF
146 ADV
147 RTN
148*LBL 11
149 FS? 01
150 XEQ 12
151 RCL 16
152 PI
153 /
154 2
155 /
156 FIX 3
157 ACX
158 " "
159 ACA
160 FC? 01
161 XEQ 16
162 RCL 50
163 FS? 01
164 FIX 2
165 FC? 01
166 FIX 3
167 RND
168 ACX
169 " "
170 ACA
171 RCL 51
172 FS? 01
173 FIX 2
174 FC? 01
175 FIX 3
176 RND
177 ACX
178 PRBUF
179 FIX 4
180 RTN
181*LBL 16
182 RCL 51
183 RCL 50
184 P-R
185 STO 50
186 X<>Y
187 STO 51
188 RTN
189*LBL 12
190 RCL 51
191 100
192 X<=Y?
193 GTO 13
194 CHS
195 X<=Y?
196 RTN
197 360
198 ST+ 51
199 RTN
200*LBL 13
201 360
202 ST- 51
203 END
CAT 1
LBL*FR
END
420 BYTES
01*LBL "NMR"
02 SF 02
03 SF 12
04 "NORM. ABCD"
05 AVIEW
06 ADV
07 CF 12
08 "ABCD COEF:"
09 AVIEW
10 ADV
11 "P.U. OF LA COEF"
12 "+S?"
13 AVIEW
14 "SF 02"
15 AVIEW
16 "SUM TO LA POLY?"
17 AVIEW
18 "SF 04"
19 AVIEW
20 "SUM TO NRM POLY"
21 "+?"
22 AVIEW
23 "CLFLGS"
24 TONE 9
25 PROMPT
26 "A"
27 AVIEW
28 RCL 00
29 STO 24
30 RCL 01
31 STO 25
32 RCL 02
33 STO 26
34 RCL 03
35 STO 27
36 FC? 04
37 XEQ 02
38 0
39 STO 28
40 STO 29
41 STO 30
42 STO 31
43 XEQ 07

```

Figure C5.3. Continued.

Program listings for “NWCALC”, “NW-D”, and “P3>LA” are given in Fig. C5.2. Listings for “FR” and “NMR” are given in Fig. C5.3.

Program “NWCALC” and Associated Programs: Registers

- R00–R15 *ABCD* parameter cubic polynomial coefficients
- R16, R17 Used by T>P3
- R18 Index for placing network functions
- R19 F_0

44 XEQ 01	85 FC? 04	123 ST/ 24
45 *B/RL*	86 XEQ 02	124 ST/ 25
46 AVIEW	87 XEQ 03	125 ST/ 26
47 RCL 04	88 XEQ 04	126 ST/ 27
48 STO 24	89 XEQ 07	127 RTN
49 RCL 05	90 XEQ 01	
50 STO 25	91 XEQ 08	128*LBL 04
51 RCL 06	92 CF 02	129 RCL 44
52 STO 26	93 CF 04	130 ST* 24
53 RCL 07	94 TONE 7	131 ST* 25
54 STO 27	95 RTN	132 ST* 26
55 FC? 04		133 ST* 27
56 XEQ 02	96*LBL 02	134 RTN
57 XEQ 03	97 FS? 02	
58 XEQ 07	98 GT0 06	135*LBL 01
59 XEQ 01	99 RCL 20	136 24.027
60 *CRG*	100 ST/ 24	137 PRREGX
61 AVIEW	101 ST/ 25	138 ADV
62 RCL 08	102 ST/ 26	139 RTN
63 STO 24	103 ST/ 27	
64 RCL 09	104 RCL 19	140*LBL 07
65 STO 25	105 ST* 25	141 RCL 24
66 RCL 10	106 ST* 27	142 ST+ 28
67 STO 26	107 X↑2	143 RCL 25
68 RCL 11	108 ST* 26	144 ST+ 29
69 STO 27	109 ST* 27	145 RCL 26
70 FC? 04	110 RTN	146 ST+ 30
71 XEQ 02		147 RCL 27
72 XEQ 04	111*LBL 06	148 ST+ 31
73 XEQ 07	112 RCL 20	149 RTN
74 XEQ 01	113 ST/ 24	
75 *DRG/RL*	114 RCL 21	150*LBL 08
76 AVIEW	115 ST/ 25	151 *TOTAL*
77 RCL 12	116 RCL 22	152 AVIEW
78 STO 24	117 ST/ 26	153 28.031
79 RCL 13	118 RCL 23	154 PRREGX
80 STO 25	119 ST/ 27	155 ADV
81 RCL 14	120 RTN	156 END
82 STO 26		
83 RCL 15	121*LBL 03	LBL*NMR
84 STO 27	122 RCL 45	END
		384 BYTES

Figure C5.3. Continued.

R20–R23	} These five groups of four registers each are used for various storage requirements, as can be seen from the examples
R24–R27	
R28–R31	
R32–R35	
R36–R39	
R40–R43	Normalized polynomial
R44, R45	R_G, R_L
R46	M , the degree to which the L_A polynomial is normalized
R47–R49	F_{\min}, F_{\max} , points per decade (converted to angular freq. and multiplier during execution)
R50, R51	Used (for loss and phase and immittances)
R52–R55	d_0, d_1, d_2, d_3 (denominator coefficients)
R56, R57	Used

C-6 PROGRAM “ABCD”: PROGRAMMABLE TWO-PORT NETWORK CALCULATOR

This program does for $ABCD$ matrices what a programmable pocket calculator does for numbers. By reading this program into program memory, the HP 41C calculator is converted into a network calculator that not only gives answers manually by keyboard operations, but can also be programmed to analyze networks. In response to a single command, it multiplies matrices (in either order), adds them, inverts matrices, and changes their sign, completing the basic four functions. In addition, it finds the resulting matrix when either the input or the output leads are permuted; it also forms $ABCD$ matrices of either shunt or series immittances.

Program “ABCD” also provides for storing and recalling matrices, as well as exchanging the register locations of pairs of matrices. As a matter of mnemonic convenience, we designate registers R00–R08 as register set RS00 and call this the working register set, or register set X. The register set beginning with R10 is designated RS10. That beginning with R20 is RS20, and similarly for RS30 and RS40. Intermediate results of calculations can be stored in any of these register sets with certain limitations to be described. Other register sets can be similarly defined up to the memory capacity of the calculator. Beyond this capacity, magnetic cards are used to store matrices. The store function copies the matrix in the working register into a register set designated by the user: to store the matrix in RS30, for example, the command is “30, XEQ ST”. Similarly, to recall the matrix in RS30 into the working register, the command is “30, XEQ RC”. Stack rules are not implemented in “ABCD”, however, so the recall function should be used with care; the matrix in the working register is lost when the recall function is executed.

An exchange command, “X-” (as used in previous programs of this appendix) exchanges the register locations of the matrix in the working register with one in a location designated by the user. To exchange the matrix in the X register with that in RS30, for example, the command is “30, XEQ X-”. This

operation is reversible. Unlike the ST and RC commands, no information is lost.

By analogy with the terminology used by the host calculator (the HP 41C), the working register set will be designated X, and RS10 will be designated Y. Two-matrix functions such as multiply and add use register sets X and Y for the input data. For the multiply function the matrix in X *pre*multiplies that in Y. To *post*multiply the matrix in X by that in Y, exchange the matrices in X and Y first. The command "10, XEQ X-" will do this, but this command is used frequently enough to warrant its own command name, "X-Y".

Several of the operations (multiply, the permuting operations, and the matrix inverse operation) use RS20 registers as scratch-pad memory. This register set should not be used to store matrices or should be used with care. This leaves RS10 (Y), RS30, and RS40 for internal matrix storage. Where more memory is required, the user can designate other register sets up to the capacity of the machine. The command "70 XEQ ST", for example, will place the matrix in X in registers R70–R78 (if they exist).

Beyond this, magnetic cards may be used for matrix storage. Command "RD" reads a matrix from a card into register set X, while "W" writes the matrix in X onto a card.

Program "ABCD" is compatible with program "CNV" previously described and with program "SP", a spanning network program to be described later. Program "ABCD" alone suffices for many network computations, as we shall see in the examples. Combined with "CNV" and "SP", essentially the complete range of two-port calculations can be done. The HP 41C becomes a powerful network calculator using these programs, one that itself is programmable.

To summarize the foregoing, the functions performed by "ABCD" sub-routines are the following:

- "ML" Premultiplies matrix in register set Y by that in X and leaves result in X.
- "PI" Permutes input leads of network whose *ABCD* matrix is in register set X. Result is left in X.
- "PO" Permutes output leads of network whose *ABCD* matrix is in X and leaves result in X.
- "DT" Finds determinant of matrix in X. Real and imaginary parts are stored in R23 and R24. This is used as a subroutine for several of the other subroutines.
- "Y" After moving the matrix in X to Y (to preserve it for future calculations), this subroutine forms the matrix of a shunt admittance in X. The admittance consists of the *series* combination of a resistance and a capacitance, to be keyed in from the keyboard in response to prompts. Where the shunt admittance is simply a conductance, a large value (e.g., 10^{12}) must be keyed in for the capacitance.
- "Z" This subroutine is the same as "Y" except that it is for a series impedance, consisting of the *parallel* combination of a resistance

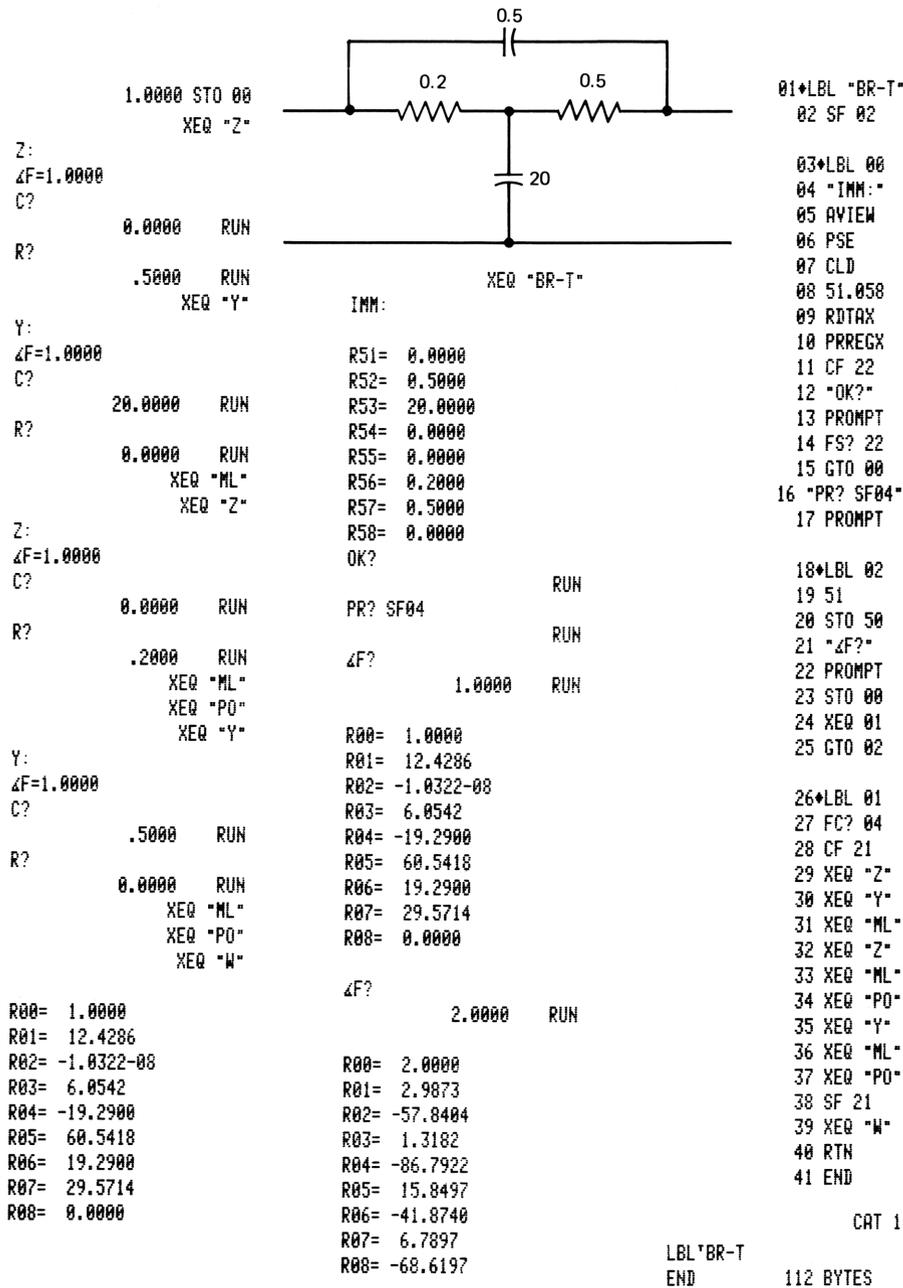
	and a capacitance. Where the series impedance consists of a pure capacitance, a large value must be keyed in for the resistance.
“AD”	Adds matrices in X and Y and leaves result in X. The matrix originally in Y is preserved in Y.
“INV”	Finds the inverse of the <i>ABCD</i> matrix in X and leaves result in X.
“CHS”	Changes the sign of the matrix in X.
“ST”	Stores (or copies) the matrix in X into the register set previously designated by the number preceding the command.
“RC”	Copies the matrix in the previously designated register set into X.
“X-Y”	Exchanges the matrices in X and Y.
“X-”	Exchanges the matrices in X and the previously designated register set.
“RD”	Reads matrix from a magnetic card into X and prints the matrix.
“W”	Prompts user to write a data card of the <i>ABCD</i> matrix in X and prints the values.

Most of the subroutines of “ABCD” can be assigned to user keys, as shown at the end of the program listing. This is convenient for manual use of the calculator and also for programming.

Unlike the other subroutines, “Y” and “Z” interrupt execution and prompt the user for values of the immittances. An alternate mode of operation is obtained by setting flag 02, which directs the calculator to find the values of *C* and *R* in registers as directed by the register value stored in register 50. This is illustrated in the next section.

Examples of the use of “ABCD” are given in Chapters 8 and 9; all these use the programmable feature to be discussed later. One simple example here serves to illustrate use of “ABCD” in its keyboard entry mode.

The bridged-T *RC* network of the examples in Fig. C6.1 is analyzed by using the program. We take the T network as a ladder and the bridging capacitor as a spanning network. The frequency at which the analysis is to be made is stored in R00. Next, execute “Z” to form the *ABCD* matrix of the series arm nearest the output. The program prompts for *C* (key in 0) and *R* (key in 0.5). Then execute “Y”; the program prompts for *C* and *R* (key in 20 and 0). Then execute “ML” to multiply the two *ABCD* matrices. Note that the shunt admittance is in register set X and the series impedance is in Y, so that the matrices are in correct order for “ML”. Continue by executing “Z” (*C*=0, *R*=0.2), and “ML”. To add the effect of the spanning capacitor, execute “PO” to permute the output leads, as discussed in Section 9.1. Then execute “Y” for the spanning network (*C*=0.5, *R*=0) and “ML”. Finally, execute “PO” to return the network to the original configuration. The *ABCD* matrix of the network is in register set X, where it can be printed or recorded on a card. The individual steps and the result are shown in Fig. C6.1. Alternatively, it can be stored in an available register set by executing “40,



```

1.0000 STO 00
  XEQ "Z"

Z:
ΔF=1.0000
C?

0.0000 RUN

R?

.5000 RUN
  XEQ "Y"

Y:
ΔF=1.0000
C?

20.0000 RUN

R?

0.0000 RUN
  XEQ "ML"
  XEQ "Z"

Z:
ΔF=1.0000
C?

0.0000 RUN

R?

.2000 RUN
  XEQ "ML"
  XEQ "PO"
  XEQ "Y"

Y:
ΔF=1.0000
C?

.5000 RUN

R?

0.0000 RUN
  XEQ "ML"
  XEQ "PO"
  XEQ "W"

R00= 1.0000
R01= 12.4286
R02= -1.0322-08
R03= 6.0542
R04= -19.2900
R05= 60.5418
R06= 19.2900
R07= 29.5714
R08= 0.0000

```

```

IMM:
R51= 0.0000
R52= 0.5000
R53= 20.0000
R54= 0.0000
R55= 0.0000
R56= 0.2000
R57= 0.5000
R58= 0.0000
OK?

RUN

PR? SF04

ΔF?

1.0000 RUN

R00= 1.0000
R01= 12.4286
R02= -1.0322-08
R03= 6.0542
R04= -19.2900
R05= 60.5418
R06= 19.2900
R07= 29.5714
R08= 0.0000

ΔF?

2.0000 RUN

R00= 2.0000
R01= 2.9873
R02= -57.8404
R03= 1.3182
R04= -86.7922
R05= 15.8497
R06= -41.8740
R07= 6.7897
R08= -68.6197

```

```

01*LBL "BR-T"
02 SF 02

03*LBL 00
04 "IMM:"
05 AVIEW
06 PSE
07 CLD
08 51.058
09 RDTAX
10 PRREGX
11 CF 22
12 "OK?"
13 PROMPT
14 FS? 22
15 GTO 00
16 "PR? SF04"
17 PROMPT

18*LBL 02
19 51
20 STO 50
21 "ΔF?"
22 PROMPT
23 STO 00
24 XEQ 01
25 GTO 02

26*LBL 01
27 FC? 04
28 CF 21
29 XEQ "Z"
30 XEQ "Y"
31 XEQ "ML"
32 XEQ "Z"
33 XEQ "ML"
34 XEQ "PO"
35 XEQ "Y"
36 XEQ "ML"
37 XEQ "PO"
38 SF 21
39 XEQ "W"
40 RTN
41 END

CAT 1

LBL"BR-T"
END
112 BYTES

```

Figure C6.1. Program "ABCD": examples, manual and programmed.

XEQ ST" to permit building up a more complex RC network or one containing transistors.

The following example shows how the previous operations can be programmed to run automatically. The manual operations done previously are contained in steps 29–37 of the program listing in column 3 (Fig. C6.1). The rest of the steps provide convenience features: listing of the immittances,

```

01*LBL "ABCD"
02*LBL "ML"
03 SREG 21
04 CLE
05 SREG 23
06 CLE
07 RCL 00
08 STO 20
09 21.1
10 STO 19
11 RCL 01
12 RCL 02
13 RCL 12
14 RCL 11
15 XEQ 01
16 RCL 03
17 RCL 04
18 RCL 16
19 RCL 15
20 XEQ 00
21 RCL 01
22 RCL 02
23 RCL 14
24 RCL 13
25 XEQ 01
26 RCL 03
27 RCL 04
28 RCL 18
29 RCL 17
30 XEQ 00
31 RCL 05
32 RCL 06
33 RCL 12
34 RCL 11
35 XEQ 01
36 RCL 07
37 RCL 08
38 RCL 16
39 RCL 15
40 XEQ 00
41 RCL 05
42 RCL 06
43 RCL 14
44 RCL 13
45 XEQ 01
46 RCL 07

47 RCL 08
48 RCL 18
49 RCL 17
50 XEQ 00
51 20
52 XEQ "X-"
53 RTN

54*LBL 01
55 RDN
56 +
57 X<>Y
58 R†
59 *
60 P-R
61 ST+ IND 19
62 X<>Y
63 ISG 19
64 ST+ IND 19
65 1
66 ST- 19
67 RTN

68*LBL 00
69 XEQ 01
70 RCL IND 19
71 ISG 19
72 RCL IND 19
73 X<>Y
74 R-P
75 1
76 ST- 19
77 RDN
78 STO IND 19
79 X<>Y
80 ISG 19
81 STO IND 19
82 ISG 19
83 RTN

84*LBL "PI"
85 XEQ "INV"
86 XEQ "PO"
87 XEQ "INV"
88 RTN

89*LBL "PO"
90 XEQ "DT"
91 RCL 08
92 RCL 07
93 P-R
94 1
95 -
96 STO 21
97 X<>Y
98 STO 22
99 RCL 02
100 RCL 01
101 P-R
102 STO 01
103 X<>Y
104 STO 02
105 RCL 21
106 RCL 23
107 -
108 ST+ 01
109 RCL 22
110 RCL 24
111 -
112 ST+ 02
113 RCL 02
114 RCL 01
115 R-P
116 STO 01
117 X<>Y
118 STO 02
119 RCL 22
120 RCL 21
121 R-P

122*LBL "R"
123 ST/ 01
124 ST/ 03
125 ST/ 05
126 ST/ 07
127 X<>Y
128 ST- 02
129 ST- 04
130 ST- 06
131 ST- 08
132 RTN

```

Figure C6.2. Program "ABCD": listing.

133*LBL "DT"	176 ISG 50	220 STO 08
134 RCL 02	177 DEG	221 RTN
135 RCL 08	178 RCL IND 50	
136 +	179 ISG 50	222*LBL "AD"
137 RCL 01	180 DEG	223 8
138 RCL 07	181 RTN	224 STO 19
139 *		225 XEQ 05
140 P-R	182*LBL 03	226 RTN
141 STO 23	183 "C?"	
142 X<>Y	184 PROMPT	227*LBL 05
143 STO 24	185 *	228 RCL IND 19
144 RCL 04	186 STO 05	229 DSE 19
145 RCL 06	187 "R?"	230 RCL IND 19
146 +	188 PROMPT	231 P-R
147 RCL 03	189 RTN	232 11
148 RCL 05		233 ST+ 19
149 *	190*LBL 04	234 RDN
150 P-R	191 XEQ "X-Y"	235 RCL IND 19
151 ST- 23	192 RCL 10	236 DSE 19
152 X<>Y	193 "ΔF="	237 RCL IND 19
153 ST- 24	194 ARCL X	238 P-R
154 RTN	195 AVIEW	239 X<>Y
	196 STO 00	240 RDN
155*LBL "Y"	197 FS? 02	241 +
156 "Y:"	198 XEQ 02	242 X<>Y
157 AVIEW	199 FC? 02	243 R↑
158 SF 05	200 XEQ 03	244 +
159 XEQ 04	201 STO 03	245 R-P
160 STO 03	202 *	246 X<>Y
161 STO 04	203 1	247 9
162 RTN	204 STO 01	248 ST- 19
	205 STO 07	249 RDN
163*LBL "Z"	206 FS? 05	250 STO IND 19
164 "Z:"	207 CHS	251 X<>Y
165 AVIEW	208 FS? 05	252 DSE 19
166 CF 05	209 X<>Y	253 STO IND 19
167 XEQ 04	210 R-P	254 DSE 19
168 STO 05	211 1/X	255 GT0 05
169 STO 06	212 ST* 03	256 RTN
170 RTN	213 ST* 05	
	214 X<>Y	257*LBL "INV"
171*LBL 02	215 CHS	258 XEQ "DT"
172 VIEW 50	216 STO 04	259 RCL 01
173 RCL IND 50	217 STO 06	260 X<> 07
174 *	218 0	261 STO 01
175 STO 05	219 STO 02	262 RCL 02

Figure C6.2. Continued.

263 X<> 00	303*LBL "X-Y"		CAT 1
264 STO 02	304 10	LBL'ABCD	
265 1	305 XEQ "X-"	LBL'ML	
266 FS?C 10	306 RTN	LBL'PI	
267 CHS		LBL'PO	
268 ST* 03	307*LBL "X-"	LBL'R	
269 ST* 05	308 STO 19	LBL'DT	
270 RCL 24	309 .008	LBL'Y	
271 RCL 23	310 STO 09	LBL'Z	
272 R-P		LBL'AD	
273 XEQ "R"	311*LBL c	LBL'INV	
274 RTN	312 RCL IND 09	LBL'CHS	
275*LBL "CHS"	313 X<> IND 19	LBL'ST	
276 -1	314 STO IND 09	LBL'RC	
277 ST* 01	315 ISG 19	LBL'X-Y	
278 ST* 03	316 DEG	LBL'X-	
279 ST* 05	317 ISG 09	LBL'U	
280 ST* 07	318 GTO c	LBL'RD	
281 RTN	319 RTN	LBL'W	
		END	640 BYTES PRKEYS
282*LBL "ST"	320*LBL "U"		
283 STO 19	321 XEQ "X-Y"	USER KEYS:	
284 .008	322 SREG 02	11 SREG	
285 STO 09	323 CLS	-11 SIZE	
286 GTO a	324 0	14 "PO"	
	325 STO 00	-14 "PI"	
287*LBL "RC"	326 1	15 "Y"	
288 1.001	327 STO 01	-15 "Z"	
289 *	328 STO 07	21 "X-Y"	
290 .008	329 RCL 10	-21 PRKEYS	
291 +	330 STO 00	-22 RND	
292 STO 09	331 RTN	-23 "CM"	
293 0	332*LBL "RD"	24 "ML"	
294 STO 19	333 .008	-24 "INV"	
	334 RDTAX	25 "AD"	
295*LBL a	335 PRREGX	-25 "CHS"	
296 RCL IND 09	336 RTN	-42 ISE	
297 STO IND 19		-51 FS?C	
298 ISG 19	337*LBL "W"	-62 PRREGX	
299 DEG	338 .008	-63 PRBUF	
300 ISG 09	339 PRREGX	-72 DEL	
301 GTO a	340 CLD	-73 WDTAX	
302 RTN	341 WDTAX	-74 RDTAX	
	342 END	-83 R-D	

Figure C6.2. Continued.

choice of printing during the run, prompting for the frequency at which the run is to be made, and calling for a card to write the results. The computation steps are just those shown in the keyboard entry example. The program is executed in the middle column of the example. Program “ABCD” is listed in Fig. C6.2.

C-7 USE OF “ABCD” AS A PROGRAMMABLE NETWORK CALCULATOR

This section describes how program “ABCD” is to be used to analyze circuits by programs written by the user. We begin by describing each step of the program “CSC” for analyzing the cascode stage shown in the example (see Fig. C7.1) at the top of the first column. After the program label, steps 2–4 prompt the user to read in the $ABCD$ matrix of the second-stage transistor from a magnetic card (we used the transistor example from program “ $E > A$ ”). At step 05 the input leads are permuted, changing transistor Q_2 to a common base stage. In step 06 the $ABCD$ parameters of this stage are transferred to register set Y (RS10). The user is then prompted to read the data card for Q_1 , the first transistor. This places the $ABCD$ matrix of the first stage in the working register. At step 10, the matrix of Q_1 premultiplies the matrix of Q_2 , completing the computation. The remaining steps label the output, print it, and prompt for writing a data card; if the result is not to be saved, press R/S. The results of the calculation are shown in the second column.

Programming is simplified by use of the user-assigned keys for the subroutines of “ABCD”. Step 05, “PI”, for example, is programmed by pressing (shift) 15 (the button labeled e^x) in user mode.

The second example, the Darlington pair, illustrates the use of the “Y” subroutine. A $0.75 \text{ k}\Omega$ resistor is connected from base to emitter of the output transistor to provide bias for the first transistor. At step 05, subroutine “Y” is called for. As seen in the calculation (in the third column), execution stops for the user to supply the value of the resistor (the parallel capacitance is zero). This done, execution proceeds with the first transistor called for. Finally, the resulting $ABCD$ matrix is printed.

Where the calculation is to be performed at several frequencies, it is tedious to key in the values for “Y” each time. As noted in Appendix C-6, the “Y” values can be stored previously in registers (e.g., R51 and R52) and 51 stored in R50. With flag 02 set, the calculation will call for the stored values rather than prompting the user to key them in.

In the third example, a program for calculating the $ABCD$ parameters of the simple current mirror and the Wilson current source is given (see Fig. C7.2). The program follows the development given in Section 9.1. In this program the diode consists of a transistor with collector and base short-circuited. The short circuit is treated as a y spanning network (we used 1 F of capacitance and zero resistance) and is included in program memory at steps 12–17. The transistor with $V_{CE} = 0.75 \text{ V}$ and $I_C = 2.0 \text{ mA}$ is read in from a card; flag 02, set by the program, causes the “Y” subroutine to apply the short circuit to the diode

```

01+LBL "CSC"
02 "XSTR 2"
03 AVIEW
04 XEQ "RD"
05 XEQ "PI"
06 XEQ "X-Y"
07 "XSTR 1"
08 AVIEW
09 XEQ "RD"
10 XEQ "ML"
11 "CSC:"
12 AVIEW
13 XEQ "W"
14 RTN
15 END

XSTR 2
R00= 2.5133
R01= 0.0154
R02= -76.1781
R03= 0.0061
R04= -145.3320
R05= 0.5846
R06= -79.7403
R07= 0.0811
R08= -94.9995

XSTR 1
R00= 2.5133
R01= 0.0154
R02= -76.1781
R03= 0.0061
R04= -145.3320
R05= 0.5846
R06= -79.7403
R07= 0.0811
R08= -94.9995

XEQ "CSC"
Q2:
R00= 0.6283
R01= 0.0033
R02= -104.9161
R03= 0.0136
R04= -171.7003
R05= 0.0459
R06= -89.1388
R07= 0.0199
R08= -110.2857

Y:
ZF=0.6283
C?
R?
.7500 RUN
Q1:
R00= 0.6283
R01= 0.0050
R02= -98.9843
R03= 0.0309
R04= -175.2716
R05= 0.0468
R06= -87.6079
R07= 0.0234
R08= -113.6491

01+LBL "DRL"
02 "Q2:"
03 AVIEW
04 XEQ "RD"
05 XEQ "Y"
06 XEQ "ML"
07 XEQ "PO"
08 XEQ "X-Y"
09 "Q1:"
10 AVIEW
11 XEQ "RD"
12 XEQ "PO"
13 XEQ "ML"
14 XEQ "PO"
15 "DRL. PR:"
16 AVIEW
17 XEQ "W"
18 END

CSC:
R00= 2.5133
R01= 0.0036
R02= -42.3333
R03= 0.0061
R04= -140.0339
R05= 0.0562
R06= 7.3769
R07= 0.0847
R08= -88.8795

DRL. PR:
R00= 0.6283
R01= 0.0083
R02= -101.3638
R03= 0.0138
R04= -170.3459
R05= 0.0460
R06= -87.4464
R07= 0.0010
R08= -69.8336

```

Figure C7.1. Programs "CSC" and "DRL" for use with "ABCD".

transistor automatically; the current mirror transistor $ABCD$ matrix postmultiplies that of the diode, giving the $ABCD$ matrix of the current mirror.

As noted in the text, the simple current mirror forms a part of the Wilson current source. The program for the Wilson source incorporates the simple current mirror program as a subroutine, and finds the characteristics by successive application of the permute operations as described in the text. The calculations for both current sources are shown at 31.6 MHz.

As a final example, we find the $ABCD$ parameters of the Design B 300 MHz amplifier previously analyzed by Program "DEV" and "AN3" of Appendix B,

```

                                XEQ "CM"
MIRROR                          01*LBL "CM"
                                02 SF 12
                                03 "MIRROR"
DIODE                            04 AVIEW
Y:                               05 CF 12
ΔF=0.1985                       06 ADV
                                07 XEQ "CK"
                                08 XEQ "M"
                                09 RTN
                                10*LBL "CK"
                                11 SF 02
                                12 1 E12
                                13 STO 51
                                14 0
                                15 STO 52
                                16 51
                                17 STO 50
                                18 STO 39
                                19 "DIODE"
                                20 AVIEW
                                21 .008
                                22 RDTAX
                                23 XEQ "PO"
                                24 XEQ "Y"
                                25 XEQ "ML"
                                26 XEQ "PO"
                                27 XEQ "X-Y"
                                28 "XSTR"
                                29 AVIEW
                                30 .008
                                31 RDTAX
                                32 XEQ "X-Y"
                                33 XEQ "ML"
                                34 "CM:"
                                35 AVIEW
                                36 RTN
                                37*LBL "WLSN"
                                38 SF 12
                                39 "WLSN CS"
                                40 AVIEW
                                41 CF 12
                                42 ADV
                                43 XEQ "CK"
                                44 XEQ "PI"
                                45 XEQ "PO"
                                46 XEQ "X-Y"
                                47 "OUT XSTR"
                                48 AVIEW
                                49 .008
                                50 RDTAX
                                51 XEQ "PI"
                                52 XEQ "X-Y"
                                53 XEQ "ML"
                                54 XEQ "PI"
                                55 "WLSN:"
                                56 AVIEW
                                57 XEQ "M"
                                58 END
                                CAT 1
                                LBL"CM
                                LBL"CK
                                LBL"WLSN
                                END
                                195 BYTES
XSTR
CM:
OUT XSTR
WLSN:
R00= 0.1985
R01= 0.0027
R02= -115.5228
R03= 0.0100
R04= -176.8837
R05= 0.1021
R06= -113.5283
R07= 1.0239
R08= -178.9336
                                XEQ "WLSN"
WLSN CS
DIODE
Y:
ΔF=0.1985
                                51.0000
                                51.0000

```

Figure C7.2. Programs "CM" and "WLSN": example and listing.

and shown in Fig. 5.19. In that program, the transistors were represented by an approximate equivalent circuit; parasitic elements were combined with the transistor equivalent circuit immittances, and delays were approximated. Here (see Figs. C7.3 and C7.4), we use the actual transistor two-port parameters, which include the delays, and we represent interstage parasitics including the biasing diodes and shunt capacitance explicitly; thus reducing the possibility of errors of estimation and of omitting potentially important circuit immittances. Also, the program is simpler for the user.

This program uses the approach suggested in Section 9.1, using output permutative feedback to incorporate the third stage and overall feedback

```

DES-B
XEQ "D-B"
T3:
R00= 0.1985
R01= 0.0016
R02= -132.8915
R03= 0.0051
R04= -176.9496
R05= 0.0456
R06= -93.6191
R07= 0.0120
R08= -147.9051
DESIGN B
ABCD PARAM:
Z F:
R00= 0.1985
ABCD:
R01= 0.0016
R02= -150.2751
R03= 2.8699-05
R04= -121.5517
R05= 0.0403
R06= -171.6121
R07= 3.8610-05
R08= -77.8539
XEQ "RQU"
QUINTIC COEF
FROM LOSS
dB? CF 01
GHZ?, SF 02
DATA ON CARDS?
SF 03
READ DATA
Z F1 = 0.1985
MAG = 0.0077
PH = -165.9275
Z F2 = 0.6283
MAG = 0.0000
PH = -136.0150
Z F3 = 1.9855
MAG = 0.0100
PH = -44.0308
a0 = -0.0076
a1 = -0.0095
a2 = -0.0049
a3 = -0.0016
a4 = -0.0003
a5 = -2.4997E-5
SCALE POLY
POLY A
5.0000 RUN
R0= -7.646E-3
R1= -9.473E-3
R2= -4.905E-3
R3= -1.614E-3
R4= -2.899E-4
R5= -2.500E-5
OK?
POLY B
b0?
1 RUN
F0B/F0A?
RUN
bM:
M?
3 RUN
bN?
.94 RUN
F0B/F0A=1.645E0
R10= 1.000E0
R11= 2.030E0
R12= 1.736E0
R13= 9.400E-1
R14= 2.778E-1
R15= 3.940E-2
T1:
R00= 0.1985
R01= 0.0049
R02= -103.1899
R03= 0.0279
R04= -178.6871
R05= 0.0613
R06= -88.7546
R07= 0.0165
R08= -131.1444
T2:
R00= 0.1985
R01= 0.0024
R02= -119.4151
R03= 0.0084
R04= -177.5521
R05= 0.0613
R06= -90.8089
R07= 0.0136
R08= -143.2454

```

Figure C7.3. Program "D-B": Design B analysis using "ABCD".

paths. In addition, it uses subroutines "Y" and "Z" to incorporate biasing and parasitic immittances and load capacitance. The immittances are stored (as Foster-form C and R pairs) in registers R51–R66, and flag 02 is set to put "Y" and "Z" in automatic mode. With one exception, the immittances are stored in order of use since "Y" and "Z" automatically advance the address of the immittance to be added. Thus, after the 0.75 k Ω bias resistor in registers R51

```

01*LBL "D-B"
02 SF 02
03 SF 12
04 *DES-B"
05 AVIEW
06 ADV
07 CF 12
08 51.066
09 *RD"
10 AVIEW
11 RDTAX
12 PRREGX
13 XEQ 02
14 XEQ 02
15 XEQ 02
16 RTN

17*LBL 02
18 *T3:"
19 AVIEW
20 .008
21 RDTAX
22 XEQ 01
23 XEQ "PO"
24 XEQ "Y"
25 XEQ "ML"
26 XEQ "PO"
27 51
28 STO 50
29 *CL"
30 AVIEW
31 XEQ "Y"
32 XEQ "X-Y"
33 XEQ "ML"
34 XEQ "X-Y"
35 *T2:"
36 AVIEW

37 .008
38 RDTAX
39 XEQ "ML"
40 XEQ 01
41 XEQ "X-Y"
42 *T1:"
43 AVIEW
44 .008
45 RDTAX
46 XEQ "ML"
47 61
48 STO 50
49 XEQ "Y"
50 XEQ "ML"
51 XEQ "PO"
52 XEQ "Y"
53 XEQ "ML"
54 XEQ "Y"
55 XEQ "ML"
56 XEQ "PO"
57 XEQ "W"
58 RTN

59*LBL 01
60 53
61 STO 50
62 XEQ "Y"
63 XEQ "ML"
64 XEQ "Z"
65 XEQ "ML"
66 XEQ "Y"
67 XEQ "ML"
68 END

CAT 1
LBL "D-B"
END
204 BYTES

```

Figure C7.4. Program "D-B": listing.

and R52 at the third stage input is accounted for by executing "Y", the series resistance of the biasing diode in registers R53 and R54 is accounted for by executing "Z". The exception is that the biasing network, identical for the second and third stages, is used twice, at label 01 in the program.

The transistor *ABCD* parameters at the proper biases are contained on cards and are read into memory as called for by prompts. The immittances are also placed on a card and read in at the beginning of program execution. (The immittance card is read only once, whereas the transistor cards are read for as many frequencies as needed.)

At the conclusion of execution at each frequency, the *ABCD* parameters of the amplifier are written onto a card for later evaluation of the network characteristics. The program output for three frequencies is shown in the example.

To compare the results obtained here with those developed in Section 5.6, the voltage loss was calculated by using eq. (8.2-3) with source and load resistances of 0.15 k Ω . (The program, not shown, is straightforward.) This was done at the three frequencies, and the results were analyzed by using program "RQU" in Appendix A, with the results shown in Fig. C7.3. The polynomial coefficients were then scaled to the same cubic coefficient as in eq. (5.6-2), the equation for which the synthesis of the earlier section was made. Although the results are close, the quartic coefficient of the more accurate analysis is somewhat larger (although still stable).

C-8 PROGRAM "SP": SPANNING NETWORK ANALYSIS (USE WITH "CNV" AND "ABCD")

This program calculates the effect of a spanning network on the $ABCD$ parameters of a network. The network parameters are in X; the k parameters (h , z , y , or g) are in register set RS30 (registers R30–R38). The program accounts for feedback, feedforward, and input and output loading imposed by the spanning network. It is intended for evaluating matrix equations developed from equivalent ladder networks, as described in Section 8.4. In particular, it performs the operations of the equation using the information presented in Table 8.2 to place the k parameters in the right position in the β , F , H , and J matrices and to give them the sign indicated in Table 8.2.

To use the program, the k parameters of the spanning network must be determined and stored in RS30. (This is done most conveniently by analyzing the $ABCD$ parameters of the spanning network using "ABCD" and converting them to k parameters using "CNV".) Then we are ready to implement the matrix equation of the equivalent ladder network.

Four operations are defined in the subroutines of "SP": "BS" adds a β matrix (see Table 8.2) to the matrix in the working register set X; "FS" adds an F matrix to X; "HS" premultiplies the matrix in X by the input loading matrix H; and finally, "JS" postmultiplies the matrix in X by the output loading matrix J. As in the case of many commands (e.g., "ST" and "RC" of "ABCD"), the call for subroutines of "SP" must be preceded by an identifying number. The number 1 indicates that the spanning network is described by A feedback or by its h parameters; the number 2 indicates B feedback, or z parameters; the number 3 indicates C feedback or y parameters; and the number 4 indicates D feedback, or g parameters. The command "1, XEQ BS" says (1) that an h matrix is stored in register set RS30 and (2) that we wish to form a β matrix from it and add it to the matrix in X. Similarly, for the command "1, XEQ FS", an F matrix is formed from the h matrix and added to X. If we wish to *subtract* the β or F matrix from the matrix in X, we precede the numeral by a minus sign, thus: "-1, XEQ BS" or "-1, XEQ FS". (Commands "HS" and "JS" are never preceded by minus signs.)

```

01*LBL "BF"
02 XEQ "Y"
03 XEQ "A-Z"
04 30
05 XEQ "ST"
06 "T1:"
07 AVIEW
08 .008
09 RDTAX
10 2
11 XEQ "JS"
12 2
13 XEQ "HS"
14 XEQ "INV"
15 2
16 XEQ "FS"
17 XEQ "INV"
18 2
19 XEQ "BS"
20 RTN

01*LBL "APR"
02 SF 02
03 1 E12
04 STO 51

05*LBL 00
06 "IMM:"
07 AVIEW
08 51.054
09 PRREGX
10 "OK?"
11 CF 22
12 PROMPT
13 FS? 22
14 GTD 00

15*LBL 01
16 "ZF?"
17 PROMPT
18 STO 00
19 1 E12
20 STO 51
21 XEQ "BF"
22 40
23 XEQ "ST"
24 XEQ "FB"
25 30
26 XEQ "ST"
27 "T2:"
28 AVIEW
29 TONE 4
30 .008
31 RDTAX
32 1
33 XEQ "JS"
34 XEQ "INV"
35 -1
36 XEQ "FS"
37 XEQ "INV"
38 -1
39 XEQ "BS"
40 XEQ "X-Y"
41 40
42 XEQ "RC"
43 XEQ "ML"
44 XEQ "INV"
45 1
46 XEQ "FS"
47 XEQ "INV"
48 1
49 XEQ "BS"
50 BEEP
51 "APR:"
52 AVIEW
53 XEQ "W"
54 RTN

55*LBL "FB"
56 51
57 STO 50
58 0
59 STO 51
60 XEQ "Z"
61 XEQ "Y"
62 XEQ "ML"
63 XEQ "A-H"
64 RTN

65*LBL "BF"
66 51
67 STO 50
68 XEQ "Y"
69 XEQ "Y"
70 XEQ "ML"
71 XEQ "A-Z"
72 30
73 XEQ "ST"
74 "T1:"
75 AVIEW
76 TONE 4
77 .008
78 RDTAX
79 2
80 XEQ "JS"
81 2
82 XEQ "HS"
83 XEQ "INV"
84 2
85 XEQ "FS"
86 XEQ "INV"
87 2
88 XEQ "BS"
89 END

LBL"APR
LBL"FB
LBL"BF
END
266 BYTES
CAT 1
    
```

Figure C8.1. Program "APR": example of program for "SP".

This procedure allows us to program equations directly as we read them. If, for example, we wish to program eq. (8.4-17):

$$T_{1B} = \beta_B + [F_B + (H_B T_1 J_B)^{-1}]^{-1}$$

we would have the program fragment shown in the example in the first column of Fig. C8.1. The first three steps label the program, form an *ABCD* matrix of an impedance in *X*, and convert this matrix to *z* parameters. Steps 04 and 05

store the z matrix in RS30. Steps 06–09 put the matrix of a transistor (e.g.) in X . Steps 10 and 11 postmultiply the transistor $ABCD$ matrix by

$$J_B = \begin{bmatrix} 1 & z \\ 0 & 1 \end{bmatrix}$$

Steps 12 and 13 premultiply the combination by H_B (equal to J_B in this example). The combination is then inverted in step 14. Steps 15 and 16 add the matrix

$$F_B = \begin{bmatrix} 0 & -z \\ 0 & 0 \end{bmatrix}$$

to the matrix in X . (See Table 8.2 for the origin of the minus sign.) The combination is then inverted, and steps 18 and 19 add the β_B matrix (the same matrix, in this case, as F_B) completing the evaluation of T_{1B} .

The transmission matrix equation for the A -feedback pair in Fig. 8.19 [eq. (8.4-14)] illustrates the subtraction process for β and F matrices:

$$T_{\text{Apr}} = \beta_A + \left[F_A + \left[T_{1B} \left(-\beta_A + \left[-F_A + (T_2 J_A)^{-1} \right]^{-1} \right) \right]^{-1} \right]^{-1}$$

in which T_{1B} is the matrix of the first stage with local B feedback given in the preceding equation. Note that H_A does not appear in the equation; it is incorporated into T_{1B} . This equation is expressed in program “APR” (see Figs. C8.1 and C8.2), starting below the “BF” program fragment in the first column of the example. Note that “APR” includes “BF” in somewhat extended form. The program uses “BF” to find T_{1B} and stores it in RS40. In this section note that “Y” is executed twice, once for each of the admittances of the resistive feedback divider. Since these are Foster-form admittances, the series capacitance must be made very large (to have no effect). After finding T_{1B} , the program calls on subroutine “FB” to calculate the $ABCD$ matrix of the resistive feedback divider. For this calculation, the series feedback resistor appears as a Foster-form *impedance*; its associated capacitance is set to zero at steps 58 and 59. The lower resistor of the divider is obtained by executing “Y”, and the two matrices are multiplied at step 62. The resulting $ABCD$ matrix is converted to an H matrix, and the main program stores the h matrix in RS30, the register set that “SP” uses for the h parameters of the spanning network. With these preliminaries taken care of, the program then evaluates the equation for T_{Apr} .

Note that as step 35 (in Fig. C8.1) a -1 appears before executing “FS”. This signals “SP” to subtract the matrix rather than adding it, as called for by the equation; similarly, at step 38 the -1 signals “SP” to subtract the feedback matrix. As before, the program simply follows the equation step by step. The

```

XEQ "APR"
INM:
R51= 1.0000+12
R52= 0.9000
R53= 1.0000+12
R54= 0.1000
OK?
      RUN
ΔF?
      .1985 RUN
Y:
ΔF=0.1985
      51.0000
Y:
ΔF=0.1985
      53.0000
T1:
Z:
ΔF=0.1985
      51.0000
Y:
ΔF=0.1985
      53.0000
T2:
APR:
R00= 0.1985
R01= 0.1015
R02= 2.5460
R03= 0.0018
R04= 41.7335
R05= 0.0035
R06= 94.7796
R07= 0.0009
R08= 89.6057

XEQ "APR"
INM:
R51= 1.0000+12
R52= 0.9000
R53= 1.0000+12
R54= 0.1000
OK?
      RUN
ΔF?
      1.9850 RUN
Y:
ΔF=1.9850
      51.0000
Y:
ΔF=1.9850
      53.0000
T1:
Z:
ΔF=1.9850
      51.0000
Y:
ΔF=1.9850
      53.0000
T2:
APR:
R00= 1.9850
R01= 0.0962
R02= 26.4722
R03= 0.0117
R04= 97.6070
R05= 0.0535
R06= 152.9833
R07= 0.0148
R08= 148.9251

XEQ "RCU"
EVAL CU COEF
dB? CF 01
GHZ? SF 02
ΔF1: 0.1985
MG: 0.1015
PH: 2.5460
ΔF2: 1.9850
MG: 0.0962
PH: 26.4722
COEF:
a0 = 0.1016
a1 = 0.0227
a2 = 0.0039
a3 = 0.0003
XEQ "N"
SCALE POLY
POLY A
DEG?
      3 RUN
R0= 1.016E-1
R1= 2.273E-2
R2= 3.925E-3
R3= 2.062E-4
OK?
      RUN
POLY B
b0?
      1 RUN
F0B/F0A?
      RUN
bM:
M?
      3 RUN
bM?
      1 RUN
F0B/F0A=7.000E0
R10= 1.000E0
R11= 1.584E0
R12= 1.937E0
R13= 1.000E0

```

Figure C8.2. Program "APR": results.

01*LBL "SP"	47 FC?C 07	93 STO 09
02*LBL "BS"	48 SF 07	94 32
03 CF 07	49 RTN	95 STO 39
04 X<0?		
05 SF 07	50*LBL "FS"	96*LBL 14
06 ABS	51 CF 07	97 ΣREG 02
07 ENTER†	52 X<0?	98 CLΣ
08 X<>Y	53 SF 07	99 CLX
09 2	54 ABS	100 STO 08
10 /	55 2	101 1
11 FRC	56 X<>Y	102 STO 07
12 X=0?	57 X<=Y?	103 STO 01
13 XEQ 13	58 XEQ 13	104 RCL IND 39
14 RDN	59 *	105 X<> IND 09
15 2	60 STO 09	106 DSE 09
16 *	61 5	107 DSE 39
17 STO 09	62 -	108 RCL IND 39
18 34	63 ABS	109 X<> IND 09
19 STO 39	64 1	110 FS? 06
	65 -	111 XEQ "X-Y"
20*LBL 12	66 RCL 09	112 XEQ "ML"
21 RCL IND 09	67 5	113 RTN
22 DSE 09	68 -	
23 RCL IND 09	69 SIGN	114*LBL "JS"
24 P-R	70 *	115 SF 06
25 X<>Y	71 3	116 STO 39
26 RCL IND 39	72 *	117 XEQ "X-Y"
27 DSE 39	73 ST- 09	118 RCL 39
28 RCL IND 39	74 36	119 2
29 P-R	75 STO 39	120 /
30 RDN	76 XEQ 12	121 FRC
31 FS? 07	77 RTN	122 4
32 CHS		123 *
33 +	78*LBL "HS"	124 4
34 X<>Y	79 CF 06	125 +
35 R†	80 STO 39	126 STO 09
36 FS? 07	81 XEQ "X-Y"	127 38
37 CHS	82 RCL 39	128 STO 39
38 +	83 ENTER†	129 XEQ 14
39 R-P	84 X<>Y	130 END
40 STO IND 09	85 2	CAT 1
41 X<>Y	86 /	
42 ISG 09	87 FRC	
43 DEG	88 2	
44 STO IND 09	89 *	
45 RTN	90 +	
	91 2	
46*LBL 13	92 +	
		LBL*SP
		LBL*BS
		LBL*FS
		LBL*HS
		LBL*JS
		END
		END
		220 BYTES
		148 BYTES

Figure C8.3. Program "SP": listing.

results, at 0.0316 and 0.316 GHz are shown for this example in Fig. C8.2, using a $95 \mu\text{m}^2$ transistor operated at 1.5 V, 1.0 mA for both stages. See Fig. C8.3 for a listing of program "SP".

To clarify these results, the polynomial coefficients of A were found by using program "RCU" in Appendix A. These coefficients were scaled to a dc loss of unity and a cubic coefficient also of unity. The first-degree coefficient, 1.584, is small relative to Butterworth, MFD, or Chebyshev polynomials, indicating that a capacitor should be placed in parallel with the feedback resistor. (Why? Or why not?)

Index

- ABCD matrix, 211ff, 233ff, 273ff, 310ff, 384
 - cascade connection, 228, 297
 - matrix manipulation, 626
 - parameter manipulation, 295, 591ff
 - parameters, definition, 214
 - parameters of bipolar transistor, 233ff, 270ff
 - sign convention, 307
- Acceptor doping density, N_A , 239, 241ff
- Active load, 336ff, 436
- Active low pass filter, 129
- Active path, 4
- Active resonator, 135, 550
- Adjacency matrix, 128
- A-feedback, 215, 274
- A-feedback amplifier, 274
- A-feedback pair, 305, 640
- All pass network, 151
- Amplifier 741, 275
- Amplitude distribution of noise, 387
- Analysis by separation, 136
- Analysis by topological manipulation, 331, 335
- Angelo, E. J., 30
- Anticausal direction of analysis, 9
- Anticausal harmonic analysis, 349
- Anticausal nature of F parameters, 221
- Anticausal two-port analysis, 224
- Antisymmetric hybrid feedback, 288
- Appropriateness of two-port descriptions, 223
- Appropriate representation, 157
- Approximation problem, 73
- Asymptotes, 17
- Asymptotic bandwidth, 94, 548
- Asymptotic cutoff frequency, 48
- Augmentation of input signal, 101
- Avalanche multiplication, 248, 255
 - effect on drive requirement, 352
 - Miller's equation, 257
 - nonlinearity, 345
- B and C feedback combination, 319
- Band gap voltage reference, 374
- Barrier, 237
- Base current shot noise, 401
- Base push-out, 260
- Base resistance thermal noise, 402
- Belevitch, V., 222
- Bessel polynomials, 78, 189, 519
- B-feedback, 215, 274, 307, 340
- Bias dependencies, bipolar transistor, 264

- Binomial loss ratio, 49
- Bipolar code, 198
- Black, H. S., 1ff, 37, 38, 128, 161
- Blakesley transformation, 297, 392
- Bode, H. W., 2ff, 158
- Bode diagram, 16, 67
- Broadband equivalent circuit, 265
- Broad-band output stage, 357
- Brush, R. K., 488
- Built-in voltage, 237
- Bulk resistances, 269
- Buried layer, 236
- Burn-out protection, 369
- Burn-out region, bipolar transistor, 253
- Butterworth polynomials, 76, 189, 519

- Calculator, 17
- Calculator analysis of Design B, 187
- Calculator programs, 58
- Canonical equation, 4, 5
- Carlin, H. J., 228, 245
- Cascade graph, 128, 226
- Cascode stage, 315, 633
- Causal analysis: of distortion, 350
 - of two-ports, 223
- Causality, 8, 9
- C-feedback, 215, 274, 292
- Chain matrix, 213
- Characteristic equation, 342
- Characteristic polynomial, 342
- Chebyshev polynomials, 80, 519
- Chunking, 183
- Circuit vector node, 216
- Class A operation, 345
- Class B: efficiency, 359
 - emitter resistors, 363
 - operation, 345
 - optimum biasing, 365
 - temperature dependence, 359
- Classic feedback equation, 164, 559
- Classification of two-port networks, 221
- Collector junction capacitance, 233, 261
 - as feedback element, 269
- Collector shot noise, 390ff, 401, 415
- Comb filter response, 163
- Common base parameters, 311ff
 - lateral pnp, 166
- Common collector parameters, 311ff
- Common emitter output stage, 345
- Common mode input admittance, 435
- Common mode input current, 434
- Common mode input signal, 432

- Common mode rejection, 340, 433ff
 - admittance, 433
 - as function of signal intensity, 453, 459
 - at high frequencies, 454ff
 - measurements, 434
 - parameters, 433
 - ratio, 433
 - submatrix, 452
- Composite matrix for grounded two-port, 311
- Conductivity modulation, 242
- Conformal mapping, 68
- Convergence, iterative synthesis, 121
- Copernicus, 156
- Core coefficients, 276
- Core parameters of bipolar transistor, 233
- Corner frequency, 17
- Cosine rolloff characteristic, 409
- Crossover distortion in class B circuits, 360ff
- Cubic polynomial roots, 484
- Current generator splitting transformation, 392
- Current mirror, 165, 323, 633
 - h. f. transmission, 325
 - transmission characteristics, 323
- Current source, 345
 - active load, 426
 - simple, 323
 - Wilson, 327

- Damping factor, 49
- Darlington pair, 318, 341, 633
- DC bias, Design B, 197
- DC level shift, 165
- DC restoration, 198
- Decision circuit, 198ff, 577
- Defect current, 188, 241
 - variation with collector current, 241
- Defect current ratio, 233, 243
- Deflation, polynomials, 487
- Deflection systems, tv, 25
- Delay, 179
 - in ABCD parameters of devices, 270
 - in common base stage, 315
 - in current mirror, 327
 - due to base resistance, 179
 - in transistors, 259ff
- Density gradient, 237
- Dependent node, 125
- Depletion region, 238, 248. *See also* Barrier
- Design A, 98ff
- Design B, 99ff, 183

Index

- with delay, 183ff, 564
- program, 532
- sensitivities, 107ff
- synthesis, 103
- two-port analysis, 321
- D-feedback, 215, 274, 340
- Differential ABCD matrix, 433
- Differential amplifier, 430
- Differential gain (or loss), 364
- Differential input signal, 432
- Differential pair, 337
 - basic constraint, 448
 - general analysis, 447
 - inverting, 337
 - noninverting, 337
- Differential two-port submatrix, 452
- Diffusion constant, 239
- Diffusion equation, 239
- Digital regenerator, 577, 585
- Diode connected transistor, 392
- Diode shot noise, 386
- Direct feedthrough, 101
- Disconnected graph, 226
- Dispersion, 157, 179
 - in media with transit time, 179
- Distortion, 4ff, 99, 354
 - analysis, 5, 11-13
 - cancellation, 354
 - as function of source impedance, 353
 - reduction, Class B, 366
 - transient, 354
- Distributed parameters, 221
- Distribution, 125
- Dominant element, 118
- Driver stage design, 345
- Dynamic crossover distortion, 369
- Dynamic nonlinearity, 24

- Early conductance, 278, 325
- Early effect, 247
 - effect on common mode rejection, 445
 - effect on input offset voltage, 442
- Early voltage, 248
- Ebers-Moll model: in forward-active operation, 244
 - in saturated operation, 253
- Efficiency: Class A, 347
 - Class B, 359
- Emitter contact resistance, 194, 251
- Emitter coupled pair, 315
- Emitter follower, 356
 - output stage, 193
- Emitter junction capacitance, 262
- Emitter resistance, r_e , 233

- Epitaxial layer, 236
- Epochs: mixed, 221
 - separate, 221
 - signal, 221
- Equivalent circuits, 217, 242
 - bipolar transistor, 242, 265ff
 - null reference, 247
- Equivalent input noise network, 398
 - correlations, 400
- Equivalent ladder networks, 100, 297
- Evaluation: of polynomials, 67ff, 500
 - of rational functions, 500
- Excitation and response, 214

- Factoring, 125
- Feedback, defined, 156
- Feedback matrix, 301
 - local, 301
 - overall, 301
- Feedback path, 4
- Feedback synthesis, 93, 546
- Feedforward, 150ff, 162, 290, 310, 336, 385
 - active load, 336
 - effect on noise, 417
 - incidental, 302
 - matrix, 302
 - in power amplifier, 155
- Fiber optic preamplifier, 408
 - at 274 mb/s, 422
- Field emission breakdown, 256
- Fleischer, P. E., 141
- Flicker noise, 388
- Fluctuation noise, 388
- Foster form RC immittances, 297
- Frequency compensation, 144, 552
- Frequency normalization, noise integration, 406

- Gardner, M., 128
- General circuit parameters, 213
- Generation of null reference matrix, 447
- Gray, P. R., 236
- Gummel number, 240, 249

- Harmonic analysis of common emitter stage: anticausal, 349
 - causal, 350
- Hewlett-Packard HP 41C calculator, 453
 - bar code, 544
 - card reader, 474
 - peripherals, 473
 - printer, 530
- Hierarchical levels, 211
 - in design process, 183, 572

- High level injection effects, 260
- Horizontal geometry, 234, 265
- HP 41-C, HP 41-CV, 473
- Hybrid feedback, 275, 276, 292
- Ideal amplifier, 22ff, 214, 228, 245, 285
 - defined, 228
 - nullor, 228
 - 2-port equivalent circuit, 214
- Ideal operational amplifier, 129
- Ideal transformer, 285, 419, 432
- Immittance control, 285
 - in simple current source, 394
- Incidence matrix, 128
- Independent node, 125
- Infinite gain amplifier, 22
- Input admittance, 277
- Input bias current, 435
- Input common mode range, 457
- Input current drive requirement, 351
- Input differential pair, 165
- Input drive requirement, 345
- Input loading, 101
- Input loading matrix, h , 305
- Input noise network, 398
- Input offset current, 435
- Input offset voltage, 376, 435, 438, 442
- Insertion loss, 287
- Integrated circuit delays, 181
- Intersymbol interference, 408
- Intrinsic silicon, 236
- Inverse, of ABCD matrix, 277
- Isolation diffusion, 236
- Isoplanar transistor, 248
- Iterative synthesis, 118
- Jacobian, 119
- Junction, metallurgical, 236
- Key concepts of bipolar transistor, 242
- Kirk effect, 260, 261
- Kuhn, Thomas, 156
- Lateral pnp transistor, 165
 - construction, 166
- Lead-interchange (permutative) feedback, 311
- Level shifter circuits, 461
- Linear approximation to delay, 173
- Linearity of Class A and Class B, 369
- Linearity with local B-feedback, 354
- Load capacitance, 58, 189
- Load line, 347
- Local feedback, 99, 109
- Loop gain, 3ff, 156, 223, 227
- Lossless feedback, 132, 286
 - hybrid, 286
 - symmetric, 287
- Loss and phase, mean and std. dev., 540
- Loss ratio, 8
- Loss variation, 566
- Loudspeaker equalizer, 153, 557
- Lumped parameters, 221
- Majority carriers, 237
- Mason, S. J., 156
- Master-slave flip-flop, 198
- Maxwell-Boltzmann statistics, 239
- Metasystem, 172
- Meyer, R. G., 236
- MFA polynomials, 76, 519
- MFD polynomials, 78, 519
- Microwave integrated circuit transistors, 271
- Microwave transistor, 236
- Minority carriers, 237
- Mitra, S. J., 245
- Modeling by polynomial coefficients, 73
- Moschytz, G. S., 155, 245
- Network loss, 277
- Newton's method, 58, 484
- Noise: in bipolar transistor, 401
 - with C feedback, 421
 - in common base stage, 419
 - in common collector stage, 418
 - in differential pair, 419
 - effect of feedback, 417
 - effect of feedforward, 417
 - equivalent conductance, 388
 - equivalent resistance, 387
 - in FETs, 397, 403
 - figure, 404
 - with lossless feedback, 419
 - with lossy feedback, 421
 - temperature, 405
 - weighting, 405
- Non-energetic feedback, 132, 421
- Norator, 245
- Normalization, frequency, 48, 475
- Nullator, 245
- Null matrix, 214
- Nullor, 245
- Null reference, 37
- Null reference equivalent circuit, 247
- Null reference matrix, 431, 432

- parameters, 439
- Nyquist, H., 2ff
- Nyquist diagram, 44, 68
 - for CFE, 167ff
 - feedback around pure delay, 163
- Nyquist frequency, 409
- Nyquist theory of stability, 70

- Observable, 152
- Observable signal, 98
- One-over-f noise, 388
 - excess noise, 388
- Operational amplifier, 129ff, 430ff, 456ff
- Operational amplifier 741, 129, 165ff, 178ff, 274, 457
- Operational transconductance amplifier, 464
- Optimum receiver design, 417
- Out-of-band stability margins, 189
- Output impedance, 277
- Output loading, 101
- Output loading matrix, 305
- Output reflection coefficient, 219
- Output stage, 344
 - single transistor, 345
- Output stage biasing, 366

- Parasitic capacitance, 181, 196, 358, 414, 464
- Passive network, 228
- Permutative feedback, 311
 - restriction on application, 322
- Pert analogy, 37
- Phase margin, 179
- Phase response, for stable and unstable systems, 68
- Pipo (C feedback), 274
- Piso (D feedback), 274
- Polynomial: coefficients from loss and phase, 510, 514, 616
 - deflation, 487
 - evaluation, 67, 500
 - performance specifications, 75
 - programs for manipulating, 473
 - scaling in frequency, 475
- Power supply isolation, 468
- Power supply rejection, 469
 - null reference matrix, 470
- Power supply signal current, 468
 - useful application, 468
- Predistortion, corrective, at input, 4, 11, 345, 349, 357
 - unbounded, 350
- Programmable network calculator, 633
- Ptolemy, 156
- Punch through, 248
- Pure delay, feedback around, 163
- Push-pull operation, 356

- Q (quality factor), 50
- Quadratic approximation to delay, 173
 - in CFE, 176
- Quadratic equation, 483
- Quadratic loss ratio, 49
- Quantized feedback, 198ff, 577
 - quadratic cutoff, 202
- Quartic performance polynomial, 189

- Rational functions, 124, 146, 279
 - evaluation, 332, 500
- Ratios of polynomials, 149
- Receiver sensitivity, 413
- Receiving node, 125
- Reciprocal formulation, 8
- Recombination, 240
- Rectification, precision, for a/d conversion, 460
- Reference condition, 37, 247, 301, 358
- Reflection coefficient, 219
 - input, 219
 - output, 219
- Regenerative feedback, 13
- Residues, 63ff, 200ff, 488, 580
- Response adjustment in fiber optic preamp, 426
- Responsivity (photodiode), 412
- Return difference, 5ff, 158
 - relation to sensitivity, 35, 36
- Reverse active operation, 251
- Riemann surface, 71
- Root locus (diagram), CFE, 168ff, 177, 529
- Roots: cubic, 483
 - by Newton's method, 488
 - quadratic, 484
 - starting point, 488, 489

- Sallen-Kay filter design, 557
- Saturation, 248, 252, 345
- Saturation current, 239
- Scattering limited velocity, 159, 237
- Sensitivities: coefficient to component, 106, 535
 - frequency dependent loss ratios, 29-35
 - loss to components, 109ff, 537
 - loss to polynomial coefficients, 87

- polynomial coefficients to device, 107, 108
- rational function to coefficients, 524
- resonator, 140
- Sensitivity, 5ff, 36
- Sequential matrix, 128
 - anticausal, 128
 - causal, 128
- Series port impedances, 281
- Settling time, 66
- Shot noise, 246, 385ff
- Shunt port admittance, 281
- Signal delay, 162
- Signal flow graphs, 125
 - of 2 port ABCD parameters, 214
- Signal intensity, 437
- Signal vector, 212
 - dependent, 213
 - independent, 213
- Single amplifier biquad, 135
- Sink node, 125
- Sipo (A feedback), 274
- Siso (B feedback), 274
- Slew limiting, 178, 353
- Slew rate requirement, 360
- Slope overload, 353
- Source impedance, effect on distortion, 353
- Source node, 125
- Spanning network, 295ff, 303, 327ff, 336ff, 638
- S parameters, 219
- Spot noise figure, 405
- Stability, 5
- Standby power, 345
- Start-up circuit, band gap reference, 377
- Statistical analysis, 188
- Step response, 63, 497
- Summation of all noise sources in network, 401
- Sum rule for sensitivities, 10, 29, 106, 104ff, 189, 207
- Sustain voltage, 258
- Symmetric feedback, 287
- Synthesis of Design B, 103
- Synthesis of transmission characteristics, 423
- “T” compensation, 144, 556
- T4M digital transmission system, 207
- Tandem connections, 274
 - b-c combination, 274
- Temperature effects, 370
- Terminating immittances, 211
- Thermal noise, 386
- Time domain analysis, 63, 81, 86, 497
 - quantized feedback, 198
- Time domain performance, 63
- Topology, 125
- Transfer coefficient, 219
 - forward, 219
 - reverse, 219
- Transfer parameters, 219
 - F parameters, 220
- Transformer feedback, 285
- Transient distortion, 354
- Transistor, 323
 - ABCD parameters, 325
 - action, 237
 - analysis, 233
 - characteristics from ABCD parameters, 314
 - diode connected, 325
 - equivalent circuit analysis, 599
 - gain mechanism, 240
 - modeling, 606
 - noise, 390
- Transitional polynomials, 79, 189, 519
- Transit time, 240, 258
 - delay, 266
 - in devices, 179
- Transmission matrix, 213, 285ff, 323
- Transmission matrix signal flow graph, 216
- Transmitting node, 125
- Transport model of bipolar transistor, 244
- Tunneling, 256
- Two-port analysis, general method, 295
- Two-port constraint, 212
- Two-port network calculator, 626
- Two-port parameters, 100
 - ABCD parameters, 213
 - conversions, 218, 592
 - equivalent circuits, 217
 - signal flow graphs, 217
 - y parameters, 100, 213
- Unilateral feedback amplifier, 289
- Unilateral network, 228
- Unitary feedback, 275
- Units, consistent set, 26
- Unity current loss frequency, f_t , 233
- Unity gain follower, 132, 299, 306
- Unity gain frequency, 17
- Unity loss time constant, 44
- Unity voltage loss time constant, 129

Index

Upper triangular matrix, 133

V_{be} vs. temperature, 370

Variability: resonator, 141
variation of loss, 112, 141

Vertical geometry, 236

Wave formulation, 219

Webster effect, 242, 260

White noise, 386

Wickes, W. C., 544

Wierstrass theorem, 73

Wilson current source, 327, 633
hf transmission, 331
stabilization, 331

Zero loss amplifier, 22-24

Zimmermann, H. J., 156

341020

TK
7835
W29
1982

Waldhauer, Fred D.
Feedback / Fred D. Waldhauer. -- New
York : Wiley, c1982.
xv, 651 p. : ill. ; 24 cm.
"A Wiley-Interscience publication."
Includes bibliographical references
and index.
ISBN 0-471-05319-8

1. Feedback (Electronics) I. Title

23 JUN 83 7739914 STAMs2 81-13104

